



**HAL**  
open science

# Study of the thermodynamics and cooperativity of protein folding by high pressure

Martin Fossat

► **To cite this version:**

Martin Fossat. Study of the thermodynamics and cooperativity of protein folding by high pressure. Other. Université Montpellier, 2016. English. NNT : 2016MONTT326 . tel-01816939

**HAL Id: tel-01816939**

**<https://theses.hal.science/tel-01816939>**

Submitted on 15 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

Pour obtenir le grade de  
Docteur

Délivré par l'Université de Montpellier

Préparée au sein de l'école doctorale  
Information Structures Systèmes

Et de l'unité de recherche  
Laboratoire Charles Coulomb

Spécialité: **Physique**

Présentée par **Martin Fossat**

**Study of the thermodynamics  
and cooperativity of protein  
folding by high pressure**

Soutenue le 15/12/2016 devant le jury composé de

Dr Andrea PARMEGGIANI	Laboratoire Charles Coulombs	Directeur
Dr Christian ROUMESTAND	Centre de Biochimie Structurale	Co-directeur
Dr Martin BLACKLEDGE	Institut de Biologie Structurale	Rapporteur
Dr Michael NILGES	Institut Pasteur	Rapporteur
Dr Carine VAN-HEIJENOORT	Institut de Chimie des Substances Naturelles	Présidente du Jury
Dr Catherine ROYER	Rensselaer Polytechnic Institute	Marraine de thèse



**Collège  
Doctoral**  
Languedoc-Roussillon





## Acknowledgments

First of all I would like to thank Catherine Royer for the opportunity she gave me by offering me this thesis, and the support she has given me throughout the years. I owe her the start of a scientific career. I would like to thank Angel Garcia for including me in his groups and sharing his precious time to answer my questions on Molecular Dynamics, and Christian Roumestand for supporting me in my summers in CBS, and for the knowledge he helped me acquire in NMR. I would also like to thank my parents and my grandmother for having educated me in the purpose of making me an independent and curious person. I would also like to thank all the people that contributed to my education, including my teachers, but in particular my Japanese host parents, Masaru and Sumiko Onishi, that have taught me a lot as a teenager and contributed greatly to my personal development. I would also like to thank Teri Hreha for being present and compassionate during the hard times of this graduate life. Finally I would like to thank the Jury members. It was an honor to present my work in front of such renowned experts.



# Contents

<b>1</b>	<b>Introduction</b>	<b>15</b>
1.1	Background . . . . .	15
1.2	Models for protein folding . . . . .	18
1.2.1	The two-state model . . . . .	18
1.2.2	The hydrophobic model . . . . .	20
1.2.3	The spin glass model . . . . .	20
1.2.4	The molten globule perspective . . . . .	22
1.2.5	Foldons and the hierarchy of protein folding . . . . .	24
1.2.6	The Ising model . . . . .	26
1.2.7	Conclusion and origin of folding pathways . . . . .	27
1.3	Introduction to thermodynamics of proteins . . . . .	30
1.4	Origin of the determinants of thermodynamic stability . . . . .	34
1.4.1	Temperature effect - entropy . . . . .	34
1.4.2	Pressure effect - the volume change upon unfolding . . . . .	40
1.4.3	Temperature-pressure cross term - Expansivity . . . . .	44
1.4.4	Effects of cosolvents on the thermodynamic stability of protein . . . . .	48
1.4.5	Conclusion on thermodynamic stability of folded proteins . . . . .	51

<b>2</b>	<b>Materials and methods</b>	<b>57</b>
2.1	Sample preparation . . . . .	57
2.1.1	PP32 . . . . .	57
2.1.2	Trp cage . . . . .	58
2.2	NMR Theory . . . . .	59
2.2.1	Reference frame and rotating frame . . . . .	61
2.2.2	Shielding effect . . . . .	62
2.2.3	Free induction decay . . . . .	62
2.2.4	Multidimensional experiments . . . . .	64
2.2.5	Chemical exchange in NMR experiments . . . . .	66
2.2.6	Experiments and analysis . . . . .	67
2.3	Molecular Dynamics simulations . . . . .	69
2.3.1	Theory . . . . .	69
2.3.2	Replica exchange . . . . .	76
2.3.3	Principal component analysis . . . . .	78
2.4	$\Phi$ -value analysis . . . . .	79
2.4.1	Activation energy . . . . .	80
2.4.2	Hollow : cavity visualization algorithm . . . . .	82
2.4.3	SMOG : Shadow contact algorithm . . . . .	82
2.4.4	Structure Based Modeling . . . . .	83
2.4.5	Pulchra . . . . .	84
2.4.6	Presentation of the calorimeter . . . . .	85

2.4.7	Data processing . . . . .	86
<b>3</b>	<b>Investigation of PP32 folding landscape through high pressure NMR</b>	<b>91</b>
3.1	Introduction . . . . .	91
3.1.1	Literature on PP32 . . . . .	91
3.1.2	Goals and preliminary study . . . . .	93
3.1.3	Deviation from two-state . . . . .	93
3.2	Conclusion and perspective . . . . .	103
3.3	Resume de l'article en francais . . . . .	104
<b>4</b>	<b>Effect of cavity mutation on the unfolding landscape of PP32</b>	<b>107</b>
4.1	Introduction . . . . .	107
4.2	Materials and methods . . . . .	107
4.3	Results and discussion . . . . .	108
4.3.1	L60A mutant . . . . .	109
4.3.2	L139A mutant . . . . .	112
4.4	Conclusion . . . . .	113
<b>5</b>	<b>Origins of the determinant of thermal expansivity through High pressure NMR</b>	<b>115</b>
5.1	Introduction . . . . .	115
5.2	Resume de l'article en francais . . . . .	126
<b>6</b>	<b>High pressure study of the tryptophan cage variant Tc5b</b>	<b>127</b>
6.1	Introduction . . . . .	127

6.1.1	NMR signal interpretation . . . . .	128
6.2	Resume de l'article en francais . . . . .	139
<b>7</b>	<b>Annex</b>	<b>141</b>
7.1	Thermodynamics in protein folding . . . . .	141
7.1.1	Detailed introduction to thermodynamics . . . . .	141
7.2	Protocols . . . . .	149
7.2.1	Transformation in competent cells . . . . .	149
7.2.2	Protein Production of PP32 . . . . .	149
7.3	Definitions . . . . .	152
7.3.1	Legendre transform . . . . .	152
7.3.2	Euler's equation . . . . .	152
7.3.3	Extent of reaction . . . . .	153
7.3.4	Conditions for a solution to be ideal . . . . .	153
7.3.5	Standard state . . . . .	153
7.3.6	Maxwell relations . . . . .	153
7.4	Supplementary information . . . . .	156
7.4.1	High-Resolution Mapping of a Repeat Protein Folding Free Energy Landscape . . . . .	156
7.4.2	Evolutionarily Conserved Pattern of Interactions in a Protein Revealed by Local Thermal Expansion Properties . . . . .	196
7.4.3	NMR and Computation Reveal a Pressure-Sensitive Folded Conformation of Trp Cage . . . . .	213
7.5	Abbreviations . . . . .	223





# List of Figures

1.1	Funnel representations of the energy landscape . . . . .	18
1.2	Phase diagram for protein folding . . . . .	21
1.3	ANS binding in the molten globule state at low pH . . . . .	22
1.4	Hydrogen exchange of the folded state hints at the existence of foldons . . . . .	24
1.5	Correlation between contact order of a protein, the kinetics of folding and the placement of the transition state . . . . .	25
1.6	Comparison between experimentally determined folding landscape of a natural repeat protein and a consensus repeat protein. . . . .	27
1.7	Decomposition of a DSC experiment. . . . .	35
1.8	Temperature dependencies of the contributions to the experimental entropy change upon protein unfolding . . . . .	37
1.9	Diagram of the Voronoi Delaunay volume decomposition method applied to folded proteins . . . . .	43
1.10	Expansivity of amino acids measured by PPC. . . . .	45
1.11	Expansivity of the folded and unfolded state as a function of temperature . . . . .	46
1.12	Correlation between expansivity change with temperature and heat capacity . . . . .	47
1.13	Artist rendering of the cell interior . . . . .	50
1.14	Chemical shift perturbations of the PP32 Y131F/D146L variant compared with wild-type. . . . .	52

1.15	Pressure temperature phase diagram of Chymotrypsinogen, Chymotrypsinogen and SNase . . . . .	54
1.16	Pressure temperature phase diagram of proteins . . . . .	55
1.17	Effect of third degree terms of the pressure temperature phase diagram of proteins	55
2.1	Reversibility of PP32 WT by 1D NMR . . . . .	58
2.2	Time scale of motion in a protein . . . . .	70
2.3	Water density as a function of temperature for different water models . . . . .	73
2.4	Probability distribution of the total potential energy for eight replicas . . . . .	77
2.5	Representation of the Hollow algorithm method. . . . .	81
2.6	Shadow Contact maps algorithm diagram . . . . .	83
2.7	Representation of the side chain steric clash resolving algorithm in PULCHRA .	85
2.8	PPC raw data . . . . .	89
3.1	Sequence of PP32 WT . . . . .	91
3.2	$\phi$ -values and protection factors of PP32 . . . . .	92
3.3	Pressure of half unfolding plotted as a function of temperature for PP32 WT . .	103
4.1	Position of the cavity creating mutation in PP32 . . . . .	108
4.2	NMR spectra comparison between PP32 and cavity enlarged mutants . . . . .	109
4.3	NMR spectra at high pressure of the L60A PP32 mutant . . . . .	110
4.4	Contact map analysis of the L139A PP32 mutant compared to the WT protein .	111
4.5	Contact map analysis of the L139A PP32 mutant compared to the WT protein .	112
4.6	Pseudo free energy diagram of the L139A PP32 mutant compared to the WT protein . . . . .	113

4.7	<i>C</i> - $\alpha$ RMSD deviation to the crystal structure vs fraction of native contacts for PP32 WT and its L139A mutant . . . . .	114
-----	--	-----



# Chapter 1

## Introduction

### 1.1 Background

Protein folding is one of the most fundamental problems in modern biology. Proteins are the functional units of organisms. Each protein is a chain of building blocks, the amino acids. There are twenty amino acids used in nature, all constituted of a similar backbone structure. The difference between amino acids lies in the properties of their side chains. Four main parameters differentiate all amino acids' side chains : their geometry (length and structure), their polarity or their acidity (capacity to be charged at certain pH) and their aromaticity.

Amino acids are zwitterions and can react with one another in a polarized manner to form a covalent bond between the amine and carboxyl group, releasing one water molecule. That covalent interaction is often referred to as a peptide bond, and is made upon t-RNA binding to the peptidyl transfer site of the mRNA ribosome, a process that is repeated for each codon of the mRNA, and stops when the ribosome dissociates at the stop codon from the mRNA and releases the newly formed polypeptide chain into the cytoplasm. To enable its function, most proteins need to fold into a complex 3-dimensional structure. To ensure a proper fold, the primary sequence, meaning the order of the different amino acids in the polypeptide chain, is designed so that the functional tri-dimensional fold of the protein is the state of lowest energy in native conditions, with the exceptions of fibrils, misfolded aggregates, and intrinsically disordered proteins (IDPs) that have a fluctuating structure even in their native state. The stability of the folded state relative to the fully unfolded, coil-like, state depends entirely on the properties and position of the amino acids that compose the protein. However, proteins have not evolved only in the direction of greater stability of the folded state. Folding cooperativity, as to achieve folding in a reasonable timescale, as well as to promote the population of some functionally relevant intermediates is also a key feature of functional proteins. In addition, the

stability of oligomers as a way to achieve function regulation through the binding constant, is also evolution driven and encoded in the primary sequence. The extraordinary amount of information necessary to obtain a folded, functional protein poses as a great, fundamental challenge that still puzzles this 70 years old field.

Understanding the link between the composition of the chain of amino acids, the primary sequence - encoded in genes -, the secondary structure formation, and the complex three-dimensional structure that enables a protein to accomplish its function is an essential step towards the goal of designing function specific synthetic proteins. However, despite considerable efforts in this domain, quantitative answers as to what parameters influence protein stability, giving rise to the native state to be the most stable conformation, remains an important goal. The number of parameters to consider renders the problem very ambiguous. In addition, the difference in stability between the states that constitute the structural ensemble is marginal considering the forces in play : the energy associated with hydrogen bonds is of the same order of magnitude as the typical difference of stability between the folded and unfolded states of a protein[1, 2]. In practice, that makes the determination of the relative stability between two states from calculation of potentials from a given structure require a enormous precision to be relevant, rendering purely theoretical approaches to the folding problem impossible without strong experimental background. To make matters worse, such experimental information is very challenging to obtain, and each method used can only yield insight to a rather small number of parameters, often with low spatial resolution, and most can only observe statistical ensembles, giving no direct insight on the mechanism of folding itself.

In an effort to overcome these limitations, the number of investigative approaches, both computational and experimental have greatly increased in the last few decades. In particular, the use of molecular dynamics for increasingly larger systems has been made possible thanks to the ever-growing computing capacity of modern computers and their cost decrease. The use of such molecular level observation tools, however, is still limited in the interpretation we can make from the data because of the clear deviations from fundamental properties (diffusion rate, density of water, bonds torsion angles or secondary structures propensity) those systems show. These biases have been studied in detail, and a wise choice of force field and water model used to answer a specific question can be sufficient to avoid artifacts. Despite these approximations, computational approaches remain a useful tool to explain experimental data by giving an insight into the molecular mechanisms present in our experiment. Other methods have also shown great improvement. For example, the multiplication of pulse sequences in NMR in the last 50 years increased greatly the toolbox for investigating the complex systems that are proteins.

One other source of improvement in the understanding of protein stability was the implementation of new denaturation methods for a variety of experimental devices. In particular, the combination of Pressure with NMR, FTIR and SAXS enables investigation of protein stability

in much greater detail, providing information on the stability along the second thermodynamic variable that contributes to stability, volume. Other methods have also emerged such as cold denaturation, the study of crowding effects in the protein folding stability, in-vivo folding or the use of optical tweezers for mechanical unfolding.

The extreme diversity of sequences present in living organisms is also a considerable challenge, but thanks to the rapidly increasing number of resolved native structures over last 20 years, the emergence of a database of different topologies with a spectrum of architectures and complexity has enabled to select proteins whose properties correspond to the feature one wishes to investigate. For example, the increased use of repeat proteins for their simple architecture has helped improve our understanding of the interactions between folding domains[3], and their role in the folding process; thus those systems prove a great model for studying energetic coupling in the tertiary structure. The design of synthetic miniproteins that fold in a timescale allowing molecular dynamics to simulate entire folding events within reasonable computing times, such as the Tryptophan-cage protein[4], allowing direct comparison of simulation and experiments, is of great interest to get a molecular insight on the folding mechanism. In addition, the possibility of identifying the conserved primary sequence positions among a wide family of proteins with conserved tertiary structure helps to identify sites that are likely to be essential for folding and protein stability. Site directed mutagenesis can be performed to change the stability of a protein, or to study the effect of specific mutations. For example, mutations from hydrophobic to ionizable residues can be used to determine the pKa of the protein core[5], and polar residues can be mutated to apolar residues to study the role of polar hydration in some thermodynamic properties of the folded state[6]. Mutations can also be used to apply the  $\phi$ -value kinetic analysis to characterize the transition state[7].

In this thesis, we will continue the investigation of the stability of proteins and the folding process using pressure denaturation. For that we will start by a description of different models used to describe protein folding and a discussion on their applications and limits. The second part of the introduction will be an introduction to protein folding thermodynamics. The last part of the introduction will be a small review of the origin of the thermodynamic quantities that drive the folded state stability in the two-state approximation. Finally, there will be a conclusion on the origin of protein stability, and a presentation of the temperature pressure phase diagram representation of proteins stability.

## 1.2 Models for protein folding

In 1969, Levinthal reported a thought experiment stating that, because of the enormous number of degrees of freedom of a polypeptidic chain, if a moderate size protein had to sample all possible conformations to find the folded state, it would result in a folding time longer than the age of the universe, whereas the observed timescale for folding of small proteins is on the order of milliseconds[8, 9]. This paradox highlights that proteins do not explore the energy landscape in a stochastic manner, but have a directed search for their energy minimum, the folded state. A number of models have emerged in order to give a phenomenological explanation to the protein folding problem that could resolve this paradox. In this section, we will discuss the implication of these models for the protein folding problem, and their relevance. Different scenarios can be imagined for protein folding, and can be represented as a folding funnel, to have a visual approach of the folding landscape, where the stability is the depth of the funnel, and the radius is the entropy of the states populated during folding, with the folded state at the center (figure 1.1). The models presented in this section are not mutually exclusive, and correspond to specific situations that may not be applicable to a given protein.

### 1.2.1 The two-state model

One of the simplest models that can be used to study protein folding is the two-state model. This model relies on the assumption that proteins fold with no intermediate in a very cooperative manner, where only two discrete states are populated: the folded and unfolded states. The two-state model suggests that all native interactions are formed at once, and thus the folding problem is in fact similar to a first order phase transition, with a disordered unfolded phase and

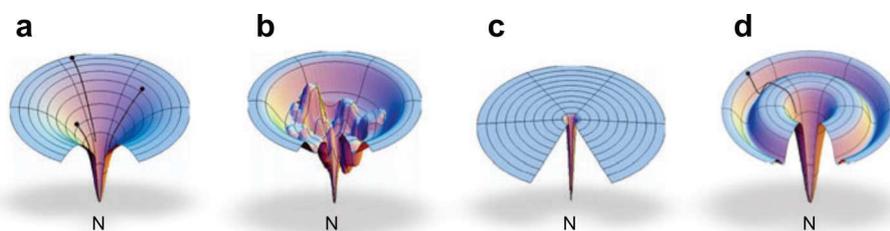


Figure 1.1: One type of energy landscape cartoon. These pictures give a sort of simplified schematic diagram, useful for illustrating a protein's partition function and density of states. (a) A smooth energy landscape for a fast folder, (b) a rugged energy landscape with kinetic traps, (c) a golf course energy landscape in which folding is dominated by diffusional conformational search, and (d) a moat landscape, where folding must pass through an obligatory intermediate. Figure taken from [10].

an ordered folded state. Thus, in the two-state model, all elements are considered to form in a single step. This is characteristic of a golf course type energy landscape (figure 1.1, panel C)[11]. Although this view is certainly not correct, the two-state model is advantageous to use for the study of protein folding because most globular, monodomain proteins will fold in an apparently two-state manner due to the lack of resolution of the methods used. In most cases, the two-state model is the only way one can extract information on the folding process. This is why, in the next section, where we will discuss the origin of the thermodynamic parameters driving protein folding, we will use a two-states approximation.

One of the major limitations of the two-state model is the fact that the nature of the unfolded state may vary depending on the method used for denaturation (pH, urea, temperature, pressure). This can result in intrinsic differences between measurements of the same thermodynamic parameters at different conditions. Therefore, the difference in volume obtained by pressure denaturation, and the difference in entropy obtained by high temperature denaturation might not actually be relative to the same unfolded ensemble. Consistent study of the residual structure of an acidic denatured state, acidic/high-temperature denatured state and urea denatured state of the Barnase globular protein by heteronuclear NMR has shown that the amount of residual structure may depend on the type of denaturation used[12]. In another example, the acid-denatured states of ribonuclease, lysozyme and chymotrypsinogen, all three common model globular proteins for the study of protein folding, have shown to retain large amounts of remaining structure, with the observation of a second Guanidinium induced transition after acidic denaturation of the protein was achieved[13, 14], thus the acidic denatured state is not fully unfolded and corresponds to an intermediate. In another study, the temperature denatured state of the  $\alpha$ -Lactalbumin globular protein was shown to result in a relatively compact but unfolded state, with no noticeable transition occurring after this state was reached, as well as a significantly different NMR spectra from both the native and Guanidinium denatured states[15, 14], underlying that in some proteins, acidic and temperature induced unfolding can lead to very different unfolded ensembles. These differences in the nature of the denatured state depending on the denaturation method should always be kept in mind, as this can result in discrepancies in the stability measurements between different methods because one of the two states used to measure the relative stability is different. Furthermore, the stability of the folded state relative to that of the unfolded state is typically extrapolated to reflect the differences that exist before the application of a denaturant, but in those more native-like conditions, the transient denatured state may be different from the one populated in the conditions used to do the extrapolation.

### 1.2.2 The hydrophobic model

The hydrophobic model is one of the first models developed for protein folding, and relies on one of the main features of protein folding, that the polypeptidic chain folds to minimize the exposure of hydrophobic residues to the aqueous solvent[16]. For decades, the molecular origin of this observation has puzzled biophysicists. A first problem appeared very rapidly with this model, because of its incapacity to explain pressure denaturation of proteins[17, 18]. This was due to the assumption made based on the model compound transfer method that the exposure of hydrophobic side chains to solvent upon unfolding would lead to a significant increase in volume. However, the estimate of that effect was shown to be an artifact due to the smaller density of packing of hydrophobic compounds compared to the packing of the core of the protein, causing a large negative change in volume upon the transfer from apolar to aqueous solvent.

The hydrophobic effect is rather pressure insensitive[19, 20], but is very temperature dependent [21], underlying its entropic nature. The origin of the energetic cost for the transfer of a hydrophobic solute from an apolar environment to an aqueous solvent is the creation of an interface with a hydrogen bond network aiming at reducing the loss of hydrogen bonds of hydrating water molecules, resulting in a positive heat capacity change upon hydration[21, 22, 23], and a negative change in entropy[24]. This result is slightly counter intuitive, since the positive heat capacity is usually the consequence of a strengthened hydrogen bond network. It suggests that the hydrogen bond network provoked by the hydration pattern formed around the hydrophobic solute has, on average, stronger hydrogen bonds. Indeed, results from Monte Carlo simulations have suggested that the average length of the water-water hydrogen bonds in the first hydration shell of an apolar solute is shorter than that of bulk solvent[23], while FTIR experiments show that the number of bent hydrogen bonds is lower in the apolar solvation shell, resulting in a more ice-like structure[25].

### 1.2.3 The spin glass model

The spin glass model is a simplified model for protein folding adapted from the physics of glasses[26]. This model describes protein folding as a three phases problem. The random coil phase, where the polypeptidic chain is extended, a folded state in which the protein is very ordered, a fluid collapsed state, in which the hydrophobic collapse has occurred, but the protein is in a highly dynamic fluid phase, and a frozen glass like state, similar to the fluid phase, but with slow dynamics and composed of a misfolded ensemble. This frozen glass state is not technically a separate phase, as it still has characteristics of a fluid phase, but has very different dynamics. The analogy with the spin-glass from physics comes from the existence of a pseudo-ordered phase that kinetics prevent from reaching the energy minimum.

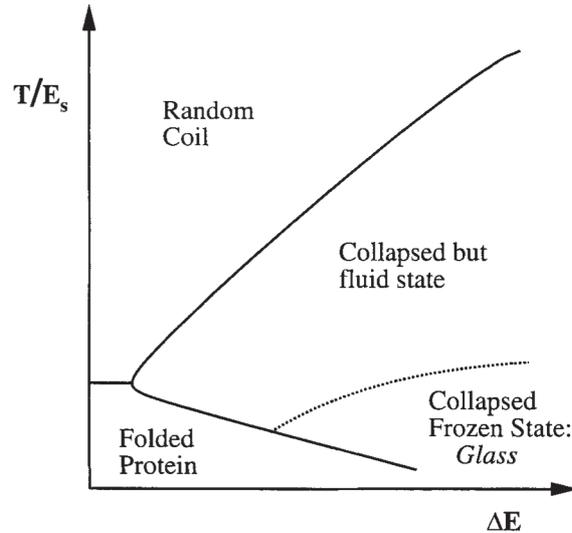


Figure 1.2: Phase diagram for a folding protein. The horizontal axis is the energy landscape roughness parameter,  $\Delta E$ . The vertical axis is the temperature divided by the stability gap  $E_s$ . The stability gap is the energy gap between the set of states with substantial structural similarity to the native state and the lowest of the states with little structural similarity to the native state. The collapse transition and the (first-order) folding transition are represented by solid lines and the (second order) glass transition is represented by a dashed line. In comparing this phase diagram with experimental phase diagrams, one must bear in mind that both  $\Delta E$  and  $E_s$  are temperature dependent because of the hydrophobic force. The average strength of the hydrophobic force could be considered as a third dimension in the phase diagram. Figure taken from [11].

In this picture, two types of transitions to the folded state can be observed (see figure 1.2). The roughness of the energy landscape, symbolized by the roughness parameter  $\Delta E$  in figure 1.2, is the determining factor of the type of transition that will occur. A rougher energy landscape is characterized by slower kinetics, and thus increased chances of reaching the glassy misfolded state. A completely flat landscape corresponds to a two-state, or first order transition (see figure 1.1, panel C) is not observed experimentally, but would correspond to the collision diffusion model, in which discrete pathways are present all the way to the folded state. The transition from the fluid state to the folded state is characterized by discrete pathways reached after the transition state [11]. Both of these phase transitions can be described in terms of transition temperature,  $T_f$  for the folding temperature, and  $T_g$  for the glass transition [27]. This first temperature depends on the average hydrophobicity of the chain as well as the average separation between the hydrophobic residues, the second is largely self averaging [11]. A good folding sequence is one that maximizes the ratio  $\frac{T_f}{T_g}$ , meaning a sequence in which the folding transition occurs before the glass transition when temperature is gradually lowered.

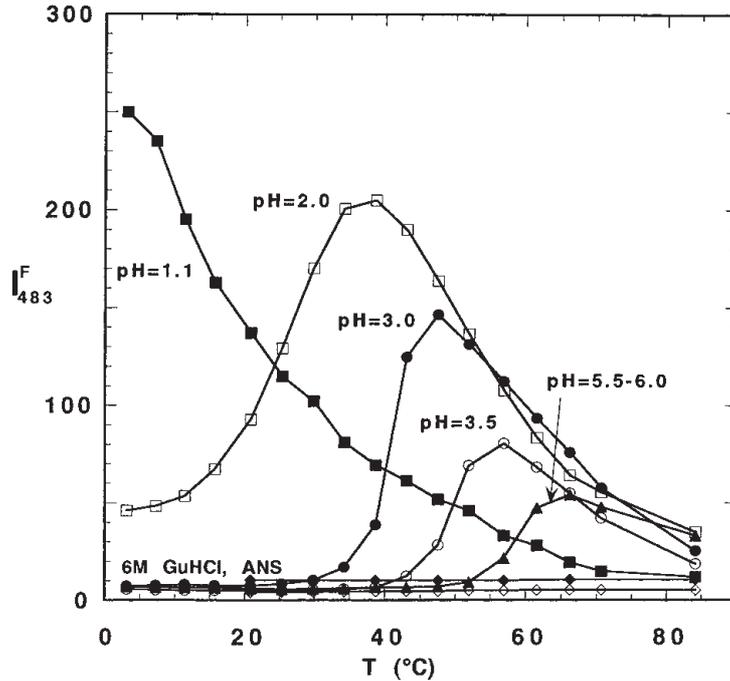


Figure 1.3: ANS fluorescence intensity at 483 nm, the emission wavelength corresponding to the bound state of the fluorescent dye,  $I_{483}^F$ , as a function of temperature in solutions of EqTxII ( $c=0.05 \text{ mg/g}_{\text{solution}}$ ) in glycine buffers at pH 1.1 (■), 2.0 (□), 3.0 (●) and 3.5 (○), in water at pH 5.5-6.0 (▲) in 6 M Gu-HCl (◆), and in any of buffer solutions with no EqTxII present (◇);  $c_{\text{ANS}} = 0.126 \text{ mg/g}_{\text{solution}}$ ,  $\lambda_{\text{exc}} = 365 \text{ nm}$ . EqTxII. Figure taken from [28]

#### 1.2.4 The molten globule perspective

The molten globule approach is a view of protein folding in which the protein has a obligatory intermediate in a collapse state, in accordance with the hydrophobic model, and similar to the fluid state described in the spin glass model. This molten-globule state was found by CD-studies to have a similar far UV spectrum to the folded state, but an aromatic region of the spectrum similar to the unfolded state, and by NMR to have amide protons protected from exchange in sites corresponding to native secondary structures [29]. This is indicative of a populated intermediate with native-like secondary structure, but no tertiary structure [30, 26]. This means that while the protein backbone has a similar entropy to that of the folded state, the side chains still have conformational freedom, leading to an intermediate state with higher entropy than the folded state, but lacking specific tertiary structures [31]. Experimental evidence of the existence of such a state has been numerous [32, 33, 34]. The molten globule is typically a metastable intermediate, although in some proteins the molten globule-like state can be stabilized by low pH, low temperature, or moderate chemical denaturant concentrations [30, 28]. This intermediate has been proposed to be a universal feature of globular protein folding [35, 36].

Interestingly, it has been reported that at low pH and low temperature, the molten globule state of the protein Equinatoxin II (EqTxII), a lipid binding protein, retains a capacity for binding 1-anilino-naphthalene-8-sulfonate (ANS), a hydrophobic probe (figure 1.3)[28]. This is an indication that, despite the relatively high disorder that remains in the molten-globule state, the structure is close enough to the tertiary structure to perform its function, or that the molten globule transiently populates the folded state often enough, allowing for measurable binding. In other proteins, the binding affinity of ANS was shown to be actually stronger to the molten globule state than for the folded state, suggesting that it might have a functional role in the binding of hydrophobic molecules[37].

As for the packing, the molten globule state is reported to be very compact, and have a similar radius as the folded state, within experimental error as determined by diffusion coefficient extracted using quasielastic light scattering in the Horse cytochrome c protein[33], or close in  $\alpha$ -Lactalbumin[38]. Thus one can assume that the molten globule has similar packing defects as the native state, but its dynamic nature suggests lower compaction.

The transition from the molten globule state to the ordered phase has been described as a highly cooperative process[36], but is not expected to be the rate limiting step for the folding reaction. Instead, the folded state appears to be in equilibrium with the molten globule state[31]. The activation volume, meaning the volume of the rate limiting state for the reaction, has been reported for Staphylococcal nuclease (SNase) to be close to that of the folded state, which leads to the interpretation that the transition state was close to a molten globule [39]. In other proteins, such as the small all- $\beta$  protein tendamistat[40], or the variant of Notch ankyrin repeat protein[41], the activation volume was reported to be even larger than the volume of the folded state[31]. These important activation volumes for unfolding suggest that the transition state lies in between the molten globule and the unfolded state, suggesting that the rate limiting step could correspond to the dehydration of the molten globule, although the level of hydration of the transition state was reported to be highly dependent on sequence[42]. Other evidence based on Hydrogen exchange NMR have emphasized that result, suggesting that the molten globule is dry, and that the rate limiting step is the dehydration of the core, with the exchange of protons being uniform among the entire protein, but occurring after the rate limiting step for unfolding for the RNase protein [43, 44, 31]. This is consistent with a short lived intermediate where the formation of the secondary structure happens during the formation of the molten globule. In some cases, experimental evidence has suggested the formation of the molten globule to be even faster than that of the secondary structure in acidic conditions[45], revealing that molten globules are diverse, and that their principal feature is the hydrophobic collapse[31]. The consideration that the rate limiting step happens before the formation of the molten globule during folding has also been reported for other proteins such as Cytochrome C[46].

These considerations taken together can help understanding the role of the molten globule

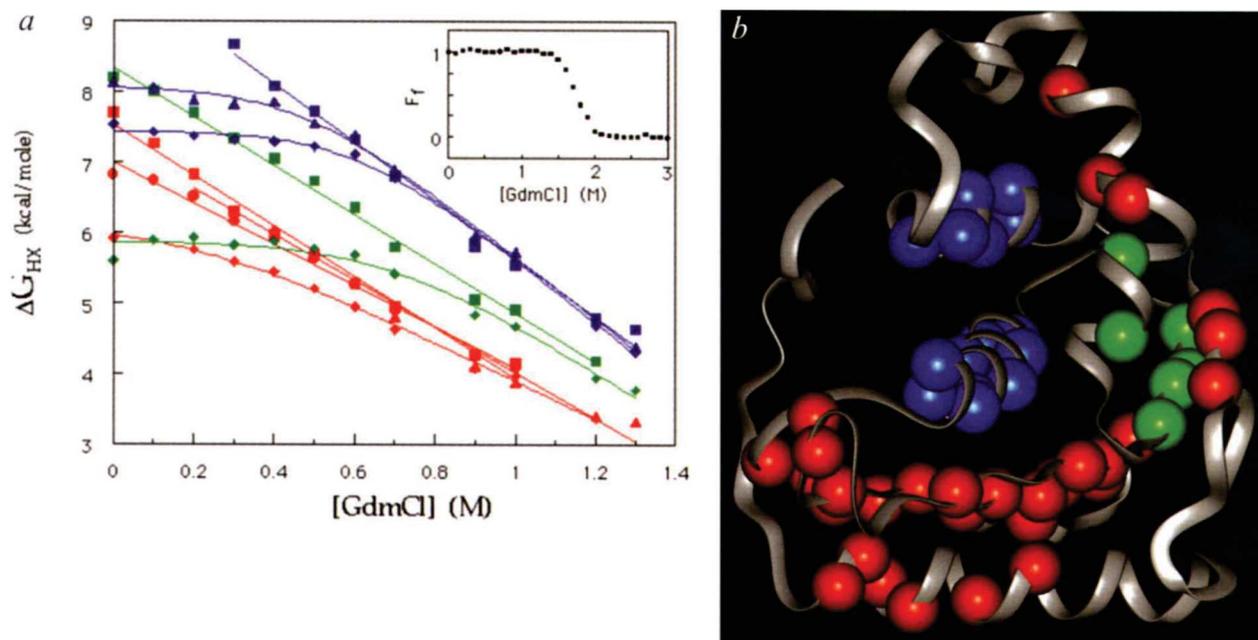


Figure 1.4: a, the free energy of Hydrogen Exchange of RNase H as a function of Guanidinium Hydrochloride concentration. Upper right corner curve shows the fraction of folded molecules as a function of Guanidinium Hydrochloride concentration determined by CD. Colors correspond to the colors of the sphere used for atom representation on the right panel b. b, the three regions with differing stabilities shown on a ribbon diagram of the RNase H crystal structure. Figure taken from [49]

state in the search of the native state. It suggests that the first transition undertaken by a protein following this model is the collapse of the chain in a compact but highly dynamic structure. This *burst phase* corresponds to the hydrophobic collapse, with formation of the secondary structure elements characteristic of a molten globule [47]. This state eventually "freezes" into an ordered native structure after conformational sampling, in order to satisfy the principle of minimal frustration [48], that is in order to form the maximum number of hydrogen bonds and reach the low energy structure characterizing the folded state.

### 1.2.5 Foldons and the hierarchy of protein folding

In globular proteins, subunits have been shown to fold independently [50], because of local variations in the folding energy of some structures with regard to others. Indeed, hydrogen exchange experiments of folded proteins have shown for globular proteins that the exposition of groups of residues to solvent were locally correlated, suggesting a greater intrinsic stability for some tertiary structure elements than for the global structure [49]. These independently folding subunits are commonly referred to as *foldons* [51, 50, 52] (figure 1.4). The presence of these foldons arises from an energetic hierarchy between the different types of structures, enabling

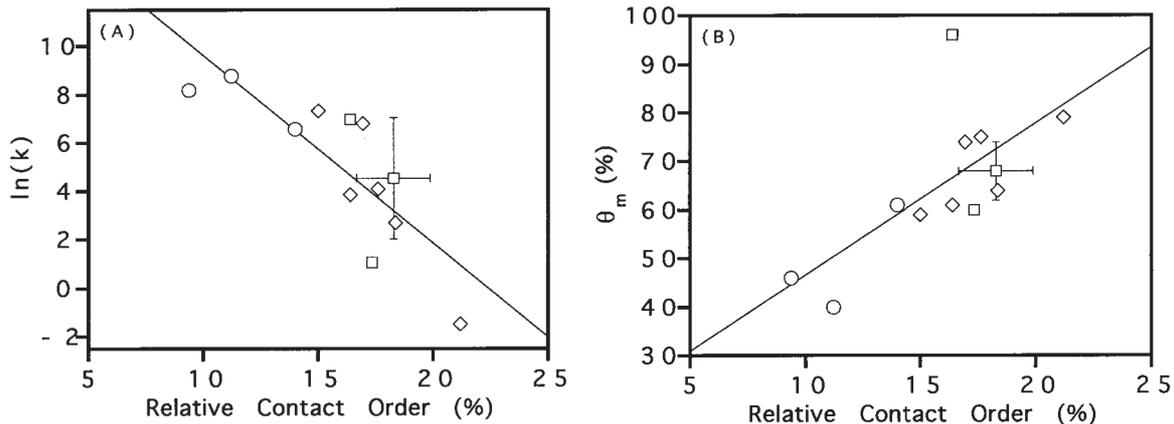


Figure 1.5: Relationship between the relative contact order of the native state and (A) the natural logarithm of the intrinsic folding rate ( $\ln(k)$ ) or (B) transition state placement for a 12 protein data set. The lines represent linear fits with correlation coefficients of 0.81 and 0.68, respectively. Circles denote helical proteins, diamonds mixed sheet-helix proteins, and squares proteins comprised predominantly of sheet structures. Figure adapted from [54]

the identification of the predominance of local versus long range interactions in the stability of a given region [53]. This is an important concept to resolve the Levinthal's paradox, because proteins' folding times have been shown to be correlated to their contact order. Contact order is the measure of the average distance (in residue position) in the primary sequence between residues that are in contact in the native structure. The relative contact order of a protein is given by the formula :

$$CO = \frac{1}{L \cdot N} \sum^N \Delta S_{i,j} \quad (1.2.1)$$

Where  $L$  is the total number of residues in a protein,  $N$  the total number of contacts,  $\Delta S_{i,j}$  the primary sequence distance between two residues  $i$  and  $j$  that are in contact in the native state [54, 55]. Using this definition, and given the intrinsic folding rate of a protein in native conditions  $k = k_F + k_U$ , we can see from experimental data that the higher the contact order is, the slower the folding will be, as seen in figure 1.5 A. In addition, there is also a correlation between the contact order of a protein and the placement of the transition state in terms of resemblance to the folded or unfolded states. The higher the contact order, the more native like the transition state (figure 1.5 B) [54, 55]. Correlations between highly conserved sequences in families of homologous proteins and early folding sites dominated by local interactions highlight that local interactions may be used to nucleate the folding process, or on the contrary, prevent the folding from going along certain folding pathways that would lead to misfolding or aggregation [56].

Taken together, two important lessons can be learned from the above results. First, fast

folding corresponds to a very gradual process, where contacts between residues that are close in the primary sequence form first, whereas more complex topologies have a more energetic transition state, probably due to the restrictions tertiary contacts force on configurational entropy, thus requiring more coordination in the formation of stabilizing native specific interactions, and resulting in slower kinetics. Second, proteins have evolved to select pairs of residues close in the native state to serve as nucleation sites. The formation of such contacts is a driving factor in the folding of subsequent structures, promoting the folding of the native local structure and preventing non-native contacts from forming. This results in smoother pathways (figure 1.1 panel A) and faster folding.

### 1.2.6 The Ising model

The Ising model, or Lenz Ising model, is a simple model for cooperativity developed by Ising for the theory of ferromagnets. In this model, the folding is described as a first order transition with simple nearest neighbor interactions[57]. The origin of the application of the Ising model to biopolymers dates back to the 1950s with the application of the model to the helix coil transition, but also to the DNA and RNA double strand formation among other phenomena[58, 59]. More recently, this model has been used to characterize the folding of linear repeat proteins. Repeat proteins are constituted of small sub-domains (the repeats) with a very conserved secondary structure content and tertiary interactions limited to intra-repeat and nearest repeat interactions[60]. Although the sequence of natural repeat proteins differs in length or content significantly from one repeat to the other, one can identify key positions corresponding to conserved amino acids properties that form the so called *consensus* of the repeats[59, 60, 61]. In addition, repeat proteins also contain two caps at each end of the protein, the role of which is to shield the interfacial region from the solvent[62]. In order to make this system ideal for the study of the energetic coupling between subunits, one can design a full consensus repeat that possesses the exact same sequence in each repeat, although a high consensus repeat can be enough for the application of the Ising model[59]. Using this sequence, one can change the number of repeats in a sequence in order to apply the Ising model[63] to extract the energetic contributions from the intrinsic energy and the interfacial energy of the repeats. The intrinsic energy is the energy of independent folding of a single repeat, and the interfacial energy corresponds to the energy difference between having two repeats folded in contact versus the energy of having these two repeats independently folded. The intrinsic energy is usually positive, with a favorable enthalpy, but strongly unfavorable entropy, and the interfacial energy negative, making the formation of a single repeat unfavorable. The extent of the difference between the two determines the level of folding cooperativity[3, 60, 64].

Lab designed full consensus repeat proteins have been shown to fold through parallel pathways [65], because the probability of any of the central repeats to fold first is somewhat equiv-

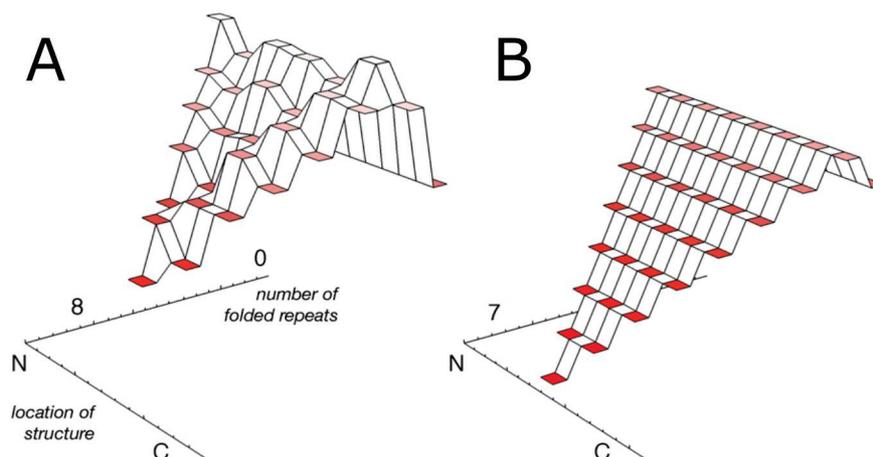


Figure 1.6: (A) Energy landscape of the Notch ankyrin domain. The energies of conformations with blocks of contiguous folded repeats, colored red according to free energies, are shown as a function of the number of folded repeats and the location of partly folded structure. (B) Energy landscape for an eight helix consensus TPR construct, using the Ising analysis of Kajander [64]. The landscapes in (A) and (B) are plotted on the same energy scale (z-axis), and in both cases the fully folded states (right-most tier) are set to zero energy. Figure from [60]

alent (if not influenced by the caps) for any of the repeats (figure 1.6 B). For natural proteins where the sequence differs among repeats, folding tends to follow a single pathway, determined by the gradient of intrinsic energy among the repeat[62](figure 1.6 A). Thus, for globular proteins, where the higher diversity in structural elements implies a broader range of stabilities for the subunits, parallel pathways are expected to be the exception[65]. This cooperative mechanism is in essence a nucleation/propagation mechanism, where the subunit with the most favorable intrinsic energy folds first, and thereby decreases the cost of folding for adjacent units, acting as a seed for folding. Thus, in natural repeat proteins, the folding rate is fast because the gradient of energy combined with the interfacial coupling direct folding in obligatory transient intermediates, thus optimizing conformation sampling. This mechanism shows how natural selection made proteins evolve not only in their ability to perform function, but also to fold within reasonable timescale with low risk of aggregation, simply by a hierarchy of the folding process, through an energetic gradient.

### 1.2.7 Conclusion and origin of folding pathways

The mechanisms presented in this section, and their diversity, highlight that there is no one way to fold a protein[66]. The mechanism for folding is expected to be strongly dependent on the topology of the native state. In addition, a single protein can have relatively independent

subdomains, thus the mechanism of folding of this subdomain might be better described by the formation of a molten globule, while the formation of the global tertiary structure may be due to interfacial interactions between those domains, following a foldon perspective on protein folding.

In Askel et al. 2011[3], it was shown by application of the Ising model to a full-consensus repeat protein that the intrinsic energy of a repeat was stabilized by enthalpy, but destabilized by entropy. In this picture, the folding of a single repeat is very energetically unfavorable, but the loss in entropy can be minimized by the formation a molten globule comprising several repeats, where side chain entropy is not lost, and forming native interactions in a single step, avoiding a strong destabilization due to the presence of a folded repeat alone. This would make the search for the folded state more efficient, because it would avoid any situation where a decrease in entropy would be realized in the same step as an increase in enthalpy, thus avoiding high energy intermediates. The resulting structure, if weakly coupled to the rest of the protein, would have all the features of a foldon : independently folded with strong intramolecular interactions, dominated by short range interactions. In that view, protein folding would consist of a lot of short range contact formation. These contacts would, by positive or negative cooperativity, restrict the conformational sampling of the rest of the chain in order to direct folding. Thus, in such proteins, the rate limiting step would likely be early, with a very unfolded-like intermediate, but would result in a smooth energetic landscape with low populated intermediates to the folded state[67], characteristic of fast folders. In such a mechanism, the folding pathways would be parallel before the formation of the first foldon, but discreet after, because directed by interfacial interactions. Strong interfacial interactions will increase the chance of folding of two subunits in a concerted manner, thus limiting the number of pathways possible.

A major factor in the determination of the folding scenario is the gradient in the formation of native specific interactions, because it acts as a guide for the collapse of the chain to form more complex structures. The relation between contact order and fast folding underlines that this principle is likely to result in the formation of local structures early, as suggested by the molten globule perspective. Thus, interactions that prevent unfavorable pathways must be formed early. Furthermore, this local stability could well be at the origin of the function as well[67], by allowing a regulation of function through local unfolding, acting like a rheostat. The scenario in which a protein would fold sequentially following an energy gradient, and thus a defined pathway, is suitable for proteins with low contact order, but less suitable for large proteins with sequence distant contacts.

A second factor comes from non-specific interactions, such as the hydrophobic contacts, that allows the population of states where sampling of native-like conformation enhance the chances of finding the energy minimum. However, states such as the fluid phase described in the spin glass theory should not be formed without guiding, because of the risk of entering

a collapsed phase that would be too far from the native structure, and could stay in a local energetic minimum, as predicted by the spin-glass theory. Thus, even if parallel pathways are thought to dominate the hydrophobic collapse, some early interactions probably prevent the most unfavorable pathways to be taken. That being said, the transition from a fluid phase to an ordered phase, to stay in the spin glass theory terminology, is expected to be a very cooperative process, thus even in this scenario, the role of interfaces between sequence distant regions is a determining factor in the formation of native structures without misfolding.

To conclude, the extent to which a protein is cooperative, and the dominating factor for the type of folding can vary in an enormous amount. To take caricatural example, the folding scenario can vary from downhill proteins, where the hierarchy of local stability appears to be the dominating factor for pathway determination, with no real cooperativity for sequence distant interactions, to fast folders, in which weak cooperativity allows the formation of pathways without significant energetic barriers, usually resulting from a domination of short distance native-specific interactions, and finally, large globular proteins which seem to show a mix of behaviors, with non-specific hydrophobic interactions and strong cooperativity between foldons appear to dominate [67].

### 1.3 Introduction to thermodynamics of proteins

In this section, we will introduce basic notion of thermodynamics used to describe the stability of proteins and introduce the pressure-temperature phase diagram. A more detailed introduction to the thermodynamic equations is available in the annex (see section 7.1.1), therefore only the main steps are going to be presented here, for the sake of simplicity. In thermodynamics, the systems are described in terms of potentials that describe the states of the system through the different contributions to its energy. The most basic potential is the internal energy, that is the sum of all the contributions to energy :

$$U = \sum_i \mu_i N_i - PV + TS \quad (1.3.1)$$

In this equation we can see that the stability is determined by the chemical potential  $\mu$  multiplied by the number of particles  $N$  of a specie  $i$  that compose the system, that is the intrinsic energy of the system due to its composition, pressure times volume, that is the work of placing the system of volume  $V$  in an environment at pressure  $P$ , and the temperature time entropy, that is the total heat that can be extracted from the system with no changes in its chemical composition. The infinitesimal variation of the internal energy is :

$$dU = \sum_i \mu_i dN_i + TdS - PdV \quad (1.3.2)$$

The total heat content of the system is called enthalpy, which is the internal energy without the work and is defined as :

$$H = U + PV \quad (1.3.3)$$

$$H = \sum_i \mu_i N_i + TS \quad (1.3.4)$$

Which infinitesimal variation is, using equation 1.3.2 :

$$dH = d(U + PV) = \sum_i \mu_i dN_i + TdS + VdP \quad (1.3.5)$$

To describe the variation in chemical composition, another potential needs to be introduced,

that is the Gibbs free energy, that describes the chemical composition of the system :

$$G = U + PV - TS \quad (1.3.6)$$

$$G = H - TS \quad (1.3.7)$$

$$G = \sum_i \mu_i N_i \quad (1.3.8)$$

By definition, at equilibrium, the variation of the Gibbs free energy is null, meaning there is no change in the chemical potential or number of particles of any specie composing the system. The variation in Gibbs energy of the system can be expressed from equation 1.3.6 :

$$\begin{aligned} dG &= dU + d(PV) - d(TS) \\ dG &= \sum_i \mu_i dN_i + VdP - SdT \end{aligned} \quad (1.3.9)$$

Or from equation 1.3.8 :

$$dG = \sum_i d\mu_i N_i + \sum_i \mu_i dN_i \quad (1.3.10)$$

Which leads to the equality :

$$\sum_i d\mu_i N_i = VdP - SdT \quad (1.3.11)$$

Which is called the Gibbs-Duhem equation[68], and is the mathematical formulation of the Le Chatellier principle, that states that the chemical potential of the species in play changes in response to changes in the conditions, thus creating a new equilibrium. For a two-state model in protein folding, a model where only two species are populated, the folded and unfolded states, the difference in Gibbs free energy between those states is expressed as :

$$\Delta G = \bar{\mu}_u N_u - \bar{\mu}_f N_f + \Delta V P - \Delta S T \quad (1.3.12)$$

Where  $\bar{\mu}_i$  is the average chemical potential of the ensemble  $i$ ,  $\Delta G = G_u - G_f$ ,  $\Delta V = V_u - V_f$  and  $\Delta S = S_u - S_f$ . Its variation can be written :

$$d\Delta G = \bar{\mu}_u dN_u - \bar{\mu}_f dN_f + \Delta V dP - \Delta S dT \quad (1.3.13)$$

For a reaction at equilibrium,  $\Delta G = 0$  because the system will readjust the number of

particles present in each state to minimize the difference<sup>1</sup>. If we define the standard conditions of our system as being  $P_0, T_0$ , we can define the difference in Gibbs free energy as :

$$\Delta G = \Delta G^0 + RT \ln Q \quad (1.3.14)$$

Where  $Q$  is the reaction quotient of the  $F \rightleftharpoons U$  reaction, and corresponds, for ideal solutions, to  $\frac{[U]}{[F]}$ , and  $\Delta G^0$  is the enthalpy difference in the standard conditions. Thus if  $\Delta G = 0$ , then :

$$\Delta G^0 = -RT \ln K = \bar{\mu}_u^0 N_u - \bar{\mu}_f^0 N_f \quad (1.3.15)$$

Where  $K = Q_{eq}$  is the reaction quotient at equilibrium. We can then write, using equation 1.3.13, that the variation of the equilibrium due to changing standard conditions is :

$$d\Delta G^0 = -dRT \ln K = \bar{\mu}_u^0 N_u - \bar{\mu}_f^0 N_f + \Delta V^0 dP - \Delta S^0 dT \quad (1.3.16)$$

And thus, the temperature dependence of the equilibrium :

$$\left( \frac{\partial \Delta G^0}{\partial P} \right)_{T,N} = -RT \left( \frac{\partial \ln K}{\partial P} \right)_{T,N} = \Delta V^0 \quad (1.3.17)$$

$$\left( \frac{\partial \Delta G^0}{\partial T} \right)_{P,N} = -RT \left( \frac{\partial \ln K}{\partial T} \right)_{P,N} = -\Delta S^0 \quad (1.3.18)$$

And thus the equilibrium at any pressure, or temperature can be determined using a linear extrapolation model, assuming that entropy and volume differences are constant over the pressure temperature interval :

$$\Delta G_{P,T_0}^0 = \Delta G_{P_0,T_0}^0 + \int_{P_0}^P \left( \frac{\partial \Delta G^0}{\partial P} \right)_{T,N} dP \quad (1.3.19)$$

$$\Delta G_{P,T_0}^0 = \Delta G_{P_0,T_0}^0 + \int_{P_0}^P \Delta V^0 dP = \Delta G_{P_0,T_0}^0 + \Delta V^0 (P - P_0) \quad (1.3.20)$$

And equivalently for temperature :

$$\Delta G_{P_0,T}^0 = \Delta G_{P_0,T_0}^0 - \int_{T_0}^T \Delta S^0 dP = \Delta G_{P_0,T_0}^0 - \Delta S^0 (T - T_0) \quad (1.3.21)$$

Additionally, entropy and volume themselves have a temperature dependence :

---

<sup>1</sup>See annex for details

$$\left(\frac{\partial^2 \Delta G}{\partial T^2}\right)_P = \frac{\Delta C_p}{T} \quad (1.3.22)$$

$$\left(\frac{\partial^2 \Delta G}{\partial P^2}\right)_T = -(\beta_u v_u - \beta_f v_f) \quad (1.3.23)$$

Where  $\Delta C_p$  is the difference in thermal expansivity between the folded and unfolded states,  $\beta = -\frac{1}{V} \left(\frac{\partial V}{\partial P}\right)_T$  is the coefficient of isothermal compressibility, and  $v_i$  is the partial volume of the ensemble  $i$ . In addition, the cross term derivative :

$$\frac{\partial^2 \Delta G}{\partial T \partial P} = \frac{\partial^2 \Delta G}{\partial P \partial T} = (\alpha_u v_u - \alpha_f v_f) \quad (1.3.24)$$

Where  $\alpha = \frac{1}{V} \frac{\partial V}{\partial T}$  is the coefficient of isobaric thermal expansivity. In the next section, we will go through each of the parameters up to the second degree dependence of the stability, and conclude by a discussion of the pressure-temperature phase diagram of proteins.

## 1.4 Origin of the determinants of thermodynamic stability

### 1.4.1 Temperature effect - entropy

Temperature is historically one of the first methods used to drive proteins out of their native state[69]. This denaturation method relies on the difference in entropy between the folded and unfolded states,  $\Delta S$ . Entropy is a complex notion that relates to the number of states that can be accessed by a system, and is thus commonly said to represent the level of disorder of a system. One can obtain the difference in entropy between the folded and unfolded states from differential scanning calorimetry (figure 1.7) using the equation :

$$\Delta S_T = \frac{\Delta H}{T} + R \ln(K) \quad (1.4.1)$$

Since the enthalpy change  $\Delta H$  is measured at the melting temperature in DSC (see figure 1.7), and that, by definition,  $\ln(K) = 0$  at the melting temperature, one can then deduce the difference in entropy between the two states as :

$$\Delta S_{T_m} = \frac{\Delta H}{T_m} \quad (1.4.2)$$

One can then determine the difference of entropy between the folded and unfolded states in native conditions by extrapolation of the measured  $\Delta C_p$  at the melting temperature, and extrapolate it[70]:

$$\Delta S_{(T)}^0 = \Delta S_{(T_m)}^0 + \Delta C_p \ln\left(\frac{T}{T_m}\right) \quad (1.4.3)$$

To better understand the origin of the difference in entropy between the folded and unfolded states, the contribution from the peptide chain conformational freedom can be separated into backbone and side-chains contributions, as they have a different temperature dependence. The difference in entropy also arises from the hydration of polar and apolar residues, and these contributions can also be separated leading to :

$$\Delta S = \Delta S_{bk} + \Delta S_{sc} + \Delta S_{hp} + \Delta S_{ha} \quad (1.4.4)$$

Where  $\Delta S_{bk}$ ,  $\Delta S_{sc}$ ,  $\Delta S_{hp}$  and  $\Delta S_{ha}$  are the contributions to the difference in entropy between the folded and unfolded states from the backbone, side chains, hydrophilic residues hydrating water molecules and apolar residues hydrating water molecules, respectively.

An experimental study of the origin of  $\Delta S$  in 1996 used considerable amount of data from the literature to extract hydration entropy contributions by using compound transfer methods[71]. In this study, the difference in hydration entropy was determined using experimentally obtained  $\Delta S$  values and determining the contributions of polar and apolar hydration using the equation :

$$\begin{aligned} \Delta S^{exp}(125^\circ C) - \Delta S^{exp}(25^\circ C) = & \Delta S_{ha}(125^\circ C) - \Delta S_{ha}(25^\circ C) + \\ & \Delta S_{hp}(125^\circ C) - \Delta S_{hp}(25^\circ C) + \\ & \Delta S_{cnf}(125^\circ C) - \Delta S_{cnf}(25^\circ C) \end{aligned} \quad (1.4.5)$$

Where the conformational entropy is  $\Delta S_{cnf} = \Delta S_{bk} + \Delta S_{sc}$ . To simplify this equation, some assumptions must be made. The first concept used in this study was the existence of a convergence temperature obtained by extrapolation of the temperature dependence of entropy[72], around  $125^\circ C$ , which is a temperature where the energy of transfer from aqueous to non-polar solution is null, meaning that the hydration effects do not contribute to the entropy, thus  $\Delta S_{ha}(125^\circ C) \simeq 0$ . An assumption was also made that the contribution of polar groups to the difference in entropy for an unfolding reaction was null across temperatures, because polar

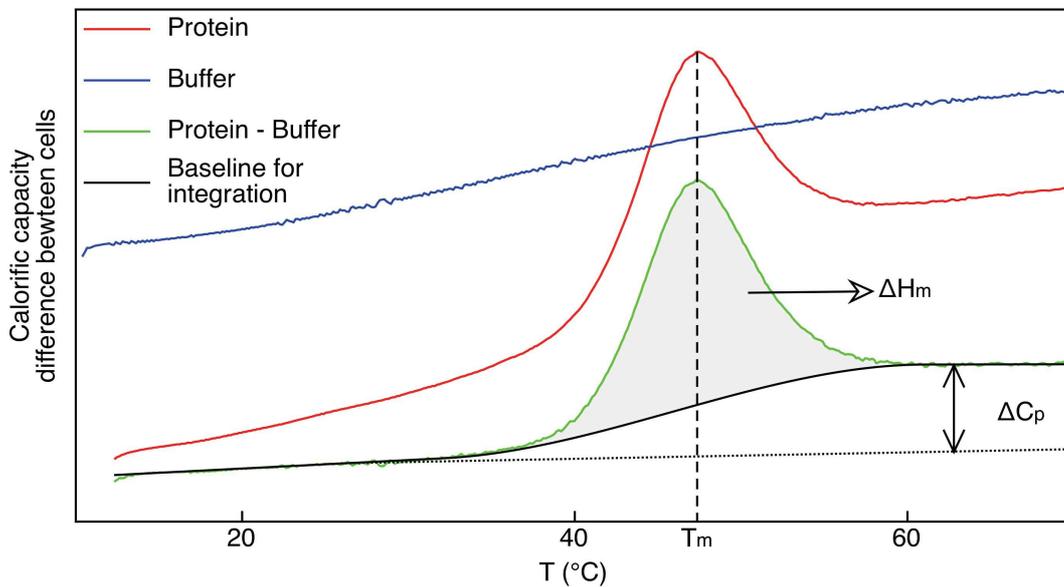


Figure 1.7: Decomposition of a DSC experiment on WT SNase. The difference in heat capacity is plotted in blue for the buffer experiment and in red for the buffer-protein experiment. Subtraction of the buffer to the buffer protein experiment, in green, yields the heat capacity of the protein. The black line is the two-state fit for the change in heat capacity. The grey area is the integration of the difference in heat capacity during the unfolding transition and thus yields the change in enthalpy  $\Delta H$  of the transition at the melting temperature  $T_m$ .

residues are always present on the surface of the protein, regardless of the state. Of course this is not totally true because of the polar backbone surface buried in the folded state. The equation then simplifies to :

$$\Delta S^{exp}(125^{\circ}C) - \Delta S^{exp}(25^{\circ}C) = -\Delta S_{ha}(25^{\circ}C) \quad (1.4.6)$$

$$(1.4.7)$$

Using that equation and previously published data, it was determined that the entropy of hydration for non-polar groups  $\Delta S_{ha}(25^{\circ}C)$  would be  $-10J.K^{-1}.mol^{-1}.res^{-1}$  for the Barnase protein and  $-15J.K^{-1}.mol^{-1}.res^{-1}$  for the Ubiquitin protein, both of which are commonly used globular proteins for the study of folding. One can also estimate the difference in entropy by rolling a probe over the surface of the native and unfolded structures of the protein, determining the polar and apolar accessible surface area in each state. The validity of this method had been tested previously[73, 74]. Using this method, and assuming that the unfolded state is in a fully extended conformation, the authors calculated the theoretical difference in hydration entropy, using previously determined data for the hydration entropy of apolar amino acids at  $25^{\circ}C$ [75], and by assuming additivity of the contributions from polar and apolar residues, using the formula :

$$\Delta S_{ap}(T) = \Delta ASA_{ap} \cdot \Delta \bar{s}_{ap} \quad (1.4.8)$$

Where  $\bar{s}_{ap}$  is the partial entropy by surface unit and is equal to  $-0.578J.K^{-1}.mol^{-1}.A^{-2}$ . This yielded a difference in apolar hydration entropy upon unfolding of  $-26J.K^{-1}.mol^{-1}.res^{-1}$  for the Barnase protein and  $-29J.K^{-1}.mol^{-1}.res^{-1}$  for Ubiquitin. After eliminating the probable sources of error that could arise from the analysis, the authors concluded that the assumption that polar groups do not contribute to the difference in entropy upon unfolding is wrong, as well as the assumption that only configurational entropy is present at the convergence temperature. To correct this, they estimated the contribution of polar groups exposure at the convergence temperature, using value for conformational entropy from literature :

$$\Delta S_{hp}(125^{\circ}C) = \Delta S^{exp}(125^{\circ}C) - \Delta S_{cnf}(125^{\circ}C) \quad (1.4.9)$$

And thus assuming temperature independence of  $\Delta S_{cnf}$  on temperature :

$$\Delta S_{hp}(125^{\circ}C) = \Delta S^{exp}(125^{\circ}C) - \Delta S_{cnf}(25^{\circ}C) \quad (1.4.10)$$

The resulting value was found to be  $-38J.K^{-1}.mol^{-1}.res^{-1}$  for Barnase and  $-35J.K^{-1}.mol^{-1}.res^{-1}$  for Ubiquitin, compared to  $-31J.K^{-1}.mol^{-1}.res^{-1}$  for Barnase and  $-29J.K^{-1}.mol^{-1}.res^{-1}$

for Ubiquitin, that was obtained for the same parameter using the additivity of residue entropy, after correcting for the presence of the buried polar surface[76]. This showed that the hydration of polar groups upon unfolding is indeed an important factor. Finally, the authors corrected the assumption of the independence of the conformation entropy to temperature by giving a dependence of  $5\text{J.K}^{-1}.\text{mol}^{-1}.\text{res}^{-1}$ . The results are shown in figure 1.8.

This method was fairly successful in the estimation of hydration entropy. Results showed that : 1/ nonpolar groups' contribution to entropy change upon protein unfolding is negative and inferior to that of polar groups at room temperature, with a positive temperature dependence (this is the major factor driving hydrophobic collapse, and thus heat denaturation), and 2/ that the polar groups' contribution to entropy change upon protein unfolding is negative at room temperature, with a negative temperature dependence[75]. Taken together, these results show how the hydration entropy of both polar and apolar residues is a stabilizing contribution for the folded state at room temperature for Ubiquitin and Barnase, but that the stabilization from nonpolar groups diminishes with temperature, which is the main contribution to thermal unfolding of proteins [71] (figure 1.8). However, one should keep in mind that a number of steps and assumptions have to be made to come to that result, and thus, that if it is probably qualitatively right, the quantitative comparison between the different contributions to the unfolding entropy should be taken cautiously. For example, one of the values used to reach this conclusion is the conformational entropy, derived from an estimation of the degrees of freedom

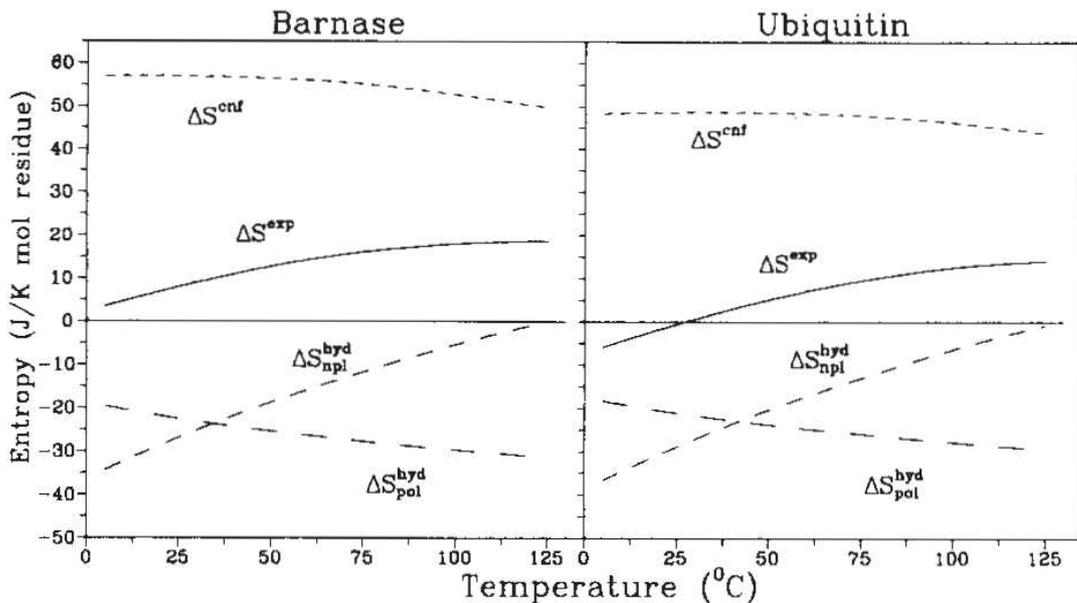


Figure 1.8: Temperature dependencies of the contributions to the experimental entropy change upon protein unfolding for Barnase (left) and Ubiquitin (right).  $\Delta S^{exp}$  is the experimentally measured entropy change.  $\Delta S^{cnf}$ ,  $\Delta S_{npl}^{hyd}$ ,  $\Delta S_{pl}^{hyd}$  correspond to the conformational, non-polar hydration and polar hydration entropy change, respectively. Figure taken from[71]

of the polypeptide chain. This quantity has latter been found to have been overestimated, due to an overestimation of the backbone entropy in the estimation used[77].

More recently, separate studies have been performed in order to get a greater insight on the contributions to conformational entropy. Using different theoretical approaches, including Molecular Dynamics and Monte Carlo simulations, estimation of the number of angles that can be populated were made by determination of the number of local energy wells associated with the 360 degree rotation around the  $\alpha C-\beta C$  torsion angle in order to estimate the conformational freedom of side chains[78]. The restriction of the dihedral angles to a smaller range upon the folding of the protein was also investigated, but was found to be negligible compared to the change due to the decrease in the number of rotamers. These studies showed a very good general agreement, and the consensus is that the restriction of the side chains' motion in the folded state had an approximate cost of  $1kcal.mol^{-1}$  per residue, or about  $0.5kcal.mol^{-1}$  per rotamer, the folding reaction being estimated to restrict from 3 to 1 the number of accessible rotamers for a side chain. This value is high and hence side chains motions seem to play an important role in protein stability.

NMR approaches have also been used to estimate the entropy of backbone conformational freedom and its loss upon folding. It was reported that the loss of conformational freedom for Ubiquitin due to backbone configurational freedom upon folding was  $1.1kcal.mol^{-1}$  per residue, and a 5 to 13 fold reduction in the number of accessible states from backbone motion. Side chain entropy loss was reported to be 0.2 to  $0.3kcal.mol^{-1}$  per residue, a very low estimate compared to previous studies, with a loss of conformational freedom of about 1.4 folds per rotamer[79]. This study suggests that the loss in side chain conformational entropy is negligible when compared to the that of the backbone.

Another approach used relaxation NMR in combination with molecular dynamics simulations to test the hypothesis that the side chains entropy of a folded protein can be estimated from the methyl motions alone, using them as a proxy for local disorder, with a method the Joshua Wand's group developed previously for the estimation of binding entropy changes[80]. This method relied on the calibration of an "entropy meter", by relating the NMR measured loss of methyl conformational freedom, through relaxation study of dynamics, to the isothermal titration calorimetry measured change in entropy upon binding. Using these data points for calibration, the method was then extended to estimate globular proteins' side chains motions. The study reported general agreement between the disorder of the methyl groups, as observed by NMR and the local motions, but noted that some site-to-site variance remained to be explained, and that further experiments were needed to confirm the validity of this approach[81]. This method could prove useful in the future for providing more accurate estimates of the folded state conformational entropy, and thus estimate its loss upon folding.

In 2004, a review was published on molecular dynamics simulation methods to estimate the

entropy from all its components[82]. The study concluded that none of the techniques considered seem to provide a good estimation of entropy change upon unfolding, both because of the enormous sampling that this estimation requires, but also simply because of the complexity of the problem, in particular because of the precision of the potentials needed for the calculation[82]. Taken together, these results highlight the lack of consensus that remains in the field on the origin of the observed entropy difference, in particular for the contribution from the conformational entropy.

#### 1.4.1.1 Heat Capacity

Heat capacity is the temperature dependence of entropy, and is a commonly measured parameter because, as opposed to entropy, heat capacity can be directly measured using Differential Scanning Calorimetry (figure 1.7).

$\Delta C_p$  of unfolding is commonly considered to be temperature independent[70] in the usual temperature range used for heat denaturation experiments. The difference in heat capacity between the folded and unfolded state is strongly correlated to the difference in accessible surface area[83]. Heat capacity is proportional to the change in entropy with temperature, and is also defined as the temperature dependence of enthalpy on temperature :

$$\left(\frac{\partial \Delta S}{\partial T}\right)_P = \frac{\Delta C_p}{T} \quad (1.4.11)$$

$$\left(\frac{\partial \Delta H}{\partial T}\right)_P = \Delta C_p \quad (1.4.12)$$

Thus, heat capacity actually is the determining factor for the curvature of the dependence of  $\Delta G$  to temperature[69]. The variation in heat capacity upon exposure to aqueous solvent of a side chain is positive for nonpolar groups and negative for polar groups[84], making the total change of heat capacity upon unfolding positive, as can be seen in figure 1.7. The gain in heat capacity upon exposure of hydrophobic groups has long been regarded as an anomaly[85]. This is a consequence of the exposure of hydrophobic surface, resulting in the creation of a water hydrogen bond network with smaller average hydrogen bond length, and thus higher temperature dependence than bulk water[23]. This is explained by the fact that the specific interactions present in the folded state are less temperature dependent than the non-specific interactions with solvent in the unfolded form, and hence that the number of hydrogen bonds varies faster in the unfolded than in the folded state, changing the number of accessible states with temperature in a more rapid manner[86].

### 1.4.1.2 Cold denaturation

Cold denaturation of proteins is a direct consequence of the curvature of the stability curve, due to the difference in heat capacity between the folded and unfolded states, resulting in a curvature in the  $\Delta G$  curve as a function of temperature, that crosses the temperature axis twice, one at high and the other at a low temperature. Because of the stability of most proteins, however, cold denaturation is rarely observed because it would happen below the freezing point of water, and hence cannot be observed without artificially lowering the stability of the protein prior to cooling. For this reason, and because it was long considered an outcast in protein folding, it is one of the least studied mechanism for protein unfolding[1]. The molecular mechanism of cold denaturation is due to the fact that at low temperature, the hydrogen bond network surrounding hydrophobic interfaces becomes more favorable enthalpically than that of bulk solvent, thus cold unfolding is associated with released heat upon hydrophobic solvation, whereas high temperature unfolding is associated with heat absorption by the hydration shell[87].

## 1.4.2 Pressure effect - the volume change upon unfolding

The most common way to drive a protein out of its native state is through the use of temperature or chemical denaturants, because it requires little special equipment. Recently, however, the development and implementation of high pressure applied to a diversity of measurement instruments has enabled a more detailed exploration of the energy landscape of proteins. The pressure unfolding of a protein is due to the difference in volume between the different states that compose its energy landscape. For a two-state model of a globular protein, the volume of the folded state is typically higher than that of the unfolded state, although the difference in volume with regard to the partial molar volume of the folded state is usually very small ( $\geq 1\%$  of the partial volume of the folded protein)[88, 18, 89]. To understand where that difference arises from, we can decompose the partial volume of a protein into :

$$V = V_{VdW} + V_v + \Delta V_{hp} + \Delta V_{ha} \quad (1.4.13)$$

Where  $V$  is the partial volume, that is the total solution volume increase upon insertion of the solute,  $V_{VdW}$  is the Van der Waals volume,  $V_v$  is the solvent excluded void volume formed by non-hydrated cavities,  $\Delta V_{hp}$  is the relative volume of the water molecules around polar residues with regard to the volume of bulk water, and  $\Delta V_{ha}$  is the relative volume of the water molecules around apolar residues with regard to the volume of bulk water. Typically, the hydration of polar residues leads to a decrease in volume because the water molecules organize into a denser pattern[90] by aligning their dipolar moments with that of the electrical field provoked by the

uneven distribution of charges or the net charge of the hydrophilic residue. In contrast, the neutral surface of hydrophobic residues leads the water molecules to organize in a pattern to minimize the loss of hydrogen bonds, which results in a lower density than bulk water[91]. From equation 1.4.13 we can look at the contributions to the difference in volume between the folded and unfolded states :

$$\Delta V_u = \Delta V_v + \Delta\Delta V_{hp} + \Delta\Delta V_{ha} \quad (1.4.14)$$

Where  $\Delta V_v$  is a negative contribution due to the presence of more excluded void volumes in the folded state than in the unfolded state,  $\Delta\Delta V_{hp}$  is negative, but low, because most of the buried surface area in the folded state is hydrophobic, although around 30% of the buried surface area is polar in globular proteins, mostly because of the amide groups[83], and  $\Delta\Delta V_{ha}$  is a positive contribution due to the larger amount of hydrophobic residues exposed in the unfolded state. Separate studies have examined the influence of each of these parameters on the total volume change.

The contribution of hydration solvent density to the total volume change upon protein unfolding was ruled out as being the major determinant for the observed unfolding volume change[92]. In this study, a natural repeat protein, the ankyrin repeat protein, was used in order to determine whether the values of  $\Delta V_u$  were correlated to the size of the protein. Engineered deletions of one, two or three repeats allowed measurement of the  $\Delta V_u$  as a function of size. The change in volume upon unfolding was shown to be uncorrelated to the nature of the amino acids exposed, by simply comparing the nature of buried amino acids in each repeat and the change in  $\Delta V_u$  upon repeat deletion. More importantly, if hydration were a major factor, the change in volume upon unfolding should be correlated with the size of the protein, because larger proteins expose more surface area upon unfolding. No such tendency was observed[92]. These results were later confirmed by a pressure perturbation calorimetry (PPC) study[93]. Most of the observed volume change was found to be associated with the central repeats. To further investigate the origin of  $\Delta V_u$ , a molecular dynamics simulation was performed in order to quantify the occurrence of solvent excluded void volumes. The procedure consisted of trying to insert a 0.11nm probe on 1000 configurations extracted from a 5ns trajectory with a time step of 5ps. If the probe could be inserted without any overlapping between the probe and the protein or solvent molecule van der Waals surface, a point was saved at the insertion coordinates. The number of successful probe insertions within 0.4nm of a grid point was used to estimate the void density. Similarly, the average number of water molecules within a 0.4nm radius of a given point was calculated to estimate the hydration density. Results showed that these central repeats had the highest void density and lower hydration among the sequence, but also the highest content in hydrophobic residues, that should contribute negatively to the observed volume change magnitude[92]. Thus, it was concluded that the major contributor to the observed change in volume upon unfolding was the presence of solvent excluded void volumes.

Further studies investigated the determinants of the volumetric properties of proteins through pressure unfolding of cavity enlarging mutants of Staphylococcal nuclease (SNase) have confirmed this result, that the major contribution of the volume change upon unfolding is due to void volumes that are present in the folded state and hydrated upon unfolding[94]. This is a very important result for the future of the field, because void volumes are oftentimes distributed inhomogeneously in the native structure. This makes pressure denaturation a very useful tool, facilitating the characterization of intermediates by selectively destabilizing the largest void volumes containing region[95], or engineering proteins with cavities in specific positions in order to destabilize specific parts and look at the effect the mutation has on the stability of other parts[96] to study cooperativity. This will be discussed in the last chapter of this thesis. This feature is unique to pressure denaturation, compared to temperature or chemical denaturation, which both work on the increase in accessible surface area upon unfolding.

### 1.4.2.1 Compressibility

The different contributions to the volume of the protein also have a pressure dependence : compressibility. The compressibility of a specie can be expressed as a isentropic compressibility or isothermal compressibility. Due to the nature of our experiments, where temperature and not entropy is kept constant, we will consider the isothermal compressibility :

$$\beta_T = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T \quad (1.4.15)$$

The characterization of the origin of the compressibility of the folded state has recently been the subject of a study by Voloshin et al.[97]. The study used full-atom simulations at different pressures of the well characterized Staphylococcal nuclease (SNase) to realize a volume decomposition, and thereby identifying the contributions to the total partial volume and their molecular origin. This method was previously successfully applied to study the thermal expansivity of the natively unfolded human islet amyloid polypeptide (hiAPP)[98]. To separate the contributions to volume, this method uses tessellation, a way of dividing the volume into easily defined cells. Each decomposition cell is defined around an atom, and delimited by a surface such that each point in the surface is equidistant to the two closest atoms. A 2D representation of the method is available in figure 1.9. In this study, the volume decomposition was applied using the radical tessellation, which allows for the determination of Voronoi cells, taking into account the radius of the atoms, as opposed to the classic Voronoi Delaunay tessellation that is only defined for discrete points. This allows for an efficient calculation of the empty volume inside each cells, and is thus necessary for quantifying the empty void volumes in a protein. The water molecules were approximated to simple spheres. Due to the method used, a different approach for decomposition was used in the study compared to the one presented in equation

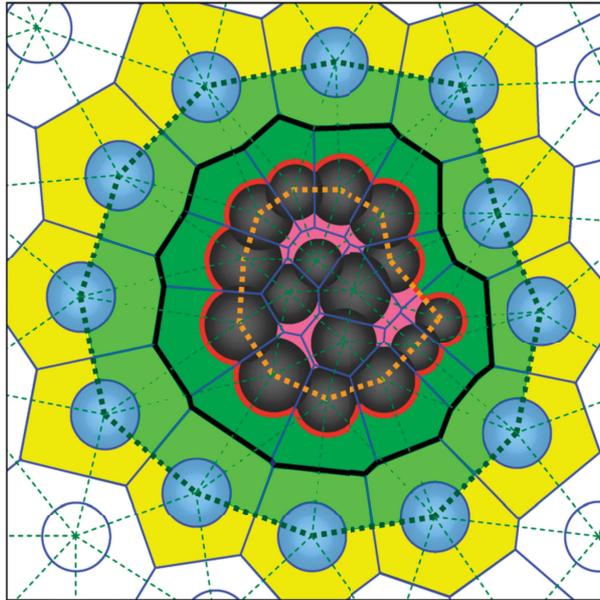


Figure 1.9: Decomposition of the volume using the Voronoi-Delaunay decomposition method. Blue lines show the limits of the Voronoi cells. Dark discs represent the solute molecules' Van der Waals volume,  $V_{VdW}^M$ . The molecular volume  $V_{VdW}^M + V_{empty}^M$  is outlined by the red line, solvent excluded void volumes,  $V_{empty}^M$ , are displayed in pink. The Voronoi volume boundary of the solute is displayed by a thick black line. The boundary volume is shown in green. It is split between a part belonging to the solvent ( $V_B^S$ , light green) and a part belonging to the solute ( $V_B^M$ , dark green). The yellow area corresponds to the hydration volume of the first hydration Voronoi shell minus the solute boundary volume  $\Delta V_{hyd} - V_B^S$ . Figure adapted from [97]

1.4.13. The apparent volume of the solute in solution was separated into the Voronoi molecular volume, and the hydration volume :

$$V_{app} = V_{Vor} + \Delta V_{hyd} \quad (1.4.16)$$

Where  $V_{Vor}$  is sum of all the volumes of the Voronoi cells constructed around the solvent atoms and  $\Delta V_{hyd}$  is the difference in volume of water molecules in the first hydration layer with regard to their volume in the bulk. We should note that no decomposition was made depending on the polarity of the residue being hydrated. However, due to the globally polar nature of the folded protein surface,  $\Delta V_{hyd}$  is negative. The use of a single hydration layer was tested, and for this system, the results were found to be equivalent if one or more hydration layers were used. The molecular Voronoi volume itself can be decomposed into the contribution from the solvent interface (here called boundary molecular volume,  $V_B^M$ ), the contribution from the Van der Waals volume of the protein ( $V_{VdW}^M$ ), and the contribution from the inner void volumes,  $V_{empty}^M$ :

$$V_{Vor} = V_B^M + V_{VdW}^M + V_{empty}^M \quad (1.4.17)$$

Additionally, a volume corresponding to the total boundary volume  $V_B = V_B^M + V_B^S$ , where  $V_B^S$  is the boundary volume on the solvent side (see figure 1.9). The conclusions of this study were that the molecular volume is the essential contribution to the decrease in volume of the protein's native state with increasing pressure. If the molecular boundary volume  $V_B^M$  was shown to decrease with pressure, it is compensated for by the decrease in the magnitude of  $\Delta V_{hyd}$ , due to the fact that bulk water is more compressible than hydrating waters, resulting in the sum  $\Delta\Delta V_{hyd} + \Delta V_B^M \simeq 0$ . In the end, since the Van der Valls volume is nearly constant in that pressure range, the compressibility of the apparent volume of a protein in solution was concluded to arise from the compression of its internal void volumes,  $V_{empty}^M$ , most likely by a slight change in packing. These findings were in good agreement with previous study of SNase compressibility made by densitometry, with a reported coefficient of isothermal compressibility of  $0.95 \cdot 10^{-5} bar^{-1}$  in the simulation study at 300K compared to  $1.1 \pm 0.2 \cdot 10^{-5} bar^{-1}$  for the densitometric measurements at 298K[89]. However, this is rather small, given that at 3000 bar, the change in volume would be of around 3%, well within the margin of error for high pressure unfolding measurement of the difference in volume.

### 1.4.3 Temperature-pressure cross term - Expansivity

Expansivity is the temperature dependence of the volume, but it is also the pressure dependence of entropy. That is because :

$$\frac{\partial^2 \Delta G}{\partial T \partial P} = \frac{\partial^2 \Delta G}{\partial P \partial T} = - \left( \frac{\partial S}{\partial P} \right)_T = \left( \frac{\partial V}{\partial T} \right)_P \quad (1.4.18)$$

$$\alpha = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right) \quad (1.4.19)$$

Where alpha is the coefficient of isobaric expansivity. Experimentally, this coefficient can be accessed directly through pressure perturbation calorimetry (PPC) measurements, by comparison with the densitometric measurements at different temperatures or indirectly with pressure denaturation at different temperatures by comparing the change in volume. However, this latter method gives the absolute difference in expansivity between the folded and unfolded state. The temperature dependence of volume due to hydration for different amino acids has been studied by PPC, after subtraction of the glycine expansivity, to account for the zwitterionic contributions to thermal expansivity[99]. These results show that hydrophilic residues have a positive coefficient of thermal expansion, that decreases with increasing temperature, while hydrophobic residues have a negative coefficient of thermal expansion. At the high temperature limit, the difference in thermal coefficients between hydrophilic and hydrophobic residues disappears because water molecules lose their interaction with the solute (figure 1.10). Hence, the remaining coefficient is independent of their nature, and very close to that of bulk solvent[100].

This phenomenon is at the origin of the so called convergence temperature, where the entropy and enthalpy changes upon unfolding become independent of the nature of the solvent, at temperatures superior to 115°C[72].

**Folded state** In a second article, Voloshin et al.[104] used the same method as in the compressibility study (see previous section and figure 1.9) on Staphylococcal nuclease, a very well characterized globular protein. This study follows a previously published similar study that was realized on the human islet amyloid polypeptide (hiAPP), a natively unfolded polypeptide, which we are going to comment on as well[98]. As for compressibility, the dependence of volume on temperature, expansivity, can equally be decomposed into the individual contributions of the partial molar volume of the solute. As a reminder :

$$V_{app} = V_{Vor} + \Delta V_{hyd} \quad (1.4.20)$$

Where  $V_{Vor}$  is the Voronoi volume of the molecule,  $\Delta V_{hyd}$  the difference in volume of the hydrating water relative to bulk solvent. Further decomposition of the Voronoi volume can be done :

$$V_{Vor} = V_B^M + V_{VdW}^M + V_{empty}^M \quad (1.4.21)$$

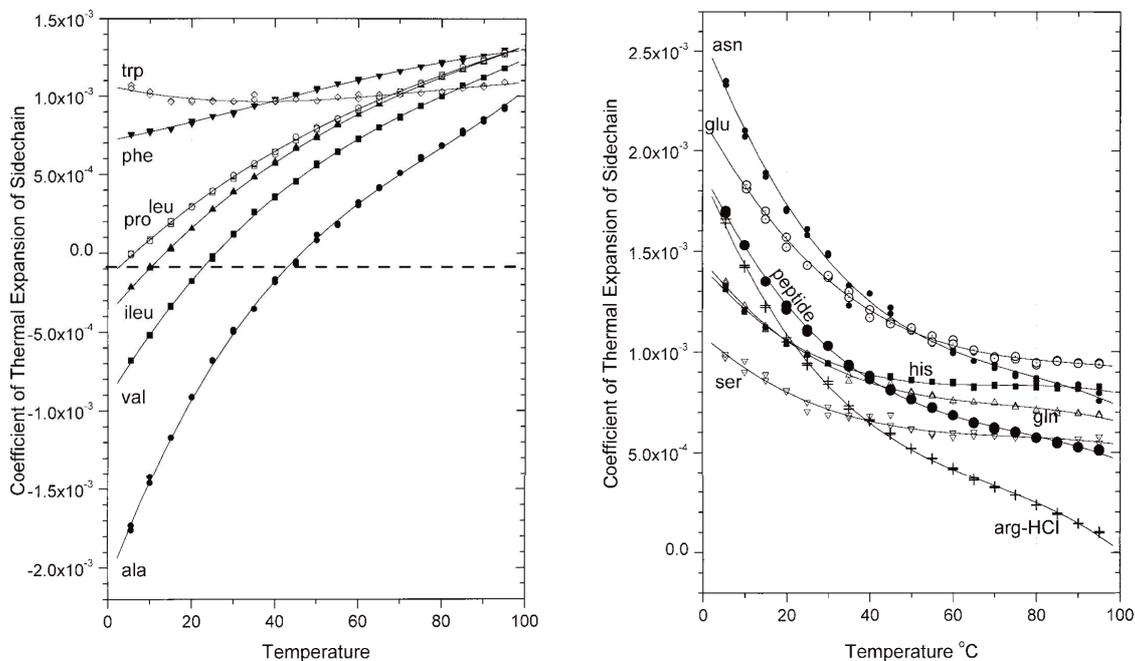


Figure 1.10: Coefficient of thermal expansion  $\bar{\alpha}$  of the partial volume for hydrophobic side chains (left). Coefficient of thermal expansion of the partial volume for hydrophilic side chains (right).

Where  $V_B^M$  is the boundary molecular volume,  $V_{VdW}^M$  the Van de Walls volume and  $V_{empty}^M$  the solvent excluded void volume (see figure 1.9). The main contribution to expansivity was shown to be the expansion of the boundary volume comprised in the space in between the first hydration shell and the solute ( $V_B^S + V_B^M$ ). Hydration effect, that is what is defined here as  $\Delta V$ , was found to be rather constant over temperature, because the thermal expansivity of bulk water is higher than that of the first hydration shell, resulting in a negative contribution to the change in volume, compensated for by the expansion of the solvent boundary volume  $V_B^S$ . Thus the total change in volume with temperature can be approximated to arise from the molecular boundary volume  $V_B^M$ . This was pointed out to be in agreement with Chalikian's view of the thermal volume[105], that is the volume due to constant reorientation and vibration of the molecules. Since the hiAPP is a rather hydrophobic unfolded polypeptide, as opposed to the folded SNase in which surface residues are essentially hydrophilic, it was pointed out that this could highlight a universal mechanism of thermal expansivity across proteins. This may be a hasty conclusion, as the study also showed that a second positive contribution to the thermal expansivity, less important in magnitude, came from the expansion of the internal solvent excluded void volumes in SNase ( $V_{empty}^M$ ).

Other direct measurements have highlighted that surface substitution of 11 polar residues to Alanine on the surface of BPTI showed no changes in the observed thermal expansivity of the folded state[6]. This seems to contradict the conclusion exposed in the Voloshin et

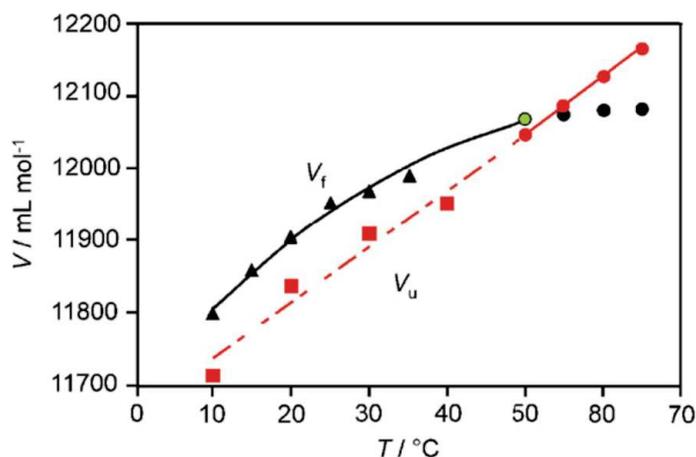


Figure 1.11: Experimentally determined volumetric properties of the folded and unfolded states of staphylococcal nuclease as a function of temperature. Black triangles and red circles are direct densitometry measurements [89], red squares are unfolded state volumes obtained by taking the volume change at each temperature determined by pressure-dependent fluorescence [101] and subtracting this from the molar volume determined by densitometry. Green point is taken from the volume change at the transition temperature from PPC measurements [102]. Figure taken from [100] after being modified from [103]

al. article that the main contribution to the thermal expansivity arises from the expansion of the boundary volume in between the solute and the solvent. Given the thermal expansion of side chain dependence on temperature (Figure 1.10), one would expect that the mutations of surface residues from polar to apolar to reduce the observed thermal expansivity. In stark contrast, single substitution to alanine from larger residues buried in SNase had drastic effects on the measured folded thermal expansivity[6]. In addition, the study showed a good negative correlation ( $R=-0.9$ ) between the folded state expansivity change with temperature of the mutants and the DSC measured difference in heat capacity between the folded and unfolded state,  $\Delta C_p$ . Assuming (reasonably) that the mutation has little effect on the unfolded state heat capacity, higher  $\Delta C_p$  are related to lower folded heat capacities[106]. Since  $\left(\frac{\partial \Delta S}{\partial T}\right)_P = \frac{\Delta C_p}{T}$ , high  $\Delta C_p$  is related to a faster change in the entropy, hence the number of accessible states, thus rapid changes in expansivity with temperature are related to more malleable proteins (Figure 1.12). Further investigation of the origin of the determinant of thermal expansivity of proteins following this work is presented later in this thesis.

Finally, it should be noted that the change in volume with temperature is not linear in the folded state, and tends become smaller with increasing temperature[100]. This is an important feature of proteins, because the change in unfolded volume being more linear in the usual laboratory temperature range, the difference in volume between the folded and unfolded states is very temperature sensitive, with lower temperature typically yielding a higher volume change, and, perhaps more interestingly, the difference of volume can theoretically be inverted a high temperature such that pressure stabilizes the folded state (Figure 1.11).

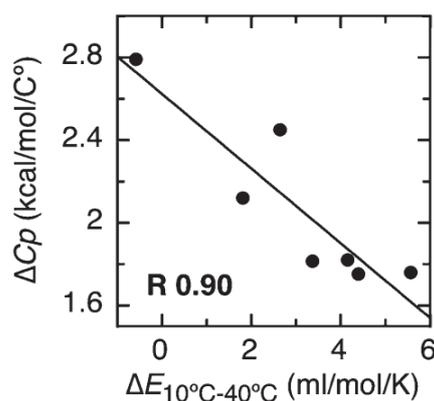


Figure 1.12: Negative correlation between  $\Delta C_p$  and the folded state molar thermal expansivity for SNase  $\Delta$ +PHS variants. Heat capacity changes are taken from DSC measurements, and  $\Delta E_{10^\circ\text{C}-40^\circ\text{C}} = E_{10^\circ\text{C}} - E_{40^\circ\text{C}}$  values for  $\Delta$ +PHS calculated from PPC experiments. Lines are linear regression fits with correlation coefficients of -0.90.

**Unfolded state** In Tsamaloukas et al. 2010[107], expansivity of the unfolded state has been studied by PPC for three different globular proteins, and results have shown that the total thermal expansivity can in fact be easily determined, both in magnitude and temperature dependence, by the partial volume weighted sum of the thermal expansivity of each individual amino acid in the sequence :

$$\alpha_{prot(T)}^u = \frac{\sum_i v_i(T)\alpha_i}{\sum_i v_i(T)} \quad (1.4.22)$$

## 1.4.4 Effects of cosolvents on the thermodynamic stability of protein

### 1.4.4.1 Chemical denaturants

**Application to protein thermodynamic studies** Chemical denaturants, typically urea or guanidinium hydrochloride, are chemicals that are used to destabilize the folded state of the protein with regard to the unfolded state. The effectiveness of chemical denaturants is defined by the *m-value*. The m-value depends on both the chemical denaturant and the protein used, and is defined as being the variation of the standard free energy change with denaturant concentration. Hence, if we monitor the change in population between two states using any kind of relevant signal, we can obtain the m-value through :

$$m = RT \left( \frac{\partial \ln K}{\partial [den]} \right)_{P,T} \quad (1.4.23)$$

This value is very well correlated to the amount of accessible surface area exposed upon unfolding, with a correlation coefficient of 0.87 for guanidinium hydrochloride and 0.84 for urea[83]. Using the linear extrapolation model[108, 109, 110], it is possible to extrapolate the obtained value to obtain the free energy difference of the protein before the addition of any denaturant :

$$\Delta G_{[den],P,T}^0 = \Delta G_{H_2O,P,T}^0 - m[den] = -RT \ln \left( K_{[den],P,T} \right) \quad (1.4.24)$$

Comparison with other determination methods has shown that this linear extrapolation method tends to give the lowest estimate of  $\Delta G_{H_2O}^0$ , but the best agreement for a single protein when both denaturants are tested[108]. However, another study has shown that the linear extrapolation model, while a good model for urea when compared to temperature denaturation, shows a significant deviation for guanidinium hydrochloride, which was attributed to the electrolyte nature of this cosolvent[109]. In the rest of the description of chemical denaturants, we will focus on urea, as it is the most commonly used, but also most studied denaturing cosolvent, and the one that was used in my work.

**Molecular mechanism** The linear extrapolation model is just one way of fitting the data, which requires no real knowledge of the mechanism in play. Historically, two main mechanisms were proposed to explain the behavior of this type of cosolvent. The first is the view that urea is a chaotropic agent, that is a substance that alters hydrogen bonding of water[111]. In this view, the loss in stability is due to the change in the relative free energy between water molecules hydrating the solute and bulk water free energy. As the concentration of the chaotropic agent increases, the gain in free energy from hydrophilic hydration diminishes, as does the loss in relative free energy for hydrophobic hydration. As a result, the cost of exposure of the hydrophobic core is decreased and the free energy difference between the folded and unfolded species becomes smaller, until the higher conformational entropy and lower volume of the unfolded state makes it more favorable. For the last decade however, the consensus has proven this view to be incorrect, as no correlation between the change in water structure by urea and the protein stability in solutions containing urea was found using calorimetry[112], while other studies failed to identify any change in the dynamic and structural properties of water[113] even at high denaturant concentrations using infrared spectroscopy[114] or simulations[115].

The second proposed mechanism, the interaction model, explains urea denaturation by direct binding of the cosolvent molecule to the protein. The increase of the surface area exposed for potential binding upon unfolding favors the unfolded state. This mechanism is now consensually recognized as being the mode of action of urea unfolding of proteins[116, 117]. In this mechanism, two interactions can be considered. The first is electrostatic and the second is Van der Waals interactions[118]. Number of studies have shown conflicting results on the role of each interaction on the stability change by urea.

In 2010, a simulation study by Canchi et al.[118] have shown that the dominating interaction driving the change in stability for Tc5b tryptophan cage variant is temperature dependent. Around room temperature (280-310K), the driving force is the Lenard-Jones interaction, whereas at higher temperature, the electrostatic Coulomb interaction becomes prevalent. This points out that the mechanism of urea denaturation is due to weak non-specific binding to the protein. The article also makes the hypothesis that given that the similar geometry of guanidinium hydrochloride and its electrolyte nature, the Coulomb interaction was expected to become dominant, while the Lenard-Jones interaction was expected to remain similar in intensity. Further analysis aiming at determining whether urea preferentially binds to side chains or backbones of proteins was performed. A full simulation study using two force fields successfully measured the interaction between those two groups, and concluded, if urea binds to both, it has a preference for side-chain binding[119]. Given the previously exposed hypothesis, the opposite may be true for guanidinium hydrochloride.

#### 1.4.4.2 Effect of crowding on stability

To account for the environment of proteins in living organisms, the effect of an environment more similar to the cell interior needs to be studied. To try and investigate those properties, the use of synthetic crowders for *in vitro* studies has permitted a first step in this direction.

One of the consequences of the effect of crowders on the protein stability is the deviation from ideality<sup>1</sup> due to the overlapping of the hydration shells resulting in a change in the activity coefficients that are no longer equivalent to the concentration of the solute[121]. This, with other experimental challenges, considerably complexifies the problem, in that the equation used to get the stability needs a non-trivial correction to account for the nonideality of the system and avoid under or overestimation of the relative concentrations, and thus the measured parameters[122]. In its molecular mechanism, crowding arises from three main contributions. The first is the overlapping of the hydration shells between different solutes, changing the stabilization of the hydrating water molecules with regard to the bulk solvent. It should be noted that in such a crowded environment as a cell interior, the very notion of bulk water as water molecules that are not coupled to any cosolute other than other water molecules loses part of its meaning. The second is non-specific interactions of cosolvent molecules with the protein that, like urea, can weakly bind different part of the protein, hence changing the stability of the state in function

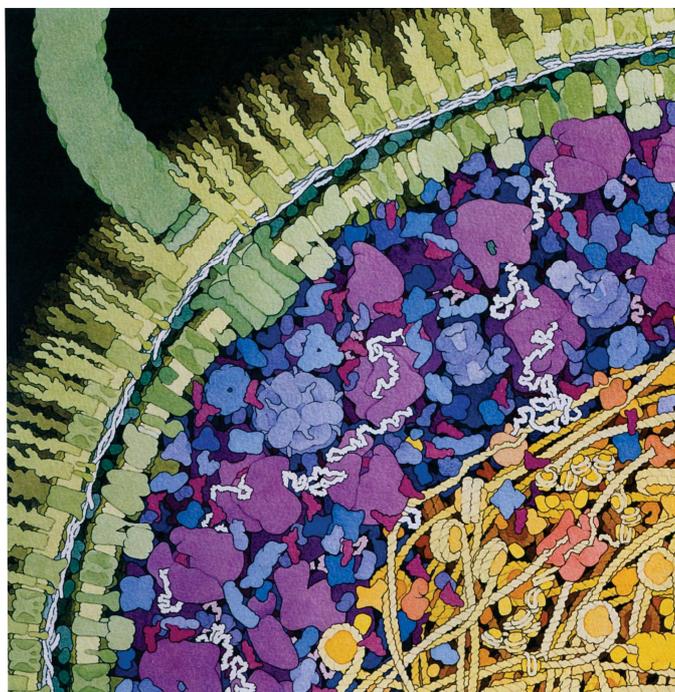


Figure 1.13: Artist rendering of the cell interior. Picture taken from [120]

---

<sup>1</sup>See 7.3.4

of the number of binding sites exposed. Due to its nonspecific nature, this contribution is likely to benefit the stability of the unfolded state, however, the effect can also arise from unfavorable nonspecific interactions with the protein, as is the case for osmolytes[123, 124, 125], and hence stabilize the folded state[126]. The third effect is the change in entropy due to the excluded volume effect decreasing the conformational entropy of the unfolded state resulting in the stabilization of the folded state with regard to the unfolded state[127].

Recently, a fourth contribution was found to have an important role in the stability of proteins in crowded environments. In this study, the stability of the folded protein was shown to depend on the net charge of the polymer used to mimic the cell interior[126, 122]. The salt concentration dependence of the destabilization, due to the screening of electromagnetic interactions by a change in the relative permittivity[128] confirmed the importance of nonspecific charge-charge interactions in the unfolded ensemble for proteins in the cell interior.

#### 1.4.5 Conclusion on thermodynamic stability of folded proteins

The origin of the stability of proteins has long puzzled biophysicists. Only in the last two decades have we started to understand the importance of the different contributions. To sum up here, we have seen that, in the general case and in lab conditions, conformational entropy and volume stabilize the unfolded state. The hydrophobic model, which has been a major concept in the protein folding community in the last 40 years, relies on the assumption that there is an entropic cost to the exposure of hydrophobic residues to solvent and furthermore, that there are nonspecific Van der Waals interactions in the folded structure between hydrophobic residues, due to the shape of the Lenard Jones potential. However, the magnitude of that effect seems to have been originally grossly overestimated[129]. The hydrophobic collapse is important in the folding process as is pointed out by models such as the spin glass model or the molten globule perspective, however, it is only one of the factors that stabilize the folded functional state relative to others. The existence of dry molten globule intermediates for some proteins underlines that excluding the solvent from the hydrophobic core is not always enough to reach the folded state. For example, in  $\alpha$ -Lactalbumin, it has been shown that a mutation preventing the formation of the two native disulfide bonds was causing the protein to collapse into a molten globule state in conditions that normally fold the protein[38]. These results confirm that the hydrophobic model alone cannot account for the stability of the folded state, and that intramolecular specific interactions play a big role in the stabilization of the folded state. On the other hand, a position dependence of the mutation effect on stability was observed, with a correlation between the number of C- $\alpha$  atoms around a residue and the loss in stability, suggesting that disruption of native state specific contacts has a major importance[130].

As another example of the importance of native specific interactions to the magnitude of

the stability of the folded state with regard to that of the unfolded state, we can look at the effect of mutations suppressing native hydrogen bonds. The stability of a  $NH \cdots O$  hydrogen bond is estimated between  $-1.58$  and  $-3.37 \text{ kcal.mol}^{-1}$  which is on the lower end, but well within the typical stability of folding observed in midsize WT proteins[2]. Indeed, the loss of a single hydrogen bond in the Y131F/D146L construct of the PP32 wild-type, while maintaining a very similar structure content as the WT protein based on the high similarity of the 2D NMR spectra, provokes a loss in stability of  $3.2 \text{ kcal.mol}^{-1}$ [62]. Even though this bond is a  $OH \cdots O$  bond, and therefore has a different energy than the  $NH \cdots O$  cited above, both are considered strong hydrogen bonds, and thus are expected to have a similar range. This example is striking because this highly conserved H-bond results from side chains that are exposed to the solvent in the folded state, which means that changes in packing, or hydration effects cannot account for the difference as they are not expected to change upon unfolding. This difference in stability is quite close to the values reported for the strength of hydrogen bonds in proteins. In addition, one must account for the structural change associated with the mutation of this very conserved hydrogen bond. As seen previously in equation 1.3.11, a change in chemical potential corresponding to the hydrogen bond would change entropy and volume, in accordance with Le Chatellier's principle, by populating a slightly different ensemble. This is shown by large changes in chemical shifts in the vicinity of the hydrogen bond in Y131F/D146L, as seen in figure 1.14.

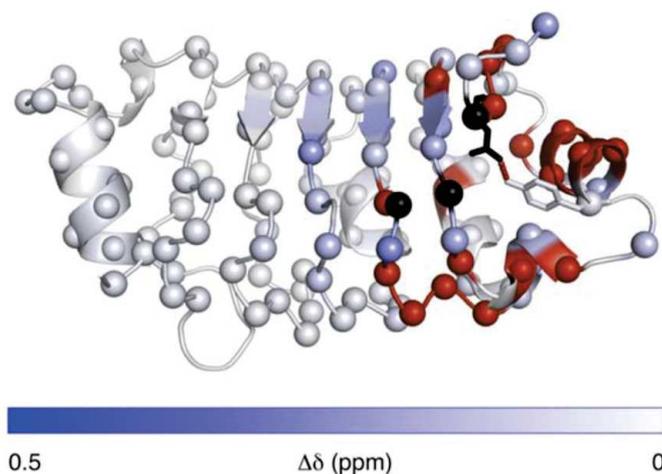


Figure 1.14: Chemical shift perturbations of the PP32 Y131F/D146L variant compared with wild-type. Y131 and D146 are represented in sticks. Spheres represent  $C_{\alpha}$ 's. Chemical shift perturbations of residues for which assignments can be transferred from PP32 to PP32 Y131F/D146L are displayed on a blue to white scale. Residues for which amide peaks disappear are displayed in black, and those that move in a crowded region of the HSQC and therefore cannot be assigned with certainty are displayed in red. Figure taken from [131]

Thus, the stability of the protein is the result of several competing factors of big magnitude that partly cancel each other out to give rise to a difference in stability of the states present in the folding landscape. Although this difference in stability is somewhat marginal, even more so if taking into account the crowding effects present in the native environment of proteins, evolution has selected interactions that leave sufficient conformational freedom for enabling actions such as binding or allosteric conformational changes, while keeping a sequence with sufficient stability heterogeneity in the structure to allow for folding in a directed manner and in a reasonable timescale, by early selection of pathways. The use of different denaturation methods is essential for the determination of the different parameters that drive protein folding. Each method can have advantages and drawbacks.

In this thesis, a major focus was put on pressure, because of the unique advantages it has. Pressure is a gentle, often fully reversible method. Pressure has very little effect on hydrophobic interactions, making the denatured state arguably closer to the denatured state that is transiently populated in native conditions. In addition, pressure also works locally on the void defects that are present in specific parts of the protein, as opposed to temperature or chemical denaturants, that work on the difference in accessible surface area between the folded and unfolded states. This region specific mechanism makes it ideal for populating intermediates, because two regions with small interfacial energetic coupling can unfold independently if one has more void volume than the other. This is not the case with temperature and urea, because the surface area exposed upon unfolding is distributed in a homogeneous manner.

#### 1.4.5.1 The pressure temperature phase diagram

The integration of equation 1.3.13 taking into account the second degree dependence of the protein stability in a two-state approximation as realized in the 1971 Hawley article is [132]:

$$\begin{aligned} \Delta G_{(P,T)} = & \Delta G_{(P_0,T_0)} + \Delta V_0(P - P_0) - \Delta S_0(T - T_0) \\ & + \frac{1}{2}\Delta\beta_0(P - P_0)^2 - \Delta C_{p,0} \left[ T \left( \ln \frac{T}{T_0} - 1 \right) + T_0 \right] + \Delta\alpha(P - P_0)(T - T_0) \end{aligned} \quad (1.4.25)$$

A full description of the steps that lead to that equation is available in the annex. This equation can be used to extrapolate the stability of a protein from one pressure-temperature to different conditions, providing a knowledge of the first and second derivatives of the stability with regard to both pressure and temperature. One should note that the second degree dependence of the thermodynamic parameters in this equation are absolute thermal expansivity and compressibility definition, contrary to the usual coefficient of isobaric thermal expansion and

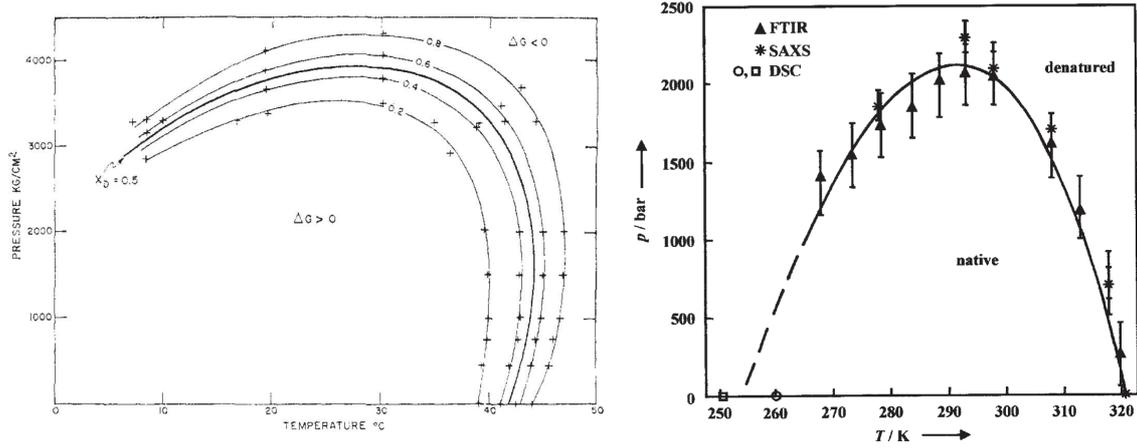


Figure 1.15: Left : Thermodynamic Transition Parameters for Chymotrypsinogen (pH 2.07) and Ribonuclease (pH 2.0) from [132]. Right : Temperature versus pressure stability diagram of SNase at pH5.5 as obtained by SAXS, FTIR and DSC measurements. Values from [133, 101, 134]. Figure from [135]

isothermal compressibility. Thus the definition in the Hawley article are :

$$\alpha^{abs} = \frac{\partial V}{\partial T} \quad (1.4.26)$$

$$\beta^{abs} = \frac{\partial V}{\partial P} \quad (1.4.27)$$

Instead of the usual :

$$\alpha = \frac{1}{V} \frac{\partial V}{\partial T} \quad (1.4.28)$$

$$\beta = -\frac{1}{V} \frac{\partial V}{\partial P} \quad (1.4.29)$$

In that regard, we can actually identify the terms presented in the introduction to thermodynamics section for the second degree dependence ( $-\beta_u v_u + \beta_f v_f$ ) to be equivalent to the  $\Delta\beta^{abs}$  term used in the Hawley article, and similarly,  $(\alpha_u v_u - \alpha_f v_f) = \Delta\alpha^{abs}$ . The elliptical shape of the diagram obtained is the result of the approximation of the  $\Delta C_p$  term at the second degree Taylor expansion series :

$$\Delta C_{p,u} \left( T \ln \left( \frac{T}{T_0} \right) - (T - T_0) \right) \simeq \frac{\Delta C_p}{2T_0} (T - T_0)^2 \quad (1.4.30)$$

And hence is an approximation around the  $T_0, P_0$  point of the real equation. If this approximation is not taken, then the ellipticity does not appear. Furthermore, for the diagram to form an ellipse, the equation needs to satisfy the condition  $(\Delta\alpha)^2 > \frac{\Delta C_p \Delta\beta}{T_0}$ . Experimental measures

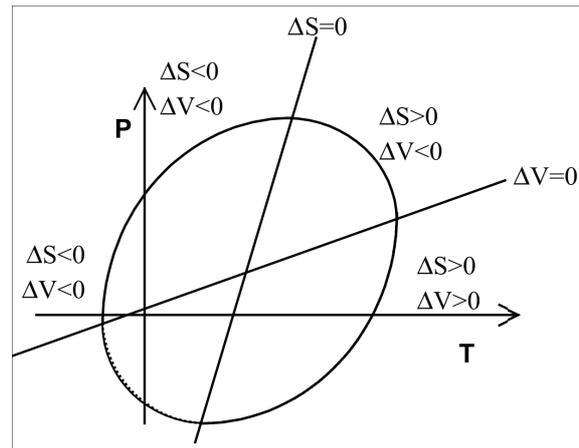


Figure 1.16: Pressure temperature phase diagram of proteins. Relative position of the  $\Delta S=0$  and  $\Delta V=0$  lines compared to the ellipse. Figure adapted from [136]

of the parameters in equation 1.4.25 have been made for some proteins and thus reveal the shape of their pressure-temperature phase diagram (figure 1.15).

The curvature resulting from that equation can be seen in figure 1.16, with the lines that correspond to the pressure and temperature where the sign of  $\Delta S$  and  $\Delta V$  change, thus underlining that the factor that drive protein stability is not constant over temperature and pressure.

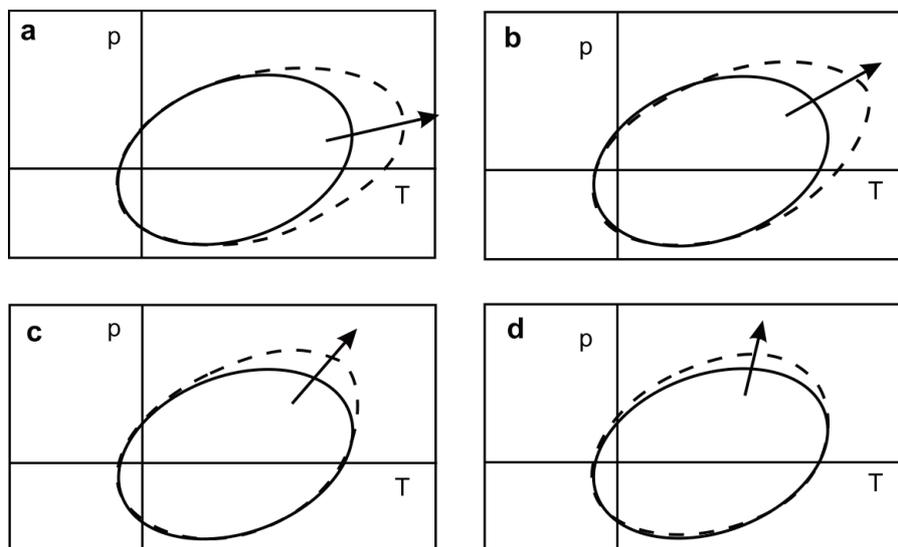


Figure 1.17: Effect of higher order terms on the shape of the elliptic diagram. Effect of terms containing (a)  $T^3$ ; (b)  $T^2p$ ; (c)  $Tp^2$ ; (d)  $p^3$ . The solid line shows the original phase diagram; the broken line shows the phase diagram if the higher order term is taken into account. The arrows show the direction of the distortion when the higher order term increases. Figure adapted from [136]

A further decomposition of the second derivative of  $\Delta G^0$  into their own dependence to pressure and temperature is possible, however, we will not go into details, because the measurement of these terms is quite difficult, and hence it is complicated to use this kind of equation to support experimental theories. However, one should keep in mind that third degree dependence of stability is likely non-null. The change of those terms would distort the pressure temperature phase diagram, as illustrated in figure 1.17.

# Chapter 2

## Materials and methods

### 2.1 Sample preparation

#### 2.1.1 PP32

##### 2.1.1.1 Protein production

Plasmids containing the protein sequence were acquired from the Barrick lab, in DH5-alpha cells. Protein expression was made by introducing the vector in BL21 Escherichia coli cells, following the protocol found in section 7.2. PP32 WT and variants were expressed using pEt24b plasmids containing a T7 promoter, a His-tag for purification and a Kanamycin resistance gene to ensure selection of bacteria containing the plasmid.

Protein expression was induced by addition of IPTG after having reached an optical density of 0.6 at a wavelength of 600nm. Proteins were then purified with a nickel column, and dialyzed into a 50mM NaCl 20mM Sodium Phosphate, 5mM DTT buffer at pH 7.8 for stocking in aliquots at -80°C. The full protocol for protein production can be found in annex (section 7.2).

##### 2.1.1.2 Sample composition

For the experiment, the buffer was replaced by a 20mM bis tris, 10mM NaCl 5mM DTT for all proteins. The WT protein solution used in chapter 3 also contained 1.4M of urea to ensure that the protein would unfold in the accessible pressure range. Because preliminary studies using tryptophan fluorescence showed aggregation under high pressure conditions, attributed to polymerization through disulfide bonds of the two Cysteines, the high concentration of DTT

proved to be necessary. Following experiments using 1D NMR showed a good reversibility at 5mM DTT (figure 2.1). Mutants presented in chapter 4 did not require the addition of any urea to unfold entirely before 2500 bar.

## 2.1.2 Trp cage

### 2.1.2.1 Protein production

The Tc5b protein used in chapter 6 was bought directly from AnaSpec in lyophilized form, due to the complexity of the synthesis of this peptide.

### 2.1.2.2 Sample composition

High pressure NMR were performed between 1-2.5 kbar at 285 K and 298 K in 100 mM d-acetic acid buffer pH 5.0 containing 7% D<sub>2</sub>O with 0, 0.5 and 1.0 M urea and 100 mM d-TrisHCl buffer pH 7.0 containing 7% D<sub>2</sub>O.

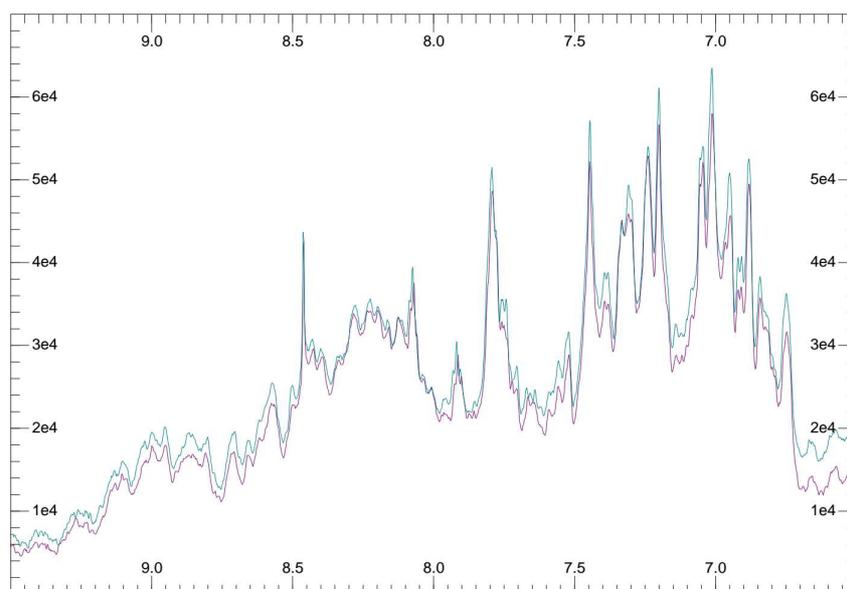


Figure 2.1: Reversibility of PP32 WT by 1D NMR. The green line corresponds to the signal before the application of pressure, the purple line represents the signal after high pressure was applied and the system was returned to atmospheric pressure

## 2.2 NMR Theory

The NMR experiments consist of a measure of the sensed magnetic fields of non zero spins of nuclei. The spin is an intrinsic property of quantum particle that is the quantum equivalent of angular momentum in classical physics. It is defined by the spin quantum number. We can determine the spin quantum number  $s$  of a nucleus based on its composition with a simple rule : nuclei that have an even number of both protons and neutrons have a null spin, nuclei that have an odd number of both protons and neutrons have a spin that is a positive integer, and all others have half integral spins[137, 138]. The quantized spin vector can be expressed from the spin quantum number through :

$$\|S\| = \hbar\sqrt{s(s+1)} \quad (2.2.1)$$

Where  $\|S\|$  is the norm of the spin vector. The projection along arbitrary direction  $i$  of which is defined as :

$$S_i = s_i\hbar \quad (2.2.2)$$

Where  $s_i$  is the secondary spin quantum number (or spin projection quantum number), ranging from  $-s$  to  $s$  in integer steps. Since the nucleus is charged, a non null spin results in the creation of a magnetic moment :

$$\vec{\mu}_n = \gamma_n\vec{S} \quad (2.2.3)$$

Where  $\vec{\mu}_n$  is the magnetic moment of the nucleus  $\gamma_n$  its gyromagnetic ratio and  $\hbar\vec{I}$  is its spin angular moment. The gyromagnetic ratio depends on the nucleus, and is the biggest factor that determines the nucleus' sensitivity to the magnetic field.

Nucleus	$\gamma_n$
$^1H$	267.513
$^2H$	41.065
$^{13}C$	67.262
$^{15}N$	-27.116
$^{17}O$	-36.264
$^{19}F$	251.662
$^{31}P$	108.291

By application of an external field  $\vec{B}$  on the magnetic moment, will become a function of the

angle with respect to the orientation of  $\vec{B}$  :

$$E_{(\theta)} = \int \Gamma_{(\theta)} d\theta = \int (\vec{\mu}_n \times \vec{B}) d\theta = -|\vec{\mu}_n| |\vec{B}| \cos(\theta) = \vec{\mu}_n \cdot \vec{B} \quad (2.2.4)$$

Where  $\Gamma_{\theta} = \vec{\mu}_n \times \vec{B}_0$ . Assuming a magnetization solely in the z-axis ( $\vec{B} = \vec{B}_z$ ), we can define an Hamiltonian function describing the system :

$$\mathcal{H} = -\mu_z B_z \quad (2.2.5)$$

In which  $\mu_z$  is the z component of the magnetic moment, according to eq 2.2.2 and eq 2.2.3

$$\mu_z = \gamma \hbar s_z \quad (2.2.6)$$

Where for a half integer spin can take the values of  $s_z = \pm 1/2$ . This leads to an hamiltonian along the z axis :

$$\mathcal{H} = -\gamma \hbar B_z s_z \quad (2.2.7)$$

From here we can define two states, a first ground state that is defined by value of  $s_z = +1/2$  in which the energy is  $E_{\alpha} = -\frac{\gamma \hbar B_z}{2}$ , and an excited state defined by value of  $s_z = -1/2$  in which he energy is  $E_{\beta} = \frac{\gamma \hbar B_z}{2}$ . We can deduce the energy difference between those states :

$$\Delta E = \gamma \hbar B_z \quad (2.2.8)$$

Using the well establish relation between energy and frequency,  $E = \hbar \omega$ , we can identify the Larmor's relation :

$$\omega = \gamma B \quad (2.2.9)$$

Where  $\omega$  is the frequency of the Larmor precession, and corresponds to the rotation of the magnetic momentum vector around the the axis of the magnetic field.  $\hbar \omega$  corresponds to the energy separating the two populations and is hence the absorption, or resonance, frequency of the nucleus. This is similar to the precession observed for a spinning top, in which the angular momentum causes the top of the sniping top to rotate horizontally around the vertical gravity field.

The difference in energy gives rise to a difference between the population of the ground and excited states :

$$\frac{N_{\alpha}}{N_{\beta}} = e^{-\frac{\gamma \hbar B_z}{k_b T}} \quad (2.2.10)$$

Which gives a very small difference in the population of each state.

By application of a Radio frequency field polarized so that the magnetic component is orthogonal to  $\vec{B}$ , we can transfer atoms from the ground state to the excited state. In addition

to the change in population of those states, the spin will also be in phase at  $t = 0$  after the end of the pulse.

### 2.2.1 Reference frame and rotating frame

Because of the very high frequency of the precession, a mathematical trick is usually performed to ease the interpretation, by introducing a rotating frame  $R'_{(x',y',z')}$  in addition to the original frame  $R_{(x,y,z)}$ . If we consider  $R'$  to be rotation at a frequency  $\Omega$  with regard to the reference frame, the rotation of a vector  $\vec{V}$  at a frequency  $\vec{\omega}_V$  in  $R$  can be expressed in  $R'$  by :

$$\left(\frac{d\vec{V}}{dt}\right)_{R'} = \left(\frac{d\vec{V}}{dt}\right)_R + \left(\frac{d\vec{u}_x}{dt}\right)_{R'} \quad (2.2.11)$$

Where  $\left(\frac{d\vec{u}_x}{dt}\right)_{R'}$  is the movement in  $R'$  of a unitary vector of the  $R$  reference frame, hence :

$$\left(\frac{d\vec{u}_x}{dt}\right)_{R'} = \vec{\Omega} \times \vec{u}_x \quad (2.2.12)$$

Hence :

$$\left(\frac{d\vec{V}}{dt}\right)_{R'} = \vec{\omega}_V \times \vec{V} + \vec{\Omega} \times \vec{u}_x \quad (2.2.13)$$

If we then replace  $\vec{V}$  by the magnetic moment  $\vec{\mu}$  :

$$\left(\frac{d\vec{\mu}}{dt}\right)_R = \vec{\omega}_L \times \vec{\mu} = \gamma \vec{B} \times \vec{\mu} \quad (2.2.14)$$

Which in  $R'$  is :

$$\left(\frac{d\vec{\mu}}{dt}\right)_{R'} = \gamma \vec{B} \times \vec{\mu} + \vec{\Omega} \times \vec{u}_x \quad (2.2.15)$$

Which is equivalent to having a rotation in the rotating frames that has a frequency of  $\gamma \vec{B}'$  that is :

$$\left(\frac{d\vec{\mu}}{dt}\right)_{R'} = \gamma \vec{B}' \times \vec{\mu} \quad (2.2.16)$$

Where  $\vec{B}' = \vec{B} + \frac{\vec{\Omega}}{\gamma}$ . By choosing a rotating frame with  $\Omega = -\gamma \vec{B}$ , we effectively cancel out the contribution for the rotation due to the Larmor precession. This is done electronically by demodulating that signal by a frequency  $\vec{\Omega} = \gamma \vec{B}_0$ , where  $\vec{B}_0$  is the strength of the field of the magnet, and  $\gamma$  the gyromagnetic ratio of the nucleus in the appropriate channel.

## 2.2.2 Shielding effect

In practice however, the sensed field of each nuclei is different from  $\vec{B}_0$ , giving rise to the difference in precession and relaxation that are at the core of the NMR signal analysis. The shift in the sensed magnetic field is due to the organization of the electrons around the nuclei, themselves having a precession movement due to the magnetic field that modify the effective field in their surroundings. Changes in the sensed magnetic field can be expressed as :

$$\vec{B} = (1 - \sigma) \vec{B}_0 \quad (2.2.17)$$

Where  $\sigma$  is the shielding of the nuclear spin, and is proportional to the density of electrons around the nucleus. The difference between the expected resonance frequency of a nucleus at the magnetic field of the NMR magnet and the measured frequency due to shielding effect is called *chemical shift*, and is the base of any NMR analysis.

From there on, we must give up the notation using magnetic moment to use the magnetization, which is the macroscopic observable corresponding to the sum of the magnetic moments :

$$M_i = \sum_j^{system} \mu_j^i \quad (2.2.18)$$

Where  $M_i$  is the component of the magnetization in the dimension  $i$  and  $\mu_j^i$  is the component of the magnetic moment in the dimension  $i$  of spin (nucleus)  $j$ .

## 2.2.3 Free induction decay

### 2.2.3.1 Relaxation time

NMR experiments rely on the measurement of two relaxation times after a perturbation. The magnetic field applied in the sample tends to align all the spins of non zero gyromagnetic ratio. For an easier understanding, we call  $z$  the direction of the magnetic field. The spins are then excited by a single phase electromagnetic field polarized along the  $xy$  plane. As a result, the spins now rotate in the  $xy$  plane with the same orientation that corresponds to the phase of the exciting wave. Two types of relaxation are then recorded at once by recording the magnetization of the sample in the  $xy$  plane.

**T1 relaxation** The first relaxation is called T1, or spin-lattice relaxation and corresponds to the loss in magnetization of the sample due to realignment of the spins with the magnetic field, that is the regain in the  $z$ -component of the magnetization. The term "lattice" is in reference

to the "lattice field", that is the magnetic field created by the nucleus in the spin system due to thermal motion. T1 relaxation is thus due the transfer of energy from the exciting radio-frequency field to the surrounding lattice, resulting in a slight increase in temperature.

**T2 relaxation** T2 relaxation, also called spin-spin relaxation or transverse relaxation, corresponds to the loss in magnetization due to the loss in coherence of the spins. The loss of coherence is due to the difference in the chemical environment of the nucleus, resulting in a difference in the precessing speed of the vector. Hence, the spins lose their coherence. This type of experiments is useful to explore the dynamics of a system. Using spin-echo ( $180^\circ$ ) pulses, one can artificially increase the resolution in the loss of phase coherence, by keeping the spins in a excited state and refocusing them. The total loss of signal is then due to the change of environment. If the signal is lost rapidly, the region associated with the signal is very dynamic, otherwise, it is more rigid. It should be noted that  $T2 \ll T1$

**T2\* relaxation** T2\* corresponds to a special type of relaxation due to the defects intrinsic to the measurement system. In most cases, it can safely be ignored.

### 2.2.3.2 Signal processing

The free induction decay signal received during the relaxation of the nucleus is a combination of both T1 and T2 relaxation times. This signal is recorded in the xy plane through two coils recording both the magnetization in the x and y plane, thus resulting in a complex signal with a phase. The analog signal then needs to be sampled into a digital form for the computer to process. Thus, the user choses the number of points needed for sampling, N, which typically ranges from 512 to 2048. In any case, it is recommended that N is a power of two, because the Fast Fourier Transform can not be performed if this is not the case[137]. The time interval between two points in the discreet FID is called the dwell time,  $\tau_{dw}$ . Thus, the total acquisition time is  $AQ = N\tau_{dw}$ . The sweep width, or spectral width, being the width of the frequency domain spectra, is defined as  $SW = \frac{1}{2\tau_{dw}}$ . The spectral resolution in the frequency domain is thus  $\frac{2SW}{N}$ .

The obtention of a spectrum is made through several operations on the FID. First, removal of the DC offset (if an offset is present), which corresponds to the offset of the FID intensity, and causes a signal to appear at 0ppm in the spectra. In a second step, the spectral resolution can be artificially increased by extending the FID. The extension of the FID can be realized using two different technics depending on the length of the FID. If the FID is longer than 3 or 4 times T2, the appropriate method for the resolution increase is called 0 filling. For shorter FIDs, of the order of magnitude of T2, the appropriate method to use is the Linear Prediction

method. One should keep in mind that the use of these methods can alter the spectrum and lead to artifacts. If the FID is cut short, a truncation artifact can be seen in the spectra. It is precisely to avoid this artifact that FID extension methods have been developed, however, if not used properly, these methods will result in artifacts themselves.

Zero filling consists in doubling the number of points in the FID by adding  $N$  zeros at its end. This is useful to enhance the quality of the spectra without the need for longer acquisition periods. This is because the resolution of the spectra in the frequency domain is equal to the inverse of the acquisition time of the FID. Zero filling is used instead of longer FID acquisition time because with an FID recorded for this long, the system is already quite relaxed and the signal is going to become very noisy. Zero filling thus increases the resolution of the spectra without having to increase the acquisition time. However, using zero filling on FIDs that are too short will result in the appearance of a wave like signal on the spectra.

If the FID is shorter, down to around  $T_2$ , the Linear Prediction algorithm should be used for the extension of the FID. Linear prediction uses extrapolation of the data to extend the FID. The value of the extrapolated points in the FID is determined by the equation :

$$x_n = \sum_i^{i=N+n} a_i x_{n-i} \quad (2.2.19)$$

Once the point  $n$  has been determined, it can be used to determine the next one, and so on until the number of points in the FID has been doubled. This approach assumes that the FID is represented by a sum of dampened functions :

$$x_n = \sum_{k=1}^K c_k e^{-\frac{n\Delta t}{T_{2,k}}} \cos(\omega_k n \Delta t + \phi_k) \quad (2.2.20)$$

The number of coefficients needed to accurately represent a signal depends on the number of resonance frequencies present in the FID. If the signal is composed of  $X$  number of different frequencies, a minimum of  $2X$  coefficients will be necessary to accurately represent the spectrum[137]. If too few coefficients are chosen, only the strongest signal will be used for linear prediction ; if too many coefficients are chosen, then the noise will be used for linear prediction resulting in the appearance of artificial peaks in the spectrum. The number of coefficients should not exceed  $N/2$ .

## 2.2.4 Multidimensional experiments

The NMR experiments can be realized in more than one dimension, by transfer of the magnetization from the hydrogen to other atoms, and letting the spin evolve on another nucleus before transferring back to the hydrogen for reading. If the delay given for the evolution in

the second nucleus is implemented gradually, one can then get a second dimension in the time domain corresponding to the evolution of the second nucleus and thus get a second dimension in the frequency domain that corresponds to the precessing frequency of the second nucleus. This use of several nuclei during the relaxation period enables to differentiate peaks that would otherwise have the same chemical shift in the hydrogen dimension. This can also be used for the assignment of NMR peaks by transferring magnetization from one residue to another, thus allowing to reveal which peaks correspond to residues that are in contact in the primary structure, because they share a frequency in at least one dimension.

#### 2.2.4.1 $^1\text{H}$ - $^{15}\text{N}$ HSQC

$^1\text{H}$ - $^{15}\text{N}$  HSQC is one of the most common experiments in NMR spectroscopy of organic molecules. This experiment consists in the transfer of the magnetization from the proton to the nitrogen isotope using an INEPT sequence (Insensitive nuclei enhanced by polarization transfer), and the system is left to evolve during a delay ( $t_1$ ). The magnetization is then transferred back onto the amide hydrogen via a retro INEPT sequence for reading. This operation is repeated with an incrementation of the  $t_1$  delay, thus providing a second dimension corresponding to the frequency in the  $^{15}\text{N}$  dimension. The presence of arginine or asparagine in the sequence of the peptide in the sample results in the appearance of doublets in the spectra, in the top-right corner, assuming a frequency window of around 10ppm to 6ppm in the proton dimension and 100ppm to 130ppm in the nitrogen dimension.

#### 2.2.4.2 $^1\text{H}$ - $^1\text{H}$ TOCSY

In the tryptophan cage study, we used the  $^1\text{H}$ - $^1\text{H}$  TOCSY experiment to monitor the changes in chemical shift in function of pressure. The  $^1\text{H}$ - $^1\text{H}$  TOCSY, for TOtal Correlation SpectroscopY, is a type of experiment where the signal is transferred from a proton to all protons in its spin system. That means magnetization is not going to be transferred only to the protons that are J-coupled (meaning coupled through chemical bonds) to the proton originally excited, but to all a proton that are coupled with one another. The resulting spectra is composed of a diagonal line that is the self correlation, and a region corresponding to the inter-correlation with the spins within the spin system. The spin system is limited to a single amino acid because of the presence of the CO atom that is not carrying any hydrogen atom, preventing the coupling of a proton of one residue to a proton of another. The use of this sequence for chemical shift change analysis in our Trp-cage study is due to the fact that, being unable to produce the sample ourselves, we did not obtain labeled sampled, thus restricting the type of experiments that can be used. The need for good precision for the chemical in the amid proton dimension makes the use of a  $^1\text{H}$ - $^1\text{H}$  TOCSY needed because the multiplicity of the peaks in

the secondary dimension allows to better distinguish residues that have similar amide proton chemical shift, thus making the spectra easier to read in crowded regions.

### 2.2.5 Chemical exchange in NMR experiments

The mixing time, which is the period used for the magnetization transfer between two groups, and the timescale of the processes observed will determine the type of exchange regime and therefore the type of signal that one will obtain. If the rate of exchange between conformations is significantly faster than the mixing time, it is a fast exchange regime. During the mixing time, the environment of the spins is changing several times between the conformations, resulting in an averaging of their frequencies. We then observe a single peak, at a position that is the weighted average of where the signal would be if the exchange regime was slow. If the rate of exchange between conformations is significantly slower than the mixing time, it is a slow exchange regime. In that case most of the spins that were excited in a given conformation have not switch to another, resulting in two distinct peaks at the position of the spectra that corresponds to the given conformation.

However, in intermediate regimes, the spins are excited in one state, and some of the spins from one conformation transfer into another, causing the normal xy magnetization to decrease faster than the normal rate. In this case, the lineshape broadens and the signal disappears because environment felt by the nucleus is very disperse, and thus results in different averages for each molecules. In practice, regimes are often not fully fast or slow. Because of that, some signal may sometimes be changed because of a slight peak broadening and become less Lorentzian in shape with the changing conditions, changing the rates of reactions. This property is sometimes used to measure the changes in rates of events with changing conditions through lineshape analysis[137, 139, 140].

In practice, the difference in precession frequency of the two states compared to the rate of chemical exchange is the factor that will determine whether the observed peak will appear as in slow or fast exchange. This difference can be denoted  $\Delta\nu = \frac{1}{2\pi}(\omega_a - \omega_b)$ , where  $\omega_i$  is the frequency of precession of the nuclei in state  $i$ , and the observed exchange rate is  $k_{ex} = k_{ab} + k_{ba}$ . The exchange regime can be determine from this two parameters. If  $k_{ex} \ll \Delta\nu$ , the system is in very slow exchange, where two sharp resonance peaks can be observed. If  $k_{ex} < \Delta\nu$  the system is in slow exchange, and two broaden resonance peaks can be seen. If  $k_{ex} \simeq \Delta\nu$ , the system is in intermediate regime, and the line shape is complex and extremely broadened. If  $k_{ex} > \Delta\nu$ , the system is in fast exchange with a single broaden visible resonance peak. Finally, if  $k_{ex} \gg \Delta\nu$ , the regime is in very fast exchange and two sharp resonance peaks can be observed[137].

## 2.2.6 Experiments and analysis

The NMR spectra presented in this work were recorded on a 600MHz brucker spectrometer using a in line high pressure ceramic tube from Daedalus Innovations. The sample were concentrated to  $500\mu M$  for the PP32 studies, 5% D<sub>2</sub>O for the locking procedure and 0.05% DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) for future referencing. The high pressure setup consists of a syringe connected to the high pressure tube by a flexible line. The syringe pushes the transmitting liquid (in our case water for the Trp-cage study, and mineral oil for the PP32 study). The sample is inserted in the high pressure tube and an interface liquid, mineral oil, is used to ensure that the sample does not diffuse into the high pressure line, in the case where it is filled with water. A special attention has to be given to limit the presence of air bubbles at the oil-sample interface. It should also be noted that the high pressure tube has an effective volume (around  $420\mu M$ ) that is inferior to the volume of a classic 5mm NMR tube due to the width of the ceramic made to be pressure resistant (the internal diameter of the high pressure tube is of 3mm). The accessible pressure range with such a tube is between 1 and 2500 bar, but can be extended to 3000 bar with special tubes.

### 2.2.6.1 Locking procedure

In solution NMR, the strength of the magnetic field needs to be adjusted to account for variation in the perceived field in the NMR sample. This is called the lock procedure, during which the magnetic field in the sample is adjusted by the addition of an additional magnetic field until the correct sensed magnetic field is reached. This added magnetic field strength is called the lockpower. The deuterium lock procedure requires the introduction of deuterium marked water ( $D_2O$ , or heavy water), usually around 5 to 10% of the total volume. In this procedure, the deuterium signal is chosen as the reference nuclei for the determination of the 0ppm frequency. In effect this is achieved by using a reference frequency generator supplying the frequency we wish to keep the precession of the deuterium atom at. This reference frequency, the absolute lock frequency, corresponds to  $\omega = -\gamma B_0$ , where  $\gamma$  is the gyromagnetic ratio of the deuterium nuclei, and  $B_0$  the reference magnetic field corresponding to the magnet. The locking procedure then consists in finding the lockpower that maximizes the excitation of the deuterium nuclei. To optimize this procedure, the signal is detected in quadrature, enabling to see if the deuterium nuclei is precessing faster or slower than the reference frequency, in order to correct the magnetic field. The nuclei is thus excited, and fed back into the lock receiver, and the lock gain is adjusted during the amplification. Lock signal can vary with temperature, and it is thus preferable to repeat the locking procedure after a change in temperature. This dependence highlights the need for referencing in order to allow comparison of spectra at different temperatures.

### 2.2.6.2 Shimming

In addition to changes in the average strength of a magnetic field, field inhomogeneities are also present in the sample. To get a good signal these inhomogeneities need to be reduced, using shim coils to make small correction along the three directions of space. During the shimming procedure, the lock signal is used as an indicator for field inhomogeneities : the more homogeneous the field is, the larger the lock signal will be.

### 2.2.6.3 Tuning and matching

The NMR probe is composed of difference channel dedicated to the transmission and reception of signal to different types of nucleus. To ensure maximum efficiency, this channels need to be adjusted. This is done by the tuning and matching of the circuits composing the different channels. The tuning and matching procedures are essential to obtain a good spectra. Tuning is the operation by which the absorption maximum is shifted in frequency, to obtain the precession frequency of the desired nuclei, and matching changes the amount of power that is reflected by the probe[137].

### 2.2.6.4 Spectral analysis

**Referencing** In order to correct the chemical shift variation due to the instrumentation and the material used, or even the solvent conditions, it is always suitable to insert DSS in the spectra in order to serve as a reference point. For example, a slightly higher solvent density at high pressure can result in a shift of the chemical shift that does not correspond to actual changes in the protein. DSS (2,2-dimethylsilapentane-5-sulfonic acid) is chemical agent that is widely used because it has a methyl chemical shift at exactly 0ppm. In applications where it is the intensity or the volume of the peak that is monitored, the referencing is not necessary, however it is always recommended, would it only be for the sake of comparison. In the case where the chemical shift is monitored, referencing becomes absolutely necessary. Referencing can be easily done and saved directly on the bruker software Topspin in the proton dimension, using 1D experiments. DSS has a easily identifiable triplet signal that should be found close to 0ppm. The values labeled as SF in the processing parameters needs to be reported in the indirect-dimensions from the proton dimension. For indirect dimensions, one should use values reported in Markley et al.[141]. The number that needs to be reported in the SF field (in bruker's software topspin) of the indirect dimension is the one obtain for the first dimension multiplied by the ratio of relative frequencies presented in table 3 of this article. The validity of the assumption that DSS signal does not depend on pressure (as it does not depend on temperature), was demonstrated by Li et al.[142, 143].

**Intensity analysis** In general, the correct way to analyze the relative populations of two states in slow exchange is through volume. However, because of deviation from an ideal Lorentzian peak lineshape, volume can yield significant noise, and thus intensity analysis is sometimes more suitable. For data analysis, we have used the CCPNMR analysis software, after processing of the spectra in topspin. The intensity of the peaks was obtained by finding the maximum values of a 2D Gaussian fit around each peak, because this method yielded better quality of unfolding profiles than either volume or Lorentzian fitting. The resulting intensities are then exported as text file, with residues in increasing order. The output corresponds to one file per titration point, therefore a script was written to transform those files into a single file, that was then read by a program written specially for this analysis, to perform fitting of the two-state model and parameters extraction, contact maps and histogram plotting. This software is simply aiming at making the procedure more intuitive and faster, using mostly the mouse to look quickly at each unfolding profile as we look the NMR spectra of the corresponding residue, to eliminate peaks that have a weird profile due to something that went wrong in the 2D fitting procedure (in which case the value of the point can be easily replaced), or because of overlap during the titration. The software also has a qualitative button to keep the memory of why the peaks was taken out of the analysis (for example because it is in the zone where the unfolding peaks appear, making its unfolding profile incomplete).

**Chemical shift analysis** For the fast exchange regime, changes in the states present in solution are reported by changes in the observed chemical shifts. That is because the peaks observed are positioned at a weighted average of the frequencies of the substates that compose the solution. Therefore, changes in the solution composition changes the influence of the substates on the observed chemical shift and this latter changes. Proton chemical shift changes are correlated to changes in average bond length, and thus can yield a significant amount of information.

## 2.3 Molecular Dynamics simulations

### 2.3.1 Theory

Molecular dynamics simulations is a tool that can be use to gain some insight on the small scale mechanisms of protein behavior. If the prediction on the structure of the protein from its primary sequence is the dream of all computational biophysicist, the reliability of the results given by molecular dynamics can only be considered with regard to experiments. This is because the models now used in MD simulation are not yet capable of a full description of the phenomena that exist on the atomic scale of the particle. This is both because of the

approximations used to enhance sampling, as well as because the potentials used do not fully reproduce the behaviors observed experimentally.

One very important thing to consider in molecular dynamic simulations is the timescale at which events occur in protein (figure 2.3). The fast motions in a protein structure is the covalent bond vibration that occurs in  $\simeq 10fs$ ; For this reason, the timestep of an all atoms simulation is usually chosen to be  $1fs$  so that it is one order of magnitude under the fastest phenomena and can describe it well. The use of constraint algorithm[144] can be used to increase the time step by giving an approximation of the behavior of the bond vibration, in order to decrease the time of the simulation and the calculation power to do it, which is a major obstacle for the MD simulation.

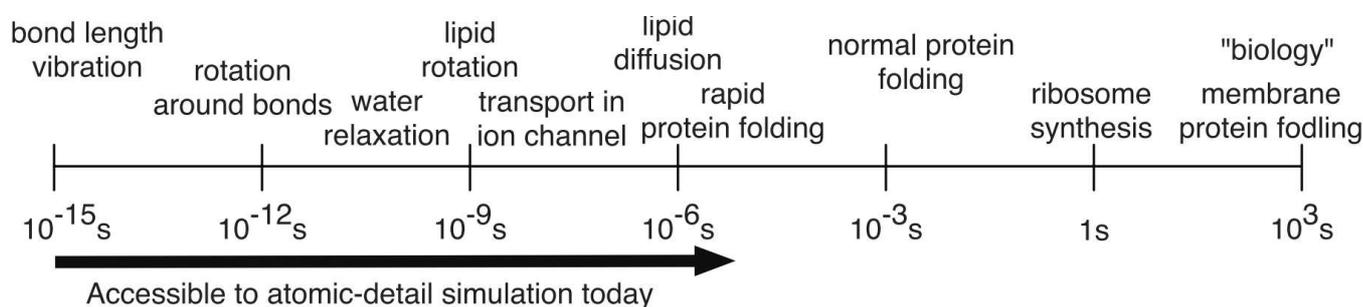


Figure 2.2: Time scale of motions in a protein. Figure from [145]

### 2.3.1.1 Run preparation steps

This describes the steps necessary to setup a simulation run with the gromacs simulation software.

**Vacuum energy minimization** Molecular dynamic simulations consist in the setting up of a box containing all the atoms of the system. The initialization of the system requires several steps in order to obtain a stable system. These steps include vacuum energy minimization, to avoid steric clashes within the protein that might arise from addition of hydrogen during the conversion of the structure file into a simulation ready file. This step uses a special integrator, the steepest descent minimization, to rapidly reach a relaxed state.

**Periodic boundary conditions setup** Once the system is relaxed, we need to set the peridodic boundary conditions (PBC) for the system. Periodic boundary condition are used to minimize the size of the system and the artifact that can arise from physical boundaries of a close system. The important parameter during the setup of a periodic boundary conditions is the distance the user chooses between the limits of the box and the closest protein atom. This distance needs to be chosen so that no interactions exist between the proteins in two periodic

images, including overlap of hydrating water. This distance is typically chosen to be equal to 1nm. Several types of boxes and periodic boundary can be set, the most commonly used for aqueous protein probably being the "dodecahedron", because it is the one that has the smallest volume for equal distance of the protein to the edge of the box.

**Solvent addition** Once the PBC have been setup properly, the solvent needs to be added to the system. In this step, the user specifies a file containing a box of water molecules of the chosen the water model. This box will be superimposed to the protein, and any water molecule from the water file that does not overlap with a protein atom will be introduced in the simulation box.

**Ions addition** One important condition for proper simulation setup is the electrostatic neutrality of the system. In this step, ions will be added to counter the charge of the protein. This is done by replacing randomly water molecules by ions. The user can chose the type of anions and cation that he desires to introduce as well as the concentration (typically 0.1M) and the software adds the ions to counter the charge of the protein and reach the desired concentration.

**Solvated system energy minimization** Added solvent and ion molecules may have caused some steric clashes, or the ions of same charges to close in space. To avoid crashes due to too fast motion of atoms, a second minimization using the steepest descent minimization algorithm is performed on the solvated system.

**Position restrain MD** Now the system biggest constraints have been dissipated by the energy minimization, but the solvent is yet to "adapt" to the protein. The solvent files used for solvent addition often are quite small with a high symmetry, and thus have a unnatural configuration. In this step, we restrain the position of the non-hydrogen atoms of the protein and let the solvent move freely until it is totally relaxed.

**Introduction of temperature coupling** In this step, we introduce a temperature coupling and let the system reached the desired temperature. This is typically done using the Berendsen thermostat, that may not reproduce a correct thermodynamic ensemble, but equilibrates faster than Nose-Hoover.

**Introduction of pressure coupling** Finally, the system needs to be relaxed to a given pressure, and this step is used to introduce the pressure coupling and letting the system relax to that pressure. In this step, the Berendsen barostat is typically used because of its fast

equilibration, even though it does not yield a correct thermodynamic ensemble as Parrinello-Rahman, which is recommended for the definitive simulation run.

### 2.3.1.2 Force fields

The force fields are used to calculate the force that apply on each atom at each time step of the simulation. For this reason the choice of the force field is extremely important and must be done carefully, as the best choice depends on the type of simulation and the parameter you wish to extract from it. There are several types of force fields, the ones being used in this thesis are all-atoms force fields, and coarse grained force field. The choice of the force field is of particular importance because different force fields reproduce physical properties with different accuracy. In particular the propensity of secondary structure can vary significantly from one force field to the other[146].

### 2.3.1.3 Water models

In addition to the choice of the force fields, the choice of the water model is equally important for explicit solvent simulations. For example, the tip4p water model that is a four sites water model, with an extra dummy atom located near the oxygen in order to better reproduce the distribution of charges in the protein. This water model is known to better reproduce the density and X-ray measured structural properties of water in the 0 to 77°C[147]. One should keep in mind however that water models are not optimized for high pressure, and thus some deviations between the simulations and experiments can arise from changes in the physical properties of water at high pressure that do not reflect a correct behavior. If this water model one of the better models developed to date, other models can be better adapted, depending on the system that is being simulated. In the study presented in this thesis concerning the Trp-cage, the water model tip3p.

### 2.3.1.4 Integration of the equations of motion

The thermostats seen previously are used to correct the equations of motion to make them closer to the experimental conditions. Here we are going to see how these equations are integrated in the MD simulation to calculate the motions of the atoms. the method used here is called the *generalized leapfrog integration*. To introduce this method we use the following writing[148]:

$$\dot{q}_i = G_{(p,q)} = \frac{\partial \mathcal{H}}{\partial p_i}, \quad \dot{p}_i = F_{(p,q)} = -\frac{\partial \mathcal{H}}{\partial q_i} \quad (2.3.1)$$

Where  $\mathcal{H}$  is the Hamiltonian describing the system, and can be expressed in the standard case as :

$$\mathcal{H} = \sum_i \frac{p_i^2}{2m_i} + U(q) \quad (2.3.2)$$

Where  $p_i$  and  $q_i$  are the canonical coordinates of the atom  $i$ , which represent its position and conjugate momentum, respectively. To describe the evolution of the system, these quantities are "updated" through a leap-frog algorithm.

The methods consist of three steps. The first step is the calculation of the momentum at half a time step. The second step uses this result to calculate the position after a full time step and recalculate the forces at the new position, then the last step move the momentum to the full time step. This algorithm can be written[148]:

$$\begin{aligned} p_{(t+\frac{1}{2}\Delta t)} &= p_{(t)} + \frac{1}{2}\Delta t \dot{p}_{(p_{(t)},q_{(t)})} \\ q_{(t+\Delta t)} &= q_{(t)} + \frac{1}{2}\Delta t \left( \dot{q}_{(p_{(t+\frac{1}{2}\Delta t)},q_{(t)})} + \dot{q}_{(p_{(t+\frac{1}{2}\Delta t)},q_{(p_{(t+\Delta t)})})} \right) \\ p_{(t+\Delta t)} &= p_{(t+\frac{1}{2}\Delta t)} + \Delta t \dot{p}_{(p_{(t+\frac{1}{2}\Delta t)},q_{(t+\Delta t)})} \end{aligned} \quad (2.3.3)$$

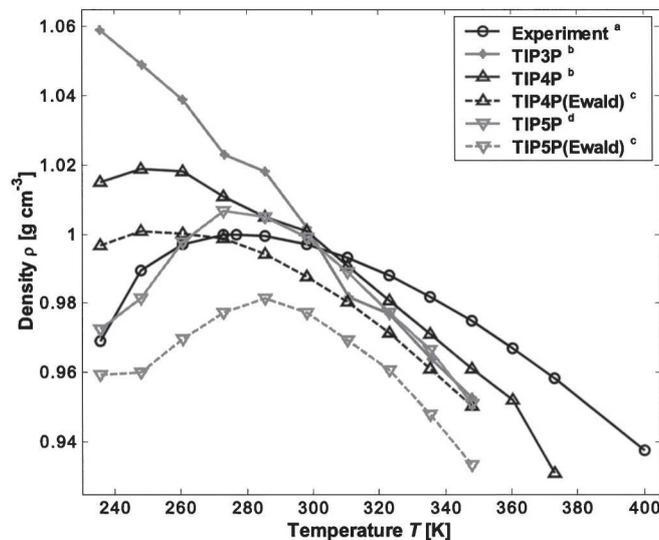


Figure 2.3: Water density as a function of temperature for different water models. Figure from [147]

### 2.3.1.5 Thermostats

One substantial problem with the MD simulation is the fact that it relies on a system that is intrinsically canonical (for the NVT simulations used in this thesis) because the periodic boundary conditions impose that there is no change in the volume or number of particles in the system. In a normal experiment, the temperature is controlled, which means a transfer of energy. Due to the constraints of the MD system, such a change in the system energy cannot be introduced in a model. To correct this, one must use thermostat. The role of these thermostats is to scale the velocity of the particles in the system in order to get closer from a correct canonical distribution observed at a given temperature.

**Berendsen thermostat** Berendsen thermostat was introduced in 1984[149]. It uses a weak coupling to a thermostatic bath, and corrects the kinetics to a defined value, suppressing the fluctuations and hence does not lead to a correct canonical approximation of the system. In this thermostat, the temperature is corrected during a  $\tau$  time scale according to :

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (2.3.4)$$

The change of temperature is realized through re-scaling of the speed vector. This thermostat should not be used for the simulation, but the short relaxation time for the temperature makes it a good thermostat for the relaxation of the system when introducing the temperature.

**Nose-Hoover** The Nose-Hoover Thermostat was first introduced by Nose in 1984 and then modified a year later by Hoover [150, 151, 152]. In this thermostat, we introduce a "heat bath" variable with its own momentum ( $p_\xi$ ) and equation of motion. That leads to an equation of motion for the particles written[153] :

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \frac{p_\xi}{Q} \frac{dr_i}{dt} \quad (2.3.5)$$

Where the equation of motion for the heat bath ( $\xi$ ) is written :

$$\frac{dp_\xi}{dt} = (T - T_0) \quad (2.3.6)$$

The  $Q$  parameter is the determinant for the strength of the coupling. This system allows for small variation in the temperature, of the system, which enables the sampling of a correct canonical ensemble. This thermostat is not implemented in the version of gromacs used in the

chapter 6 of this thesis for replica exchange.

**Langevin Dynamics** Langevin dynamics are special type of thermostat designed for stochastic integration. In this approach, the acceleration of a particle at an instant  $t$  is model by the equation

$$m\ddot{R}_{(t)} = -\nabla U_{(R)} - \gamma m\dot{R} + C_{(t)}\sqrt{2\gamma mk_b T} \quad (2.3.7)$$

Where  $m$  is the mass of the particle,  $R$  its position in space,  $U$  the potential energy at position  $R$ ,  $\gamma$  the drag coefficient,  $T$  the temperature, and  $C$  a Gaussian distributed random number. In this equation, the drag is representing the viscosity of the solvent, slowing the particle down, and the square root term with the random number introduce stochastic motion to the particle. The choice of the value of the drag determines to what extent the system has a "memory" of the past steps. If the drag is chosen to be high, the speed of the particle will depend exclusively on the stochastic term, however, if it is low, the speed of the particle will be strongly correlated from step to step. This is the thermostat that was used in the study presented in this thesis, for both the Structure Based Modeling and all-atom simulations.

### 2.3.1.6 Barostats

To maintain pressure to a stable value in MD simulations, barostat can be introduced. There are two main types of barostats used for pressure control in MD simulations. The Berendsen barostat is equivalent to its thermostat counterpart, and is typically used for equilibration of the pressure during the preparation of the system. The second, more accurate is the Parrinello-Rahman. We will not go into details into these thermostats as none of the work presented here used a pressure coupling, the implicit solvent simulation do not require pressure coupling, and the all-atoms simulations presented in the Tryptophan paper are in NVT ensemble, meaning that it is the volume, and not the pressure that is being kept constant.

**Berendsen barostat** The Berendsen barostat is the simplest barostat used for pressure control in molecular dynamic simulations. It uses the same equation for the equilibration of pressure as its thermostat counterpart does for temperature. In this algorithm however it is not the intensity of the speed vector that is rescaled in order to obtain a correct pressure, but the length of the box vector that are changed, meaning the distance between all atoms is rescaled to obtain the correct pressure[145].

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau} \quad (2.3.8)$$

As for the Berendsen thermostat, this barostat does not reproduce a correct NPT ensemble, but the average pressure is correct and the equilibration fast, and it is thus suitable for the equilibration of pressure during the steps where the system is prepared for the run.

**Parrinello-Rahman barostat** The Parrinello-Rahman barostat uses a more complex approach to better reproduce the NPT ensemble in the simulation. This approach is similar to the Nose Hoover approach for temperature coupling, but in this case, the size of the box vectors is re-scaled to such that :

$$\frac{d^2b}{dt^2} = VW^{-1}b'^{-1} (P - P_{ref}) \quad (2.3.9)$$

Where  $V$  is the volume of the box,  $W$  is a matrix parameter that determines the strength of the coupling, and  $b$  is the matrix representing the box vectors[154]. The matrices  $P$  and  $P_{ref}$  are the current and reference pressures, respectively. This is used to modify the equation of motion through :

$$\frac{d^2r_i}{dt^2} = \frac{F_i}{m_i} - M \frac{dr_i}{dt} \quad (2.3.10)$$

Where :

$$M = b^{-1} \left[ b \frac{db'}{dt} + \frac{db}{dt} b' \right] b'^{-1} \quad (2.3.11)$$

### 2.3.2 Replica exchange

To increase the speed at which one can explore the folding landscape of a protein, Sugita et al.[155] adapted an algorithm called the replica exchange molecular dynamics (REMD) method from the parallel tempering Monte Carlo method[156, 157]. In this algorithm, a system is simulated at different temperatures (including very high temperatures) in order to increase the probability of crossing high energy barriers. The temperatures are such that each replica has a distribution of potential energy that is overlapping that of the following temperature. This overlap is defined as to generate an acceptance ratio. The bigger the overlap, the more probable the exchange. In order to guarantee a correct sampling, one must choose temperatures such that the overlap in energy distribution is constant for all neighboring replicas (figure 2.4). During the simulation, the algorithm will attempt exchange at specific time intervals between two neighboring replicas. The time interval is typically thought of as needed to be sufficient to

allow no auto-correlation in a structure between two attempts. However, it has been shown that shorter interval times are beneficial because they increase the sampling of the conformational space without affecting the canonical distribution of the ensemble[158]. The probability of exchange is determined by the acceptance criteria :

$$P_{(1\leftrightarrow 2)} = \min \left( \exp \left[ \left( \frac{1}{k_B T_1} - \frac{1}{k_B T_2} \right) (U_1 - U_2) \right], 1 \right) \quad (2.3.12)$$

Where  $U$  is the instantaneous potential of each replica, and  $T$  their temperature. This equation defines the Metropolis criterion for parallel tempering. Looking at the equation we can notice that since the temperature is constant with relatively low variations, only the difference in potential energy has an influence to change the probability of exchange of two replica between two tests. The higher the difference of potential energy, the lower the exchange probability. However, the criterion is designed so that lower temperature replicas that have a higher potential energy than their upper neighbors will automatically exchange. After exchange, the velocities are scaled by  $\left(\frac{T_1}{T_2}\right)^{\pm 0.5}$  where the sign depends on the direction of the exchange, and a neighbor search is performed the next step. To avoid artifacts due to the exchange of all neighbors at the same time. This is because if an exchange is performed between replica 1, and replica 2, then the exchange probability not only depends on the energies of replicas 2 and 3, but also on the energy of replica 1. To avoid this problem, the gromacs algorithm performs exchange only for half the neighbors at a time such that an exchange attempt is performed for example at a timestep of 1000ps between replica 1 and 2, and replica 3 and 4 etc. . and exchange attempts between replicas 2 and 3 or replicas 5 and 6 will be performed at a timestep of 2000ps[154].

Replica exchange methods can be implemented with a way to control pressure, however that was not implemented in our study[159].

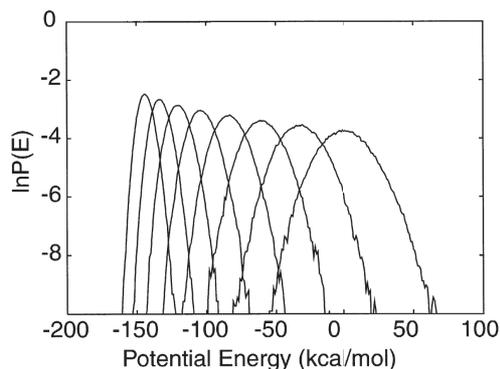


Figure 2.4: The canonical probability distributions of the total potential energy of Met-enkephalin obtained from the replica-exchange MD simulation at the eight temperatures. The distributions correspond to the following temperatures from left to right : 200, 239, 286, 342, 409, 489, 585, and 700 K. [155]

### 2.3.2.1 Simulation conditions for the Trp-cage paper

The tryptophan cage variant Tc5b was simulated using a capped version (sequence Ac-NLYIQWLKDGGPSSGRPPPS-Nme) with charged Lysine, Arginine and Aspartic acid side chains. One  $\text{Na}^+$  and two  $\text{Cl}^-$  ions were added to the 2635 TIP3P water molecules[160] to neutralize the system. Electrostatic interactions were modeled using Particle Mesh Edwald (PME) integration method with a cubic  $36 \times 36 \times 36$  grid and a Van der Walls integration cut-off of at 1.0nm. A grid size of 0.12nm was used for the PME. The equations of motion were integrated using the stochastic dynamics with a coupling time of 1ps, and using a 2fs time frame. Simulations are done using GROMACS and the ff99SB force field[161]. REMD simulations are done at constant volume in a cubic box of 4.42 nm corresponding to the volume of the system at pressure of 1atm, and temperature of 300K, obtained from a 10ns NPT (constant particle number, pressure and temperature) simulation. We simulated 40 systems with temperatures 280.0, 284.1, 288.2, 292.4, 296.7, 301.1, 305.6, 310.2, 314.9, 319.7, 324.6, 329.6, 334.7, 340.0, 345.4, 351.0, 356.6, 362.5, 368.4, 374.6, 380.9, 387.3, 394.0, 400.8, 407.8, 415.1, 422.5, 430.1, 438.0, 446.0, 454.3, 462.8, 471.6, 480.6, 489.8, 499.3, 509.0, 519.0, 529.2, and 539.7 K. Temperatures were selected such that an exchange rate of 0.15 is obtained. Simulations were extended to  $1\mu\text{s}$  per replica, and the last  $0.5\mu\text{s}$  were used for analysis. We also simulated the system at high average pressure over the same temperatures (in a cubic box of 4.176 nm). The volume was reduced such that the average pressure at 310 K is 400 MPa. The initial state of the system was obtained from a 10 ns NPT simulation. The REMD simulation was extended for  $1\mu\text{s}$  per replica and the last  $0.5\mu\text{s}$  are used for calculating averages[162].

### 2.3.3 Principal component analysis

Principal component analysis is a way of extracting correlations in the position of atoms in the protein. Using gromacs, the user needs to input a trajectory using the command `gmx covar` which will output the eigen vectors of the largest eigen values, after fit to a reference structure. In this case the eigen values correspond to the amplitude of the collective motions (here motion is relative to the reference structure), while the eigenvector is the direction of that motion. The user then uses the analysis program `g ana eig` with the eigenvalue and eigenvectors files as an input. Note the files are organized in order of decreasing eigenvalues, meaning the largest collective motions are first in the files. The user then just has to specify how many principal components he wants to have as an output, by using the option `-filter`. The output can be visualized with a normal protein visualization software and can be used to determine the coordinate that suits the best the type of motion that one wishes to visualize. In the procedure used, the C- $\alpha$  the correlated structure are organized by their occurrence probability. This allows for easier finding of good reaction coordinate that separate the population in distinct states.

### 2.3.3.1 Use in the Trp-cage study

In the tryptophan cage study, PCA was used in order to find a reaction coordinate that would allow to separate distinct populations, after secondary structure analysis revealed that the population that we could distinguish using Rg as a second reaction coordinate were actually a mix of states with different structural properties. The first eigenvector corresponding to the largest eigenvalue corresponded to an increase distance between the glycine 10 and serine 14 residues C<sub>α</sub> atoms and allowed to clearly distinguish two structurally different ensembles in state 2. Consequently, this distance was chosen as a secondary reaction coordinate.

## 2.4 Φ-value analysis

Φ-value is a quantity often used to characterize the nature of the transition state of a two-state protein. It corresponds to the relative energy changes due to mutation. Giving that the stability of the transition state has an effect on the global kinetics of the unfolding process, we use the difference kinetic of the process to characterize the effect of the mutation on the structure of the transition state.

$$\Phi = \frac{\Delta G_W^{TS-U} - \Delta G_M^{TS-U}}{\Delta G_W^{F-U} - \Delta G_M^{F-U}} = \frac{\Delta \Delta G^{TS-U}}{\Delta \Delta G^{F-U}} \quad (2.4.1)$$

Where  $\Delta G^{TS-U}$  represents the energy difference between the transition state and the unfolded state for the wild type (*W*) or mutant (*M*) protein, and  $\Delta G^{F-U}$  the energy difference between the folded and unfolded state. A qualitative explanation of this value is that when a mutation is made, the structure around the mutation is affected in terms of energy. What is important is the relative change of energy of the folded and transition states in regard to the unfolded state. If the mutation affects the stability of both the folded and transition state, it shows that the structure of the mutated residue is the same as in the folded structure, if not, there is a change in this structure.

If Φ-value is close to one, the mutation affects the stability of the folded and transition state equally which means that the transition state involves no change in the structure to which the mutated residue belongs.

If Φ-value is close to zero, it then means that the mutation affects the transition state more than it does the folded state. In this case, it means that the structure to which belongs the mutated residues is, at least partially, unfolded in the transition state.

Thought it has become a very useful tool in the characterization of the transition state ensemble, there are several limitations to this approach. First, it cannot differentiate multi-pathways transition from a partially unfolded structure for values between 0 and 1. The model also makes the assumptions that the unfolding pathways are not altered by the mutation. For this reason, the mutations made must be conservative, meaning the global properties of the mutated residues must be close to the wild type residue. At last, it assumes that the energy difference between mutated protein and wild type protein is large enough, otherwise the unfolding of the structure cannot be deduced, and that the mutation does not stabilize the structure to which it belongs. This strong assumption make the interpretation of this value more hazardous, hence a lot of mutants must be made for the validation of a transition state.

### 2.4.1 Activation energy

In this section, the activation energy described as if it was accessed through pressure jump kinetic experiments. The activation energies are equivalent to the activation volume multiplied by the pressure to which the jump is made. Activation volume is a determinant factor for protein unfolding. The sum of the absolute value of folding and unfolding activation volume is equal to the absolute total volume change of the unfolding reaction, as activation volume represent the difference of volume of the protein with regard to the total volume change upon the reaction of unfolding.

$$\begin{aligned}\Delta V_u &= \Delta V_u^\ddagger - \Delta V_f^\ddagger \\ \Delta V_f &= \Delta V_f^\ddagger - \Delta V_u^\ddagger\end{aligned}\tag{2.4.2}$$

An activation volume close to the volume of the folded state means that the volume of the rate limiting transition state is closer to that of the folded state. Activation volume is a determinant parameter for the folding an unfolding rate under pressure. The folding rate of this reaction writes :

$$\begin{aligned}k_u &= k_u^0 \exp\left(\frac{-P\Delta V_u^\ddagger}{RT}\right) \\ k_f &= k_f^0 \exp\left(\frac{-P\Delta V_f^\ddagger}{RT}\right)\end{aligned}\tag{2.4.3}$$

Hence, the activation volume can be deduced from the kinetics of the folding/unfolding reaction. In practice thought, one cannot directly access the individual rates as they are both merged into one single observable which means the decrease of the observable follows

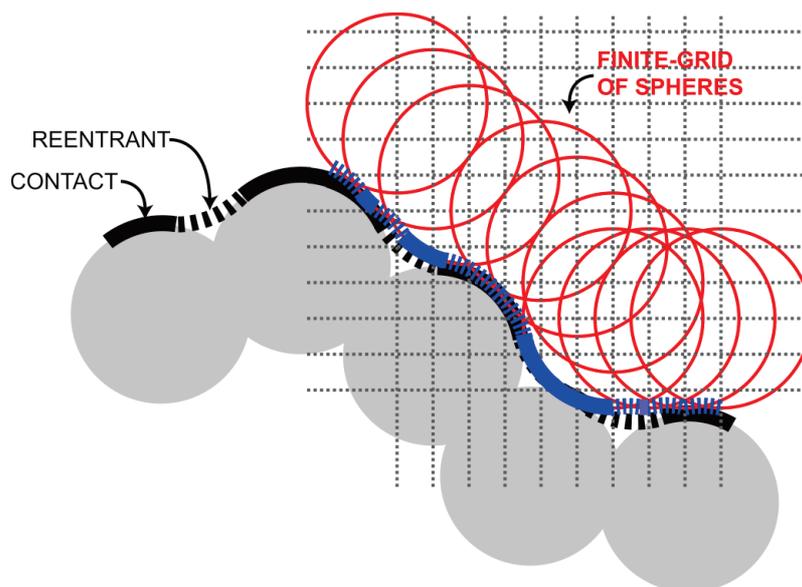


Figure 2.5: Representation of the Hollow algorithm method. The finite-grid approximation in HOLLOW results in a surface (blue) that closely approximates the molecular surface. Figure taken from [163]

$\exp(t(k_u + k_f))$ . The characteristic time  $\tau$  for this exponential is then  $\tau = \frac{1}{k_u + k_f}$ . To access the standard reaction rate, we use the plot of  $\ln \tau$  over the unfolding pressure range. This gives us a chevron plot in the mean pressure of unfolding is the top of the plot. From this plot we can extract the standard reaction rates that are necessary to the determination of the activation volumes of the folding and unfolding reaction, as :

$$\begin{aligned} \lim_{P \rightarrow 1} \frac{1}{k_u + k_f} &= \frac{1}{k_f} \\ \lim_{P \rightarrow \infty} \frac{1}{k_u + k_f} &= \frac{1}{k_u} \end{aligned} \quad (2.4.4)$$

Because  $K_u = \frac{k_u}{k_f} = \frac{[U]}{[F]}$  goes from 0 to  $\infty$ , when the mid-point is passed for long enough,  $k_u$  becomes negligible compared to  $k_f$ . That means the far end of the chevron plot is driven by  $k_u$  and the very beginning by  $k_f$ . This allows the deduction of the number of configurations of the transition state ensemble, assuming each configuration has a different volume, and thus a different rate. A similar procedure can be made for any denaturation method, like temperature or urea.

### 2.4.2 Hollow : cavity visualization algorithm

To visualize cavities, the software Hollow was [163] was used. This software is a simple python program to be used in combination with pymol in order to make a visualization of the void spaces inside a protein. The original purpose of this software was to apply its algorithm to ions channels. It uses a simple procedure to generate good approximations of the position of the cavities. The algorithm identifies void volumes by creating a grid of 0.5Å resolution around the protein. In order to refine the protein's surface, a surface probe of 8Å is placed at every point of the grid. If the probe can be placed at a point in the grid without entering in contact with one of the atoms of the protein, the point is defined as outside. Once all the points have been tested, the second step does a similar work with a smaller probe only on the grid points that have been defined as inside. The size of the second probe is typically set to 1.2Å, which is an approximation of the radius of water. After the points defining void volumes have been defined, the software outputs a pdb file that is to be loaded on pymol. The file contains a list of points that correspond to points in the grid where the probe could fit. From there the user select the groups of points he wishes to visualize. Some of the clusters may not be desirable because they contain an open path to the exterior of the protein, or because they are too small and that small fluctuations will likely change their shapes and volumes, therefore, a user usually selects the most significant clusters only. From there we simply generate a surface on each of the clusters selected. If necessary, the surface can be smoothed for better rendering.

### 2.4.3 SMOG : Shadow contact algorithm

The determination of the folded contact map is a crucially important step for structure based modeling simulations preparation. In the PP32 study, we have used the Shadow contact map algorithm from SMOG webserver[164]. This algorithm is more suitable to define native contacts than simple cut-off algorithm, because it incorporates a way to discard contact between atoms that are within the cut-off but hidden from one another by another atom. That procedure enables to only account for "real" contacts, that have an actual, physical contact between them in the folded state. Although the presence of folded specific bonding is not taken into account, this approximation is usually sufficient for the implementation of structure based modelling algorithms[164].

---

<sup>1</sup>Because in any exponential  $\exp(Ax)$ , the characteristic time is  $\frac{1}{A}$

## 2.4.4 Structure Based Modeling

Structure based modeling is a fast way of simulating folding events through biasing the simulation by making the folded state from a known structure, have the lowest energy. This is realized by associating potentials to the native contacts. These potentials are made inversely proportional in magnitude to the distance they have in the native structure. In our study we use a  $C_\alpha$  coarse grain model for the protein. In this model, each of the amino acids is represented by a single bead, centered around the position of the  $C_\alpha$  in the protein. The algorithm also exists for all-atoms models, and in its newer version can also support amide nitrogen coarse grain model[165]. The structure based modeling algorithms are implicit solvent model, meaning that no solvent molecule is physically represented in the simulation. Thus, the introduction of temperature has to come from stochastic dynamics (also called Langevin dynamics, see chapter 2).

### 2.4.4.1 Reduced units

As a consequence of the implicit solvation and the simplification assumed in a structure based model, the scales for the basic units are said to be reduced, in order for them to be self-consistent. In the SBM simulations used here, the length scale from the pdb structure are

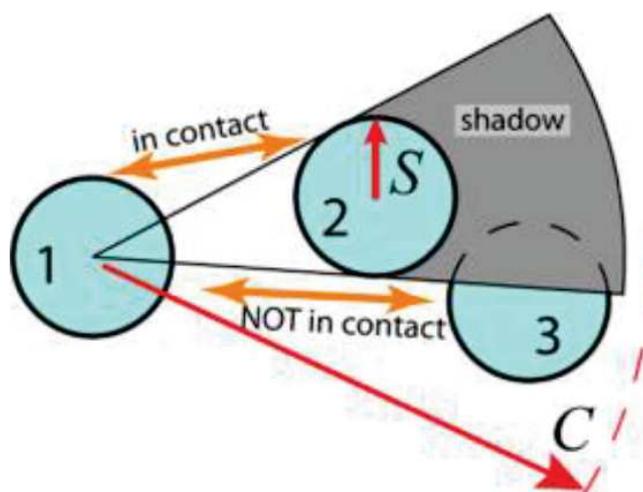


Figure 2.6: The Shadow contact map screening geometry. Only atoms within the cutoff distance  $C$  are considered. Atoms 1 and 2 are in contact because they are within  $C$  and have no intervening atom. To check if atoms 1 and 3 are in contact, one checks if atom 2 shadows atom 1 from atom 3. The three atoms are viewed in the plane, and all atoms are given the same shadowing radius  $S$ . Since a light shining from the center of atom 1 causes a shadow to be cast on atom 3, atoms 1 and 3 are not in contact. Figure taken from [164]

kept in nm, but the time, mass and energy scales are "free". As a result, for example, the timestep used for the potential calculation is of 0.0005 ps instead of 0.002 typically used in classic all atom simulations, because the time units are not "real", and thus the "real" unit is longer than specified in ps in the mdp file.

Temperature is also not expressed in "real" units in the SBM simulation. Therefore, determination of the correct temperature to reproduce the ensemble present in the data is up to the user appreciation. In order to get as close as possible to the data used for the introduction of the constraint, temperature was determined by trying to match, within 3%, the proportion of frames that contain less than <sup>1</sup> of native contacts.

### 2.4.5 Pulchra

To get a better visualization of the intermediates in the folding landscape, we use a reconstruction algorithm called pulchra to map the protein back to an all-atom representation[166]. This software is specifically designed to go back from a coarse grain representation of a protein to a full-atom model for simulations. The procedure of the software consists in a step by step reconstruction of the missing backbone and sidechain atoms. The first step is a steepest-descent minimization to adjust the position of the  $C_\alpha$  atoms. The backbone nitrogen and carbonyl groups are then added. Optionally, an optimization of the hydrogen bond pattern of the backbone can be realized, using the definition found in the DSSP[167] program and a minimization of the energy of the  $C - O \cdots H - N$  hydrogen bonds using the formula :

$$E_{HB} = \frac{q_1 q_2}{r_{ON} + r_{CH} - r_{OH} - r_{CN}} \cdot 332 \quad (2.4.5)$$

Where  $q_1 = 0.42e$  and  $q_2 = 0.20e$ , with  $e$  being the electron charge unit and  $r_{XY}$  is the distance between the atoms X and Y in angstroms, and  $E_{HB}$  the energy of the hydrogen bond in kcal.mol<sup>-1</sup>. The backbone hydrogens are thus reconstructed at this stage. The side chain heavy atoms are then reconstructed. The system is then optimized for side chain steric clashes. This is done by classifying each side chain by order of the number of steric clashes they have, a steric clash being defined as a distance between two atoms is less than 2Å. The algorithm then iterates through the side chains in that order testing rotamers from an internal library to minimize the number of steric clashes. If steric clashes are still present, the  $C_\alpha$ - $C_M$  vector, where  $C_M$  is the center of mass of the side chain, is rotated, as illustrated on figure 2.7. Finally, the side chain hydrogens are added. Each of the steps of this procedure is optional. Some of the steps were skipped, because the output was only needed for visualization. This resulted in some residues not having a formally correct backbone representation.

---

<sup>1</sup>Check again the actual number

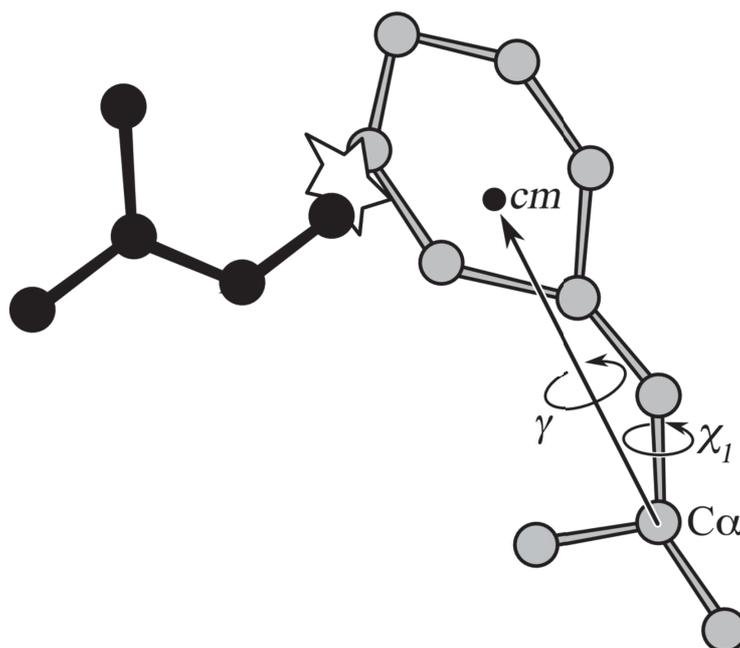


Figure 2.7: Optimization of side chain positions. The side chain is rotated by an angle  $\gamma$  around a  $C_\alpha$ - $C_M$  vector until the total number of clashes with other heavy atoms is minimized. Additionally, the  $\chi_1$  angle is calculated and tested against the allowed range to exclude nonphysical side chain conformations. Figure taken from [166]

## 2.4.6 Presentation of the calorimeter

In this section an introduction to differential scanning calorimetry and pressure perturbation calorimetry is made. The material used for both these experiments is essentially the same, a VP-DSC micro-calorimeter from Microcal[168], only a pressure controller is added for the pressure perturbation calorimetry.

The device is composed of two cells in which are placed the reference solution (buffer) in one, and the sample solution in the other. Both cells are surrounded by a thermic shield making them adiabatic. The temperature inside them is controlled by a computer in an independent way. In both experiments, the computer is used to re-equilibrate the temperature in cells. It is the difference of power used for this re-equilibration is recorded, and used as the output data of the experiments. The integration of this power over time yields the heat consumed or released by the sample[169].

All samples were prepared from lyophilized protein, at least two dialyses of 400mL were made in a  $\text{Na}_2\text{HPO}_4$  of 50mM concentration at pH7, and were centrifuged at low temperature to avoid aggregates before and after dialyses. For both experiments, all solutions must be

degassed before put in the cells so that no air bubble appear during the heating process. The protein's concentration must be measured directly from the syringe used to introduce the sample and entered into the software to calculate the molar heat capacity of the protein.

#### 2.4.6.1 Differential Scanning Calorimetry

In Differential Scanning Calorimetry, the temperature is raised in a continuous way, and the power needed for the active compensation of temperature of the sample cell with regards to the reference is recorded. A pressure cap applying 1.8 bar in both cell is used to lower the risk of formation of air bubbles during the heating process. In order to avoid artifacts from the buffer and/or the machine, a first run is done in which both cells are filled with buffer solution. For the second run, a solution containing protein in a concentration range of  $60\mu\text{M}$  to  $90\mu\text{M}$  is put into the sample cell, while the reference cell is still filled with buffer solution. The scan rate is of  $1^\circ\text{C}/\text{min}$ .

#### 2.4.6.2 Pressure Perturbation Calorimetry

For Pressure Perturbation Calorimetry, a computer controlled pressure controller is added to the device instead of the pressure applying cap[170]. This controller allows instant pressure jumps of 5 bars in both cells at the same time. In this experiments temperature equilibrium is reach in both cells, then the software applies a pressure jump. Heat is then given or taken to reach equilibrium at the same temperature than before the pressure jump. The difference of heat flow between the two cells is then recorded during the set re-equilibration time. The pressure then returns to ambient pressure, and the compensation process is recorded again. When this last equilibrium is reach, the software heats or cools both cells to the next desired temperature and this the same process is repeated for as many temperatures as the experiment requires. The concentration of protein required for the sample run of this experiment is between 5 mg/ml to 6 mg/ml. For all PPC experiments, a DSC scan is previously done to know the position of unfolding peak. The temperature steps are of  $2\text{-}3^\circ\text{C}$  outside the unfolding peak, and  $1^\circ\text{C}$  in the peak, for a better resolution during the unfolding transition.

#### 2.4.7 Data processing

Both experiments a treated with the microcal software "origin".

### 2.4.7.1 Differential Scanning Calorimetry

Three parameters are accessible with this technique. The melting temperature  $T_m$ , the difference of heat capacity between the folded and unfolded state  $\Delta C_p$ , and the total heat consumed by the unfolding,  $\Delta H_m$ .

The output of the DSC experiment is on the difference of heat flow between the sample cell and the reference cell ( $\Delta_{cells}dQ$ ) required for the active temperature compensation, being the actual heat flow consumed by the solute. The user indicates the concentration of solute of the sample in the cell, from which the software deduces the molar calorific capacity of the protein from :

$$\begin{aligned} dQ &= [P]V_{cell}C_p dT & (2.4.6) \\ \Rightarrow C_p &= \frac{dQ}{[P]V_{cell}dT} \end{aligned}$$

The difference of heat capacity is simply given by the difference of value of the plateau before and after the unfolding transition :

$$\Delta C_p = C_p^u - C_p^f \quad (2.4.7)$$

The melting temperature is the value of the maximum heat capacity of the unfolding peak, and  $\Delta H_m$  is the total energy consumed by mole unit during the unfolding transition, and is the result of the integration of  $C_p$  over the limits of the unfolding transition :

$$\Delta H_m = \int_{T_{trans}} C_p dT \quad (2.4.8)$$

To obtain these parameters, the output from the buffer-buffer experiment is subtracted from the output of the sample-buffer. A "cubic connect" is done to set the baseline for integration of the unfolding peak. A Gaussian function is then fitted to the data by a linear regression algorithm, and integration of the area of this Gaussian yields the total heat consumed by the unfolding,  $\Delta H_m$ , while the position of the center of the Gaussian is the melting temperature  $T_m$  (figure 1.7).

### 2.4.7.2 Pressure Perturbation Calorimetry

The output of a PPC experiment consist of a function of the heat flow versus time (figure 2.8). In practice, both cells have a heat compensation due to the fact that both their calorific capacities changes with the application of the pressure jump. However, as in DSC experiments, the recorded heat flow is relative, and corresponds to the difference of power given to the sample cell, comparing to the reference cell. The total energy consumed by the studied specie is then obtain by integrating the heat flow difference over the re-equilibration time ( $\Delta_{cell}\Delta Q$ ). From this value, a coefficient of thermal expansion is deduced using the following equations[170] :

$$dS = \frac{dQ_{rev}}{T} \quad (2.4.9)$$

$$\left(\frac{\partial Q_{rev}}{\partial P}\right)_T = T \left(\frac{\partial S}{\partial P}\right)_T \quad (2.4.10)$$

According to the Maxwell relation  $\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P$ , and as the coefficient of thermal expansion is written  $\alpha_V = \frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_P$ , we can write :

$$\left(\frac{\partial Q_{rev}}{\partial P}\right)_T = -T \left(\frac{\partial V}{\partial T}\right)_P = -TV\alpha \quad (2.4.11)$$

Then we can integrate  $dQ_{rev}$  on pressure :

$$\begin{aligned} \int_P dQ_{rev} &= - \int_P TV\alpha dP \\ \Delta Q &= -\alpha TV \Delta P \\ \Rightarrow \alpha &= -\frac{\Delta Q}{TV_{cell}\Delta P} \end{aligned} \quad (2.4.12)$$

For the determination of the actual contribution of the protein in the solvent, the contributions from the solute (protein) and the solvent must be differentiated. For this purpose, two sets of experiment must be made. In the first, the solvent of the protein is compared to pure water, and its coefficient of thermal expansion is deduced. The second experiment is the actual measurement of the thermal expansion of the protein, comparing the protein containing sample to the solvent. For both experiments, a prior water-water or buffer-buffer run is made to subtract the effects due to the calorimeter itself.

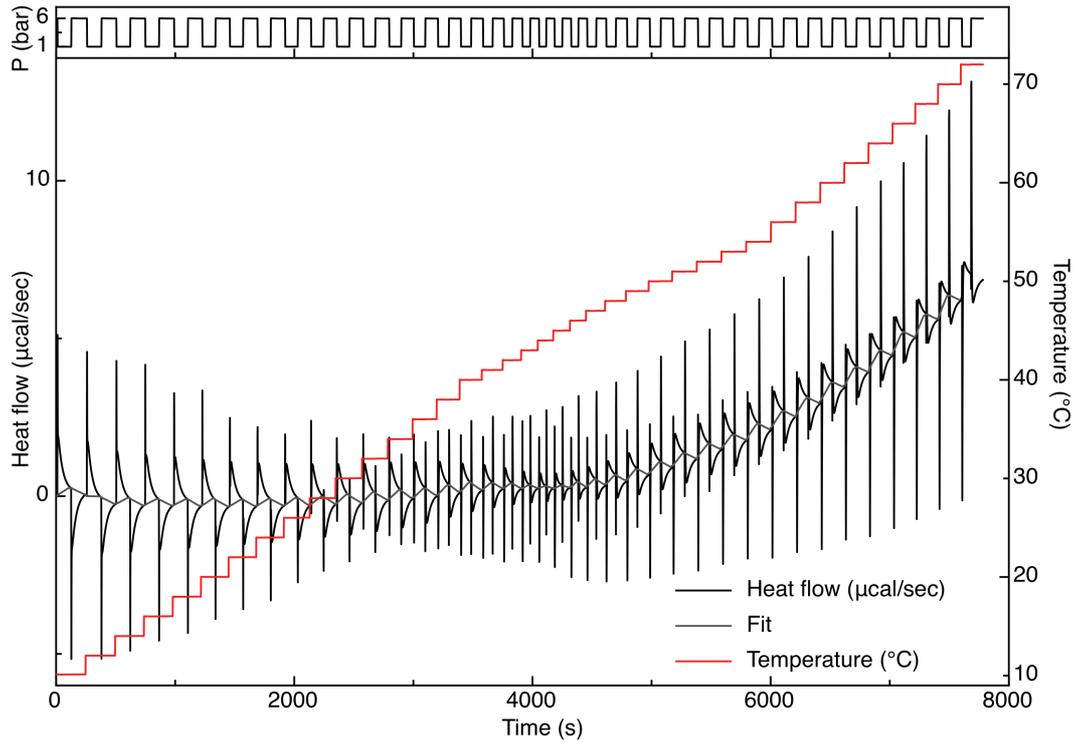


Figure 2.8: The above graph shows the pressure jump versus the time, the red line shows the evolution of temperature (right axis) and the black line is heat flow difference between the two cells (left axis).  $\Delta_{cell}\Delta Q$  is obtained by integrating the area between the heat flow difference and the fit (gray line). This graph was made from I92A-L125A  $\Delta$ +PHS mutant

As the coefficient of thermal expansion is strongly dependent on the number of moles of its species, and as the volume occupied by the protein in the sample cell is occupied by the same volume of solvent in the reference cell, a correction must be made. The total volume is written  $V_{tot} = m_0V_0 + m_s\bar{V}_s$ , where  $V_0$  is the specific volume of the solvent and  $V_s$  the partial specific volume of the solute :

$$\Delta Q = -Tm_s\bar{V}_s(\bar{\alpha}_s - \alpha_0)\Delta P \quad (2.4.13)$$

From which the solute coefficient of thermal expansion is extracted :

$$\bar{\alpha}_s = \alpha_0 - \left( \frac{\Delta Q}{Tm_s\bar{V}_s\Delta P} \right) \quad (2.4.14)$$

For experiments where the buffer is not pure water, as it is the case in ours, the  $\alpha_0$  of the buffer must be extracted before the sample containing the solute of interest is measured.

At last, PPC measurements give access to the change in volume upon the temperature of

unfolding. To obtain this data, a simple integration of the coefficient of thermal expansion on the temperature interval of unfolding determined by DSC is made :

$$\frac{\Delta\bar{V}_s}{\bar{V}_s} = \int_{T_0}^{T_e} \bar{\alpha}_s dT \quad (2.4.15)$$

In most studies, it is not the coefficient of thermal expansion that is displayed, but the molar expansivity,  $E$  :

$$\begin{aligned} E &= \bar{\alpha}_s \bar{V}_s \\ \Delta\bar{V}_s &= \int_{T_0}^{T_e} E dT \end{aligned} \quad (2.4.16)$$

# Chapter 3

## Investigation of PP32 folding landscape through high pressure NMR

### 3.1 Introduction

#### 3.1.1 Literature on PP32

PP32 (Anp32a) is a natural human tumor suppressor repeat protein constituted of 5 repeats and two caps[171]. The protein belong to the class of the Leucine Rich Repeat proteins (LRR), characterized by a consensus sequence with conserved leucines[171]. In this protein, both the

Sequence Name	Sequence	Residues
<b>N-CAP</b>	MEMGRRIHLELRNRTPSD	1-18
<b>Repeat 1</b>	VKELVLD.NSRSNEGK.....LEGLTDEFEE	19-43
<b>Repeat 2</b>	LEFLSTI.NVGLTS.....IANLPK.LNK	44-65
<b>Repeat 3</b>	LKKLELSDNR.VSGG.....LEVLAEKCPN	66-89
<b>Repeat 4</b>	LTHLNLSGNK.IKDLST....IEPLKK.LEN	90-114
<b>Repeat 5</b>	LKSLDLF.NCEVTNLNDYRENVFKL...LPQ	115-141
<b>C-CAP</b>	LYYL <b>DGYDRDDKE</b>	142-154

Figure 3.1: Sequence alignment of the five LRRs of hAnp32 1-154. Conserved hydrophobic residues and asparagines are highlighted in cyan. The conserved hydrogen bond donor and acceptor Y131 and D146 are bolded. The five C-terminal residues included in this study (residues 150-154) are highlighted in yellow. Construct hAnp32A 1-149 is missing the residues highlighted in yellow. Construct hAnp32A 1-145 is missing residues in the red box.

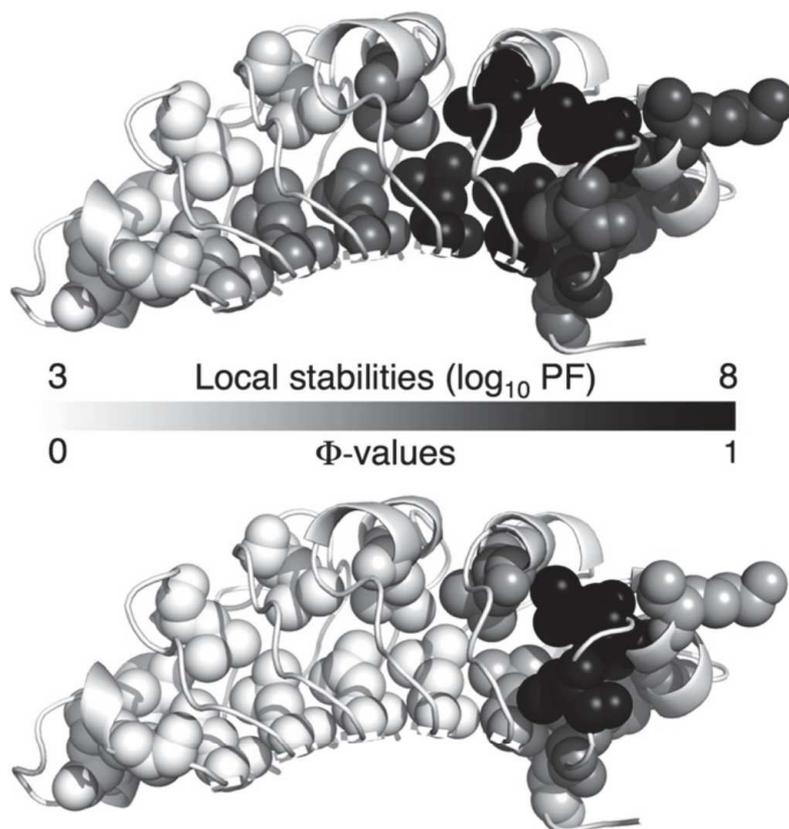


Figure 3.2: Ribbon representation, with  $\phi$ -value substitutions shown as spheres. Residues are shaded from white to black, with black having the highest local stability (Upper) and  $\phi$ -value (Lower). For direct comparison, only sites with both  $\phi$ -value and local stability data are shaded. Based on the coincidence of high protection factors and  $\phi$ -values, folding is initiated at the most stable region of PP32.

length and secondary structure content of the repeats differ from repeat to repeat (figure 3.1), however conserved asparagines, leucines or isoleucines constitute the consensus of this natural repeat protein. The C-terminal part of the protein is also characterized by a conserved hydrogen bond between the C-cap and the last repeat of the protein.

Previous studies by Dao et al. in the Barrick lab had revealed a very polarized landscape, with a very stable C-terminal part[131, 62]. HX exchange of the WT protein showed that the protection factor form a gradient from the fast exchanged N-terminal residues to the very high protection of the C-terminal. Furthermore, a kinetic study of 20 variants with mutations all across the tertiary structure revealed a very unfolded like transition state, with  $\phi$  values close to one located exclusively in the C-terminal last two repeats and C-cap (figure 3.2).

### 3.1.2 Goals and preliminary study

Study of PP32 were realized in collaboration with the Barrick lab at Johns Hopkins University. The intent of the study was to see if we could have a good representation of the folding landscape through NMR by using pressure. PP32 WT has a natural cavity that main part is situated around near the middle of the protein. A preliminary study through fluorescence of its added tryptophan, located just after the C-terminal cap, and 1D NMR had shown that PP32 unfolds in the accessible pressure range (1-2500bar) at relatively low urea concentration (1.4M). The Barrick lab had already assigned the spectra of the 2D HSQC, showing a very good dispersion of the peaks with little overlap. The analysis was mostly done by developing a program for peaks peaking, fitting, and data visualization. The second part of the work consisted in adapting a method developed by Julien Roche[94] for the determination of a folding landscape from constrained structure based modeling simulations. This method was modified in order for the ensemble present in the simulations to better reflect that of the NMR experiment. The method proved to be much more adapted to PP32, because of the significant deviation from two-states this protein revealed, as opposed to the SNase protein of the Roche paper.

### 3.1.3 Deviation from two-state

In this study, the use of pressure of PP32 has shown to result in a non two-state unfolding. However, due the lack of intermediate specific signal, the global deviation from two-state was determined by the local application of a two-state model to the residue specific information provided by two dimensional NMR. Significant changes in both the slope of the unfolding curve ( $\Delta V$ ) and the stability ( $\Delta G^0 = P_{\frac{1}{2}} \Delta V$ ) reveal a significant deviation from a two-states behavior, which would correspond to having all curves to be exactly equivalent, within the margin of error.

# High-Resolution Mapping of a Repeat Protein Folding Free Energy Landscape

Martin J. Fossat,<sup>1</sup> Thuy P. Dao,<sup>2</sup> Kelly Jenkins,<sup>1</sup> Mariano Dellarole,<sup>3</sup> Yinshan Yang,<sup>4</sup> Scott A. McCallum,<sup>5</sup> Angel E. Garcia,<sup>6</sup> Doug Barrick,<sup>2</sup> Christian Roumestand,<sup>4</sup> and Catherine A. Royer<sup>1,\*</sup>

<sup>1</sup>Biological Sciences, Rensselaer Polytechnic Institute, Troy, New York; <sup>2</sup>Department of Biophysics, Johns Hopkins University, Baltimore, Maryland; <sup>3</sup>Unité de Virologie Structurale, Centre National de la Recherche Scientifique UMR 3569, Institut Pasteur, Paris, France; <sup>4</sup>Centre de Biochimie Structurale, Centre National de la Recherche Scientifique UMR 5048, Institut National de la Santé et de la Recherche Médicale, Université de Montpellier, France; <sup>5</sup>Center for Biotechnology and Interdisciplinary Studies and <sup>6</sup>Department of Physics, Rensselaer Polytechnic Institute, Troy, New York

**ABSTRACT** A complete description of the pathways and mechanisms of protein folding requires a detailed structural and energetic characterization of the conformational ensemble along the entire folding reaction coordinate. Simulations can provide this level of insight for small proteins. In contrast, with the exception of hydrogen exchange, which does not monitor folding directly, experimental studies of protein folding have not yielded such structural and energetic detail. NMR can provide residue specific atomic level structural information, but its implementation in protein folding studies using chemical or temperature perturbation is problematic. Here we present a highly detailed structural and energetic map of the entire folding landscape of the leucine-rich repeat protein, pp32 (Anp32), obtained by combining pressure-dependent site-specific <sup>1</sup>H-<sup>15</sup>N HSQC data with coarse-grained molecular dynamics simulations. The results obtained using this equilibrium approach demonstrate that the main barrier to folding of pp32 is quite broad and lies near the unfolded state, with structure apparent only in the C-terminal region. Significant deviation from two-state unfolding under pressure reveals an intermediate on the folded side of the main barrier in which the N-terminal region is disordered. A nonlinear temperature dependence of the population of this intermediate suggests a large heat capacity change associated with its formation. The combination of pressure, which favors the population of folding intermediates relative to chemical denaturants; NMR, which allows their observation; and constrained structure-based simulations yield unparalleled insight into protein folding mechanisms.

## INTRODUCTION

A major challenge in the quest for detailed characterization of protein folding mechanisms is the difficulty of obtaining high-resolution experimental structural information in a site-specific manner across the entire protein sequence and along the folding reaction coordinate. Most folding studies monitor tryptophan fluorescence, which is sensitive to changes in the local structure, or circular dichroism, which provides global information about the secondary structure. Although site-specific information can be obtained from Fourier transform infrared spectroscopy or fluorescence resonance energy transfer probes, such studies require many separate labeled samples and experiments. Moreover their structural resolution is limited (1,2). High-resolution

structural methods such as two-dimensional (2D) NMR, yield site-specific information, but NMR is difficult to implement with high concentrations of chemical denaturant or high temperature. H-D exchange (3–8) by NMR or mass spectrometry has been used to obtain residue-level detail in folding reactions, but these are not real-time techniques and do not monitor the folding/unfolding reaction directly.

Hydrostatic pressure leads to the unfolding of proteins because the molar volume of the unfolded state is smaller than that of the folded state. This decrease in volume arises primarily because the folded structure contains solvent-excluded void volume that is largely eliminated upon unfolding (9,10). This mechanism of pressure perturbation, by which pressure acts on specific nonhomogeneous structural features of the folded state, leads to a higher probability of populating partially folded structures than unfolding by temperature or denaturants, the efficacy of which depends homogeneously on the amount of surface area exposed in the unfolded state. Advances in pressure cell technology (11) have made it possible to routinely perform

Submitted May 23, 2016, and accepted for publication August 26, 2016.

\*Correspondence: [royerc@rpi.edu](mailto:royerc@rpi.edu)

Angel E. Garcia's current address is Center for Non-linear Science, Los Alamos National Laboratory, Los Alamos, New Mexico

Editor: Elizabeth Komives.

<http://dx.doi.org/10.1016/j.bpj.2016.08.027>

© 2016 Biophysical Society.

multidimensional NMR measurements under pressure in a straightforward manner, allowing for real-time observation of reversible unfolding at nearly every residue in the protein. Hence, the combination of high pressure and NMR permits both the population and the direct observation of folding intermediates, providing the highly detailed description of protein folding pathways that is required for understanding folding reactions.

Most small globular proteins generally fold relatively cooperatively in chemical denaturation studies, although intermediates have been detected in some cases (12–14). Unfortunately, the complex tertiary structure of globular proteins, with their numerous sequence distant contacts, renders the quantitative analysis of folding cooperativity extremely challenging. In contrast, the linear topological interactions implicated in repeat proteins are much more tractable. The N-terminal domain of the pp32 protein consists of five leucine-rich repeats (LRR) (15) (Fig. 1) stabilized by capping motifs on both the N- and C-termini (16). Urea-induced unfolding profiles of pp32 exhibited two-state behavior with very high apparent cooperativity (17), although kinetics experiments revealed the transient population of an intermediate (17) and  $\phi$ -value analysis allowed characterization of the barrier to folding (18). Three folding units (or foldons (8)) were defined in these previous urea unfolding studies (namely, the unstable N-Cap-repeat 1-repeat 2 region, the more stable central repeats 3 and 4, and the highly stable region comprised of repeat 5 and the C-cap). This C-terminal region was found to be ordered at the folding barrier; hence the most stable region is also the first to fold.

We demonstrate here that equilibrium high-pressure NMR provides a highly detailed structural and energetic description of the folding/unfolding reaction of this model protein. Coarse-grained structure-based simulations constrained by the NMR data allowed calculation of the structures present and pseudo-free energy profiles of the pp32 ensemble as a function of both pressure and temperature.

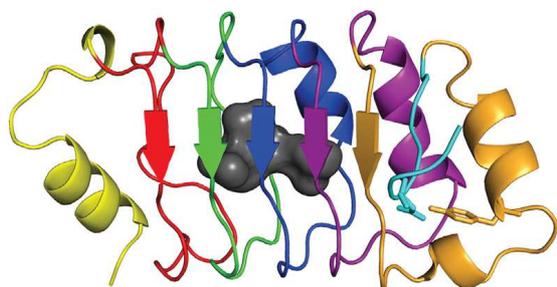


FIGURE 1 Ribbon diagram of the crystal structure of the leucine-rich repeat protein pp32 (PDB: 2JE0) (16). The N-terminal cap is shown in yellow, the first repeat is colored in red, the second in green, the third in blue, the fourth in purple, the fifth in brown, and the C-terminal cap in cyan. The major cavity is displayed in gray. The cavity was visualized using the software HOLLOW (34) with a grid of 0.25 Å and a probe radius of 1.2 Å. Rendered using the PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC. To see this figure in color, go online.

We found that the cooperativity of pp32 unfolding was highest at 293 K, although an intermediate could be populated and characterized at ~900 bar. This intermediate exhibits disorder in the N-terminal region (N-cap and repeats 1 and 2) consistent with previous kinetics measurements (18). The pseudo-free energy profiles revealed that the protein ensemble at the main folding barrier displays a fraction of native contacts,  $Q$ , close to that of the unfolded state. The structures of the transition state ensemble exhibit collapse, and organization in the C-terminal repeat and the capping motif consistent as well, with results from urea  $\phi$ -value analysis (18). A more pronounced deviation from two-state behavior was observed upon increasing or decreasing the temperature, reflecting a significant heat capacity change associated with the formation of the major intermediate.

## MATERIALS AND METHODS

### Protein production and purification

The leucine-rich repeat domain of pp32 (Anp32) was expressed in *Escherichia coli* in minimal media with  $^{15}\text{N}$ -labeled  $\text{NH}_4\text{Cl}$  and purified as previously described in Dao et al. (17,18). The construct included the additional peptide RDDKEWLEHHHHHHH at its C-terminus to provide a His-tag for purification and a tryptophan residue to aid in the determination of protein concentration.

### High-pressure NMR

To permit unfolding of the protein within our accessible pressure range (1–2500 bar), 1.4 M urea was added to the solution. Previous peak assignments of the pp32 LRR domain NMR peaks was confirmed using  $^{15}\text{N}$ -labeled protein, followed by a urea titration to ensure reliable identification of the peaks at the urea concentrations used in the high pressure experiments.

NMR spectra were acquired using a 600 MHz AVANCE III spectrometer (Bruker, Billerica, MA), with a Broad Band inverse probe equipped with z-gradients (Bruker), and using a standard 5 mm O.D. ceramic tube from Daedalus Innovations (Aston, PA). Hydrostatic pressure was applied to the sample directly within the magnet using the Xtreme Syringe Pump (Daedalus Innovations). 2D [ $^1\text{H}$ - $^{15}\text{N}$ ] HSQC spectra were recorded in steps of 200 bar, with a 10 min relaxation time after every pressure change, to allow the protein to reach full equilibrium. Relaxation times for the folding/unfolding reaction (<1 min) were previously obtained by high-pressure fluorescence of the C-terminal tryptophan. Hence, equilibration at each pressure for 10 min largely ensured that equilibrium was reached before data acquisition. Data sets were acquired in this manner for four temperatures (288, 293, 298, and 303 K).

Spectral analysis was performed using CCPNMR Analysis software (19). Maximum peak heights were obtained from 2D Lorentzian fitting. Peaks that crossed other peaks during the pressure titration were eliminated from the analysis. The heights of the remaining peaks were fitted with a two-state unfolding model,

$$S_{\text{obs}} = \frac{S_u + S_f e^{\frac{-\Delta G_f^0 - P\Delta V_f}{RT}}}{1 + e^{\frac{-\Delta G_f^0 - P\Delta V_f}{RT}}}, \quad (1)$$

for the residue-specific apparent equilibrium volume change ( $\Delta V_f$ ) and free energy at atmospheric pressure ( $\Delta G_f^0$ ) of folding using the nonlinear

least-square fitting method from the Scipy Python library (20), from which values were extracted. The low- and high-pressure plateau values were floating parameters in the fit, and the data and fitted values were normalized (after the fit) using these plateau values to yield plots of fraction folded as a function of pressure. Floating the plateau values allows us to take into consideration the experimental error in the high-pressure plateau (rather than assigning it to be zero). Consequently, the high-pressure plateau values can be slightly lower or slightly higher than zero. This procedure also allows for the possibility that the apparent fraction folded at atmospheric pressure for a given residue may not necessarily be 100%. Histograms of apparent  $\Delta V_f$  and  $\Delta G_f^0$  values were fitted to Gaussian distributions.

## Contact maps

Native contact maps were obtained by using the web-server SMOG Shadow contact map (21) with a threshold of 6 Å around the  $C_\alpha$  of each residue, using the pp32 crystal structure (PDB: 2JE0) (16). Using the geometric mean, rather than the joint probability as previously done (10), ensures the correct unfolding profile in the case of two-state unfolding. Probability values were plotted as a heat map in a contact map dot plot. Histograms of the contact maps were obtained by counting each residue involved in the repeat to which it belongs.

## Gō-model simulations

Gō-model (22) simulations rely on information about the lowest energy (native) state to bias the simulation. Using residue-specific fractional contacts obtained from high-pressure NMR, a supplementary bias on the topology was introduced to simulate the conformational ensemble as a function of pressure. The pp32 protein construct used in our studies contained C-terminal residues not present in the crystal structure. Consequently, we used PyMol (Molecular Graphics System, Version 1.8 Schrödinger, LLC) to add those residues to the crystal structure. A nanosecond relaxation of the full construct was performed

using all-atom simulation with an Amber-99sb force field (23) and TIP4P-ew explicit solvent (24) to ensure that no steric clashes were present. The resulting structure was submitted to the SMOG web-server (25) to build a C- $\alpha$  structure-based model (26), using the previously determined SMOG shadow contact map (21). For each pressure/temperature simulated, experimental bias was introduced via 300 parallel simulations in which the contacts present defining the folded state were randomly deleted based on the experimentally determined probability of these contacts being formed, as described in the text. As a result, a specific contact had the same probability of occurring across the 300 topology files as it did in the experimental data. Periodic boundary conditions were used in the simulations, and the size of the box was set to 50 nm in each direction to ensure no self-interaction with the periodic image, even in a fully unfolded state. Frames belonging to a certain  $Q$  range were clustered and collectively analyzed using a Gromacs 454 gmxc cluster command. The obtained configuration was that which exhibited the smallest average root mean-square deviation (RMSD) to all other structures in its cluster. Selected configurations were reconstructed to a full atom model using the software Pulchra (27). Further information concerning the structure-based modeling can be found in the [Supporting Material](#).

## RESULTS

### NMR-detected high pressure unfolding of pp32

$^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$  labeled pp32 were acquired as a function of pressure at four temperatures (Figs. 2, A–C, and S2–S4, A–C, in the [Supporting Material](#)). All resolved backbone amide resonances (~100 total) exhibited pressure-dependent shifts of the native-state resonance frequencies due to compression (28). In addition, we observed a decrease in overall intensity of each native state peak as a function of pressure. Concomitantly we observed an increase in the

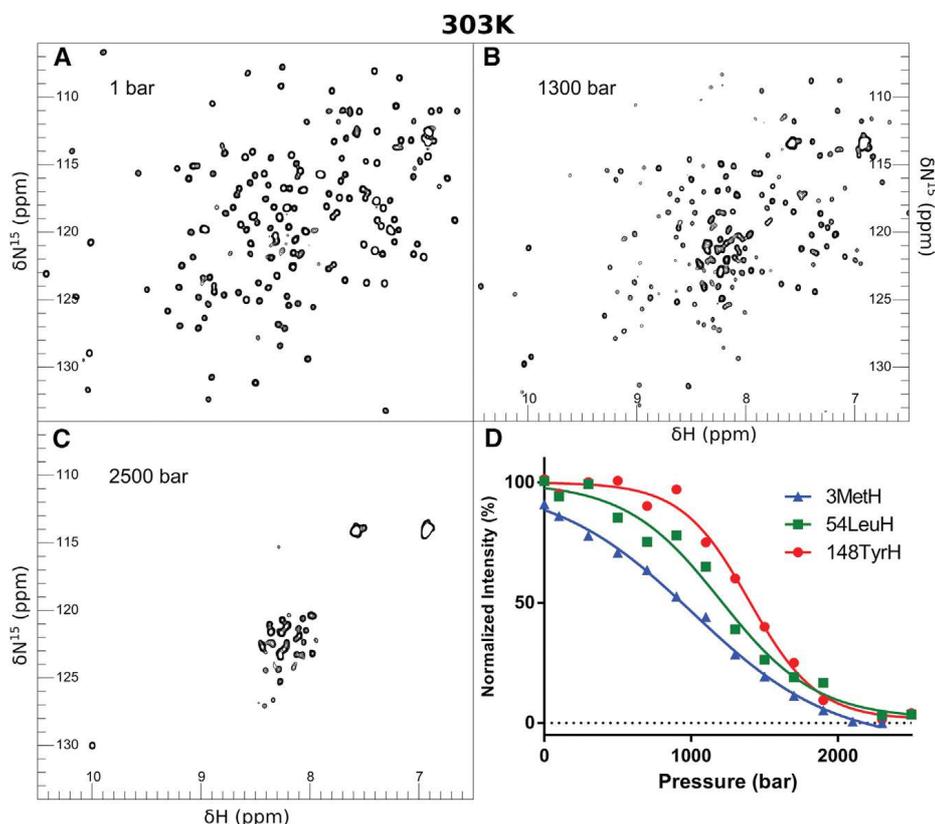


FIGURE 2 NMR detected high pressure unfolding of pp32 at 303 K and 1.4 M urea. (A–C) Examples of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra at different pressures as indicated. (D) Examples of three residues exhibiting distinct unfolding profiles. To see this figure in color, go online.

intensity of the peaks centered between 7.7 and 8.7 ppm in the proton dimension due to population of the unfolded state. Moreover, no broadening of the native state peaks was observed. Both the folded and unfolded state peaks of the tryptophan indole NH resonance (~10 ppm in the proton dimension) are visible at intermediate pressures. These observations indicate that each residue of pp32 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 at all four temperatures individually to a two-state pressure-induced unfolding model as described in Materials and Methods (Figs. 2 D, S2–S4 D, and S5–S8), yielding residue specific values for the apparent volume change ( $\Delta V_f$ ) and apparent free energy ( $\Delta G_f$ ) of folding (Tables S1–S4 in the Supporting Material).

The two-state model was adequate for all of the individual unfolding profiles, taken separately. However, distinct unfolding profiles for different residues were observed (for example, Figs. 2 D and S2–S4 D), demonstrating clear deviation from two state-behavior. Loss of intensity for a given

cross peak at pressures below that of the main unfolding transition and/or retention of intensity at pressures above the midpoint relative to other residues, broadens the apparent transition. This is due to the population of one or more partially folded intermediates exhibiting disorder at that residue. Fitting these broader transitions to a two-state unfolding model results in lower values for both the apparent volume change and the apparent free energy of unfolding. This leads to a distribution of apparent  $\Delta V_f$  and apparent  $\Delta G_f$  values at each temperature tested (Fig. 3). These distributions were narrowest at 293 K, approaching the experimental uncertainty. The experimental uncertainty in apparent  $\Delta V_f$  and apparent  $\Delta G_f$  from the fits of the individual unfolding profiles for each residue was 12.7 mL/mol and 0.382 kcal/mol at 293 K, and very similar for 298 K (12.9 mL/mol and 0.40 kcal/mol), and 303 K (12.2 mL/mol and 0.37 kcal/mol), while these values were slightly higher for 288 K (16.0 mL/mol and 0.45 kcal/mol). While the experimental error was not temperature-dependent, the parameter distributions broadened significantly upon either increasing or decreasing temperature. The average apparent stability for each repeat calculated from the individual residue apparent  $\Delta G_f$  values (Fig. S9) increases nonrandomly from the N- to the C-terminus at 303

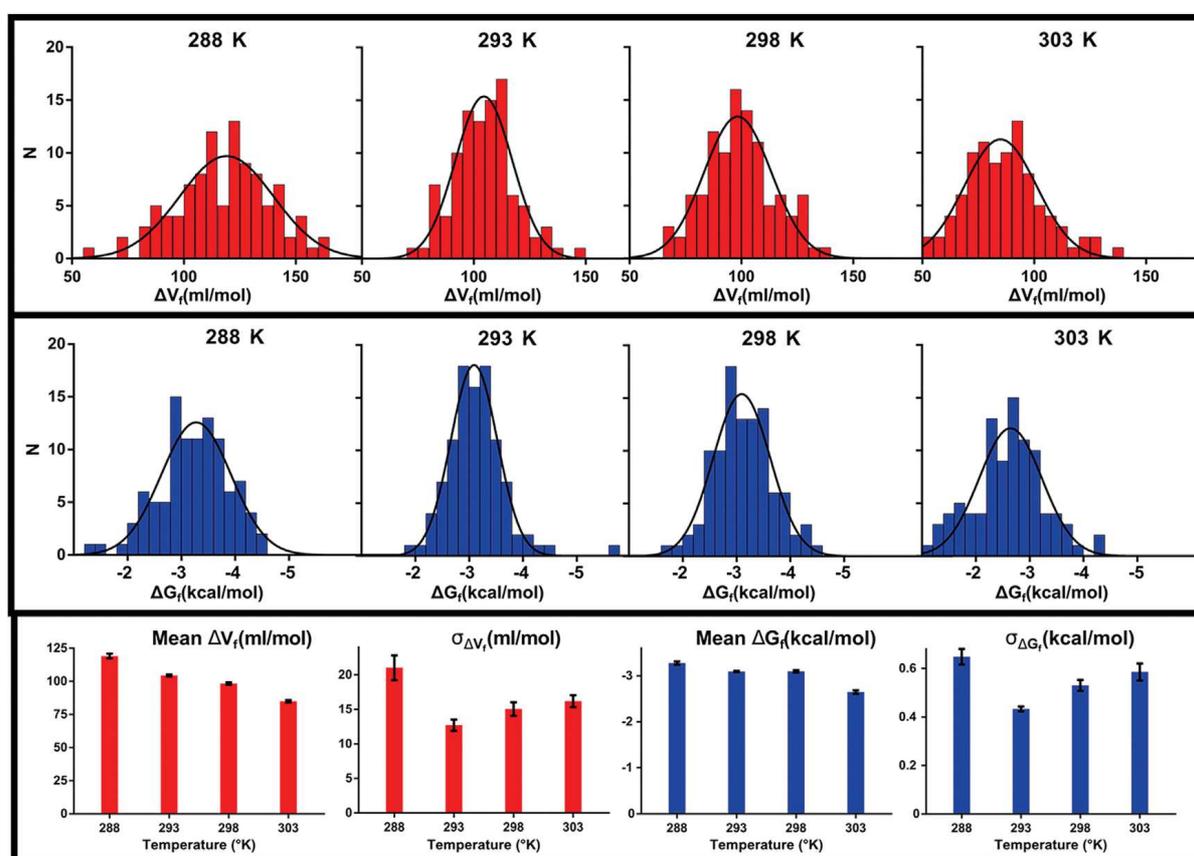


FIGURE 3 Distribution of the apparent  $\Delta V_f$  and  $\Delta G_f^0$  values (top and middle panels, respectively) obtained from fits of the residue-specific pressure unfolding profiles at each temperature as indicated. The lines in the top and middle panels represent a Gaussian fits to each parameter distribution. The bottom panels provide the means and SDs of the distributions at each temperature; the error bars represent the SD of the fitted Gaussian distributions. To see this figure in color, go online.

K, and the difference between the stability of the N-Cap and C-Cap regions is significant at this temperature. The average value of the apparent  $\Delta V_f$  distributions decreases with increasing temperature, as expected from the lower thermal expansion of the folded versus unfolded state (29–31), which is related to the protein's intramolecular interaction network (32).

### Fractional contact maps

To visualize more precisely which regions of the protein become disordered at intermediate pressures, we constructed fractional contact maps (10). We define the probability of contact for any pair of residues,  $P_{i,j}$ , as the geometric mean of the fractional probability that each of the two residues is in the folded state at a given pressure:

$$P_{(i,j)} = \sqrt{P_i \cdot P_j}, \quad (2)$$

where  $P_i$  or  $P_j$  is the ratio of the fitted intensity value of the HSQC resonance for residue,  $i$  or  $j$ , at a given pressure relative to the fitted plateau intensity at atmospheric pressure:

$$P_i = \frac{I_i}{I_{i0}}. \quad (3)$$

In the fractional contact maps calculated at 900 bar and at all four temperatures (Fig. 4), significant heterogeneity in the stability of the different regions of the protein was

apparent. Fractional contact values can be seen to increase from the N- to the C-terminus. Histograms of fractional contacts were constructed for these four conditions of temperature and pressure (Fig. 5) and for the entire pressure unfolding profile at all four temperatures (Fig. S10, A–D), coloring the contacts in each repeat according to the color scheme in Fig. 1. Contact distributions were found to be asymmetric from the N- to C-terminus, with the distribution broadening to lower contact values very significantly at temperatures above and below 293 K. A bimodal distribution is apparent at 303 K, with considerable disruption of the N-cap and the first two repeats compared to the rest of the protein. Even at 293 K, the N-terminal residues exhibit slightly lower contact values than their C-terminal counterparts.

### Structural and energetic features of the pp32 folding landscape

Because we observe distinct pressure unfolding profiles for different residues and regions of pp32, we consider that at any given temperature and pressure, the protein populates an ensemble of conformations that include the folded and unfolded state, but also one or more intermediate states, the number of which we do not know a priori. Thus, rather than using an a priori three-state model, for example, we sought to use the NMR data to obtain an unbiased structural and energetic map of the folding landscape of pp32. We note that our approach using constrained structure-based

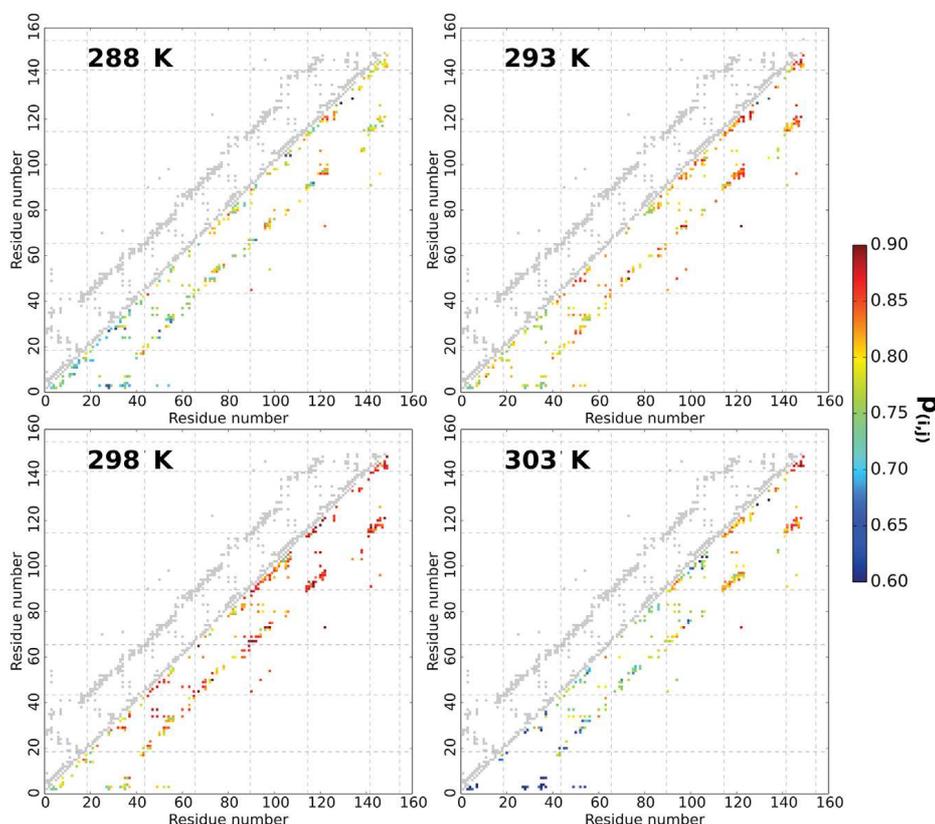


FIGURE 4 Fractional contact maps for pp32 calculated from the NMR HSQC peak intensity as described in the text at 900 bar, 1.4 M urea, and at different temperatures as indicated. Fraction represents the probability that a contact is formed in the ensemble. The color scale is the same for all temperatures. Gray dots above the diagonal correspond to native contacts calculated from the crystal structure (PDB: 2JE0), colored dots below the diagonal correspond to contact probabilities calculated from the NMR data as described in the text. The full color scale corresponds to fractional contacts between 60 and 90%. To see this figure in color, go online.

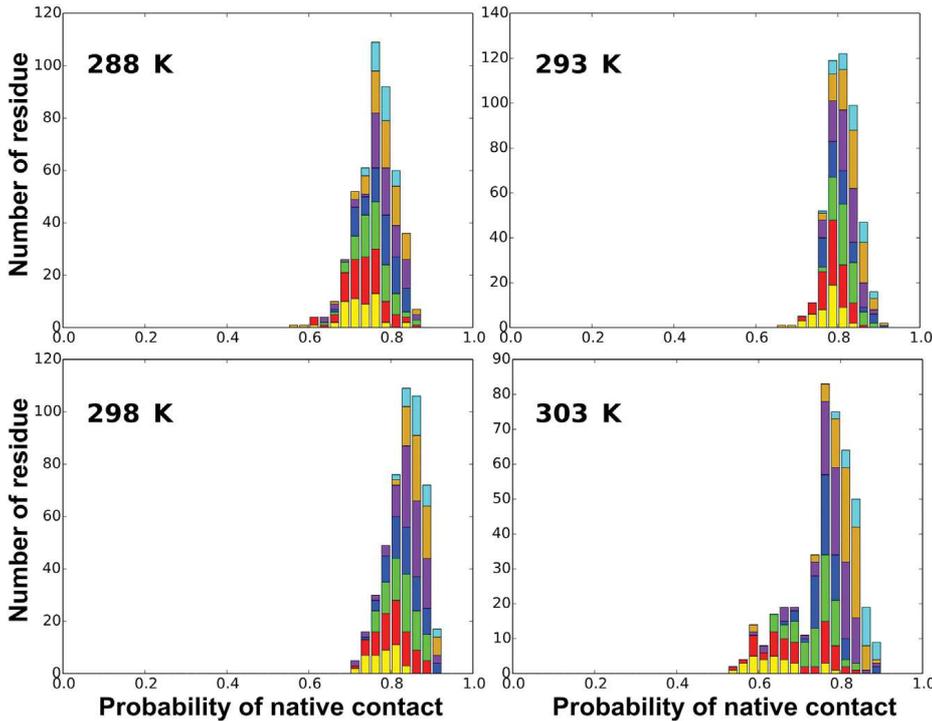


FIGURE 5 Histograms of the probability of native contact at 900 bar and 1.4 M urea at all four tested temperatures, as listed. The histogram is colored for each repeat and cap using the color code in Fig. 1. Each residue contributing to a contact is counted in the repeat it belongs to, such that interfacial contacts are counted once in each repeat, and those involved in a residue pair within a repeat are counted twice in the repeat. To see this figure in color, go online.

modeling is not restricted to repeat proteins as is the one-dimensional Ising model, which has been used by one of us to analyze repeat protein folding (33). Instead, it is general to all protein topologies and relies on the NMR data to parse the stabilities, rather than defining them a priori. Hence, the experimental fractional contact maps obtained from the pressure-dependent HSQC data were employed as a constraint to further characterize the conformational ensemble along an order parameter related to the degree of folding of the protein. Structure-based coarse-grained models were used to simulate the conformational ensemble of the protein at the pressure and temperatures of interest. The percentage of native contacts,  $Q$ , is the natural order parameter for such models. A value of  $Q = 0\%$  refers to states that do not contain any native contacts, while  $Q = 100\%$  refers to states in which all native contacts are formed. The experimental fractional contacts for a given set of conditions were modeled probabilistically as an ensemble of  $K$  contact maps whose average reproduces the experimental contact map. That is,

$$\langle C_{ij} \rangle = \frac{1}{K} \sum_{l=1}^K C_{ij}^{(l)}, \quad (4)$$

where  $C_{ij}^{(l)}$  indicates quenched (i.e., frozen) contact maps, and  $\langle C_{ij} \rangle$  is the experimental fractional contact map. The quenched contact maps were obtained by randomly selecting native contacts as present or not, in such a way that the average probability of a given native contact satisfied the measured probabilities. Given that the probability for each native contact in the experimental fractional contact

map is known, we drew a random number uniformly distributed between 0 and 1. If the number was higher than the known probability for the contact, then it was deleted from the quenched contact map, otherwise it was kept. Contacts involving unresolved residues were set at the average contact value for the particular temperature and pressure tested. Following the methods used in spin glasses, we define the partition function for the system by

$$Z^K = \prod_{l=1}^K Z^{(l)}(\{C_{ij}^{(l)}\}), \quad (5)$$

where  $Z^{(l)}(\{C_{ij}^{(l)}\})$  is the partition function of the quenched contact map,  $C_{ij}^{(l)}$ .

For each quenched contact map we built a structure-based model and solved for the energy landscape of the quenched system by molecular dynamics simulations. The potential of mean force as a function of  $Q$ ,  $\Delta F(Q)$ , was built for each quenched contact. The free energy landscape for the ensemble was defined as the average of the quenched potential of mean force,

$$\Delta F(Q) = \frac{1}{K} \sum_{l=1}^K \Delta F^{(l)}(Q). \quad (6)$$

From the fractional contacts at each pressure and temperature of interest,  $K = 300$  separate contact lists were generated. The number of contact maps,  $K$ , was chosen to ensure convergence of the average potential of mean force,  $\Delta F(Q)$ . Coarse-grained  $C\alpha$  molecular dynamics simulations were

carried out for 100 ns on all 300 contact lists for data obtained at the four temperatures at 900 bar, as this was the pressure at which the most experimental heterogeneity was apparent. Taking the last 50 ns of each simulation yielded  $\sim 30,000,000$  configurations for each condition. The resultant conformational ensembles were mapped back to all-atom representations (see Materials and Methods).

Heat maps at 900 bar and all four temperatures (Fig. 6) of the RMSD of all configurations relative to the native state structure versus the percentage of native contacts ( $Q$ ) reveal the population of the folded, unfolded, and intermediate states at a given temperature and pressure, as well as the folding/unfolding transition region. For each condition, the number of configurations exhibiting a given fraction of native contacts was used to calculate a pseudo-free energy profile using the relation,  $G \propto -\ln(N)$  (Figs. 7 and S11–S13). The average structure of the major intermediate, E, at  $Q \sim 60\%$ , along the folding pathway was determined by cluster analysis to be ordered in repeats 3–5 and the C-cap, whereas the N-Cap and repeats 1 and 2 are disordered. The pseudo free-energy landscape also reveals a broad folding barrier near the unfolded state, regions (B–D), with a percentage folded for the transition state ensemble,  $Q_T$ , between 35 and 50% ( $Q_U = \sim 30\%$ ), and accretion of structure from the C- toward the N-terminus. It is notable that in the case of pp32, the structural and energetic properties of the folding barrier can be obtained from equilibrium data using our modeling approach. This is because the major folding and unfolding barriers, while significant, are not extremely high. In previous studies of high-pressure NMR and constrained modeling of staphylococcal nuclease (10), the barrier was too high for the simulations to provide any structural information about the transition state. The temperature dependence of the stabilities of the various states on the landscape cannot be extracted from the con-

strained modeling at a single pressure because the overall stability at 900 bar is not identical for the different temperatures. However, we do observe subtle temperature-dependent changes in the shape of the landscape. For example, the folding barrier is much narrower and the transition state is closer to the unfolded state at 303 K compared to the other temperatures. The intermediate basin is also much more shallow at 303 K, and at its most stable point both at 288 and 303 K, it exhibits fewer native contacts than at 293 K. Fractional contact maps (Figs. S14–S17) for the average structures (A–E) obtained from the resulting ensembles are consistent with the representative structures shown in Figs. 7 and S11–S13.

## DISCUSSION

Equilibrium high-pressure NMR coupled with constrained coarse-grained structure-based molecular modeling has been used to obtain quantitative and highly detailed structural and energetic maps of the folding of a model repeat protein, pp32, in the p-T plane. The structures and relative stabilities of the conformations revealed using this approach are entirely consistent with the structural and energetic information about the major folding intermediate of pp32 and the folding transition state, both derived from a kinetic  $\phi$ -value analysis of a large number of pp32 mutants and from amide hydrogen exchange studies (18). For example, the structural ensemble at the folding barrier that we find by our constrained modeling approach exhibits order only in the C-cap and C-terminal repeat. This is exactly the conclusion reached from the denaturant-based  $\phi$ -value analysis cited above. The structural ensemble obtained here for the major intermediate, with a largely disordered N-cap and repeat 1, and partial order in repeats 2–4, is consistent with the rather large  $m$  value for its folding to the native state and

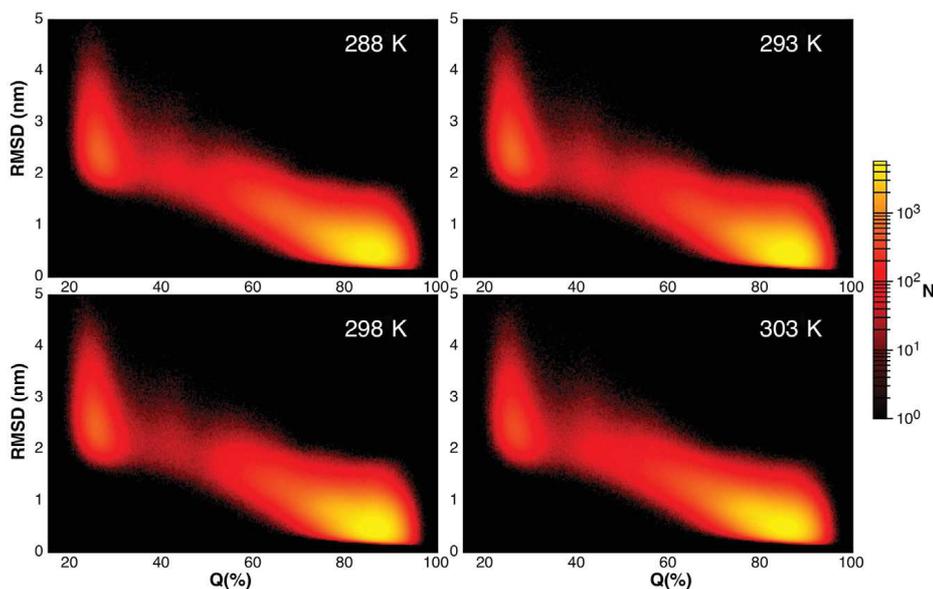


FIGURE 6 Heat maps of RMSD versus  $Q$  (percentage of native contacts) for conformational ensembles of pp32 obtained from the coarse-grained simulations at 298, 293, 298, and 303 K in 1.4 M urea. Note the heat map scale is logarithmic. To see this figure in color, go online.

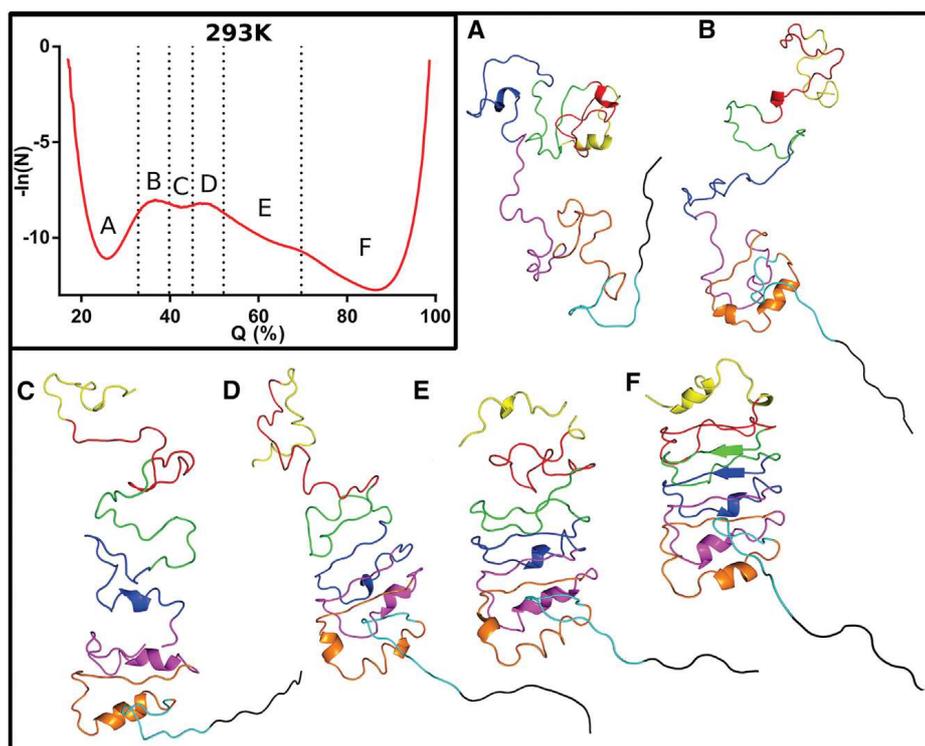


FIGURE 7 (Inset) Pseudo-free energy landscape as extracted from the simulations using the data obtained at 293 K. The displayed structures are reconstructed from the free energy regions indicated by the corresponding letters (A–F). Representative structures were obtained as described in the Materials and Methods. Structures are colored as in Fig. 1. Similar figures for other temperatures can be found in the [Supporting Material](#). To see this figure in color, go online.

the range of local stabilities observed in the native state hydrogen exchange measurements. These equilibrium high-pressure NMR studies provide significantly greater structural and energetic detail as a result of observables at nearly every residue in the protein. Moreover, combining these experimental observables with simulation yields the level of structural and energetic insight required to understand protein folding.

In contrast to the clear deviations from two-state behavior observed here under pressure, equilibrium urea-induced unfolding transitions of pp32 were highly cooperative (17). This difference in behavior reflects fundamental differences between the driving forces for pressure and urea-induced unfolding. Denaturants equally affect all regions of a protein via approximately homogeneous interaction propensity with the protein surface in the unfolded state. In contrast, pressure works to decrease volume primarily via the elimination of the solvent excluded void volume present in the folded structure (9,10). Thus, the global structural consequences of pressure depend upon the coupling between the most pressure-sensitive, least well-packed regions and the rest of the protein.

The pp32 LRR domain exhibits a large internal and hydrophobic cavity in the regions of repeats 2 and 3 (Fig. 1) that would be exposed to solvent in the major folding intermediate in which the N-terminus is disordered. Although this cavity could presumably be exposed as well upon disruption of the central core or the C-terminus, it is the N-terminus that is the least stable region of pp32 (18). Hence, this region is preferentially destabilized by pressure. The structural properties of this major intermediate are

consistent with that detected transiently in denaturant kinetics experiments, but is populated at equilibrium under pressure. Because of the local nature of pressure effects, the folding landscape observed in pressure experiments may more closely reflect the reaction as it occurs under native conditions. The nonlinear temperature dependence of the population of the major folding intermediate is consistent with a significant heat capacity change and exposure of nonpolar surface area upon disruption of the N-terminal region.

The equilibrium high-pressure NMR and constrained modeling studies presented here provide a map of unprecedented detail of the structures and relative stabilities of the conformations populated in the pp32 ensemble along its folding pathway. Apparent cooperativity in protein folding can result from equivalent stabilities and responses to perturbation for all domains or regions of a proteins' structure, or, alternatively, from strong interactions between domains coupled to their individual marginal stability in isolation. These straightforward equilibrium high-pressure NMR studies on pp32 confirm the existence of a hierarchy of three folding units (or foldons (8)) in the structure of this model repeat protein, namely the unstable N-Cap-repeat 1-repeat-2 region, the more stable central repeats 3 and 4 and the highly stable region comprised of repeat 5 and the C-cap. This hierarchy in structural stability was not apparent in equilibrium unfolding studies using urea as a denaturant, and could only be inferred from kinetic studies on multiple mutants (17,18). It suggests that the apparent cooperativity in folding of pp32 is defined by the relative local stabilities of these three regions.

## SUPPORTING MATERIAL

Supporting Materials and Methods, four tables, and 17 figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(16\)30749-4](http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)30749-4).

## AUTHOR CONTRIBUTIONS

M.J.F. carried out experiments and analysis, wrote analysis software, carried out simulations, worked on interpretation, and wrote the article; T.P.D. prepared samples, worked on interpretation, and wrote the article; K.J. carried out experiments and analysis and participated in interpretation; M.D. acquired data and participated in analysis; Y.Y. confirmed and corrected NMR assignments; S.A.M. assisted with NMR acquisition and analysis; A.E.G. assisted with molecular dynamics simulations and interpretations and wrote the article; D.B. contributed to conception of work and interpretations, and wrote the article; C.R. assisted in NMR acquisition and analysis, and wrote the article; and C.A.R. supervised and coordinated the project, conceived experiments and analysis, participated in the interpretation, and wrote the article.

## ACKNOWLEDGMENTS

The work was supported by National Science Foundation grants No. MCB1514575 and No. MCB1050966 to C.A.R. and A.E.G., respectively; from National Institutes of Health grant No. GM068462 to D.B.; and from French Infrastructure for Integrated Structural Biology grant No. ANR-10-INSB-05-01 to C.R.

## REFERENCES

- Ziv, G., and G. Haran. 2009. Protein folding, protein collapse, and Tanford's transfer model: lessons from single-molecule FRET. *J. Am. Chem. Soc.* 131:2942–2947.
- Pirchi, M., G. Ziv, ..., G. Haran. 2011. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat. Commun.* 2:493.
- Roder, H. 1989. Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol.* 176:446–473.
- Beasty, A. M., and C. R. Matthews. 1985. Characterization of an early intermediate in the folding of the  $\alpha$ -subunit of tryptophan synthase by hydrogen exchange measurement. *Biochemistry.* 24:3547–3553.
- Miranker, A., S. E. Radford, ..., C. M. Dobson. 1991. Demonstration by NMR of folding domains in lysozyme. *Nature.* 349:633–636.
- Hughson, F. M., D. Barrick, and R. L. Baldwin. 1991. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry.* 30:4113–4118.
- Hu, W., B. T. Walters, ..., S. W. Englander. 2013. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. USA.* 110:7684–7689.
- Bédard, S., L. C. Mayne, ..., S. W. Englander. 2008. The foldon substructure of staphylococcal nuclease. *J. Mol. Biol.* 376:1142–1154.
- Rouget, J. B., T. Aksel, ..., C. A. Royer. 2011. Size and sequence and the volume change of protein folding. *J. Am. Chem. Soc.* 133:6020–6027.
- Roche, J., J. A. Caro, ..., C. A. Royer. 2012. Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. USA.* 109:6945–6950.
- Peterson, R. W., and A. J. Wand. 2005. Self-contained high-pressure cell, apparatus, and procedure for the preparation of encapsulated proteins dissolved in low viscosity fluids for nuclear magnetic resonance spectroscopy. *Rev. Sci. Instrum.* 76:094101.
- Jackson, S. E. 1998. How do small single-domain proteins fold? *Fold. Des.* 3:R81–R91.
- Chan, H. S., S. Bromberg, and K. A. Dill. 1995. Models of cooperativity in protein folding. *Philos. Trans. R. Soc. B Biol. Sci.* 348:61–70.
- Muñoz, V., L. A. Campos, and M. Sadqi. 2016. Limited cooperativity in protein folding. *Curr. Opin. Struct. Biol.* 36:58–66.
- Kajava, A. V. 1998. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* 277:519–527.
- Huyton, T., and C. Wolberger. 2007. The crystal structure of the tumor suppressor protein pp32 (Anp32a): structural insights into Anp32 family of proteins. *Protein Sci.* 16:1308–1315.
- Dao, T. P., A. Majumdar, and D. Barrick. 2014. Capping motifs stabilize the leucine-rich repeat protein PP32 and rigidify adjacent repeats. *Protein Sci.* 23:801–811.
- Dao, T. P., A. Majumdar, and D. Barrick. 2015. Highly polarized C-terminal transition state of the leucine-rich repeat domain of PP32 is governed by local stability. *Proc. Natl. Acad. Sci. USA.* 112:E2298–E2306.
- Vranken, W. F., W. Boucher, ..., E. D. Laue. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* 59:687–696.
- Jones, E., T. Oliphant, and P. Peterson. 2001. SciPy: Open Source Scientific Tools for Python. <http://www.scipy.org/>.
- Noel, J. K., P. C. Whitford, and J. N. Onuchic. 2012. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B.* 116:8692–8702.
- Taketomi, H., Y. Ueda, and N. Gö. 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.
- Hornak, V., R. Abel, ..., C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.
- Horn, H. W., W. C. Swope, ..., T. Head-Gordon. 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* 120:9665–9678.
- Noel, J. K., P. C. Whitford, ..., J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 38:W657–W661.
- Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
- Rotkiewicz, P., and J. Skolnick. 2008. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* 29:1460–1465.
- Kitahara, R., K. Hata, ..., K. Akasaka. 2013. Pressure-induced chemical shifts as probes for conformational fluctuations in proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* 71:35–58.
- Dellarole, M., K. Kobayashi, ..., C. A. Royer. 2013. Probing the physical determinants of thermal expansion of folded proteins. *J. Phys. Chem. B.* 117:12742–12749.
- Royer, C., and R. Winter. 2011. Protein hydration and volumetric properties. *Curr. Opin. Colloid Interface Sci.* 16:568–571.
- Tsamaloukas, A. D., N. K. Pyzocha, and G. I. Makhatadze. 2010. Pressure perturbation calorimetry of unfolded proteins. *J. Phys. Chem. B.* 114:16166–16170.
- Dellarole, M., J. A. Caro, ..., C. Roumestand. 2015. Evolutionarily conserved pattern of interactions in a protein revealed by local thermal expansion properties. *J. Am. Chem. Soc.* 137:9354–9362.
- Street, T. O., C. M. Bradley, and D. Barrick. 2007. Predicting coupling limits from an experimentally determined energy landscape. *Proc. Natl. Acad. Sci. USA.* 104:4907–4912.
- Ho, B. K., and F. Gruswitz. 2008. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* 8:49.

## 3.2 Conclusion and perspective

One interesting feature revealed in this article is the fact that the level of cooperativity is temperature dependent, and that cooperativity of folding is disrupted for temperature both higher and lower than 293K. This is a direct consequence of the change in the parameters that drive protein stability with changing temperature, as exposed in the introduction figure 1.16. This means that the parameters driving stability in PP32 are not the same for all repeats, thus have a different response to changing temperature. Indeed, we can see from figure 3.3 that the pressure of half unfolding, meaning the pressure at which the intensities are in average for all resolved residues half of their original intensity. This figure shows that in our experiments, we are crossing the  $\Delta S = 0$  line at a temperature of around 298K.

This study was able to make a simple description of the folding landscape of PP32. Using the residue specific information of 2D NMR combined with the stabilization of otherwise short lived intermediates through pressure denaturation, it showed a simple way of visualizing the differences in unfolded proportion in the different parts of the protein. Using the method of constrained structure based modeling to reproduce the structural ensemble of the protein in solution is a promising technique for visualization of intermediates and folding pathways. This step however could still be perfected, and will probably be enhanced in future work. However, this technique was in very good agreement with the finding of the Barrick lab concerning the stable C-terminal part, but with simple equilibrium experiments, as opposed to very time consuming  $\phi$ -value analysis requiring kinetics analysis of a number of mutants. This method will be implemented in our lab to some of the mutants tested in the Barrick lab, with preliminary

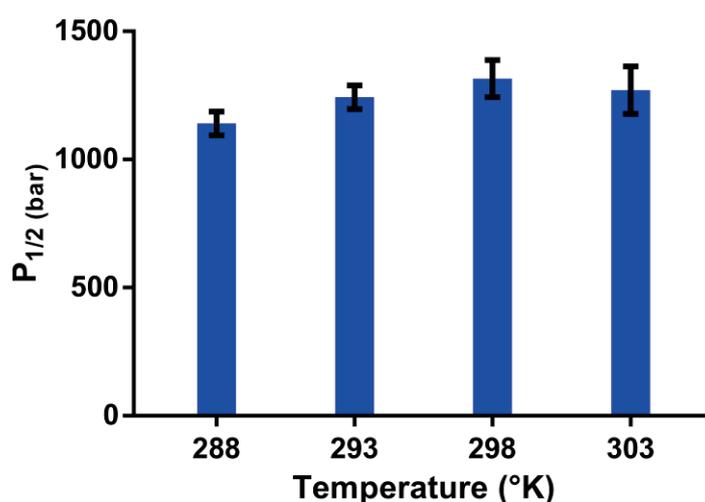


Figure 3.3: Pressure of half unfolding plotted as a function of temperature for PP32 WT. The histogram bars represent the mean value, the error represents the standard deviation of  $P_{\frac{1}{2}}$  among all the residues resolved in the NMR spectra

results presented in the next chapter.

### 3.3 Resume de l'article en francais

#### Etude du paysage de repliement de PP32

PP32 est une protéine dite répétition riche en leucine (LRR en anglais) qui est composée de cinq répétitions et de deux « caps ». Les protéines à répétitions sont idéales pour l'étude de la coopérativité de repliement à cause de leur haut niveau d'ordre de contact, c'est-à-dire leur manque d'interactions entre des résidus qui sont éloignés dans la structure tertiaire de la protéine. Cette étude suit une caractérisation complète de la protéine par dénaturation par l'urée réalisée par le laboratoire de Doug Barrick, à l'Université de John Hopkins aux USA. Le but de cette collaboration est d'utiliser les avantages uniques à la dénaturation par pression des protéines afin de peupler et d'observer des intermédiaires de réaction métastables. La dénaturation par pression est un outil unique pour l'étude du repliement des protéines à cause du mécanisme par lequel elle déstabilise l'état plié de la protéine. La pression favorise les états qui minimisent le volume en solution. Le plus petit volume de l'état déplié des protéines due à la présence de volume vide dans l'état plié et le facteur déterminant du dépliement par pression. Puisque ces volumes vides sont typiquement distribués de façon non-homogène dans la structure native de la protéine, l'effet de la pression peut être très local. Dans la protéine PP32, le volume vide le plus large est situé environ au milieu de la protéine. Puisque la stabilité de la protéine est très polarisée comme l'a démontré le laboratoire de Doug Barrick, nous espérons que cette inhomogénéité dans la distribution des volumes vides nous aidera à peupler des intermédiaires de réaction. De fait, l'utilisation de profils de dépliements résidu par résidu obtenue par RMN-2D à haute pression a montré une déviation significative de niveau de repliement de la partie N-terminal à la partie C-terminal de la protéine. Afin de souligner cette déviation, nous avons utilisé les profils de repliement afin de construire une carte de contact fractionnelle pour visualiser les différences de réponse à la pression d'un bout à l'autre de la protéine. Pour cela, nous avons assigné la probabilité de contact entre deux résidus comme étant la moyenne géométrique des fractions de signal des deux résidus faisant partie du contact dans l'état natif. De plus, nous avons utilisé ces informations sur la fraction de contact pour contraindre des simulations dites « structure based modelling », c'est-à-dire basées sur l'état natif, à une pression et température données. Cette procédure a permis l'obtention d'un paysage de repliement et la visualisation de l'ensemble structural présent dans les données expérimentales. De plus, cette thèse présente des résultats préliminaires de la même procédure utilisée sur des mutants de cette protéine. Ces mutations, réalisées à des bouts opposés de la protéine, créent ou élargissent le volume vide de l'état plié. Ces mutations ont pour effet de fortement moduler le paysage de repliement. Ces résultats confirment que

l'utilisation de la pression pour l'étude du repliement des protéines est un outil idéal pour l'étude des intermédiaires de réactions et de la coopérativité de repliement.



# Chapter 4

## Effect of cavity mutation on the unfolding landscape of PP32

### 4.1 Introduction

The next step in the study of PP32 is without a doubt the investigation of the unfolding landscape under pressure using cavity mutants. The unique advantage of pressure denaturation to have a site specific action in function of the presence of the void volumes makes for a perfect tool to study the disruption in apparent cooperativity upon void volumes creating or enlarging mutations. In this short chapter, we will go through early results obtained recently in the lab concerning cavity creating mutations in different repeats, using data that were collected and analyzed by me or Kelly Jenkins, a grad student from our lab. The mutants used in this study were created and intensively studied by Thuy Dao in the Doug Barrick lab[131, 62]. Result from previous studies using kinetic experiments and  $\phi$ -value analysis. A short description of  $\phi$ -value analysis is available in the material and methods chapter 2. The two mutations introduced are both mutations from Leucine to Alanine, which means the hydrophobic character of the residue is conserved, but its volume diminishes. These mutations are at position 60 and 139, as displayed on figure 4.1.

### 4.2 Materials and methods

The material and method for this study is in all ways similar to that presented in the previous chapter and the corresponding article. The NMR spectra of the mutants revealed to be very similar to that of the Wild Type protein (figure 4.3). The assignment of the mutants spectra

was realized by comparing  $^{15}\text{N}$ -NOESY-HSQC and  $^{15}\text{N}$ -TOCSY-HSQC to that of the WT. As opposed to WT, these mutants did not need the addition of urea to unfold in the pressure range available experimentally. The solution conditions for the NMR experiments for both mutants were 10mM NaCl, 20mM Bis-Tris, 5mM DTT at pH 6.8, as for the WT experiment.

### 4.3 Results and discussion

The spectra for both NMR high pressure unfolding of the two mutants presented here resulted in very different effects due to their position.

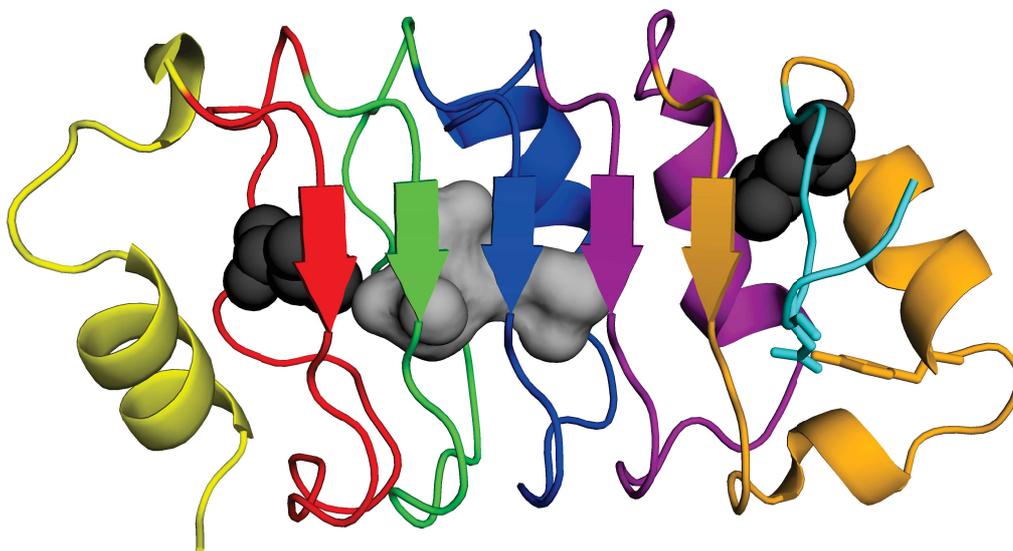


Figure 4.1: The leucine rich repeat protein PP32. The cavity, calculated using the software Hollow, with a grid resolution of  $0.25\text{\AA}$  and a probe size of  $1.2\text{\AA}$  is displayed in light gray. The backbone is colored by repeat. Mutation sites L60 and L139 are colored in dark grey

### 4.3.1 L60A mutant

The mutations appear to strongly modulate the energy landscape. The L60A mutation is destabilizing a region of the protein that is already very pressure sensitive in the WT study. The creation of additional volume in that region enhances the effects of pressure in a drastic way. In this mutant, some folded peaks do not disappear even at the highest pressure permitted by our experimental setup (2500bar). Furthermore, some of those peaks correspond to an intermediate in slow exchange with both the folded and unfolded state such that they gradually appear out of the unfolded region from mid-range pressure. The peaks remaining on the spectra have a smaller intensity than the fully folded state at atmospheric pressure, suggesting the population of the intermediate is only a fraction of the protein present in solution. These new peaks have

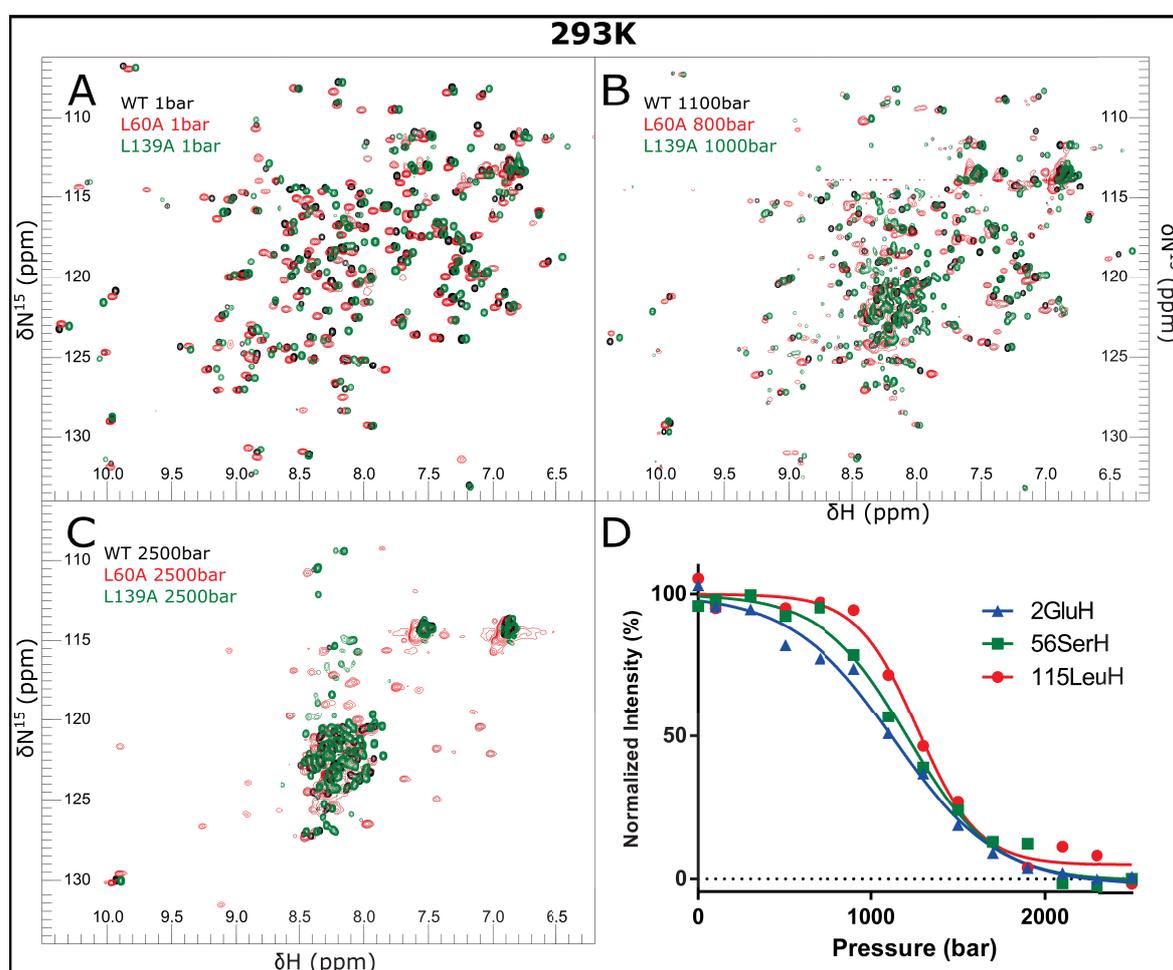


Figure 4.2: Pressure denaturation of PP32 WT (black) and mutants L60A (red) and L139A (green) at 293°K as observed by NMR. Panels A, B and C show examples of spectra at atmospheric pressure, intermediate pressure and high pressure. Panel D shows examples of pressure denaturation curves from WT and corresponding fits for residues yielding different apparent thermodynamic parameters

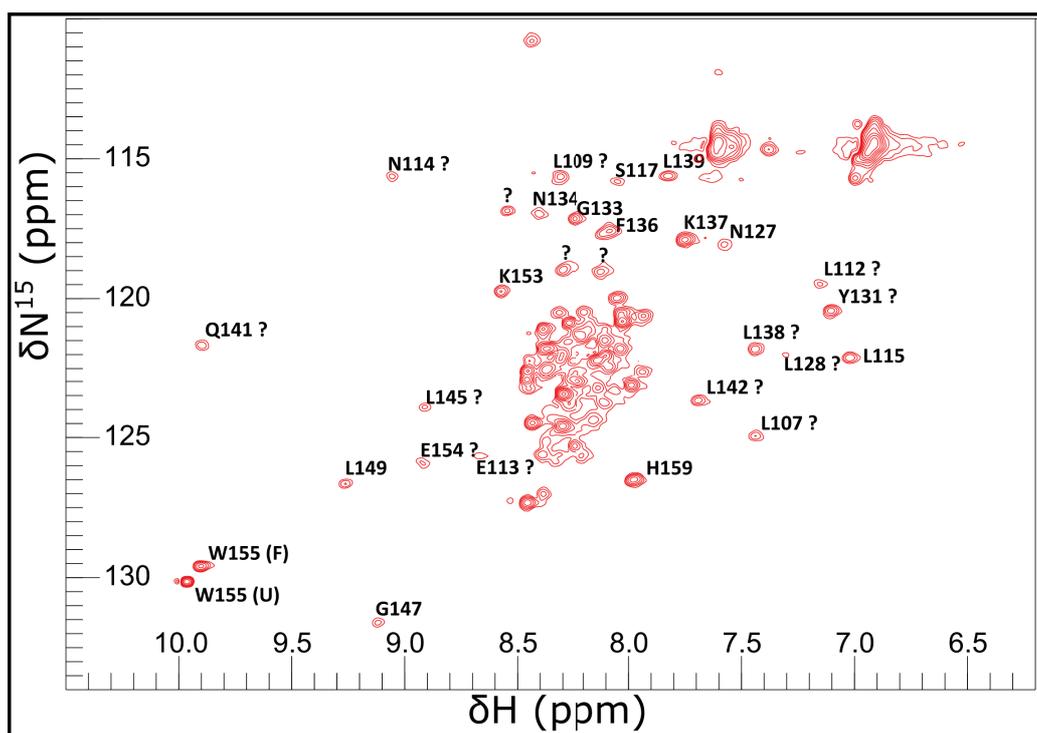


Figure 4.3: Spectra of L60A at 2500bar 293K. Residual peaks give hints of the presence of a stable intermediate. Due to slow exchange of the intermediate with the unfolded and folded states, some peaks disappear and reappear in a new positions away from the unfolded region. Peaks that appeared in an intermediate are labeled with a "?" if the peak appeared in an alternate position with their expected identities, without if they could be followed during the entire titration

not been formally assigned, but one would expect the intermediate to have similar chemical shift than the wild type PP32 for the residues that are still folded in this new ensemble. Thus, the comparison of the position with those of the fully folded state allows to speculate on the identity of those states (figure 4.3). All the peaks found in the last pressure correspond to residue numbers higher than 107, meaning they are part of the two last repeats or the C-terminal cap (Purple, orange and cyan on figure 4.1). The application of the fractional contact maps method as well as the constrained structure based modeling was not realized for this protein, because it would necessitate clear identification of the peaks present in the intermediate, as well as adaptation of the procedure used for the WT protein to take the intensity of those peaks into account as being part of a folded structure.

The finding of an intermediate in slow exchange at high pressure for the L60A mutant is indicative of a significant energy barrier for the folding of the last two repeats of PP32. This intermediate is a very good confirmation that the experimental approach for the future of the work I started on the WT protein is likely to succeed in helping to populate and observe otherwise invisible intermediates. The energy barrier between this intermediate and the fully

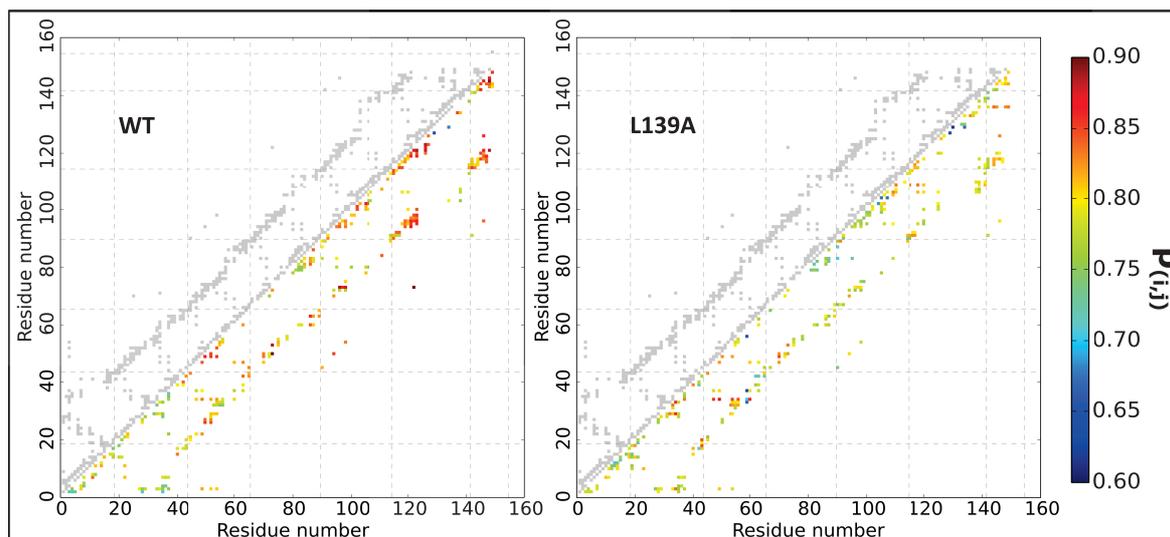


Figure 4.4: Fractional contact maps for PP32 WT (left) and its L139A mutant (right) at 900bar 293K. Grey dots represent crystal structure contacts, color dots represent the probability of contact between two residues displayed in the colormap (right). Grey dashed lines represent the limit of the repeats and caps in the primary structure.

unfolded state is what one would expect to be the rate limiting step for the folding of PP32. PP32 folding has been shown to be dominated by the gradient of local stability, with the C-terminal part being the most stable part of the protein. The folding of this protein is thus likely to be rather smooth once the C-terminal part is folded. This first intermediate could also be qualified as the most stable foldon. This is a perfect example of the complexity of protein folding and the limitations of the application of models for folding in a strict manner, because the presence of an energy barrier resulting in the slow exchange of this intermediate would suggest a two-states behavior for the C-terminal part of the protein, while the rest of the protein varies significantly from a two-states behavior. This intermediate was not observed in the WT, most likely because the WT is too stable to allow unfolding in the pressure range experimentally available, thus it required the addition of urea that probably participated in the destabilization of the C-terminal part. Thus the L60A mutation introduced here combined to the use of pressure participated in the destabilization of the protein in a very position dependent way, underlining the advantage of pressure over chemical denaturants. One should note that the fractional contact map of the WT at 303K does reveal a superior folded fraction for the two last repeats and C-terminal cap, suggesting that this part is indeed visibly more stable even in the WT protein.

### 4.3.2 L139A mutant

The L139A mutation results in the creation of a cavity in the C-cap part of the protein (Figure 4.1), which is the most stable part, as demonstrated by  $\Phi$ -value analysis and Hydrogen exchange NMR[62]. This mutant resulted in a fully unfolded spectra at high pressure (figure 4.3). Contact map analysis of the mutant compared to that of the WT protein is displayed in figure 4.4, and a histogram of the fractional contact is displayed in figure 4.5. The application of the structure based modeling constrained by the NMR data was realized, following the method introduced in the previous chapter. Comparison of the folding landscape extracted from the simulations between the WT protein and the L139A mutant is displayed in figure 4.6, and C- $\alpha$  RMSDs to the crystal structure is displayed in figure 4.7.

The contact map analysis of the L139A mutant revealed a more homogeneous distribution of unfolding profiles and thus a higher apparent cooperativity (figure 4.4 and 4.5). This is the result of a more homogeneous distribution in energy among the repeat, by the introduction of the cavity in the most stable part of the WT protein. The structure based modeling of this mutant revealed a more populated transition ensemble than the WT, but a similar shape of the pseudo free energy profile (figure 4.6). Looking at the RMSD distribution as a function of the number of native contacts, we can see that the folded state is less distributed toward higher RMSD, and thus that this mutation seems to result in a protein that can accommodate less perturbations without unfolding. Furthermore, the mutant is significantly less stable than the

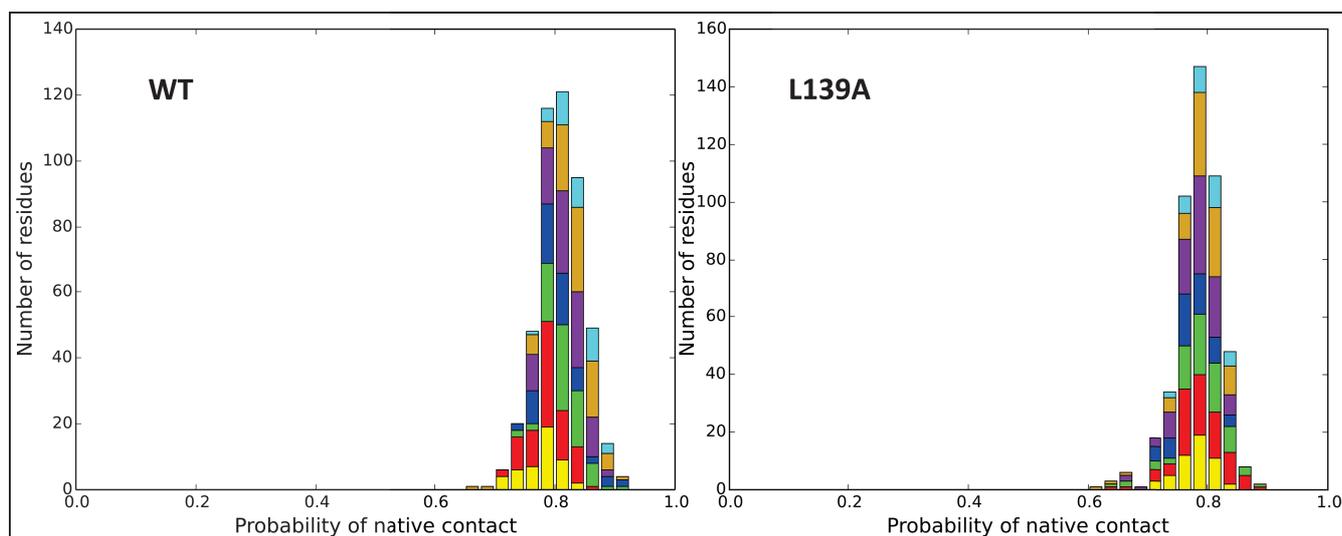


Figure 4.5: Histogram of probability of contact distribution for PP32 WT (left) and its L139A mutant (right) at 900bar 293K. Colors correspond to the color used in figure 4.1. Each contact is counted in a repeat if at least one of the residues forming it are in this repeat. Thus contact corresponding to interfaces are counted once in each repeat, and intra-repeat contacts twice in the same repeat

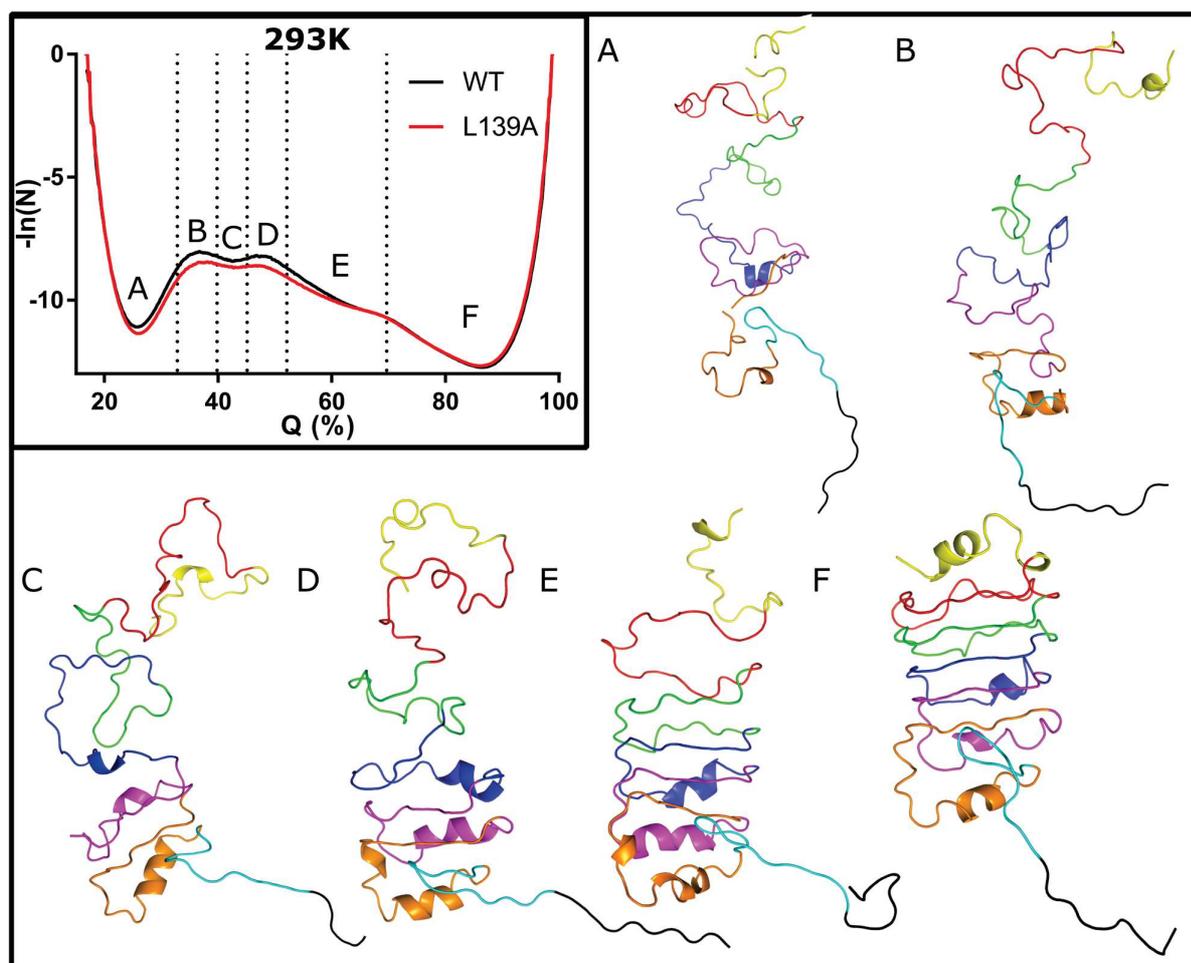


Figure 4.6: Pseudo free energy diagram (top left) for PP32 WT (black) and its L139A mutant (red) as extracted from Structure Based Modeling simulations using 900bar 293K data. Dotted lines represent the edge of the coordinate used for structure determination. Structure shown are extracted from the L139A simulations and correspond to the structure with the smallest average RMSD to all other structure in the region pointed to by the corresponding letter

WT protein, as it does not require urea to unfold in the accessible pressure range.

## 4.4 Conclusion

The preliminary results exposed here show that the high pressure denaturation in combination with repeat proteins can be a very useful tool to observe intermediates. The introduction of mutations and the modulation of their folding landscape by change in local stabilities reveal the importance in the gradient of energy in the folding process, and is a promising approach to further study protein folding. Characterization of the transition ensemble by pressure jump

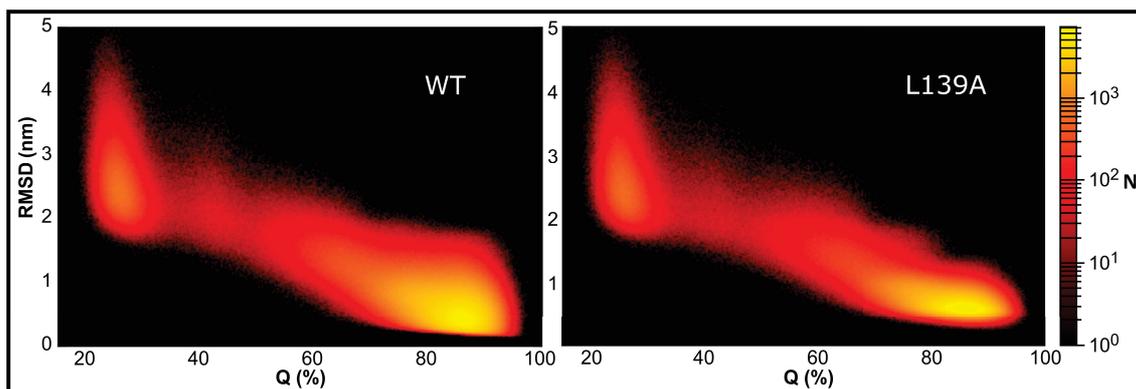


Figure 4.7: RMSD based on crystal structure as a function of number of native contacts, for PP32 WT (left) and its L139A mutant (right) as extracted from structure based modeling simulations using 900bar 293K data. Heatmap scale is logarithmic.

fluorescence spectroscopy of the C-terminal tryptophan, in order to characterize the transition ensemble of the folding reaction is the next step to realize a full characterization of the folding landscape of this protein, and is being pursued currently in the lab.

# Chapter 5

## Origins of the determinant of thermal expansivity through High pressure NMR

### 5.1 Introduction

In the article by *Rouget et al.*, briefly exposed in the introduction, the finding that hydration of surface residue is not a primary determinant of the thermal expansivity of folded proteins has led to more questions than answers. However, the paper ended on a opening note stating that the enthalpy fluctuations of the folded state are linked to the intrinsic network of interactions that stabilizes the folded state. These latter must place constraints on the folded state thermal expansivity. We have already stated in the introduction that the low heat capacity changes upon unfolding are correlated to higher folded state heat capacity, and thus that this correlation indicates that a rapid change in the number of accessible states is correlated with rapid changes in the expansivity of the folded state. This behavior was expected to result from the structural constraint of the folded state, and thus one would expect that proteins with stronger intramolecular bonds would have a lower expansivity. Thus the next step was logically to test the correlation between the structural constraints and the observed thermal expansivity.

The analysis leading to the conclusion of the article presented here that the thermal expansivity of the folded state arises from structural constraints, and thus that low expansivity is correlated to the presence of strong intramolecular hydrogen bonds is based on a study from Baxter and Williamson[172]. The authors of that study reported that the temperature dependence of the amide proton chemical shift change with temperature is linear, and is the results of increased thermal motion of the protein. Temperature coefficients that were inferior to

-4.5ppm.K<sup>-1</sup> were interpreted as being the result of the amid proton being a donor in a hydrogen bond. Although the study did mention that other amide protons that do not display that property may also be hydrogen bonded, a high number of amide protons that did display such a property was interpreted as being reflective of higher constraints on the thermal motions of the protein. This interpretation is made more credible by the fact that all mutants used in the present study have been characterized by crystallography, and displayed an equivalent tertiary structure, thus the increase in the number of hydrogen bonds is unlikely to reflect a change in conformation. Using a similar cut-off, the study presented here uses this definition to correlate a lower thermal expansivity of a protein to a higher number of hydrogen bonds that satisfied this definition.

# Evolutionarily Conserved Pattern of Interactions in a Protein Revealed by Local Thermal Expansion Properties

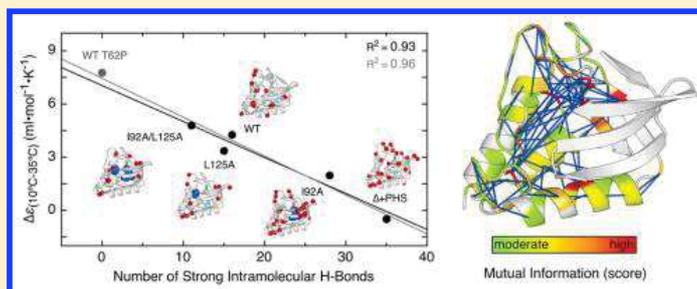
Mariano Dellarole,<sup>†,§</sup> Jose A. Caro,<sup>‡,#</sup> Julien Roche,<sup>†,||</sup> Martin Fossat,<sup>†,⊥</sup> Philippe Barthe,<sup>†</sup> Bertrand García-Moreno E.,<sup>‡</sup> Catherine A. Royer,<sup>\*,†,⊥</sup> and Christian Roumestand<sup>\*,†</sup>

<sup>†</sup>Centre de Biochimie Structurale, CNRS UMR5048, INSERM U554, Université Montpellier 1, 29 rue de Navacelles, Montpellier, France 34090

<sup>‡</sup>T. C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 N. Charles St., Baltimore, Maryland 21218, United States

## Supporting Information

**ABSTRACT:** The way in which the network of intramolecular interactions determines the cooperative folding and conformational dynamics of a protein remains poorly understood. High-pressure NMR spectroscopy is uniquely suited to examine this problem because it combines the site-specific resolution of the NMR experiments with the local character of pressure perturbations. Here we report on the temperature dependence of the site-specific volumetric properties of various forms of staphylococcal nuclease (SNase), including three variants with engineered internal cavities, as measured with high-pressure NMR spectroscopy. The strong temperature dependence of pressure-induced unfolding arises from poorly understood differences in thermal expansion between the folded and unfolded states. A significant inverse correlation was observed between the global thermal expansion of the folded proteins and the number of strong intramolecular hydrogen bonds, as determined by the temperature coefficient of the backbone amide chemical shifts. Comparison of the identity of these strong H-bonds with the co-evolution of pairs of residues in the SNase protein family suggests that the architecture of the interactions detected in the NMR experiments could be linked to a functional aspect of the protein. Moreover, the temperature dependence of the residue-specific volume changes of unfolding yielded residue-specific differences in expansivity and revealed how mutations impact intramolecular interaction patterns. These results show that intramolecular interactions in the folded states of proteins impose constraints against thermal expansion and that, hence, knowledge of site-specific thermal expansivity offers insight into the patterns of strong intramolecular interactions and other local determinants of protein stability, cooperativity, and potentially also of function.



## INTRODUCTION

The folded states of proteins are marginally stable relative to their unfolded states. The stability and conformational landscape of each protein has evolved to enable function appropriate to the organism and environment in which the protein functions. A fundamental understanding of how the sequence of amino acids determines the folding and functional properties of a protein requires detailed characterization of conformational landscapes and the effects of mutations thereon. This usually involves perturbation of the protein in a controlled fashion to attempt to identify the conformational states that constitute the ensemble under a variety of conditions.

Folding equilibria can be perturbed by changing the fundamental thermodynamic variables, temperature and pressure, by changing a fundamental physiological variable, pH, or by the addition of chemical compounds known to either destabilize or stabilize the folded state. Temperature, pH, and chemical denaturants have been widely used to probe the physical mechanism and sequence determinants of protein folding and stability. The mechanism of action of temperature and of chemical denaturants is governed by the amount of

surface area that is exposed to either solvent or denaturant upon unfolding.<sup>1</sup> pH effects are governed by differences in  $pK_a$  values of ionizable groups in the different conformational states in the ensemble.<sup>2</sup>

Pressure has been used also to unfold proteins, although to a much lesser extent, and until recently its mechanism of action was not well-understood. Differences in solvent density related to the hydration of surface area exposed by unfolding were thought to contribute significantly to the volume change upon unfolding. Such effects, like those of denaturants, should scale with the size of the protein. In contrast to this expectation, the deletion of 3 out of 7 repeats of the ankyrin repeat domain of the Notch receptor did not decrease the magnitude of the volume change for folding,  $\Delta V_f$ . Indeed, deletion of the first two repeats actually increased  $\Delta V_f$ .<sup>3</sup> These studies demonstrate the lack of significant contribution of hydration effects to  $\Delta V_f$ . This may arise from compensation of effects of different sign from the polar backbone and hydrophobic moieties. In contrast

Received: April 26, 2015

Published: July 2, 2015

to the lack of effect of the size of the protein, single amino acid substitutions that created cavities in the interior of a globular protein, staphylococcal nuclease (SNase), could double  $\Delta V_f$ .<sup>4</sup> Together our results led to the conclusion that the internal solvent-excluded void volume or packing defects in the folded state is the major contributing factor to the magnitude of pressure effects on thermodynamic stability.<sup>3,4</sup>

Packing defects are local features, specific to the structure of each protein. Thus, the effects of pressure on stability are exerted locally, and hence, the extent to which pressure disrupts the structure globally depends on the internal network of interactions unique to the fold of the individual protein. Characterization of the volumetric properties of proteins with the site-specific resolution afforded by NMR spectroscopy should provide insight into the structural basis and sequence determinants of stability and folding cooperativity.

Since Bridgman first reported on the effects of pressure on a protein,<sup>5</sup> it has been known that volume changes for unfolding are strongly dependent on temperature. Protein stability diagrams in the pressure–temperature plane, assuming cooperative two-state equilibrium between folded and unfolded states and energetically equivalent unfolded states, can be described by an ellipse,<sup>6–9</sup> where the first- and second-order parameters,  $\Delta H$ ,  $\Delta S$ ,  $\Delta C_p$ ,  $\Delta V$ ,  $\Delta\kappa'$ , and  $\Delta\epsilon$  correspond to the differences in molar enthalpy, entropy, heat capacity, volume, compressibility ( $d\Delta V/dp$ ) and thermal expansivity ( $d\Delta V/dT$ ) between folded and unfolded states, and where  $T_o$  and  $p_o$  are the chosen reference points in temperature and pressure (298 K and 1 bar).

$$\begin{aligned} \Delta G(p, T) = & \Delta G^o - \Delta C_p \left[ T \left( \ln \frac{T}{T_o} - 1 \right) + T \right] - \Delta S(T - T_o) \\ & + \Delta V(p - p_o) - \frac{1}{2} \Delta\kappa' (p - p_o)^2 \\ & - \Delta\epsilon(T - T_o)(p - p_o) \end{aligned} \quad (1)$$

In contrast to the rather small differences in compressibility,  $\Delta\kappa'$  (more than 10-fold smaller than the  $\Delta V$  and within the uncertainty of our measurement),<sup>8</sup> the differences in thermal expansivity,  $\Delta\epsilon$ , between folded and unfolded states of proteins are significant relative to  $\Delta V$  and are the cause of the strong temperature dependence of pressure effects. Indeed, the magnitude of  $\Delta V$  for unfolding of proteins decreases significantly with increasing temperature and can even change sign. The molar thermal expansivity of the unfolded states of proteins can be modeled empirically based on sequence alone, assuming additivity of the expansivity values of the individual amino acid residues.<sup>10,11</sup> On the other hand, the molecular contributions to the thermal expansion of folded states are not known. Substitution of a single amino acid in the core of a protein can change the molar expansivity of their folded state significantly (up to 3-fold), whereas significant changes to the polarity at the protein–water interface make very little difference.<sup>12,13</sup> Earlier studies using the temperature dependence of the intrinsic tryptophan fluorescence anisotropy<sup>14–16</sup> suggest that the intramolecular interactions in the folded state may act as constraints against thermal expansion. Hence, characterization of thermal expansivity with site-specific resolution could yield significant insight into the organization of intramolecular interactions of the folded protein and how they depend on sequence.

To probe systematically the molecular determinants of the thermal expansion of a protein, we examined the temperature

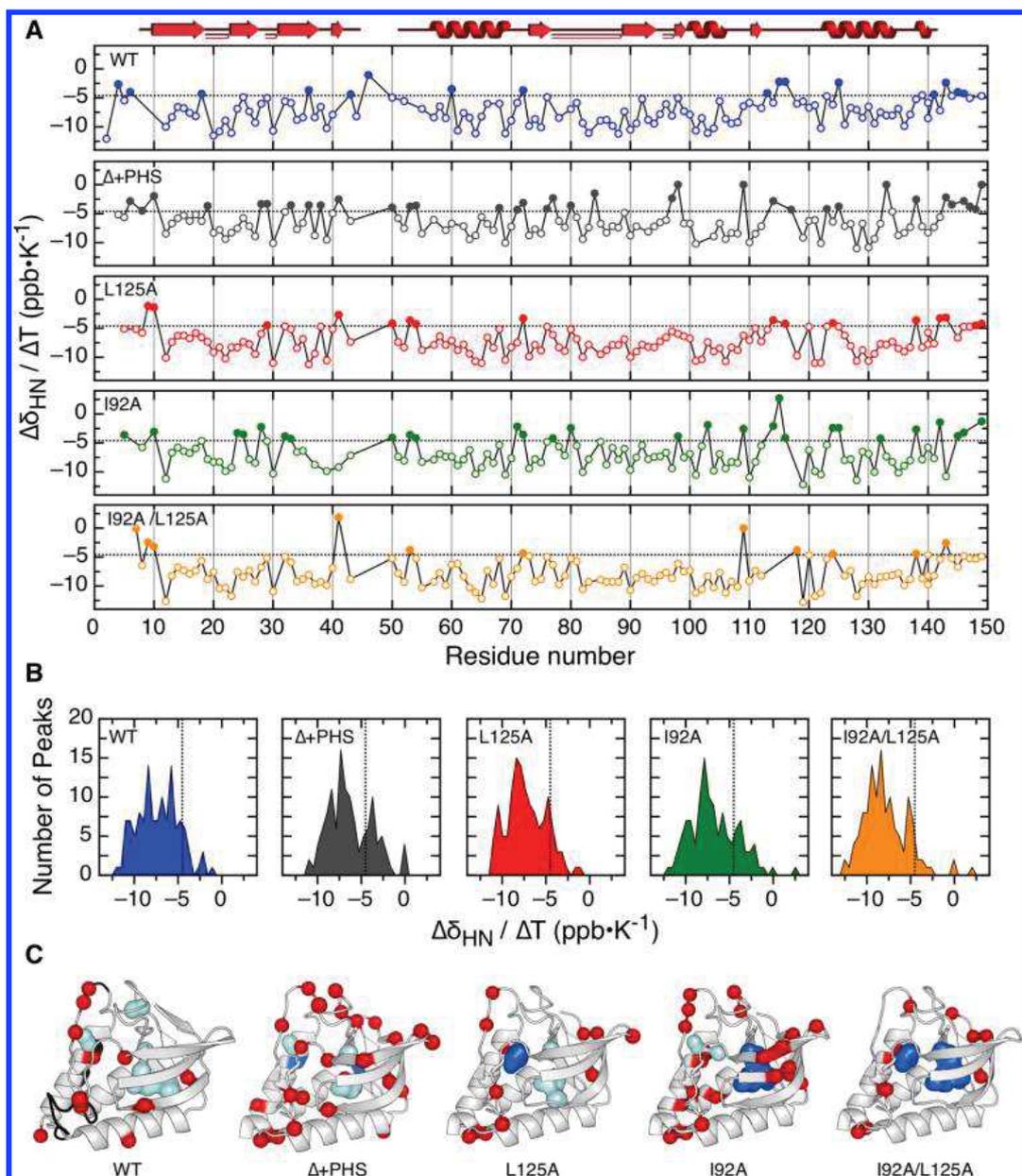
dependence of the volumetric properties SNase using the highly stable form of this protein known as  $\Delta$ +PHS, and variants of  $\Delta$ +PHS in which cavities were engineered by substitution of core residues by alanine. Some of these variants were used previously to demonstrate the importance of packing defects in the magnitude of the pressure effect.<sup>4</sup> In these prior studies, the three-dimensional (3-D) structures of the single substitution cavity variants were determined by X-ray crystallography, and the existence of the expected cavities was confirmed. The structure of a double variant (I92A/L125A) can be found in Supporting Information. Negligible rearrangements in the structures of the cavity containing variants were observed relative to the reference protein. Moreover, no penetration of water molecules into these cavities could be inferred from the electron densities, although this does not rule out penetration and the presence of transient or disordered water molecules in the cavities. Likewise, NMR chemical shift perturbations by the mutations were minimal,<sup>17</sup> particularly for the variants bearing cavities in the OB-fold region of the protein.

In the present work, the temperature response of the folded states of several of these cavity containing variants and the pressure-induced unfolding at different temperatures was monitored using 2-D NMR spectroscopy. Results from these experiments were compared with volume changes at the folding transition temperature,  $T_m$ , and with the expansivity values of their folded states at low temperature, measured previously by pressure perturbation calorimetry (PPC).<sup>12</sup> The temperature dependence of the amide proton chemical shifts revealed a strong inverse correlation between the number of amide resonances exhibiting particularly small temperature coefficients, which are equated with strong intramolecular hydrogen bonds (H-bonds),<sup>18</sup> and the thermal expansion of the folded states determined by PPC. The specific patterns of H-bonds and the perturbations affected by mutations compared with the co-evolution of pairs of residues in the SNase family suggest how the protein's sequence evolved with respect to function.

## MATERIALS AND METHODS

**NMR.** All proteins were produced as described earlier.<sup>4</sup> Uniform <sup>15</sup>N labeling was obtained from overexpression of recombinant protein in *Escherichia coli* grown in M9 medium containing <sup>15</sup>NH<sub>4</sub>Cl as the sole nitrogen source, as described for SNase previously.<sup>19</sup> Uniformly <sup>15</sup>N-labeled  $\Delta$ +PHS SNase and its variants with either I92A, L125A, or I92A/L125A substitutions were dissolved at approximately 1 mM concentration in 300  $\mu$ L 50 mM Tris buffer at pH7. True wild-type SNase was dissolved at a similar concentration but in 50 mM Bis-Tris buffer at pH5.5. 10% of D<sub>2</sub>O was added for the lock procedure. In all experiments the <sup>1</sup>H carrier was centered on the water resonance, and a WATERGATE sequence<sup>20,21</sup> was incorporated to suppress solvent resonances. All NMR spectra were processed and analyzed with GIFA.<sup>22</sup> <sup>1</sup>H and <sup>15</sup>N resonance assignments were available for the wild-type SNase and for the  $\Delta$ +PHS protein and its variants with either I92A or L125A.<sup>4</sup> Amide resonances of the I92A/L125A variant were assigned at atmospheric pressure from 3D [<sup>1</sup>H,<sup>15</sup>N] NOESY-HSQC (mixing time 150 ms) and 3D [<sup>1</sup>H,<sup>15</sup>N] TOCSY-HSQC (isotropic mixing 60 ms) double-resonance experiments<sup>23,24</sup> recorded on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm Z-gradient <sup>1</sup>H–<sup>13</sup>C–<sup>15</sup>N cryogenic probe, using the standard sequential procedure. <sup>1</sup>H chemical shifts were directly referenced to the methyl resonance of DSS, and <sup>15</sup>N chemical shifts were referenced indirectly to the absolute frequency ratios <sup>15</sup>N/<sup>1</sup>H = 0.101329118.

Variable pressure experiments were recorded at four different temperatures (288, 293, 298, and 303 K) on a 600 MHz Bruker Avance III spectrometer equipped with a 5 mm Z-gradient <sup>1</sup>H-X



**Figure 1.** Intramolecular H-bonds variations among SNase proteins. (A) Amide proton temperature coefficients of WT,  $\Delta$ +PHS SNase, and of cavity enlarging variants with L125A, I92A, and I92A/L125A substitutions. Values above the solid line ( $-4.5 \text{ ppb/K}^{-1}$ ) are represented with filled symbols. (B) Histogram of  $\Delta\delta_{\text{NH}}/\Delta T$  for each protein. The solid line corresponds to  $-4.5 \text{ ppb/K}^{-1}$ ; see Materials and Methods. (C) Structures of WT,  $\Delta$ +PHS, and variants with substitutions L125A, I92A or I92A/L125A. Locations of amides with  $\Delta\delta_{\text{NH}}/\Delta T$  values above  $-4.5 \text{ ppb/K}^{-1}$  are red spheres. Residues that were deleted or mutated from WT to engineer  $\Delta$ +PHS SNase are colored black. Residues L125 and I92 in  $\Delta$ +PHS are colored blue. The molecular surface of the internal void volumes is shown in (light blue) and highlighted in dark blue when additional void is present due to mutation.

double-resonance broadband inverse (BBI) probe. Commercial zirconia ceramic high-pressure NMR tubes connected to an Xtreme 60 syringe automatic pump (Daedalus Innovations, Philadelphia, PA) were used to vary the pressure in the 1–2500 bar range. Fourteen 2D [<sup>1</sup>H, <sup>15</sup>N] HSQC spectra were recorded at variable pressure to monitor protein unfolding. Subdenaturing 1.8, 0.5, and 0.75 M guanidinium chloride concentrations were added to the NMR sample to achieve the complete denaturation of  $\Delta$ +PHS and variants with L125A and I92A, respectively, in the pressure range accessible by the instrumentation. The <sup>15</sup>N/<sup>1</sup>H cross-peak maximal intensities were used to construct the fractional intensity vs pressure plots for each residue. Cross peaks were picked using a box size of 0.3/0.03 ppm (<sup>15</sup>N/<sup>1</sup>H) with the PARIS algorithm<sup>25</sup> included in the GIFA software. Cross peak line width changes were minimal and uniform. Thermal coefficients were calculated as the slope of linear regression fits of <sup>1</sup>H chemical shifts recorded as a function of temperature at 1 bar.<sup>18</sup> Only residues with thermal coefficients calculated from linear regressions with  $R^2 > 0.95$

were considered (a total of 622 amides). Apparent residue-specific  $\Delta V$  were obtained as described previously.<sup>4,26</sup> Briefly, the HSQC peak intensity pressure profiles were fit to a two-state model, for the  $\Delta G^\circ$  and  $\Delta V^\circ$  values for unfolding (and the asymptotic intensity values) at each temperature, assuming that the difference in compressibility between folded and unfolded states was negligible. The residue-specific apparent folded-state expansion corresponds to the slope of the thermal dependence of residue-specific apparent  $\Delta V$  calculated from linear regression fits. For more detail on our previously published data analysis procedures see Supporting Information. Only residues with  $\Delta V$  for more than three temperatures and with linear regression coefficients,  $R^2 > 0.7$  were considered (a total of 477 amides with an  $R^2$  median of  $0.94 \pm 0.13$ ). The magnitude distribution of the residue-specific apparent folded-state expansion was independently normalized among variants and reported on the corresponding crystal structures via an in-house Python tool. Experimental data were analyzed using

ProFit (QuantumSoft) and Prism (Graphpad) software packages and plotted using ProFit (QuantumSoft).

**Equilibrium Thermodynamics.** Differential scanning calorimetry (DSC), PPC, and high-pressure fluorescence experiments were performed as described earlier.<sup>12,27</sup> Refer to Figures S4 and S5 for details.

**Sequence Analysis.** Sequences of 3872 staphylococcal nuclease homologues obtained from the NCBI PFAM database (PF00565)<sup>28</sup> were reduced to 3196 nonredundant sequences using Duplicate Finder Java standalone application. A subgroup of 1976 homologues sequences of amino acid length equal or higher than 100 were selected using Jalview v 1.5.0 to discard partial and or incomplete sequences. The selected subgroup was aligned using the online multiple sequence alignment (MSA) tool kalign.<sup>29</sup> Residual co-evolution among the MSA was estimated using mutual information (MI) using MISTIC online source.<sup>30</sup> The network of co-evolved pairs having MI score >9 were represented on 1SNc pdb structure using MISTIC interactive network view tool.<sup>30</sup> Refer to Figure S7 for MI statistical analysis details.

## RESULTS

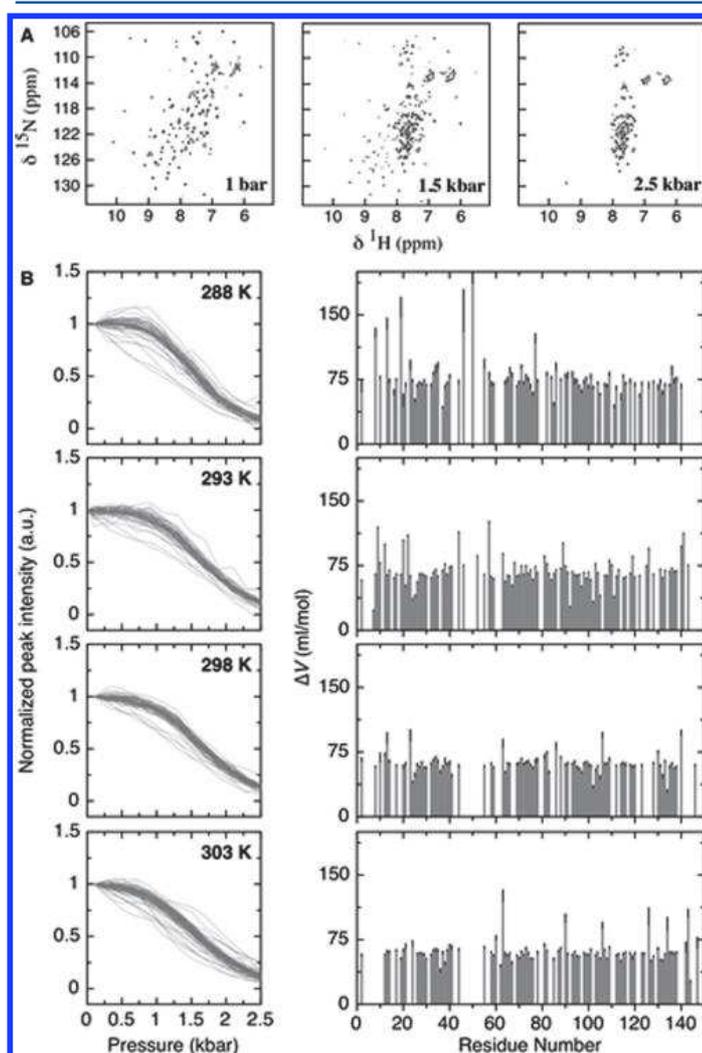
**Temperature Dependence of Amide Proton Chemical Shifts.** HSQC spectra were collected over a range of temperatures for wild-type (WT) and  $\Delta$ +PHS SNase and for the  $\Delta$ +PHS L125A,  $\Delta$ +PHS I92A, and  $\Delta$ +PHS L125A/I92A variants. The crystal structures of these proteins are nearly identical except for size and location of the internal cavities<sup>4</sup> and Table S1. The differences between the structures of WT and the  $\Delta$ +PHS protein are limited to the substitutions (G50F, F51N, P117G, H124L, S128A) and the deletion of the  $\Omega$ -loop (44–49) used to engineer the  $\Delta$ +PHS variants. The structure of SNase consists of an OB-fold domain (5 stranded  $\beta$ -barrel and  $\alpha$ -helix 1), a C-terminal  $\alpha$ -helical domain (helix 3), and an interfacial domain with several loops and  $\alpha$ -helix 2.

HSQC peak assignments of the backbone amides for all the variants have been previously reported,<sup>4</sup> except for the double variant (Figure S1).

As noted by Baxter and Williamson,<sup>18</sup> amide proton chemical shifts depend linearly upon temperature below the temperature of unfolding, and different amide protons in protein structures shift to different extents upon heating. This is the result of increased thermal motions. Small temperature coefficients ( $\Delta\delta_{\text{NH}}/\Delta T$ ) for the chemical shifts of particular amide protons (more positive than  $-4.5$  ppb/K) were interpreted by these authors as arising from constraints against expansion due to strong intramolecular hydrogen bonding for those residues. We used the histogram of  $\Delta\delta_{\text{NH}}/\Delta T$  for all the residues of the SNase variants to establish the strong H-bond cutoff (Figure S2) at  $-4.5$  ppb/K, as was done by Baxter and Williamson.<sup>18</sup> The temperature coefficients of the amide proton chemical shifts for the SNase variants (Figure 1A,B) revealed an increase in the number of the presumably strong intramolecular H-bonds in the  $\Delta$ +PHS variant, compared to WT SNase. Interactions were reinforced particularly in the region linking the C-terminal helix to the core of the protein, where the  $\Omega$ -loop was deleted in the  $\Delta$ +PHS variant (Figure 1C). The destabilizing cavity creating substitutions I92A and L125A in the  $\Delta$ +PHS protein lead to a nearly identical loss in global stability relative to the  $\Delta$ +PHS reference protein,<sup>4</sup> yet they exhibited very different perturbations to their H-bond pattern. Interestingly, the I92A substitution, deep in the core of the protein, had only a moderate effect on the number and distribution of strong H-bonds. In fact, the H-bonds in the core  $\beta$ -barrel region appear to have been reinforced by the enlargement of the cavity (Figure 1C). In contrast, the

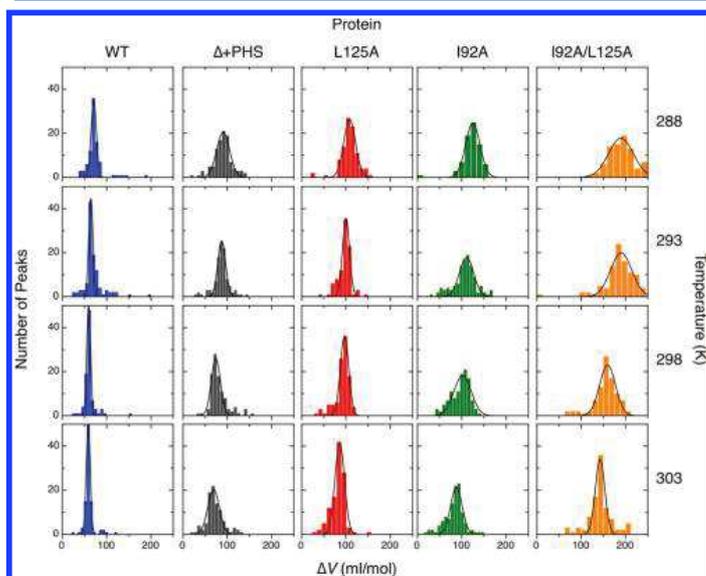
L125A substitution leads to a much larger decrease in the number of strong intramolecular H-bonds compared to the  $\Delta$ +PHS reference protein, with a pattern resembling that of WT SNase. These observations are consistent with the previously reported chemical shift perturbations between the reference protein and its I92A and L125A variants.<sup>17</sup> In those studies, very small chemical shift perturbations were observed for the I92A mutation, whereas considerable and long-range perturbations were apparent in the HSQC spectra of the L125A variant, relative to the reference protein. The I92A/L125A variant exhibits only a few residual amide protons with small temperature coefficients, indicating only a few residual strong H-bonds remain in the structure of this highly destabilized variant.

**Temperature Dependence of Pressure-Induced Unfolding.** The pressure-induced unfolding of WT SNase was characterized at four temperatures by observing the pressure dependent decrease in the intensity of the folded-state HSQC resonances (Figure 2). Each of the over 100 residue-specific pressure unfolding profiles at each temperature was fit to a two-state unfolding model to recover residue-specific apparent values for  $\Delta G_f^\circ(T)$  and  $\Delta V_f^\circ(T)$ , the free energy and volume changes for folding, respectively. The distributions for the



**Figure 2.** Temperature dependence of pressure induced unfolding of WT SNase monitored by HSQC peak intensity. (A) Representative HSQC spectra recorded at 1 bar, 1.5 kbar, and 2.5 kbar and 293 K. (B) Normalized intensity profiles and fitted  $\Delta V_f$  values of individual amides as a function of pressure recorded at 288, 293, 298, and 303 K.

apparent  $\Delta V_f^\circ(T)$  values (Figure 3) were fairly narrow for the WT protein, close to the uncertainty of the measurement,



**Figure 3.** Temperature dependence of the distribution of  $\Delta V_f$  values among SNase protein and cavity enlarging mutants. Bins of  $\Delta V_f$  values of SNase WT, the highly stable  $\Delta$ +PHS, and the corresponding cavity enlarging variants L125A, I92A, and I92A/L125A from all 2D  $^{15}\text{N}$ - $^1\text{H}$  HSQC fitted peaks recorded at 288, 293, 298, and 300 K. The line is a fit to a Gaussian function.

although a few outliers are apparent in the distributions, and some broadening occurred at low temperature. This indicates that equilibrium pressure unfolding of WT SNase can be modeled accurately as a two-state transition. It is known that the SNase WT folding mechanism follows a “foldon” type scheme, with the  $\beta$ -strand 5 and  $\alpha$ -helix 2 of the OB-fold representing the most stable and first folding unit.<sup>31</sup> A kinetic intermediate after the major folding barrier involves an ordered OB-fold region and disorder in the interface and much of the C-terminal helix.<sup>32</sup> We have observed both the foldon behavior and the population of this intermediate in p-jump NMR<sup>33</sup> and in pressure-dependent hydrogen-exchange experiments.<sup>26</sup> However, the equilibrium population of these intermediates under pressure is negligible for WT SNase, leading to the narrow distribution of the over 100 apparent volume changes obtained from the high-pressure NMR curves and allowing analysis according to a two-state model.

NMR-detected pressure-induced unfolding of the  $\Delta$ +PHS variant (Figure S3A) yielded site-specific unfolding profiles and site specific apparent values of  $\Delta G_f^\circ(T)$  and  $\Delta V_f^\circ(T)$  at each temperature (Figure 3). The distributions of the residue-specific apparent  $\Delta V_f^\circ(T)$  for the  $\Delta$ +PHS variant (Figure 3) were somewhat broader than for the WT SNase. At the higher temperatures they were rather asymmetric. This heterogeneity in the recovered parameters is consistent with some departure from a two-state transition for the highly stable  $\Delta$ +PHS variant, as previously reported.<sup>4</sup> Small apparent values for  $\Delta V_f^\circ$  obtained from the fits of the pressure dependence of the HSQC peak intensity of a given residue are indicative of partial unfolding involving that residue. Nonetheless, the distributions of apparent volume changes remain reasonably narrow, allowing us to consider the peak of the distribution to be a reasonable estimation of the thermodynamic  $\Delta V_f^\circ(T)$ . For both proteins the average value of  $\Delta V_f^\circ(T)$  decreased as a function of increasing temperature (Table 1). This is a general phenomenon observed for all proteins and is due to the smaller thermal expansivity of the folded state relative to that of the unfolded state, although deviations from two-state behavior can contribute as well to the temperature-dependent decrease in  $\Delta V_f^\circ(T)$ ; thermal expansion of the folded state of the  $\Delta$ +PHS variant is smaller than that of the WT SNase, as previously observed directly by PPC.<sup>12</sup>

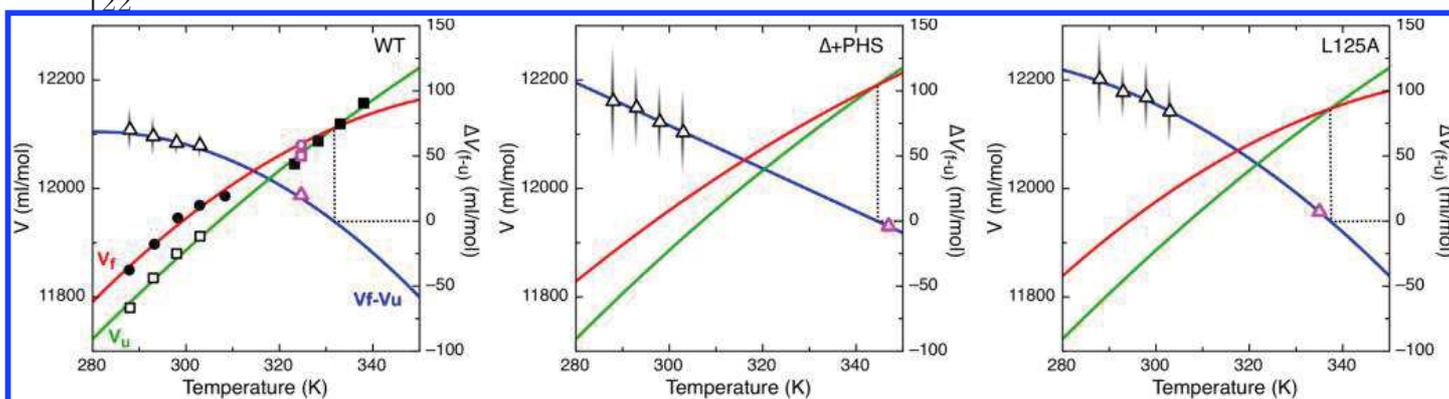
The difference in apparent expansivity between the folded and unfolded states,  $\Delta\epsilon$ , calculated from the pressure dependence of the average volume change of unfolding, was twice as large for the  $\Delta$ +PHS variant than for WT SNase. Because the differences in sequence are relatively modest, the unfolded-state expansivities are expected to be very similar for these two proteins. Hence the differences in  $\Delta\epsilon$  derived from the HSQC experiments must arise from a much smaller average folded-state expansivity for the  $\Delta$ +PHS variant.

The pressure-induced unfolding of the three cavity-containing variants of the  $\Delta$ +PHS reference protein was also monitored by 2D-NMR HSQC (Figures S3B–D) and Trp fluorescence (Figure S4) at four temperatures. The profiles were fit for apparent values of  $\Delta G_f^\circ(T)$  and  $\Delta V_f^\circ(T)$  at each temperature. The distributions of the NMR-detected site-specific  $\Delta V_f^\circ$  values at each temperature (Figure 3) all shifted to lower values as a function of increasing temperature. The  $\Delta V_f^\circ$  distributions for the L125A variant were relatively narrow and symmetric, although some broadening indicates minor deviation from two-state behavior. Nonetheless it was not unreasonable in this case to assume that the peak of the

**Table 1. Volume Changes upon Folding  $\Delta V_f$  at Different Temperatures and Their Temperature Dependence,  $\Delta\epsilon^a$**

protein	parameter	temperature (K)			
		288	293	298	303
SNase WT	$\Delta V_f$ (ml/mol)	70 (12)	65 (11)	60 (7)	58 (7)
	$\Delta\epsilon$ (ml/mol/K)			−0.86 (0.67)	
$\Delta$ PHS	$\Delta V_f$ (ml/mol)	93 (18)	87 (13)	76 (14)	68 (16)
	$\Delta\epsilon$ (ml/mol/K)			−1.98 (0.95)	
I92A	$\Delta V_f$ (ml/mol)	126 (16)	109 (20)	100 (17)	86 (18)
	$\Delta\epsilon$ (ml/mol/K)			−2.37 (1.76)	
L125A	$\Delta V_f$ (ml/mol)	109 (15)	99 (11)	95 (12)	84 (11)
	$\Delta\epsilon$ (ml/mol/K)			−1.52 (0.93)	
I92A/L125A	$\Delta V_f$ (ml/mol)	195 (27)	186 (22)	158 (20)	142 (19)
	$\Delta\epsilon$ (ml/mol/K)			−4.47 (2.11)	

<sup>a</sup>A linear temperature dependence of the  $\Delta V_f$  over the range probed is assumed. Uncertainties are given in parentheses for each parameter value.



**Figure 4.** Temperature dependence of volumetric properties of SNase variants. Left axis, partial molar volumes of folded (red line) and unfolded (green line) states of SNase variants folded (black circles). For WT SNase,  $V_u$  (black closed squares) is taken from densitometry measurements of WT SNase<sup>37</sup> at high temperature and extrapolates to the  $V_u$  values (black open squares) obtained from subtracting  $\Delta V_f$  (NMR) from  $V_f$  (filled black circles), measured also previously by densitometry at low temperatures. The  $V_f$  value for WT SNase at  $T_m$  (pink open circle) was obtained by adding  $\Delta V_f$  (pink open triangle) obtained at  $T_m$  by PPC to the  $V_u$  value (pink open square) obtained at  $T_m$  by interpolation of the  $V_u$  values from densitometry at high temperature. For  $\Delta$ +PHS and L125A,  $V_u$  is assumed to be identical to WT and  $V_f$  is calculated by adding to that the present  $\Delta V_{\text{NMR}}$  values. Right axis:  $\Delta V_{\text{NMR}}$  (black open triangles) and  $\Delta V_{\text{PPC}}$  (pink open triangles) of WT,  $\Delta$ +PHS, and L125A. Blue line is a quadratic fit to  $V_f - V_u$  for WT or L125A, and a linear fit of  $\Delta V$  values for  $\Delta$ +PHS. Vertical shadows correspond to the distributions of site specific  $\Delta V_{\text{NMR}}$  values.

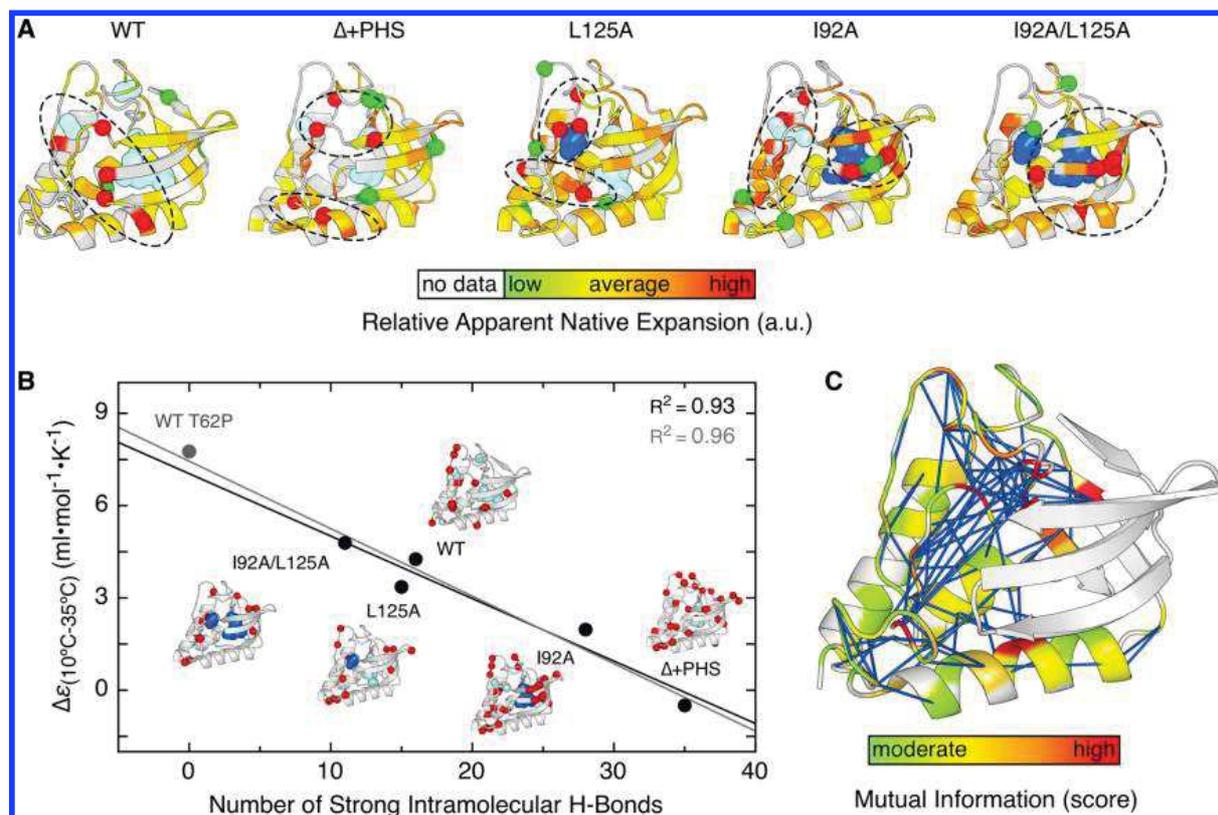
distribution represents a reasonable approximation of the true thermodynamic volume change. The average apparent difference in expansivity,  $\Delta\epsilon$ , calculated from the temperature dependence of the average  $\Delta V_f^o$  (Table 1) was intermediate between that observed for the WT and the  $\Delta$ +PHS reference protein, indicating that the folded-state expansivity of L125A is larger than that of  $\Delta$ +PHS, yet smaller than that of WT SNase.

In contrast to the near two-state behavior of WT SNase, the L125A variant, and  $\Delta$ +PHS, the variants with I92A and I92A/L125A substitutions both exhibited very broad  $\Delta V_f^o$  distributions, with those for the I92A variant at intermediate temperatures exhibiting bimodal character. Small apparent values for  $\Delta V_f^o$  obtained from the fits of the pressure dependence of the HSQC peak intensity of a given residue are indicative of partial unfolding involving that residue. The positions of these residues in the I92A structure (Figure S5) show that partial unfolding occurs in the C-terminal helix and its interface to the OB-fold domain, in addition to some disruption in the vicinity of the I92A substitution. This folding intermediate of SNase and several of its variants, disrupted in the C-terminal helix, has been reported previously.<sup>26,34–36</sup>

**Thermal Expansion and Volumetric Profiles.** Complete volumetric profiles of the folded and unfolded states of the SNase variants can be derived from the temperature dependence of the averages of the NMR based  $\Delta V_f^o$  values (Figure 4), as done previously for WT SNase based on  $\Delta V_f^o(T)$  values obtained from fluorescence detected unfolding.<sup>13</sup> This was possible only for those variants with quasi-two-state behavior (WT,  $\Delta$ +PHS, and L125A), for which the peaks of the  $\Delta V_f^o(T)$  distributions are assumed to approximate the thermodynamic values reasonably well. The black triangles in Figure 4 with dispersed points represent the distribution of values of  $\Delta V_f^o(T)$  obtained from the present NMR experiments. The values of the volume of the folded WT SNase,  $V_f$ , were measured directly by densitometry as a function of temperature and reported previously.<sup>37</sup> The values for the volume of the unfolded WT SNase,  $V_u$ , below 320 K were calculated by subtracting the average value of  $\Delta V_f^o(T)$  from  $V_f$ . The values of the molar volume of unfolded WT SNase above 320 K were measured previously directly by densitometry,<sup>37</sup> since WT SNase unfolds at  $\sim 325$  K. The pink triangle represents the value of  $\Delta V_f(T_m)$ ,

the volume change for folding at the transition temperature, obtained directly from PPC measurements<sup>12,38</sup> and Figure S6. Similar plots were constructed for  $\Delta$ +PHS and its L125A variant, assuming by convention that the unfolded-state volume is the same for all proteins. Different unfolded-state volumes would simply shift both curves up or down on the y-axis relative to WT SNase. We exclude the possibility of residual volume in the unfolded states of the variants obtained by pressure denaturation because pressure, by nature, favors the state of least volume.

Over the limited temperature range for which pressure-induced unfolding profiles were measured by NMR, the temperature dependence of  $\Delta V_f^o(T)$  ( $=\Delta\epsilon$ ) is approximately linear for all three variants. However, taking into account the volume change obtained from PPC, the temperature dependence of  $\Delta V_f^o$  deviates from linearity over a broader temperature range for WT and the L125A variant, although not for the  $\Delta$ +PHS protein. Deviation from linearity for  $\Delta V_f^o(T)$  is not surprising since the expansivities for the folded states of WT and L125A decrease significantly with increasing temperature. In contrast, the expansivity of the folded state of  $\Delta$ +PHS is rather low and constant over a broad temperature range,<sup>13</sup> and given that the expansivities of unfolded states are not strongly temperature dependent,<sup>39</sup> the linearity observed for  $\Delta V_f^o(T)$  of the hyper-stable variant is expected. The apparent expansivity of the unfolded states of the  $\Delta$ +PHS protein and of its L125A variant was considered to be the same as for the WT protein. This assumption is supported by several lines of evidence. First, identical expansivities of the unfolded states of these variants were measured directly by pressure perturbation calorimetry at temperatures above the unfolding transition.<sup>12</sup> Second, the zero point temperature for the  $\Delta V_f^o(T)$  plots for the three variants corresponds well to the crossover temperature of the  $V_f$  and  $V_u$  curves. That is the point above which pressure would favor folding rather than unfolding (dotted lines in Figure 4). Finally, the expansivity values for the folded state (slope of the red lines in Figure 4) deduced based on the assumption of equivalent unfolded-state expansivities and the  $\Delta V_f^o(T)$  values from the high-pressure NMR experiments ( $\sim 6$  mL/(mol·K) for  $\Delta$ +PHS and between  $\sim 8$  mL/(mol·K) at low temperature and 4 mL/(mol·K) at high temperature for the L125A variant) are in



**Figure 5.** Correlation of native expansion with intramolecular interactions. (A) Cartoon representation of the thermal dependence of site specific folded-state expansivity values, see Materials and Methods, of WT and  $\Delta$ +PHS SNase and of cavity enlarging variants with I92A, L125A and I92A/L125A substitutions. (B) Negative correlation between folded-state expansivity<sup>12</sup> obtained from PPC<sup>12</sup> and the number of strong intramolecular H-bonds derived from Figure 1. Gray and black lines are linear regression fits including or excluding SNase WT T62P data, respectively. Cartoons highlight in red spheres the backbone amide positions of residues implicated in strong intramolecular H-bonds. The molecular surface of the internal void volumes is shown in light blue and highlighted in dark blue when the size is altered by the mutation. (C) Mutual information interaction network of SNase family with MI values  $>6.5$  (significant) (low = 9; high = 26.5) represented as straight yellow sticks and colored according to their score. PDB codes for SNase,  $\Delta$ +PHS, I92A, L125A, and I92A/L125A proteins correspond to 1SNC, 3BDC, 3MEH, 3NXW, and 4DGZ, respectively.

reasonable agreement with the expansivities of the folded state measured for the  $\Delta$ +PHS and L125A variants directly by pressure perturbation calorimetry (5 mL/(mol·K) for  $\Delta$ +PHS and between 11 and 7 mL/(mol·K) for L125A.<sup>12</sup>

## DISCUSSION

Pressure-dependent NMR spectroscopy has the potential to contribute unprecedented site-specific structural insight concerning folding mechanisms and structural origins of stability and cooperativity under a variety of solution conditions (temperature, denaturant, pH). Assuming quasi-two-state behavior and that the expansion of the unfolded form is the same across all variants, one can interpret the temperature dependence of the site specific  $\Delta V_f^\circ$  (which provide residue specific values for  $\Delta\epsilon$ ) as a reflection of the local apparent thermal expansivity ( $\Delta\epsilon$ ) in the folded state. High values for differences in expansion between the folded and unfolded states correspond to low values of expansion of the folded state.

We showed previously by PPC that the global thermal expansivity of WT SNase in the folded state at 283 K is rather high, 16 mL/mol·K, approximately 3-fold higher than that of the  $\Delta$ +PHS variant, 5.8 mL/mol·K. The L125A and I92A variants have intermediate expansivity at 283, 12.2, and 11.5 mL/mol·K, respectively,<sup>12</sup> while the double mutant increases to 13.8 mL/mol·K (Figure S6), nearly that of WT SNase.

From the present high-pressure NMR data (Figure 5A) we observe that locally, WT SNase exhibits the highest thermal expansion values for residues at the interface between the C-

terminal helix 3 and the core of the protein and the nearby helix 1, all in the vicinity of the flexible  $\Omega$ -loop. The deletions and substitutions used to engineer the highly stable  $\Delta$ +PHS variant not only lower the overall thermal expansion 3-fold but they also disrupt the interfacial pattern of residues with the highest expansivity. For the L125A variant, the pattern of residues with the highest expansivity resembles that of WT SNase, with the exception of helix-1, indicating that creation of a cavity at the interface between helix-3 and the core by the L-to-A substitution disrupts interactions that constrain the  $\Delta$ +PHS protein against expansion. The strong deviation from two-state behavior of the I92A and I92A/L125A variants precluded establishing this type of correlation between structure and thermodynamics. However, the pattern of “apparent expansivity” for these variants, which is distributed throughout the structure, is consistent with temperature-dependent population of multiple intermediate states and a complex folding landscape, as previously shown for the I92A variant.<sup>4</sup>

The difference in thermal expansion between 285 and 308 K (5–35 °C),  $\Delta\epsilon_{10-35}$ , provides another good global measure of folded-state expansivity for all the SNase variants.<sup>12,13</sup> (Figure S6). We find a strong inverse correlation between the thermodynamic expansivity of the folded state of the variants and the number of strong intramolecular H-bonds in their structure, as deduced from the temperature dependence of the amide proton chemical shifts (Figure 5B). Including results for a T62P variant of WT SNase, which is constitutively unfolded in water, assuming it contains no strong H-bonds, reinforces

this negative correlation. These observations support the notion that these strong intramolecular H-bonds impose local constraints against thermal volumetric expansion; the stronger the interactions, the lower the expansivity.

Multiple sequence alignment of the SNase family of proteins and extraction of the mutual information (MI) network (Figures 5C and S7) revealed a large number of co-evolving residues at the interface between the core OB-fold domain and the C-terminal helix, whereas co-evolution within the OB fold domain was more limited. The strongest MI scores linked residues in the three interfacial loops to each other and to the N-terminus of helix-1. Interestingly, these are the same regions for which the stabilizing deletions and substitutions used to engineer the  $\Delta$ +PHS protein reinforce H-bonding, compared to WT SNase. These inverse correlations between H-bonding and co-evolution suggest that SNase has evolved to retain considerable flexibility in and around its active site (near the  $\Omega$ -loop). This flexibility is evident as significant local thermal expansion in these regions, measured by high-pressure NMR spectroscopy.

Data about co-evolving residues are often used to infer physical contact between two residues. In this case, the mutual information informs on a more complex requirement for the appropriate balance of interaction and dynamics. Indeed, in addition to flexibility, certain key interactions such as the active site clamp between D77-T120 have been conserved,<sup>40</sup> although here D77 only exhibits low native-state expansivity in the context of the  $\Delta$ +PHS variant. The notion that these patterns of conserved flexibility levels are key to tuning the functional properties of the protein is supported by the fact that the mutual information is strongest in and around the substrate binding site at the interface between the two subdomains (Figure S6).

## CONCLUSIONS

Two-dimensional NMR spectroscopy was used to examine structural origins of the temperature dependence of the volumetric properties of some proteins. The thermal expansivity of the protein appears to be anticorrelated with the presence of strong intramolecular H-bonds. Stabilizing substitutions that locally reinforce H-bonding impose constraints against thermal expansion, whereas local disruption of interactions via cavity creating mutations increases the thermal expansion of the native state in a site-specific manner. The effects of these perturbations to the hydrogen-bonding patterns of protein correlate well with the network of mutual information obtained from examination of the co-evolution of sequence within this protein family. This analysis revealed a link between the high thermal expansivity near the active site of this enzyme, with the known functional requirements for conformational flexibility in this region. This study shows how the characterization of volume and volumetric expansion with the site-specific resolution afforded by NMR spectroscopy can yield detailed information about intramolecular interactions essential for folding and function.

## ASSOCIATED CONTENT

### Supporting Information

Figure S1, Residue Specific assignments for backbone amide group of I92A/L125A double mutant. Figure S2, Histogram of  $\Delta\delta\text{NH}/\Delta T$  for all measured protein variants. Figure S3, Thermal dependences of pressure induced unfolding of  $\Delta$ +PHS protein and the corresponding variants L125A, I92A,

I92A/L125A as followed by 2D  $^{15}\text{N}$ - $^1\text{H}$  HSQC. Figure S4, Pressure unfolding curves followed by the center of spectral mass of Trp fluorescence of I92A/L125A protein recorded at 288, 293, 298, 303, and 308 K. Figure S5, Representation of the residues of I92A which exhibit low volume changes. Figure S6, DSC and PPC molar expansivity vs temperature calorimetric profiles of I92A, L125A, and I92A/L125A proteins. Figure S7, Co-evolving network of the SNase protein family. Figure S8, Temperature dependence of SNase WT protein  $\Delta V$  as calculated by NMR, fluorescence, and PPC. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b04320.

## AUTHOR INFORMATION

### Corresponding Authors

\*royerc@rpi.edu

\*christian.roumestand@cbs.cnrs.fr

### Present Addresses

<sup>§</sup>Structural Virology Department, Pasteur Institute, 28 rue du Dr. Roux, Paris, France 75015

<sup>||</sup>Laboratory of Physical Chemistry, National Institutes of Health, Rockville Pike, Bethesda, Maryland 20892, United States

<sup>†</sup>Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, New York 12218, United States

<sup>#</sup>904 Stellar-Chance Labs Department of Biochemistry & Biophysics University of Pennsylvania Perelman School of Medicine 422 Curie Blvd. Philadelphia, PA 19104-6059, United States

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by grants from the Agence National pour la Recherche PiriBio no. 09-455024 to CAR, from the National Science Foundation (MCB-0743422) to BGME and from the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01 to the Center for Structural Biochemistry in Montpellier, France.

## REFERENCES

- (1) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138.
- (2) Bell-Upp, P.; Robinson, A. C.; Whitten, S. T.; Wheeler, E. L.; Lin, J.; Stites, W. E.; Garcia-Moreno E., B. *Biophys. Chem.* **2011**, *159*, 217.
- (3) Rouget, J. B.; Aksel, T.; Roche, J.; Saldana, J. L.; Garcia, A. E.; Barrick, D.; Royer, C. A. *J. Am. Chem. Soc.* **2011**, *133*, 6020.
- (4) Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; Garcia-Moreno E., B.; Royer, C. A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 6945.
- (5) Bridgman, P. W. *J. Biol. Chem.* **1914**, *19*, 511.
- (6) Brandts, J. F.; Oliveira, R. J.; Westort, C. *Biochemistry* **1970**, *9*, 1038.
- (7) Meersman, F.; Smeller, L.; Heremans, K. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 346.
- (8) Winter, R.; Dzwolak, W. *Philos. Trans. R. Soc., A* **2005**, *363*, 537.
- (9) Zipp, A.; Kauzmann, W. *Biochemistry* **1973**, *12*, 4217.
- (10) Lin, L. N.; Brandts, J. F.; Brandts, J. M.; Plotnikov, V. *Anal. Biochem.* **2002**, *302*, 144.
- (11) Tsamaloukas, A. D.; Pyzocha, N. K.; Makhatadze, G. I. *J. Phys. Chem. B* **2010**, *114*, 16166.
- (12) Dellarole, M.; Kobayashi, K.; Rouget, J. B.; Caro, J. A.; Roche, J.; Islam, M. M.; Garcia-Moreno E., B.; Kuroda, Y.; Royer, C. A. *J. Phys. Chem. B* **2013**, *117*, 12742.

- (13) Mitra, L.; Rouget, J. B.; Garcia-Moreno E., B.; Royer, C. A.; Winter, R. *ChemPhysChem* **2008**, *9*, 2715.
- (14) Rholam, M.; Scarlata, S.; Weber, G. *Biochemistry* **1984**, *23*, 6793.
- (15) Scarlata, S.; Rholam, M.; Weber, G. *Biochemistry* **1984**, *23*, 6789.
- (16) Weber, G.; Scarlata, S.; Rholam, M. *Biochemistry* **1984**, *23*, 6785.
- (17) Roche, J.; Caro, J. A.; Dellarole, M.; Guca, E.; Royer, C. A.; Garcia-Moreno E., B.; Garcia, A. E.; Roumestand, C. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1069.
- (18) Baxter, N. J.; Williamson, M. P. *J. Biomol. NMR* **1997**, *9*, 359.
- (19) Castaneda, C. A.; Fitch, C. A.; Majumdar, A.; Khangulov, V.; Schlessman, J. L.; Garcia-Moreno E., B. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 570.
- (20) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661.
- (21) Sklenar, V. *Basic Life Sci.* **1990**, *56*, 63.
- (22) Pons, J. L.; Malliavin, T. E.; Delsuc, M. A. *J. Biomol. NMR* **1996**, *8*, 445.
- (23) Bax, A.; Pochapsky, S. S. *J. Magn. Reson.* **1992**, *99*, 638.
- (24) Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1989**, *28*, 6150.
- (25) Stoven, V.; Mikou, A.; Lallemand, J.-Y.; Piveteau, D.; Guittet, E. *J. Magn. Reson.* **1989**, *82*, 163.
- (26) Roche, J.; Dellarole, M.; Caro, J. A.; Guca, E.; Norberto, D. R.; Yang, Y.; Garcia, A. E.; Roumestand, C.; Garcia-Moreno E., B.; Royer, C. A. *Biochemistry* **2012**, *51*, 9535.
- (27) Dellarole, M.; Royer, C. A. *Methods Mol. Biol.* **2014**, *1076*, 53.
- (28) Finn, R. D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; Sonnhammer, E. L.; Tate, J.; Punta, M. *Nucleic Acids Res.* **2014**, *42*, D222–D230.
- (29) Lassmann, T.; Sonnhammer, E. L. *BMC Bioinf.* **2005**, *6*, 298.
- (30) Simonetti, F. L.; Teppa, E.; Chernomoretz, A.; Nielsen, M.; Marino, B. C. *Nucleic Acids Res.* **2013**, *41*, W8.
- (31) Bedard, S.; Mayne, L. C.; Peterson, R. W.; Wand, A. J.; Englander, S. W. *J. Mol. Biol.* **2008**, *376*, 1142.
- (32) Maki, K.; Cheng, H.; Dolgikh, D. A.; Roder, H. *J. Mol. Biol.* **2007**, *368*, 244.
- (33) Roche, J.; Dellarole, M.; Caro, J. A.; Norberto, D. R.; Garcia, A. E.; Garcia-Moreno E., B.; Roumestand, C.; Royer, C. A. *J. Am. Chem. Soc.* **2013**, *135*, 14610.
- (34) Jacobs, M. D.; Fox, R. O. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 449.
- (35) Kalnin, N. N.; Kuwajima, K. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 163.
- (36) Walkenhorst, W. F.; Green, S. M.; Roder, H. *Biochemistry* **1997**, *36*, 5795.
- (37) Seemann, H.; Winter, R.; Royer, C. A. *J. Mol. Biol.* **2001**, *307*, 1091.
- (38) Mitra, L.; Smolin, N.; Ravindra, R.; Royer, C.; Winter, R. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1249.
- (39) Schweiker, K. L.; Makhatadze, G. I. *Methods Enzymol.* **2009**, *466*, 527.
- (40) Carra, J. H.; Anderson, E. A.; Privalov, P. L. *Biochemistry* **1994**, *33*, 10842.

## 5.2 Resume de l'article en francais

### **Etude de l'origine de l'expansivité thermique des protéines dans leur état plié**

La seconde étude présentée dans cette thèse concerne l'expansivité thermique des protéines dans leur état plié par RMN haute pression. Cette étude suit des résultats publiés précédemment par notre équipe, liant l'expansivité thermique à la plasticité de l'état plier des protéines, après avoir éliminé l'hydratation des résidus de surface de la protéine comme déterminant majeur de l'expansivité thermique. Cette étude utilise une corrélation établie auparavant entre la faible dépendance du signal RMN à la température et la présence de liaisons hydrogène dans l'état plié. Nous utilisons cette propriété afin de relever un lien entre le nombre de pics RMN qui satisfassent le critère pour les liaisons hydrogène afin d'estimer le nombre de liaisons hydrogène fortes dans la protéine. Les résultats montrent qu'il existe une corrélation inverse entre la présence de nombreuses liaisons hydrogène fortes et l'expansivité thermique de l'état plier. De plus, la coévolution des résidus révèle que la séquence de la protéine SNase a évolué pour maintenir un certain niveau de flexibilité autour de son site actif, en accord avec l'opinion grandissante que la flexibilité des protéines est nécessaire pour leurs fonctions.

# Chapter 6

## High pressure study of the tryptophan cage variant Tc5b

### 6.1 Introduction

Tryptophan cage is one of the most studied protein in the protein folding field. It is a particularly valuable tool in Molecular Dynamics, because of its very short folding time scale. The ability to obtain a large number of folding/unfolding events makes it ideal to test experimental hypothesis. Using a combination of methods to test the variation in stability under various conditions and compare results with simulation predictions is important in the protein folding field to be able to converge experiments and simulations and thus enable visualization of the molecular mechanisms at the origin of the observed signal, as was realized here. Tryptophan cage is one of the fastest folding proteins, and its folded structure has been resolved by all atomic simulations from a fully extended state even before the NMR structure was determined[173].

This study was realized in collaboration with Angel Garcia, whose group has studied tryptophan-cage variants in numerous studies, including the effect of urea[118, 119], or other cosolvents[174, 124] of capping of the C and N terminal end of the protein[175], the determination of the pressure-temperature phase diagram[176, 177, 162], and the effects of sequence variation on the folding of tryptophan cage[177]. In this work, our post doc Soichiro Kitazawa realized the NMR acquisition and data treatment, while I worked on the data analysis of the molecular dynamic simulations, the interpretation of the NMR data, and wrote the article.

### 6.1.1 NMR signal interpretation

The interpretation of the NMR signal change with pressure in different solution condition is based on the work of Dr Kazuyuki Akasaka, one of the leading experts and pioneer in the high pressure NMR field. In his work on the interpretation of the change in chemical shift with increasing pressure, he exposed the origin of the amide proton signal dependence on pressure. In his interpretation, a linear dependence of the amide hydrogen chemical shift with increasing pressure was related to the compressibility of individual N-H covalent bonds[178, 179, 180]. A non-linear chemical shift change however was reported to have two possible origins. The first is linked to the creation of pressure induced excited states, meaning states where the compression of the bond is no longer linear, and results in a small conformational change. A second possible interpretation is the change in population between two states that are present in equilibrium in the experimental conditions and that have different compressibilities. Thus the shift in the population due to differential volumes between those states results in a overall compressibility change with increasing pressure and thus in a non-linear signal[178]. To differentiate between these two effects, one can thus use different solution conditions, and compare the change in chemical shift. If the non linear chemical shift dependence is also dependent on solution condition, it is unlikely that the effect results in a single state with pressure dependent compressibilities, because this would require this state to have a solution dependent compressibility, which is hardly believable, given the covalent nature of the bonds involved in the change in signal. Thus such changes are interpreted as resulting from a different equilibrium between states due to the different solution conditions, resulting in a different weight of the compressibilities of the species in play depending on their populations. This method is of course limited to states that are in fast exchange, and that have a different compressibility.

# NMR and Computation Reveal a Pressure-Sensitive Folded Conformation of Trp-Cage

Soichiro Kitazawa,<sup>†,#</sup> Martin J. Fossat,<sup>†,||,#</sup> Scott A. McCallum,<sup>§</sup> Angel E. Garcia,<sup>\*,‡,⊥</sup> and Catherine A. Royer<sup>\*,†</sup>

<sup>†</sup>Biological Sciences, Rensselaer Polytechnic Institute, Troy, New York

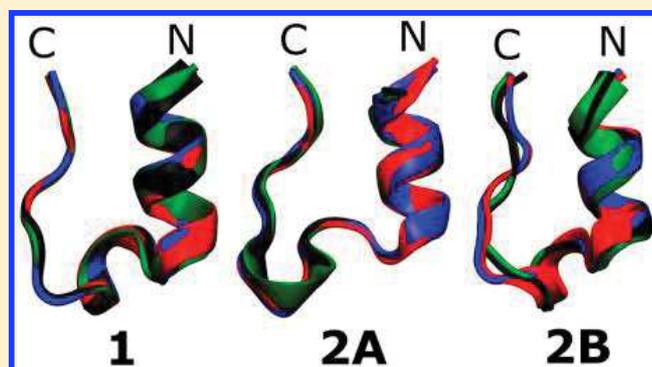
<sup>‡</sup>Department of Physics, Rensselaer Polytechnic Institute, Troy, New York

<sup>§</sup>Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York

<sup>||</sup>Laboratoire Charles Coulomb UMR 5221 CNRS-UM, Montpellier, France

## Supporting Information

**ABSTRACT:** Beyond defining the structure and stability of folded states of proteins, primary amino acid sequences determine all of the features of their conformational landscapes. Characterizing how sequence modulates the population of protein excited states or folding pathways requires atomic level detailed structural and energetic information. Such insight is essential for improving protein design strategies, as well as for interpreting protein evolution. Here, high pressure NMR and molecular dynamics simulations were combined to probe the conformational landscape of a small model protein, the tryptophan cage variant, Tc5b. Pressure effects on protein conformation are based on volume differences between states, providing a subtle continuous variable for perturbing conformations. 2D proton TOCSY spectra of Tc5b were acquired as a function of pressure at different temperature, pH, and urea concentration. In contrast to urea and pH which lead to unfolding of Tc5b, pressure resulted in modulation of the structures that are populated within the folded state basin. The results of molecular dynamics simulations on Tc5b displayed remarkable agreement with the NMR data. Principal component analysis identified two structural subensembles in the folded state basin, one of which was strongly destabilized by pressure. The pressure-dependent structural perturbations observed by NMR coincided precisely with the changes in secondary structure associated with the shifting populations in the folded state basin observed in the simulations. These results highlight the deep structural insight afforded by pressure perturbation in conjunction with high resolution experimental and advanced computational tools.



## INTRODUCTION

Detailed structural and energetic mapping of protein free energy landscapes requires appropriate combination of experiment and computation. The overlap between experiment and computation has increased in recent years due to faster data acquisition methodologies and more powerful computational approaches.<sup>1</sup> In the past decade, numerous computational studies of the designed 20-residue Trp-cage<sup>2,3</sup> peptide have been performed, largely because its folding time scale allows for a full sampling of folding/unfolding events within a reasonable computation time window.<sup>4–10</sup> In fact, the first structural model of Trp-cage was determined via all atom simulations, and correctly predicted the structure later confirmed by NMR.<sup>11</sup> This small peptide of sequence NLYIQWLKDGPGSSGRPPPS has been characterized experimentally,<sup>12</sup> in particular through NMR studies.<sup>13,14</sup> Prior computational and NMR studies on Trp-cage have focused on validating force fields<sup>9,15,16</sup> and solvation models,<sup>17</sup> and studying the effects of protonation,<sup>18</sup> capping,<sup>5</sup> kosmotropic and chaotropic agents,<sup>14,19–24</sup> and sequence variation.<sup>3,25</sup> Trp-cage has served as a very useful

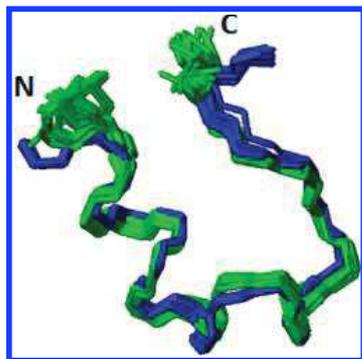
model for benchmarking molecular dynamics force fields.<sup>26</sup> The Tc5b Trp-cage variant is composed of an  $\alpha$ -helix (residues 1–8), a  $\beta$ -turn (involving a H-bond between residues 8 and 10), a 3–10 helix (residues 11–14), and a turn (residues 15–16) followed by a poly-proline segment (residues 17–20)<sup>2</sup> (Figure 1).

Pressure has emerged as a very useful approach for exploring protein conformational landscapes. Its effects are due to differences in molar volume between conformational states, with less ordered states generally presenting smaller molar volumes.<sup>27,28</sup> Recent experimental and computational studies have demonstrated the predominant role of solvent excluded void volumes in the sign and magnitude of volume changes associated with pressure-induced unfolding.<sup>29–31</sup> In addition to complete unfolding of proteins, pressure leads to the population of intermediate and excited states more readily

Received: November 23, 2016

Revised: January 17, 2017

Published: January 20, 2017



**Figure 1.** Representation of the backbone structure of Tc5b. Superposition of the best structural models obtained in this work (blue) as described in [Materials and Methods](#) and those previously reported (green) (PDB 1L2Y<sup>2</sup>).

than other perturbations due to the nonuniform distribution of solvent excluded volume in folded protein structures.<sup>31,32</sup> The folded state of Trp-cage exhibits minimal solvent excluded volume because the core of the protein is partially exposed to solvent. Moreover, the volume differences associated with disruption of helices are small, on the order of 1–2 mL/mol, and are difficult to detect accurately.<sup>33–35</sup> Thus, we did not expect that pressure would induce unfolding of Tc5b below the accessible pressure maximum of 300 MPa for high pressure NMR. Rather, pressure was used here to explore the conformational properties of Tc5b in its folded state basin.

We have investigated the consequences of pressure perturbation of Tc5b structure by combining high pressure (HP) 2D proton TOCSY NMR experiments with analysis of exhaustive molecular dynamics simulations. Experimentally, we observed that pressure resulted in modulation of the structures populated within the folded state basin. Principal component analysis of the MD simulations revealed two structural subensembles in the folded state basin, one of which was strongly destabilized by pressure.<sup>9,36–38</sup> The pressure-dependent structural perturbations observed by NMR were found to match precisely the regions implicated in the conformational transitions observed in the simulations.

## MATERIALS AND METHODS

**Structure Determination.** For structure determination, the protein solution was adjusted to a concentration of 4.2 mM in 100 mM AcOH buffer pH 5.0 containing 10% D<sub>2</sub>O at 282 K. For signal assignments, NMR experiments were performed at 282 K (pH 5.0) on an Avance III spectrometer (Bruker BioSpin Co., <sup>1</sup>H, 800 MHz) with a HCN cryo-probe. All proton signals were assigned by using 2D-NOESY, 2D-TOCSY, and HMBC experiments. <sup>1</sup>H chemical shifts were referenced to the methyl signal of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Data processing was done with TopSpin (Bruker BioSpin Co.) and NMRPipe.<sup>39</sup> Signals were assigned using NMRViewJ.<sup>40</sup> Structure calculations were performed with CYANA version 2.1<sup>41</sup> or NOE assignment and CNS 1.3<sup>42,43</sup> for structure refinement with distance constraints, which were obtained from 2D-NOESY experiments with a mixing time of 100 ms.

**NMR Measurements.** NMR experiments were performed on Avance III-600 (<sup>1</sup>H; 600 MHz) and Avance II-800 (<sup>1</sup>H; 800 MHz) spectrometers (Bruker BioSpin Co.). Assignment of all backbone and side chain atoms was performed by the analysis of proton correlation in 2D-TOCSY (mixing time, 80 ms) and 2D-NOESY (mixing time, 120 ms). Urea titration NMR

experiments were performed by collecting a series of 2D TOCSY spectra in steps of 0.75 M urea between 0 and 6 M urea at 298 K. pH-titration experiments were performed in 7% D<sub>2</sub>O and 93% water between pH 2.3 and 9.4. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid; 1 mM) was included as a pH-independent internal reference (0 ppm for <sup>1</sup>H). High pressure NMR experiments were performed between 1 and 2.5 kbar at 285 and 298 K in 100 mM *d*-acetic acid buffer pH 5.0 containing 7% D<sub>2</sub>O with 0, 0.5, and 1.0 M urea or in 100 mM *d*-Tris–HCl buffer pH 7.0 containing 7% D<sub>2</sub>O. The measurements were carried out using an Avance III-600 spectrometer (Bruker BioSpin Co.) equipped with a ceramic pressure resistant cell connected to an automated pressure pump (Daedalus Innovations).<sup>44</sup> Data was processed with the Topspin 2.6 and NMRPipe.<sup>39</sup> Data analysis was carried out by the programs of NMRView<sup>40</sup> and Kujira.<sup>45</sup>

**Simulations.** The tryptophan cage variant Tc5b was previously simulated<sup>9</sup> using a capped version (sequence Ac-NLYIQWLKDG GPSSGRPPPS-Nme) with charged lysine, arginine, and aspartic acid side chains. One Na<sup>+</sup> and two Cl<sup>−</sup> ions were added to the 2635 TIP3P water molecules<sup>46</sup> to neutralize the system. Electrostatic interactions were modeled using the particle mesh Ewald (PME) integration method with a cubic 36 × 36 × 36 grid and a van der Waals integration cutoff at 1.0 nm. A grid size of 0.12 nm was used for the PME. The equations of motion were integrated using the stochastic dynamics with a coupling time of 1 ps and using a 2 fs time frame. Simulations were performed using GROMACS and the ff99SB force field.<sup>47</sup> REMD simulations are done at constant volume in a cubic box of 4.42 nm corresponding to the volume of the system at a pressure of 1 atm and a temperature of 300 K, obtained from a 10 ns NPT (constant particle number, pressure, and temperature) simulation. We simulated 40 systems with temperatures of 280.0, 284.1, 288.2, 292.4, 296.7, 301.1, 305.6, 310.2, 314.9, 319.7, 324.6, 329.6, 334.7, 340.0, 345.4, 351.0, 356.6, 362.5, 368.4, 374.6, 380.9, 387.3, 394.0, 400.8, 407.8, 415.1, 422.5, 430.1, 438.0, 446.0, 454.3, 462.8, 471.6, 480.6, 489.8, 499.3, 509.0, 519.0, 529.2, and 539.7 K. Temperatures were selected such that an exchange rate of 0.15 was obtained. Simulations were extended to 1 μs per replica, and the last 0.5 μs were used for analysis. We also simulated the system at high average pressure over the same temperatures (in a cubic box of 4.176 nm), in order to get a large effective pressure difference between the two systems. The volume was reduced such that the average pressure at 310 K is 400 MPa. The initial state of the system was obtained from a 10 ns NPT simulation. The REMD simulation was extended for 1 μs per replica, and the last 0.5 μs are used for calculating averages.

In this study, we carried out further detailed analyses of these trajectories to compare the results with experiments done under similar conditions. Analysis was performed on ensembles obtained at temperatures in the range of those used for the NMR experiments for both densities, resulting in a large effective difference in pressure between the high and low density runs (Table 1). Here we use the C<sub>α</sub> RMSD distance from an experimentally determined structure of Tc5b (PDB code 1L2Y<sup>2</sup>) to distinguish folded (RMSD < 0.23 nm) and unfolded (RMSD > 0.23 nm) states.

In order to obtain an informative second reaction coordinate for characterization of Tc5b structures, principal component analysis was performed on the configurations in the ensembles. To ensure that all states were indeed present at atmospheric

**Table 1. Temperatures and Pressures of the Simulation Results Analyzed**

	pressure (bar)	temperature (K)
high density–high temperature	3876 ± 284	297.6
high density–low temperature	3575 ± 279	285
low density–high temperature	−500 ± 233	297.6
low density–low temperature	−676 ± 229	285

pressure and were not the result of negative pressures, replicas yielding an average pressure of 0.02 bar (closest to 1 bar) and a temperature of 330.5 K were analyzed as well. Single linkage clustering analysis was performed using a cutoff of 0.1 nm. For state 1, cluster analysis was repeated with a smaller cutoff (0.05 nm) in order to obtain more precision in the determination of the heterogeneity of this state under different conditions. Secondary structure propensity was determined, using a dictionary of secondary structure,<sup>48</sup> for states 1, 2A, and 2B. Block averaging (50 ns blocks) was performed to ensure that the states of interest were populated throughout the simulation and did not correspond to unique events. The proportion of folded proteins ( $C_{\alpha}$ RMSD < 0.23 nm) as well as the population of the state of interest were found to be similar for each block, confirming that the system was equilibrated (Figure S1).

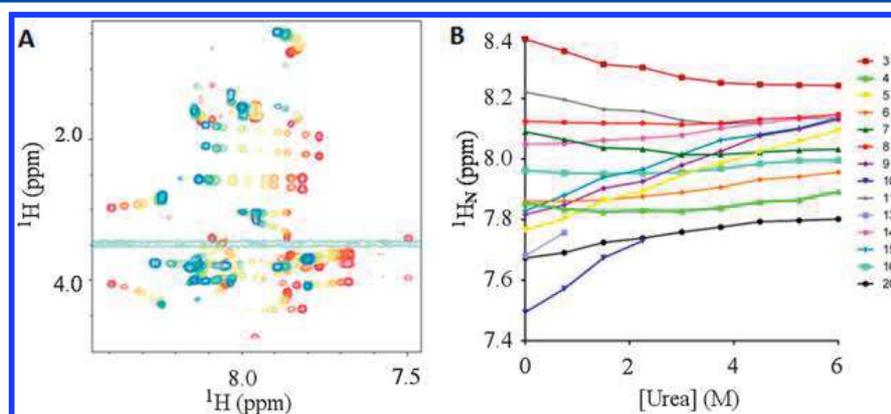
## RESULTS

The NMR structural models of Tc5b were calculated from proton and natural abundance <sup>13</sup>C NMR data (NOESY, TOCSY, HMBC) and yielded structures nearly identical to those previously determined by NMR (PDB code 1L2Y<sup>2</sup>) (Figure 1).

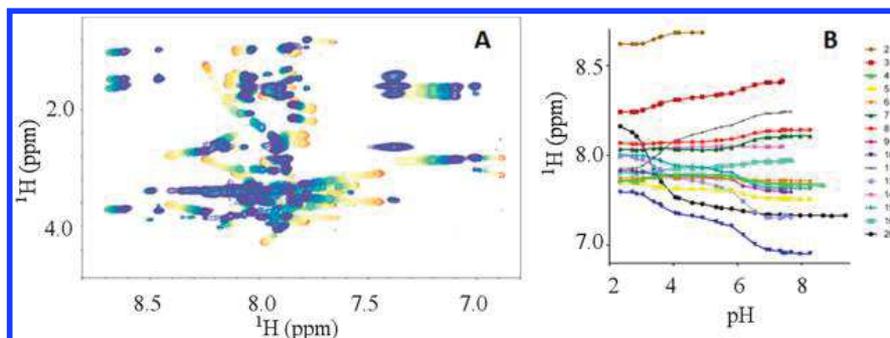
**The Urea Dependence of the Tc5b Amide Proton TOCSY Spectrum.** To evaluate structural changes in Tc5b brought about by changing solution conditions, we monitored the amide region of the 2D <sup>1</sup>H TOCSY spectrum (Figure 2). Urea titration resulted in changes in chemical shift with no significant intensity loss for most peaks, indicating a system in fast exchange with its unfolded state. Only residues 10, 11, and 13 exhibited losses in peak height due to peak broadening, indicating that these peaks transitioned to a slower exchange regime at a urea concentration of 2.25, 4.5, and 0.75 M, respectively, for these resonances. It was previously reported that a change in exchange regime occurred at high urea concentration, resulting in the appearance of doubled peaks for some residues, the existence of which was proposed to be due

to proline 12 cis–trans isomerization.<sup>14</sup> Although this particular behavior was not observed in the present study, final chemical shift values at 6 M urea were consistent with an unfolded state, confirming the results of previous studies.<sup>14,20</sup>

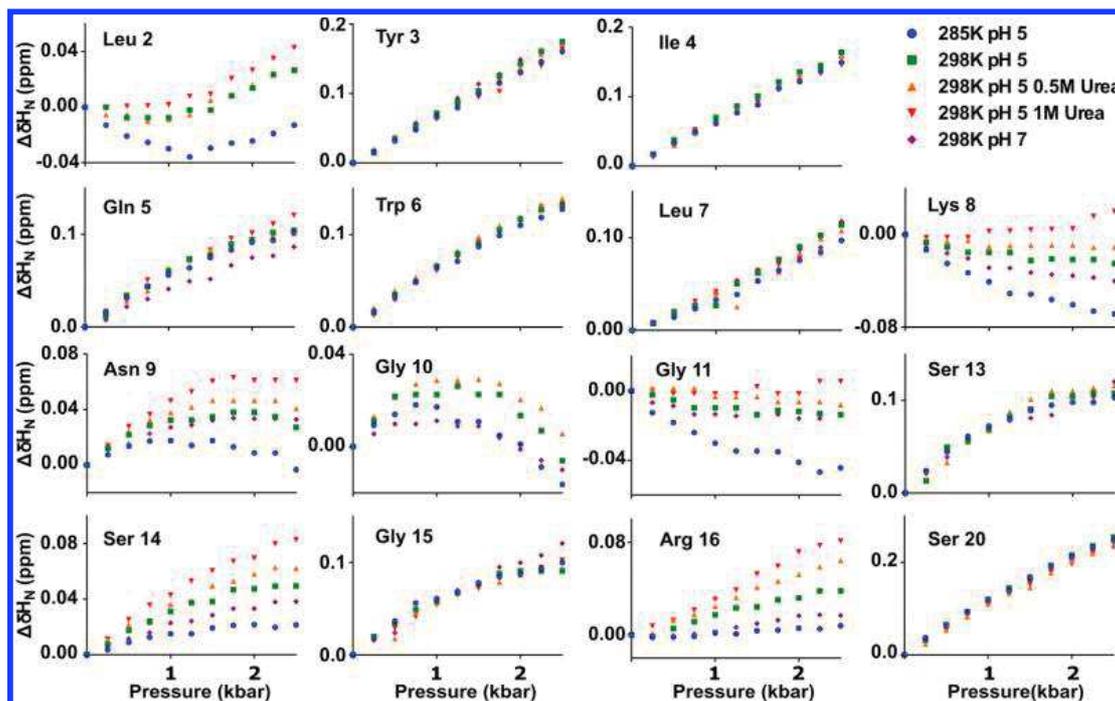
**The pH Dependence of the Tc5b Amide Proton TOCSY Spectrum.** Tc5b shows a strong pH dependence of its TOCSY spectrum, with very low spectral dispersion at low pH (<3), suggesting a random coil conformation (Figure 3). The system appears to undergo a sharp transition between pH 3 and 4.5 with an increase in the signal dispersion. Another transition is apparent in the pH 6–7 range, resulting in an even greater spectral dispersion. The first transition between pH 3 and 4.5 affects most residues (3, 4, 5, 9, 10, 11, 13, 15, and 16) in a cooperative manner. It has been amply demonstrated that chemical shifts can be perturbed upon protonation or deprotonation of rather distant residues, as through space effects are far from negligible.<sup>49–52</sup> Deprotonation of the carboxyl group of the side chain of aspartate 9 is not likely to be responsible for this transition, as the change in chemical shift for the amide proton for this residue was rather small in the pH range of 3–4. However, we note that the amplitude of the observed transitions is proportional to the chemical shift difference at that site. The absence of an observed transition at a particular site does not necessarily imply that it is not involved in a pH-dependent conformational change. Nonetheless, it is more likely that this transition is due to the deprotonation of the carboxyl group of the C-terminal residue, serine 20, since its amide proton resonance undergoes a large change in chemical shift in this pH range. It was reported previously that capping of the C-terminal carboxyl group decreased the stability for the Tc10b variant.<sup>3</sup> The second transition, between pH 6 and 7, is observed via changes in amide chemical shift for all residues, except residues tryptophan 6, serine 14, and serine 20. It is unclear what is responsible for this transition, leading to the fully folded state. Deprotonation of arginine 16 or lysine 8 would require perturbation of their pK<sub>a</sub> values by more than 3–7 pH units from their normal values.<sup>53</sup> The transition could arise from protonation of the aspartate 9–arginine 16 salt bridge or a potentially stabilizing glutamine 5–aspartate 9 interaction, observed previously in  $\alpha$ -helices.<sup>54</sup> In any case, the spectral dispersion increases significantly at each transition, indicating that both transitions reflect the population of a more organized, folded ensemble, in agreement with the previously reported difference of stability between pH 2.5 and 7 of a large number of tested Trp-cage variants, including Tc5b.<sup>3</sup> It is



**Figure 2.** Effects of urea on the Tc5b proton TOCSY spectrum. (A) Urea effects on the amide region of the H-TOCSY spectrum of Tc5b. Low urea concentration is displayed in red, and high urea concentration is displayed in blue. (B) Chemical shift variation per residue as a function of urea concentration; a color code corresponding to residue number is indicated in the panel on the right.



**Figure 3.** Effect of pH on the Tc5b proton TOCSY spectrum. (A) Amide region of the H-TOCSY spectra of Tc5b as a function of pH. Low pH is displayed in blue, and high pH is displayed in red. (B) Chemical shift variation per residues as a function of pH. The color code corresponding to residue number is indicated in the panel on the right.



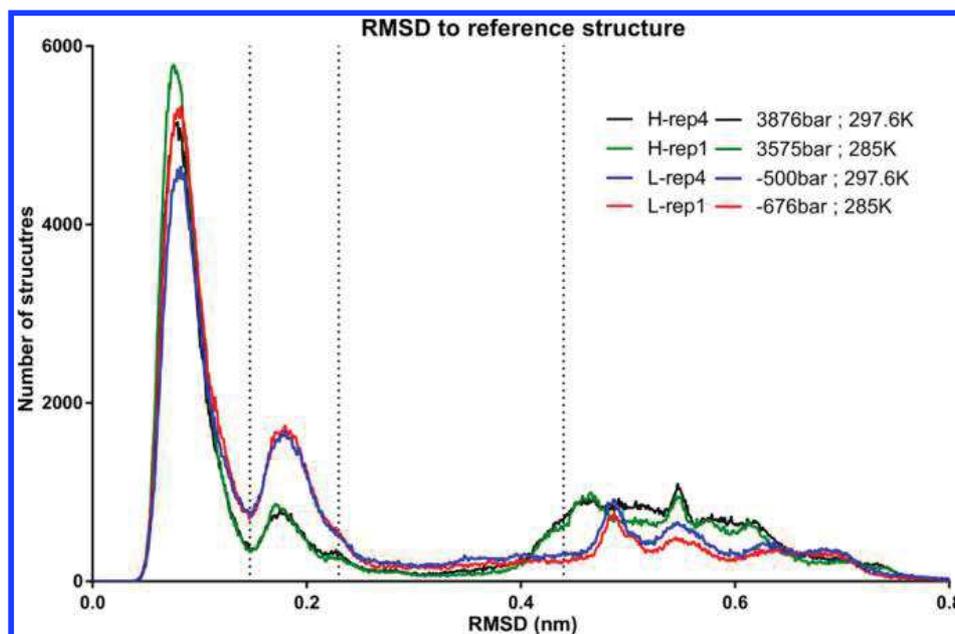
**Figure 4.** Relative chemical shift changes as a function of pressure for each amide proton resonance of Tc5b. Residue type and number are displayed on top of each graph. The color code is shown in the top right corner.

interesting to note that the tryptophan amide proton resonance shows no change over the entire pH range tested, despite the fact that the mini-protein folds in this range.

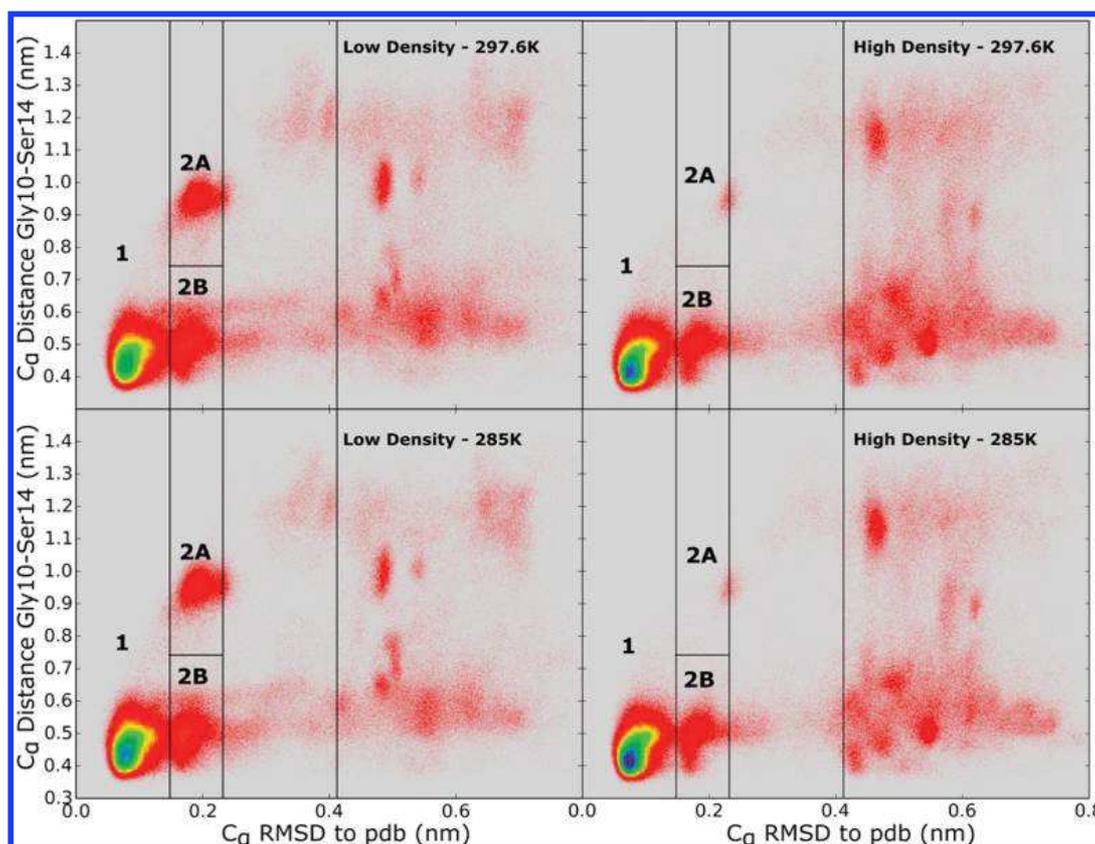
**Pressure Dependence of the Tc5b Amide Proton TOCSY Spectrum.** The pressure dependence of the NMR TOCSY spectrum of Tc5b was measured at different pH values and temperatures and in the absence and presence of low concentrations of urea (Figures S1–S3). Under all conditions, the amide region of the proton TOCSY spectrum shows a small increase in the spectral dispersion with increasing pressure (Figure 4). Although downfield shifts, associated with intramolecular compressibility,<sup>55</sup> are observed with increasing pressure as expected, the behavior of the peaks is heterogeneous, indicating a change of conformational state with no clear change in the overall degree of structure. Three distinct categories of pressure-dependent behavior of the amide proton chemical shifts were observed. Residues 3, 4, 6, 7, and 20 show a nearly linear pressure dependence of the chemical shift which does not change with changing solution conditions. Residues 5, 13, and 15 show a curvature in the chemical shift pressure dependence but little or no sensitivity to changes in the solution composition and temperature. Residue 16 exhibits a

linear pressure dependence in the chemical shift which is modified in amplitude by the solution conditions. Finally, residues 2, 8, 9, 10, 11, and 14 exhibit a nonlinear chemical shift pressure dependence that is also strongly dependent on the solution conditions. As observed in the pH titrations, the pressure dependence of the chemical shift of the amide proton of the tryptophan residue of Tc5b is linear and very insensitive to solution conditions.

A linear pressure dependence of chemical shift is generally considered to arise from intramolecular compression and increased hydration upon increasing pressure with no significant change in conformation.<sup>55</sup> Curvature of the chemical shift upon application of pressure to a system in fast exchange is commonly interpreted as being due to the population of pressure induced folded-like excited states.<sup>55</sup> Finally, large effects of solution composition on the pressure-dependent chemical shift have been interpreted as arising from pressure and solution condition dependence of the population of pre-existing states.<sup>55</sup> These states are stabilized or destabilized by changes in pH, temperature, or denaturant concentration based on the difference in accessible surface area (ASA) and/or



**Figure 5.** Analysis of MD simulation results as a function of pressure and temperature, as indicated in the panel. Distributions of the  $C_{\alpha}$ -RMSD with respect to the PDB coordinates are plotted for four conditions of pressure and temperature. Red and blue lines correspond to low density simulations at 285 and 297.6 K, respectively, and green and black lines correspond to high density simulations at 285 and 297.6 K, respectively. Dotted lines show the boundaries chosen for RMSD based discrimination of structures at RMSDs of 0.147, 0.23, and 0.44 nm.



**Figure 6.** Analysis of results of MD simulations off Tc5b as a function of pressure and temperature. Histograms of the  $C_{\alpha}$  Gly10H<sub>N</sub>–Ser15H<sub>N</sub> distance are plotted vs  $C_{\alpha}$ RMSD with respect to the PDB coordinates (PDB 1L2Y<sup>2</sup>) at low (left) and high (right) density and at two temperatures (top, 297.6 K; bottom, 285 K). Black lines represent the boundaries chosen for RMSD based discrimination of structures at RMSDs of 0.147, 0.23, and 0.44 nm for the vertical lines and Gly10H<sub>N</sub>–Ser15H<sub>N</sub> distance of 0.75 nm for the horizontal lines.

protonation state. The effects of pressure on the population of these states depend on their volumetric properties.

**Analysis of the Pressure Dependence of Replica Exchange Simulations of Tc5b.** The changes in the chemical shifts of the backbone amide protons of Tc5b as a function of pH, temperature, urea, and pressure indicated

significant shifts in the conformational ensemble as a function of solution conditions. To gain structural insight into these spectral changes, we analyzed the configurations obtained from previously reported replica exchange molecular dynamics simulations of a capped version of Tc5b at different pressures and temperatures.<sup>5</sup> The different populations present in the

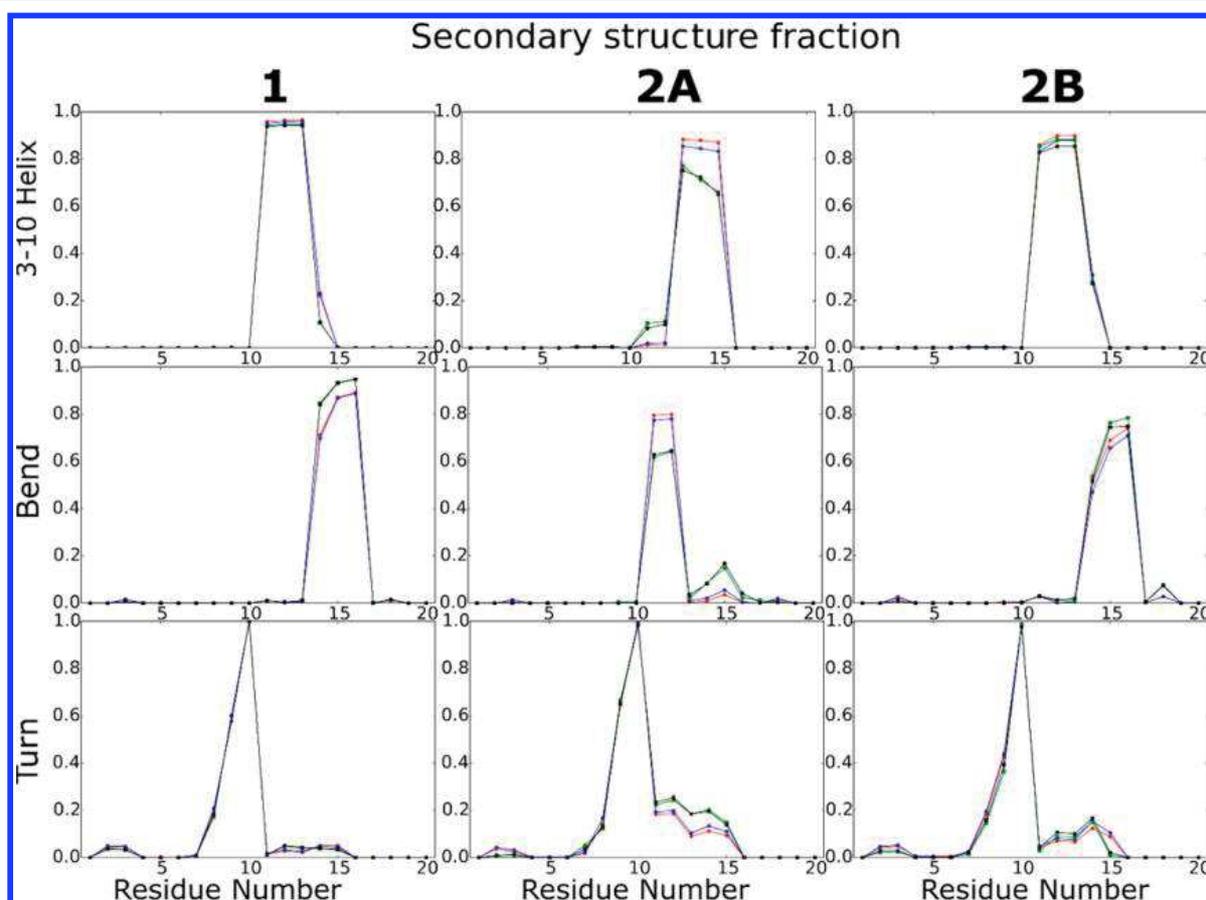
Table 2. Percentage of Sampled Conformations Found in Each State in the Simulations

	low density - rep1	low density - rep4	high density - rep1	high density - rep4
1	51.5	45.4	51.7	46.2
2A	7.2	6.3	0.3	0.4
2B	12.0	12.4	8.0	8.0
3	7.3	8.8	0.3	0.4
4	21.9	27.0	35.8	40.2

Table 3. Percentage of Structure Found in the Main Cluster of Each State<sup>a</sup>

	low density - rep1	low density - rep4	high density - rep1	high density - rep4
1	99.9	100	99.9	99.9
2A	98.6	97	65.8	94.8
2B	98.6	98.6	94.5	99.5

<sup>a</sup>State 2A of the high density rep1 shows lower values that can be explained by the presence of a second cluster with a population of 25% that differs only from the main cluster in the unfolding of leucine 2.



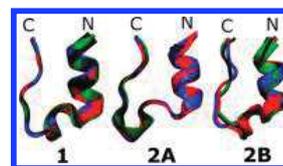
**Figure 7.** Comparison of secondary structure for Tc5b as a function of temperature and pressure. Plots represent the fraction of secondary structure content by residue for each state as defined in Figure 6, for each type of secondary structure, 3–10 helix, bend, and turn (indicated on the left). Points and line are colored by conditions, with blue, red, black, and green representing low density 297.6 K, low density 285 K, high density 297.6 K, and high density 285 K.

simulated ensembles are apparent in plots of the population distributions obtained at 285 and 297.6 K and at high and low density (Table 1) as a function of the RMSD from the first structure deposited in the PDB file (1L2Y<sup>2</sup>) (Figure 5). A broad distribution of unfolded Tc5b at RMSD > 0.23 nm is observed. This unfolded region of the RMSD distribution shows two distinguishable states. State 3 corresponds to a sparsely populated intermediate state ensemble between the folded ensemble (RMSD < 0.23 nm) and the unfolded ensemble, state 4 (RMSD > 0.44 nm). State 3 is more populated at low pressure, while state 4 is more populated at high pressure, indicating that pressure increases the coopera-

tivity of the (un)folding reaction for Tc5b. The population of both states is enhanced at higher temperature. These observations are consistent with previous analysis of these data sets<sup>9</sup> in which it was determined that Tc5b would unfold under pressure, regardless of the temperature. However, due to the small calculated volume change for unfolding (−1.94 mL·mol<sup>−1</sup>), the simulations did not predict significant population of the unfolded state below 5 kbar, a pressure which is beyond our experimentally accessible pressure range. Hence, in our comparison of the experimental and computational results, we focus on pressure effects on the folded state ensemble of Tc5b.

A broad distribution, within the folded state manifold, exhibiting an RMSD below 0.23 nm is clearly evident in Figure 5 under all conditions. The amplitude of this peak is strongly pressure-dependent. To gain further structural insight into the properties of this ensemble, principle component analysis (PCA) was carried out for all four conditions examined. The first eigenvector of the covariance matrix revealed a strong variance of the distance between the  $C_{\alpha}$  atoms of residues 10 and 14 which define the beginning and the end of the small 3–10 helix linking the  $\alpha$ -helix and the poly-proline helix of Tc5b (Figure 1). Consequently, the distance between the  $C_{\alpha}$  of these two residues was chosen as a second reaction coordinate. A heat map (Figure 6) based on the PCA of the configurations reveals that, in addition to the fully folded conformation (state 1), the peak near 0.2 RMSD in Figure 6 is composed of two states (2A and 2B), which differ in the distance between the  $C_{\alpha}$  atom of residues 10 and 14. This distance is 0.4–0.6 nm for state 2B, while that of state 2A is larger (0.8–1.1 nm). Heat maps obtained from block averaging (50 ns) of the  $C_{\alpha}$ RMSD vs the distance between residues 10 and 14  $C_{\alpha}$  atoms in the low density replica confirmed that states A and B were present in all cases (Figure S4). We also confirmed that these states were present at a density corresponding to a pressure of 0 bar (Figure S5). The fully folded state 1 was relatively insensitive to pressure (Table 2), exhibiting only a sharpening of the population peak at high density, with a slightly lower root mean square fluctuation among the structures (Figure S6). The two other peaks, states 2A and B (0.147 nm < RMSD < 0.23 nm) both show a decrease in population with increasing pressure. However, state 2A was an extremely pressure sensitive state with almost no population in the high density data sets (Table 2).

Single linkage clustering analysis was performed for all states obtained from the PCA for all four conditions to assess the structural diversity of these ensembles. Cluster analysis of state 2B yielded a consistent structure throughout the data set and revealed a difference in structure for residues 15–20, including a break of the Asp9–Arg16 salt bridge at high pressure, as opposed to both states 1 and 2A, which retain this salt bridge. The secondary structural organization of states 1, 2A, and 2B at both temperatures and pressures (Figure 7) reveals that state 1 is very similar to previously reported folded structures for Tc5b. A small effect of pressure is observed on the propensity of bends in state 1 (residues 14–16). State 2B is characterized by a slightly higher propensity for residue 2 to be in a coil rather than in an  $\alpha$ -helical state compared to state 1, and also exhibits a slightly lower content of bend and 3–10 helix, compared to state 1, and a slight increase in the propensity for residue 16 to be in a coil-like conformation. The turn structure at residue 10 is present in 100% of structures in all three states. Most interestingly, the highly pressure-sensitive state 2A shows an inversion in the position of the 3–10 helix (residues 13–15) and the bend (residues 11–12) compared to states 1 and 2B (3–10 helix - residues 11–13, bend - residues 14–16). The structures of the centroid of the main clusters for states 1, 2A, and 2B are displayed in Figure 8, and the proportion of structures that belong to the main cluster for each state is given in Table 3. While the positions of secondary structural elements in state 2A are swapped compared to states 1 and 2B, the total secondary structural content remains constant (Figure S7). Interestingly, the strong pressure destabilization of state 2A suggests that the succession of turn and 3–10 helix, found in



**Figure 8.** Main centroid structure for each state of the Tc5b native state ensemble. Overlapped structures represent the different simulation conditions. Colors correspond to the conditions, with blue, red, black, and green representing low density 297.6 K, low density 285 K, high density 297.6 K, and high density 285 K.

states 1 and 2B, achieves better packing than the presence of a bend after the turn of residue 10, as observed in state 2A.

Radial distribution functions of the water molecule oxygen atom around the NH of each residue in each state were averaged for all conditions and displayed by state in order to show the difference in the hydration between the structures among states 1, 2A, and 2B (Supplementary Figure S8). Significant differences in the RDF are observed for residues 11–16, precisely where the secondary structural swapping occurs. Hydration appears to be favored for residue 14 in state 2A and for residues 13 and 15 in state 1. These changes in hydration can be explained by the increased exposure of the amide proton of residue 14 in the 3–10 helix of state 2A, whereas residues 13 and 15 are more exposed when the bend and helix positions are swapped in state 1 relative to state 2A (Figure S9).

#### Comparison between Experiment and Simulations.

The results presented here show a remarkable agreement between the effects of pressure on Tc5b observed in the chemical shifts of amide proton TOCSY NMR and in all atom replica exchange simulations. Qualitatively, the pressure-dependent variation in the relative population of states 1 and 2A revealed by the simulations provides a structural interpretation for the solution-dependent nonlinear pressure response of several amide proton resonances. Together these observations indicate an important pressure-dependent shift in conformational equilibrium between pre-existing states. These changes were most apparent for residues 8–15. Residues 11–15 are those found in the analysis of the simulations to be implicated in the switched positions of the bend and 3–10 helix between states 1 and 2B with state 2A, the latter of which is destabilized by pressure. Residues 8–10 exhibit no significant conformational changes between states 1 and 2A in the simulations, despite a clear nonlinear pressure dependence of their chemical shifts. However, it should be noted that those three residues are involved in the  $\beta$ -turn structure (with a hydrogen bond between residues 10, donor, and 8, receptor). Moreover, the proximity of residues 8–10 to the rather large conformational changes between states 1 and 2A may render their chemical shifts sensitive to such changes. A similar conformation of Tc5b has been recently reported in another simulation study,<sup>56</sup> and experimental evidence suggests the existence of a folding intermediate implicating residues 11–15,<sup>36,57</sup> consistent with possible conformational heterogeneity of this region. The nonlinear pressure dependence of the chemical shift of residue 2 is likely the result of pressure-dependent fraying of the N-terminus of the  $\alpha$ -helix, as can be seen in state 2B, which exhibits a small pressure-dependent decrease in helix propensity. Due to very small volume changes associated with helix–coil transitions,<sup>33–35</sup> the effects of pressure on total helical content are quite small. Moreover, small differences between experiment and simulations at the N-

and C-termini of the protein might be expected, since the variant used here for the MD simulations was capped at its termini. While the simulations predict an increase in the unfolded population, the pressures attained in the simulations are significantly higher than those in the experiments (3500 bar vs 2500 bar). A small change in unfolded state population at 2500 bar would not likely result in a clear variation in chemical shifts, which at this pressure is dominated by changes within the folded basin.

## CONCLUSION

The combination of residue specific high-pressure NMR data with molecular dynamics simulations conducted at both high and low density allowed for the detailed characterization of structural changes related to the pressure dependence of the Tc5b conformation. The pressure-dependent changes observed in both experiment and simulation are dominated by the change in population within the folded ensemble. The folded state basin of Tc5b was shown to be composed of at least three species. One of these folded states, state 2A, while retaining the same total amount of each secondary structural element present in the ground state, exhibits a swap in their relative positions. Moreover, this alternate folded state of Tc5b is very pressure sensitive. Detection of a significant proportion of an excited folded state with distinct organization of secondary structural elements is essential for understanding the structural and energetic properties of this model peptide. State 2A, because it does not exhibit changes in total secondary structural content, might be hidden in common reaction coordinates and NMR observables.

The site-specific information obtained in the present studies provides deep structural insight into the conformational landscape of this model protein. The results demonstrate that the combination of high pressure with atomic resolution approaches such as NMR and molecular dynamics simulations allows for the population and detailed structural characterization of excited conformational states of proteins. Further investigation of the nature and presence of the swapped 3–10 helix Tc5b state under a broader range of conditions should be very helpful for benchmarking of molecular dynamics force fields and for better understanding the role of hydration in protein stability and dynamics. Moreover, understanding how sequence modulates the probability of population of such low lying excited states has implications for understanding functional evolution and for protein design.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.6b11810.

Figure S1, pressure titration spectra for the 298 K pH 7 conditions; Figure S2, pressure titration spectra for the 285 K pH 5 conditions; Figure S3, pressure titration spectra for the 282 K pH 5 conditions; Figure S4, block averaging of the low density replica 1 simulation run; Figure S5,  $C_{\alpha}$  distance Gly10H<sub>N</sub>–Ser15H<sub>N</sub> vs  $C_{\alpha}$ RMSD to PDB structure (PDB 1L2Y) for replica 11, yielding an average pressure of 0.02 bar and a temperature of 330 K; Figure S6, root mean square fluctuation for state 1; Figure S7, fraction of secondary structure content by residue for each state, as delimited in Figure 6; Figure S8, radial distribution function of water oxygen around H<sub>N</sub> of

each residue averaged over all data sets for states 1 and 2A; Figure S9, comparison of the positions of the amide hydrogen between state 1 and state 2A for residues 13, 14, 15, and 16 (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

\*Phone: (505) 667-3883. E-mail: [agarcia@lanl.gov](mailto:agarcia@lanl.gov).

\*Phone: (518) 276-3796. E-mail: [royerc@rpi.edu](mailto:royerc@rpi.edu).

### ORCID

Angel E. Garcia: 0000-0002-0912-0960

Catherine A. Royer: 0000-0002-2670-3391

### Present Address

<sup>†</sup>A.E.G.: Center for Nonlinear Studies, CNLS MS B258, Los Alamos National Laboratory, Los Alamos, NM 87545.

### Author Contributions

<sup>#</sup>S.K., M.J.F.: These authors contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by grants MCB 105966 and MCB 1514575 from the National Science Foundation to A.E.G. and C.A.R., respectively, and by the U.S. Department of Energy LDRD program at LANL.

## REFERENCES

- (1) Prigozhin, M. B.; Gruebele, M. Microsecond Folding Experiments and Simulations: A Match Is Made. *Phys. Chem. Chem. Phys.* **2013**, *15* (10), 3372–3388.
- (2) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9* (6), 425–430.
- (3) Barua, B.; Lin, J. C.; Williams, V. D.; Kummeler, P.; Neidigh, J. W.; Andersen, N. H. The Trp-Cage: Optimizing the Stability of a Globular Miniprotein. *Protein Eng., Des. Sel.* **2008**, *21* (3), 171–185.
- (4) Pitera, J. W.; Swope, W. Understanding Folding and Design: Replica-Exchange Simulations of Trp-Cage” Miniproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (13), 7587–7592.
- (5) English, C. A.; García, A. E. Charged Termini on the Trp-Cage Roughen the Folding Energy Landscape. *J. Phys. Chem. B* **2015**, *119* (25), 7874–7881.
- (6) Paschek, D.; Hempel, S.; Garcia, A. E. Computing the Stability Diagram of the Trp-Cage Miniprotein. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (46), 17754–17759.
- (7) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. Ab Initio Folding Simulation of the Trp-Cage Mini-Protein Approaches NMR Resolution. *J. Mol. Biol.* **2003**, *327* (3), 711–717.
- (8) Snow, C. D.; Zagrovic, B.; Pande, V. S. The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2002**, *124* (49), 14548–14549.
- (9) Day, R.; Paschek, D.; Garcia, A. E. Microsecond Simulations of the Folding/ Unfolding Thermodynamics of the Trp-Cage Miniprotein. *Proteins: Struct., Funct., Genet.* **2010**, *78* (8), 1889–1899.
- (10) Paschek, D.; Nymeyer, H.; García, A. E. Replica Exchange Simulation of Reversible Folding/unfolding of the Trp-Cage Miniprotein in Explicit Solvent: On the Structure and Possible Role of Internal Water. *J. Struct. Biol.* **2007**, *157* (3), 524–533.
- (11) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124* (38), 11258–11259.
- (12) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 Ms. *J. Am. Chem. Soc.* **2002**, *124* (44), 12952–12953.
- (13) Mok, K. H.; Kuhn, L. T.; Goetz, M.; Day, I. J.; Lin, J. C.; Andersen, N. H.; Hore, P. J. A Pre-Existing Hydrophobic Collapse in

the Unfolded State of an Ultrafast Folding Protein. *Nature* **2007**, *447* (7140), 106–109.

(14) Rogne, P.; Ozdowdy, P.; Richter, C.; Saxena, K.; Schwalbe, H.; Kuhn, L. T. Atomic-Level Structure Characterization of an Ultrafast Folding Mini-Protein Denatured State. *PLoS One* **2012**, *7* (7), e41301.

(15) Matthes, D.; De Groot, B. L. Secondary Structure Propensities in Peptide Folding Simulations: A Systematic Comparison of Molecular Mechanics Interaction Schemes. *Biophys. J.* **2009**, *97* (2), 599–608.

(16) Duan, L.; Mei, Y.; Li, Y.; Zhang, Q.; Zhang, D.; Zhang, J. Z. Simulation of the Thermodynamics of Folding and Unfolding of the Trp-Cage Mini-Protein TCSb Using Different Combinations of Force Fields and Solvation Models. *Sci. China: Chem.* **2010**, *53* (1), 196–201.

(17) Juraszek, J.; Bolhuis, P. G. Sampling the Multiple Folding Mechanisms of Trp-Cage in Explicit Solvent. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (43), 15859–15864.

(18) Jimenez-Cruz, C. A.; Makhatazde, G. I.; Garcia, A. E. Protonation/deprotonation Effects on the Stability of the Trp-Cage Mini-protein. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 17056.

(19) Gattin, Z.; Riniker, S.; Hore, P. J.; Mok, K. H.; Van Gunsteren, W. F. Temperature and Urea Induced Denaturation of the TRP-Cage Mini Protein TCSb: A Simulation Study Consistent with Experimental Observations. *Protein Sci.* **2009**, *18* (10), 2090–2099.

(20) Wafer, L. N. R.; Streicher, W. W.; Makhatazde, G. I. Thermodynamics of the Trp-Cage Mini-protein Unfolding in Urea. *Proteins: Struct., Funct., Genet.* **2010**, *78* (6), 1376–1381.

(21) Canchi, D. R.; García, A. E. Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273–293.

(22) Moelbert, S.; Normand, B.; De Los Rios, P. Kosmotropes and Chaotropes: Modelling Preferential Exclusion, Binding and Aggregate Stability. *Biophys. Chem.* **2004**, *112* (1), 45–57.

(23) Canchi, D. R.; Jayasimha, P.; Rao, D. C.; Makhatazde, G. I.; Garcia, A. E. Molecular Mechanism for the Preferential Exclusion of Osmolytes from Protein Surfaces. *Biophys. J.* **2013**, *104* (2), 189a.

(24) Canchi, D. R.; Paschek, D.; García, A. E. Equilibrium Study of Protein Denaturation by Urea. *J. Am. Chem. Soc.* **2010**, *132* (7), 2338–2344.

(25) English, C. A.; García, A. E. Folding and Unfolding Thermodynamics of the TC10b Trp-Cage Mini-protein. *Phys. Chem. Chem. Phys.* **2014**, *16* (7), 2748.

(26) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5* (8), e1000452.

(27) Royer, C. A. Revisiting Volume Changes in Pressure-Induced Protein Unfolding. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **2002**, *1595* (1–2), 201–209.

(28) Akasaka, K.; Kitahara, R.; Kamatari, Y. O. Exploring the Folding Energy Landscape with Pressure. *Arch. Biochem. Biophys.* **2013**, *531* (1–2), 110–115.

(29) Rouget, J. Size and Sequence and the Volume Change of Protein Folding. *J. Am. Chem. Soc.* **2011**, *133* (15), 6020–6027.

(30) Chen, C. R.; Makhatazde, G. I. ProteinVolume: Calculating Molecular van Der Waals and Void Volumes in Proteins. *BMC Bioinf.* **2015**, *16* (1), 101.

(31) Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; Garcia-Moreno E, B.; Royer, C. A. Cavities Determine the Pressure Unfolding of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (18), 6945–6950.

(32) Neumaier, S.; Kiefhaber, T. Redefining the Dry Molten Globule State of Proteins. *J. Mol. Biol.* **2014**, *426* (13), 2520–2528.

(33) Paschek, D.; Gnanakaran, S.; Garcia, A. E. Simulations of the Pressure and Temperature Unfolding of an Alpha-Helical Peptide. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6765–6770.

(34) Imamura, H.; Kato, M. Effect of Pressure on Helix-Coil Transition of an Alanine-Based Peptide: An FTIR Study. *Proteins: Struct., Funct., Genet.* **2009**, *75* (4), 911–918.

(35) Imamura, H.; Isogai, Y.; Takekiyo, T.; Kato, M. Effect of Pressure on the Secondary Structure of Coiled Coil Peptide GCN4-p1. *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804* (1), 193–198.

(36) Culik, R. M.; Serrano, A. L.; Bunagan, M. R.; Gai, F. Achieving Secondary Structural Resolution in Kinetic Measurements of Protein Folding: A Case Study of the Folding Mechanism of Trp-Cage. *Angew. Chem., Int. Ed.* **2011**, *50* (46), 10884–10887.

(37) Culik, R. M.; Annavarapu, S.; Nanda, V.; Gai, F. Using D-Amino Acids to Delineate the Mechanism of Protein Folding: Application to Trp-Cage. *Chem. Phys.* **2013**, *422*, 131–134.

(38) Bunagan, M. R.; Yang, X.; Saven, J. G.; Gai, F. Ultrafast Folding of a Computationally Designed Trp-Cage Mutant: Trp 2-Cage. *J. Phys. Chem. B* **2006**, *110* (8), 3759–3763.

(39) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6* (3), 277–293.

(40) Johnson, B. A.; Blevins, R. A. NMR View: A Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR* **1994**, *4* (5), 603–614.

(41) Herrmann, T.; Güntert, P.; Wüthrich, K. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J. Mol. Biol.* **2002**, *319* (1), 209–227.

(42) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; et al. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54* (5), 905–921.

(43) Brünger, A. T. Version 1.2 of the Crystallography and NMR System. *Nat. Protoc.* **2007**, *2* (11), 2728–2733.

(44) Peterson, R. W.; Nucci, N. V.; Wand, A. J. Modification of Encapsulation Pressure of Reverse Micelles in Liquid Ethane. *J. Magn. Reson.* **2011**, *212* (1), 229–233.

(45) Kobayashi, N.; Iwahara, J.; Koshiba, S.; Tomizawa, T.; Tochio, N.; Güntert, P.; Kigawa, T.; Yokoyama, S. KUJIRA, a Package of Integrated Modules for Systematic and Interactive Analysis of NMR Data Directed to High-Throughput NMR Structure Studies. *J. Biomol. NMR* **2007**, *39* (1), 31–52.

(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926.

(47) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 712–725.

(48) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577–2637.

(49) Buckingham, A. D. Chemical Shifts in the Nuclear Magnetic Resonance Spectra of Molecules Containing Polar Groups. *Can. J. Chem.* **1960**, *38* (6630), 300–307.

(50) Boyd, J.; Domene, C.; Redfield, C.; Ferraro, M. B.; Lazzaretti, P.; Uni, V.; Aires, D. B.; V, C. U.; Pab, I.; Aires, B.; et al. Calculation of Dipole-Shielding Polarizabilities ( $\sigma_{R\gamma I}$ ): The Influence of Uniform Electric Field Effects on the Shielding of Backbone Nuclei in Proteins. *J. Am. Chem. Soc.* **2003**, *125*, 9556–9557.

(51) Kukic, P.; Farrell, D.; McIntosh, L. P.; García-Moreno E, B.; Jensen, K. S.; Toleikis, Z.; Teilmann, K.; Nielsen, J. E. Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations. *J. Am. Chem. Soc.* **2013**, *135* (45), 16968–16976.

(52) Hass, M. A. S.; Mulder, F. A. A. Contemporary NMR Studies of Protein Electrostatics. *Ann. Rev. Biophys.* **2015**, *44*, 53–75.

(53) Fitch, C. A.; Platzer, G.; Okon, M.; E, B. G.; McIntosh, L. P. Arginine: Its *pK* a Value Revisited **2015**, *24*, 752–761.

(54) Huyghues-Despointes, B. M.; Klingler, T. M.; Baldwin, R. L. Measuring the Strength of Side-Chain Hydrogen Bonds in Peptide Helices: The Gln<sub>Asp</sub> (I, I + 4) Interaction. *Biochemistry* **1995**, *34* (41), 13267–13271.

138

(55) Kitahara, R.; Hata, K.; Li, H.; Williamson, M. P.; Akasaka, K. Pressure-Induced Chemical Shifts as Probes for Conformational Fluctuations in Proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *71*, 35–58.

(56) Kim, S. B.; Palmer, J. C.; Debenedetti, P. G. Computational Investigation of Cold Denaturation in the Trp-Cage Miniprotein. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (32), 8991–8996.

(57) Rovó, P.; Stráner, P.; Láng, A.; Bartha, I.; Huszár, K.; Nyitray, L.; Perczel, A. Structural Insights into the Trp-Cage Folding Intermediate Formation. *Chem. - Eur. J.* **2013**, *19* (8), 2628–2640.

## 6.2 Resume de l'article en francais

### Etude sous pression de la protéine modèle tryptophan-cage

La dernière étude présentée dans cette thèse est une étude de la stabilité de la variante de la mini protéine tryptophan-cage Tc5b. Cette protéine est l'une des protéines les plus étudiées dans le domaine de l'étude du repliement des protéines. Il s'agit d'une protéine faite par un laboratoire à partir d'une séquence plus large. Elle se plie extrêmement vite, ce qui permet de simuler son repliement entièrement dans des délais de computation relativement courts. Ceci fait d'elle une protéine modèle idéale afin d'étudier les effets des champs de forces et modèles d'eau utilisés en simulation, mais aussi afin de comparer des résultats expérimentaux avec les simulations, permettant ainsi d'interpréter des résultats expérimentaux avec des mécanismes moléculaires visibles dans les simulations. Ce rôle de protéine modèle pour les simulations, l'obtention de données expérimentales pour cette protéine est très importante pour permettre une meilleure compréhension des mécanismes de repliement des protéines. Dans cette étude par RMN, nous avons utilisé une autre propriété de la pression sur les protéines afin de déterminer les propriétés des résidus et leur dépendance à la pression, en utilisant les propriétés des variations de chemical shifts des peaks RMN en fonction de la pression. En plus de cela, une titration complète utilisant de l'urée et le pH a également été réalisée afin de mieux comprendre les mécanismes de stabilisation de cette protéine. L'importance de cette étude est due à l'utilisation de trajectoires de simulation moléculaires tout-atomes, en combinaison avec les profils par résidu de la titration par haute pression en RMN. Grâce aux simulations, nous sommes parvenus à identifier un sous état de l'ensemble plier de la protéine qui ne correspond pas à la structure RMN de la littérature avec une forte dépendance à la pression. Nous avons donc pu conclure que les effets de la pression sur cette protéine sont quasi exclusivement liés à cet état, et que les effets sur cet état expliquent remarquablement bien les données.



# Chapter 7

## Annex

### 7.1 Thermodynamics in protein folding

#### 7.1.1 Detailed introduction to thermodynamics

##### 7.1.1.1 Definition of the equilibrium

The classic statistical approach to investigate protein properties is through the determination of the thermodynamics stability. For this analysis, the environment of the protein is gradually changed from a native like environment to a less favorable environment, and a signal depending on the level of order in the structure (Rg, NMR peaks shift or intensity, CD absorption, fluorescence signal etc...) is monitored to see the effect of the environment changes on the equilibrium between the different states. For the sake of simplicity, the folding reaction is most often described as a two states equilibrium between two ensembles :

$$\begin{aligned}G &= U + PV - TS \\G &= H - TS \\G &= \sum_i \mu_i N_i\end{aligned}\tag{7.1.1}$$

Where  $\mu_i$  is the chemical potential from the specie  $i$  in the, and  $N$  the number of particles in that state. This quantity is the total partial Gibbs energy, and is a reflection of the entropy volume and interactions (Van der Walls interaction, hydrogen bonds, salt bridges ect...) in play

for the state considered. The variation of  $G$  is :

$$dG = dU + d(PV) - d(TS)$$

$$dG = \sum_i \mu_i dN_i + TdS - PdV + VdP + PdV - SdT - TdS \quad (7.1.2)$$

$$dG = \sum_i \mu_i dN_i + VdP - SdT \quad (7.1.3)$$

$$(7.1.4)$$

The free energy of the system comprising this two species is denoted :

$$G = \bar{\mu}_f n_f + \bar{\mu}_u n_u \quad (7.1.5)$$

To reflect the fact that the quantities exposed here reflect an ensemble rather than a single defined state, the chemical potential is denoted  $\bar{\mu}$ , and is an average value over the ensemble. To be correct this equation must account for the solute molecule and all the interacting water<sup>1</sup>, in order to respect the conditions for ideality<sup>2</sup>. If we change the composition of the solution by advancing the reaction towards on the unfolded state of a mol, we get the new Gibbs potential[181]:

$$G + \left( \frac{\partial G}{\partial n_u} \right)_{P,T,n_f} dn + \left( \frac{\partial G}{\partial n_f} \right)_{P,T,n_u} dn = \bar{\mu}_f (n_f + dn_f) + \bar{\mu}_u (n_u + dn_u) \quad (7.1.6)$$

$$G + \left( \frac{\partial G}{\partial n} \right)_{P,T} dn = \bar{\mu}_f (n_f + \nu_f d\xi) + \bar{\mu}_u (n_u + \nu_u d\xi) \quad (7.1.7)$$

In which  $d\xi$  is the extent of the reaction<sup>3</sup> in mol, and is related to the change in population of each specie  $i$  by  $\nu_i$ , the stoichiometric coefficient, meaning  $\nu_i d\xi = dn_i$ . The infinitesimal change of  $G$  is :

$$dG = \bar{\mu}_f (\nu_f d\xi) + \bar{\mu}_u (\nu_u d\xi) \quad (7.1.8)$$

---

<sup>1</sup>That is the maximum number of interacting water the solute can have, so the total number of particle stays constant.

<sup>2</sup>See section 7.3.4

<sup>3</sup>See 7.3.3

Therefore :

$$\left(\frac{dG}{d\xi}\right)_{P,T,\{n_i\}} = \bar{\mu}_f\nu_f + \bar{\mu}_u\nu_u \quad (7.1.9)$$

In the case of a stoichiometric reaction with two species, like an unfolding reaction where  $f$  gets transformed into  $u$ ,  $\nu_u = -\nu_f = 1$ , and hence :

$$\left(\frac{dG}{d\xi}\right)_{P,T,\{n_i\}} = \bar{\mu}_u - \bar{\mu}_f \quad (7.1.10)$$

Which, in which  $\bar{\mu}_i$  molar free energy of the ensemble  $i$ . From this we get :

$$\left(\frac{dG}{d\xi}\right)_{P,T,\{n_i\}} = \Delta_u G \quad (7.1.11)$$

Where  $\Delta_u G$  is the difference of molar free energy between the folded and unfolded ensembles. This relation indicates how the change in free energy of the solution evolves faster, for a mole of advancement, as the difference of energy between the species involve in the reaction is bigger. The condition for equilibrium is[182]:

$$\left(\frac{dG}{d\xi}\right)_{P,T,\{n_i\}} = 0 \quad (7.1.12)$$

Which is equivalent to saying that equilibrium corresponds to the bottom well of the  $G_{(\xi)}$  function, meaning it is the extent of the reaction for which  $G$  is minimized. At equilibrium,  $\Delta_u G = 0$  which means that the new equilibrium is found so that the number of particles makes the sum the smallest, which is determined by the change in chemical potentials with changes in conditions. We can then express  $\Delta_u G$  differently by replacing the terms of chemical potential with their definitions  $\bar{\mu}_i = \bar{\mu}_i^0 + RT \ln \bar{\alpha}_i$ , where alpha is the chemical activity of specie  $i$ , which in ideal solution is equivalent to the concentration :

$$\begin{aligned}
\Delta_u G &= \bar{\mu}_f \nu_f + \bar{\mu}_u \nu_u \\
\Delta_u G &= \left( \bar{\mu}_u^0 + RT \ln(\bar{\alpha}_u) \right) \nu_u + \left( \bar{\mu}_f^0 + RT \ln(\bar{\alpha}_f) \right) \nu_f \\
\Delta_u G &= \bar{\mu}_u^0 \nu_u + RT \nu_u \ln(\bar{\alpha}_u) + \bar{\mu}_f^0 \nu_f + RT \nu_f \ln(\bar{\alpha}_f) \\
\Delta_u G &= \bar{\mu}_u^0 \nu_u + \bar{\mu}_f^0 \nu_f + RT \left( \ln(\bar{\alpha}_u^{\nu_u}) + \ln(\bar{\alpha}_f^{\nu_f}) \right) \\
\Delta_u G &= \bar{\mu}_u^0 \nu_u + \bar{\mu}_f^0 \nu_f + RT \ln(\bar{\alpha}_u^{\nu_u} \times \bar{\alpha}_f^{\nu_f})
\end{aligned} \tag{7.1.13}$$

From there we can easily identify the the reaction quotient  $Q_u = \bar{\alpha}_u^{\nu_u} \times \bar{\alpha}_f^{\nu_f}$ . In addition we can identify the standard enthalpy change  $\Delta_u G^0 = \bar{\mu}_u^0 \nu_u + \bar{\mu}_f^0 \nu_f$ , leaving us with :

$$\Delta_u G = \Delta_u G^0 + RT \ln(Q_u) \tag{7.1.14}$$

And thus we can deduce from equation 7.1.12 that :

$$\Delta_u G^0 = -RT \ln(K_u) \tag{7.1.15}$$

Where  $K = Q_{eq}$ , the equilibrium constant is the reaction quotient at equilibrium.

### 7.1.1.2 Variation of equilibrium

#### First degree dependence

This equilibrium is defined for each pressure temperature conditions and changes as the values of the chemical potentials varies with those conditions. Now let us consider the variation of  $\Delta G$ . We have seen in eq.7.1.2 that the variation of  $G$  using the Legendre transform was  $dG = \sum_i \mu_i dN_i + VdP - SdT$ . Using the definition of  $G$  from eq.7.1.5 and the Euler relation<sup>1</sup>, we can state that :

$$G = \sum_i n_i \left( \frac{\partial G}{\partial n_i} \right)_{T,P,\{n_{j \neq i}\}} = \sum_i \bar{\mu}_i n_i \tag{7.1.16}$$

And introduce :

---

<sup>1</sup>See annex eq.7.3.7

$$\begin{aligned}
dG &= \sum_i \left( \frac{\partial (\bar{\mu}_i n_i)}{\partial n_i} \right)_{\bar{\mu}_i} dn_i + \sum_i \left( \frac{\partial (\bar{\mu}_i n_i)}{\partial \bar{\mu}_i} \right)_{n_i} d\bar{\mu}_i \\
dG &= \sum_i \mu_i dn_i + \sum_i n_i d\bar{\mu}_i
\end{aligned} \tag{7.1.17}$$

Hence :

$$\begin{aligned}
\sum_i \bar{\mu}_i dn_i + \sum_i n_i d\bar{\mu}_i &= \sum_i \bar{\mu}_i dn_i + V dP - S dT \\
\sum_i n_i d\bar{\mu}_i &= V dP - S dT
\end{aligned} \tag{7.1.18}$$

Which is called the Gibbs-Duhem equation, and shows that the of the partial molar quantities that compose  $G$  evolve with changing conditions. This is in essence the mathematical formulation of the Le Chatellier's principle<sup>1</sup>. We can then write :

$$\left( \frac{\partial G}{\partial P} \right)_{T, \{n_i\}} = \sum_i \left( \frac{\partial \bar{\mu}_i}{\partial P} \right)_{T, n_i} n_i = V \tag{7.1.19}$$

$$\sum_i \bar{v}_i n_i = V \tag{7.1.20}$$

We can now separate the contribution from eq.7.1.1 and look at how the standard chemical potential varies with changing conditions and define their pressure dependence according to eq.7.1.20 :

$$\left( \frac{\partial \bar{\mu}_u^0}{\partial P} \right)_T = \bar{v}_u^0 \tag{7.1.21}$$

$$\left( \frac{\partial \bar{\mu}_f^0}{\partial P} \right)_T = \bar{v}_f^0 \tag{7.1.22}$$

And look at there evolution with pressure :

$$\bar{\mu}_u^0 + \left( \frac{\partial \bar{\mu}_u^0}{\partial P} \right)_{T, \{n_i\}} dP = \bar{\mu}_u^0 + \bar{v}_u^0 dP \tag{7.1.23}$$

$$\bar{\mu}_f^0 + \left( \frac{\partial \bar{\mu}_f^0}{\partial P} \right)_{T, \{n_i\}} dP = \bar{\mu}_f^0 + \bar{v}_f^0 dP \tag{7.1.24}$$

Where  $\bar{v}_f^0$  and  $\bar{v}_u^0$  are the partial molar volume occupied by the unfolded and folded states,

---

<sup>1</sup>Interestingly, Le Chatellier was the translator of Gibbs work in french language.

respectively. From there we can replace the chemical shift from eq.7.1.13 and obtain :

$$\Delta_u G^0 + \left( \frac{\partial \Delta_u G^0}{\partial P} \right)_{T, \{N_i\}} dP = \bar{\mu}_f^0 \nu_f + \bar{\mu}_u^0 \nu_u + \bar{v}_f^0 \nu_f dP + \bar{v}_u^0 \nu_u dP \quad (7.1.25)$$

And hence the change in the difference in standard Gibbs free energy between the folded and unfolded state between a pressure  $P_0$  and a pressure  $P$  :

$$\Delta_u G^0 + \int_{P_0}^P \left( \frac{\partial \Delta_u G^0}{\partial P} \right)_{T, \{N_i\}} dP = \bar{\mu}_f^0 \nu_f + \bar{\mu}_u^0 \nu_u + \int_{P_0}^P \bar{v}_f^0 \nu_f dP + \int_{P_0}^P \bar{v}_u^0 \nu_u dP \quad (7.1.26)$$

$$(7.1.27)$$

Giving us the evolution of the difference in molar free energy between the folded and unfolded specie with pressure. If we replace the stoichiometric coefficient with  $\nu_u = -\nu_f = 1$ .

$$\Delta_u G^0 + \int_{P_0}^P \left( \frac{\partial \Delta_u G^0}{\partial P} \right)_{T, \{N_i\}} dP = \bar{\mu}_u^0 - \bar{\mu}_f^0 + \int_{P_0}^P (\bar{v}_u^0 - \bar{v}_f^0) dP \quad (7.1.28)$$

$$\Delta_u G^0 + \int_{P_0}^P \left( \frac{\partial \Delta_u G^0}{\partial P} \right)_{T, \{N_i\}} dP = \Delta_u \bar{\mu} + \int_{P_0}^P (\Delta_u \bar{v}) dP \quad (7.1.29)$$

Hence if one state has a greater volume than the other, it will see its entropic cost raise faster. If we relate to the equilibrium constant we get :

$$\Delta_u G_{(P)}^0 = \Delta_u G_{P_0}^0 + \int_{P_0}^P \left( \frac{\partial \Delta_u G^0}{\partial P} \right)_{T, \{N_i\}} dP = -RT \ln (K_{P_0}) + \int_{P_0}^P (\Delta_u \bar{v}) dP \quad (7.1.30)$$

$$\Delta_u G_{(P)}^0 = -RT \ln (K_{P_0}) + \Delta_u \bar{v} (P_0 - P) \quad (7.1.31)$$

$$\Delta_u G_{(P)}^0 = -RT \ln (K_P) \quad (7.1.32)$$

Where  $\Delta_u G_P^0$  if the molar free energy difference between the species in play, and  $K_P$  the equilibrium constant at pressure  $P$ . Equivalently for temperature, the same operations would lead to :

$$\Delta_u G_{(T)}^0 = -RT \ln (K_{T_0}) + \Delta_u \bar{s} (T_0 - T) \quad (7.1.33)$$

$$\Delta_u G_{(T)}^0 = -RT \ln (K_T) \quad (7.1.34)$$

Where  $\Delta \bar{s}$  is the partial molar entropy difference between the folded and unfolded state.

### Second degree dependence

However, those relations need to be corrected to account for the fact that entropy and volume both have a pressure and a temperature dependence. Let's start with the pressure dependence of the volume at constant temperature, compressibility, that is defined[68]:

$$\beta_T = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T \quad (7.1.35)$$

Hence we can define the volume at any pressure P as :

$$\begin{aligned} \Delta \bar{v}_{u(P)} &= \Delta \bar{v}_{u(P_0)} + \int_{P_0}^P \left( \frac{\partial \Delta \bar{v}_u}{\partial P} \right)_T dP \\ \Delta \bar{v}_{u(P)} &= \Delta \bar{v}_{u(P_0)} + \int_{P_0}^P \left( -\bar{v}_{u(P_0)} \beta_u + \bar{v}_{f(P_0)} \beta_f \right) dP \\ \Delta \bar{v}_{u(P)} &= \Delta \bar{v}_{u(P_0)} + \left( -\bar{v}_{u(P_0)} \beta_u + \bar{v}_{f(P_0)} \beta_f \right) (P - P_0) \end{aligned} \quad (7.1.36)$$

Equivalently for the variation of volume on temperature, expansivity, can be defined as :

$$\alpha_P = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_P \quad (7.1.37)$$

And so :

$$\begin{aligned} \Delta \bar{v}_{u(T)} &= \Delta \bar{v}_{u(T_0, P_0)} + \int_{T_0}^T \left( \frac{\partial \Delta \bar{v}_{u(T_0, P_0)}}{\partial T} \right)_P dT \\ \Delta \bar{v}_{u(T)} &= \Delta \bar{v}_{u(T_0, P_0)} + \int_{T_0}^T \left( \bar{v}_{u(T_0, P_0)} \alpha_u - \bar{v}_{f(T_0, P_0)} \alpha_f \right) dT \\ \Delta \bar{v}_{u(T)} &= \Delta \bar{v}_{u(T_0, P_0)} + \left( \bar{v}_{u(T_0, P_0)} \alpha_u - \bar{v}_{f(T_0, P_0)} \alpha_f \right) (T - T_0) \end{aligned} \quad (7.1.38)$$

Equivalently for the entropy :

$$C_{p,u} = T \left( \frac{\partial \bar{s}_u}{\partial T} \right)_P \quad (7.1.39)$$

And therefore :

$$\begin{aligned} \Delta \bar{s}_{u(T)} &= \Delta \bar{s}_{u(T_0, P_0)} + \int_{T_0}^T \left( \frac{\partial \Delta \bar{s}_{u(T_0, P_0)}}{\partial T} \right)_P dT \\ \Delta \bar{s}_{u(T)} &= \Delta \bar{s}_{u(T_0)} + \int_{T_0}^T \frac{1}{T} (C_{p,u} - C_{p,f}) dT \\ \Delta \bar{s}_{u(T)} &= \bar{s}_{u(T_0)} + \Delta C_{p,u} \ln \left( \frac{T}{T_0} \right) \end{aligned} \quad (7.1.40)$$

Finally, for the entropy dependence on pressure, using the Maxwell relations  $-\left(\frac{\partial S}{\partial P}\right)_T = \left(\frac{\partial V}{\partial T}\right)_P$ <sup>1</sup>, we can establish that :

$$\alpha_P = -\frac{1}{V} \left( \frac{\partial S}{\partial P} \right)_T \quad (7.1.41)$$

And for two states :

$$\begin{aligned} \Delta \bar{s}_{u(P)} &= \Delta \bar{s}_{u(T_0, P_0)} + \int_{P_0}^P \left( \frac{\partial \Delta \bar{s}_{u(T_0, P_0)}}{\partial P} \right)_T dP \\ \Delta \bar{s}_{u(P)} &= \Delta \bar{s}_{u(T_0)} + \int_{P_0}^P (-\alpha_u \bar{v}_u + \alpha_f \bar{v}_f) dP \\ \Delta \bar{s}_{u(P)} &= \Delta \bar{s}_{u(T_0)} + (-\alpha_u \bar{v}_u + \alpha_f \bar{v}_f) (P - P_0) \end{aligned} \quad (7.1.42)$$

Using these relation, we can establish the variation of  $\Delta G^0$  as a function of both pressure and temperature using :

$$\Delta G_{(T,P)}^0 = \Delta G_{(T_0, P_0)}^0 + \int_{P_0}^P \left( \frac{\partial \Delta G_u^0}{\partial P} \right)_T dP + \int_{T_0}^T \left( \frac{\partial \Delta G_u^0}{\partial T} \right)_P dT \quad (7.1.43)$$

$$\begin{aligned} \Delta G_{(T,P)}^0 &= \Delta G_{(T_0, P_0)}^0 + \\ &\int_{P_0}^P \left( \Delta \bar{v}_u^0 + (-\beta_u v_u + \beta_f v_f) (P - P_0) + (\alpha_u v_u - \alpha_f v_f) (T - T_0) \right) dP - \\ &\int_{T_0}^T \left( \Delta \bar{s}_u^0 + (-\alpha_u v_u + \alpha_f v_f) (P - P_0) + \Delta C_{p,u} \ln \left( \frac{T}{T_0} \right) \right) dT \end{aligned} \quad (7.1.44)$$

$$(7.1.45)$$

$$\begin{aligned} \Delta G_{(T,P)}^0 &= \Delta G_{(T_0, P_0)}^0 + \Delta v_u^0 (P - P_0) - \Delta \bar{s}_u^0 (T - T_0) + \\ &\frac{(-\beta_u v_u + \beta_f v_f)}{2} P^2 - \frac{(-\beta_u v_u + \beta_f v_f)}{2} P_0^2 - (-\beta_u v_u + \beta_f v_f) P_0 (P - P_0) + \\ &(\alpha_u v_u - \alpha_f v_f) (T - T_0) (P - P_0) - \\ &(-\alpha_u v_u + \alpha_f v_f) (P - P_0) (T - T_0) - \\ &\Delta C_{p,u} \left( T \ln \left( \frac{T}{T_0} \right) - (T - T_0) \right) \end{aligned} \quad (7.1.46)$$

$$(7.1.47)$$

$$\begin{aligned} \Delta G_{(T,P)}^0 &= \Delta G_{(T_0, P_0)}^0 + \Delta v_u^0 (P - P_0) - \Delta \bar{s}_u^0 (T - T_0) + \frac{(-\beta_u v_u + \beta_f v_f)}{2} (P - P_0)^2 + \\ &2(\alpha_u v_u - \alpha_f v_f) (T - T_0) (P - P_0) - \Delta C_{p,u} \left( T \ln \left( \frac{T}{T_0} \right) - (T - T_0) \right) \end{aligned} \quad (7.1.48)$$

---

<sup>1</sup>See equations 7.3.10

## 7.2 Protocols

### 7.2.1 Transformation in competent cells

Take your DH5- $\alpha$  containing the plasmid from the  $-80^{\circ}\text{C}$  fridge. Use a miniprep kit and carefully follow the instructions to extract the plasmids. Check the plamid concentration using the DNA option of the nanodrop. The (?) ratio should be less that 2. If it is too high, dilute your sample and measure concentration again. You should typically obtain concentrations of around 100ng/ml.

For this range of concentration, use  $1\text{-}2\mu\text{L}$  of plasmid solution and put in in BL21 cells on ice and wait 30min.

Apply a heat shock to open the cells membrane by putting the tubes in a  $42^{\circ}\text{C}$  water bath for precisely 30s, then put them back on ice for two minutes (don't wait too long !). Add  $250\mu\text{L}$  of glycerol containing SOC (Super Optimal Broth) and incubate at  $37^{\circ}\text{C}$  for 1H.

Spread  $50\mu\text{L}$  of the cells on a antibiotic containing plate and incubate overnight.

Pick a single colony and grow it overnight.

Prepare papers containing all the information (variant, type of cell, antibiotic resistance and date) to be attached to the eppendorff. For each variant, use 3 sterile eppendorff and use put the papers you have printed on it. Take  $500\mu\text{L}$  of the cell solution and  $150\mu\text{L}$  of sterile 50% glycerol stock. Put the eppendorff in the  $-80^{\circ}\text{C}$ .

### 7.2.2 Protein Production of PP32

In this section, the protocol of protein production for the PP32 protein is described. This protocol was adapted from the one that I was taught from the Barrick lab during a visit in the in John Hopkins University. If  $^{15}\text{N}$  isotope labeling is required, one should simply use a M9 reduced media for the cell growth, and follow the subsequent steps normally.

### 7.2.2.1 Buffer preparation

**Growth media** Use the premixed LB powder following the instructions on the label. Add 1 mL of 30mg/mL kanamycin for each liter of growth media.

**Buffer A** Prepare a buffer 20mM Sodium Phosphate, 500mM NaCl 0.1mM TCEP 25mM imidazol pH 7.4. You should typically need 100ml per liter of growth media, but be sure to have some spare.

**Buffer B** Same as buffer A, with 250mM imidazol. You should need less than 100ml, but make sure you have some spare.

**Dialysis Buffer** Prepare a buffer containing 150mM NaCl 20mM Sodium Phosphate, 5mM DTT pH 7.8. You will need at least 2-3 liter, but as you can never have too much, prepare a lot. You might notice that the reducing agent and its concentrations are different than for the purification buffers. That is because our high pressure experiments have shown DTT to work much better than TCEP to avoid aggregation, but such a high concentration will have a reducing effect on the nickel column, preventing it from working.

### 7.2.2.2 Cell Growth

The growth of the cells in this protocol is in 3 steps. In the first step, we grow and select a strain containing the expression vector from a glycerol stock. The second is the growth of the cell strictly speaking : we produce a solution of cell to a certain concentration. We then turn on the expression vector for the third step, which is the protein production.

These steps are long, but not very time consuming, so one should just plan his week carefully before starting a preparation.

**Step 1 ( $\simeq 12H$ )** Take the glycerol stock and unfreeze it on ice. Spread a few microliter on a LB+Kanamycine petri dish. Grow cells overnight (not more than 16H) at 37°C.

**Step 2 ( $\simeq 12H$ )** Pick up a single colony, and introduce it in a tube containing 10mL LB and kana per liter of growth you want. Grow 12H at 37°C with shaker.

Put the solution in 1L of LB and kanamycin at 37°C, and test the optical density until you obtain a value of around 0.6 at 600nm. Proceed fast to step 3.

**Step 3**( $\simeq 12 - 16H$ ) Add 1mL of 500mM IPTG, and put the solution at around 20°C (to avoid protein aggregation) for 12 to 16 hours.

After this step is completed, spin the cells down in polycarbonate tubes ( $\simeq 7000g$ ), get rid of the excess liquid, and freeze the pellet. **The production process can be interrupted here.**

### 7.2.2.3 Protein purification

Resuspend cells into 50ml buffer A, add 100 $\mu$ L of 10mg/mL lysozyme and 1 protease inhibitor tablet (EDTA free).

Freeze and thaw the solution three times.

Add 1mL of 1mg/mL DNase, 100 $\mu$ L of  $MgCl_2$ . Wait 30min.

Spin cells down (25000-30000rpm for 15min).

Load the supernatant onto Ni<sup>2+</sup> column. Wash with 50 mL of buffer A. Elute with 50 mL of buffer B.

Run gel (12-15%) using the washing liquid and check the presence of your protein in the eluting solution.

Dialyze protein into the dialysis buffer (4°C).

## 7.3 Definitions

### 7.3.1 Legendre transform

Legendre transform is a mathematical property for convex functions. Here, we will focus on the behavior of differential using Legendre transform only, because of their relevance to the derivation of thermodynamic potentials. Given a function  $f(x, y)$ , with a derivative  $f'(x, y) = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = udx + vdy$ , then we can define  $g = f - ux$ , and from there :

$$dg = df - d(ux) \quad (7.3.1)$$

$$dg = udx + vdy - xdu - udx \quad (7.3.2)$$

$$dg = vdy - xdu \quad (7.3.3)$$

$$(7.3.4)$$

### 7.3.2 Euler's equation

Given a homogeneous and of first degree function representing an integral quantity[68]:

$$Y(T, P, \{\lambda n_i\}) = \lambda Y(T, P, \{n_i\}) \quad (7.3.5)$$

Then :

$$\sum_i \left( \frac{\partial Y}{\partial \lambda n_i} \right)_{T, P, \{n_j \neq i\}} \left( \frac{\partial \lambda n_i}{\partial \lambda} \right) = Y \quad (7.3.6)$$

If we set  $\lambda = 1$ , then we get :

$$Y = \sum_i n_i \left( \frac{\partial Y}{\partial n_i} \right)_{T, P, \{n_j \neq i\}} \quad (7.3.7)$$

Where  $\left( \frac{\partial Y}{\partial n_i} \right)_{T, P, \{n_j \neq i\}}$  is the partial molar quantity of specie  $i$  in the system. This equation shows that the integral quantity  $Y$  can be expressed as a sum of its components. Using that relation, one has to be careful that  $Y$  is still homogeneous and of first degree along the parameters used for derivation. In particular, the use of function having  $S$  (entropy) or  $V$  (volume) as a natural variable are not of first degree along derivation of the molar amount of substance  $i$ , because these variables are extensive, and hence depend on the amount of substance themselves. Thus application of the previous equation will not verify the Euler's identity, and the

derivation  $\left(\frac{\partial T}{\partial n_i}\right)_{T,P,\{n_{j \neq i}\}}$  of such a quantity is not a partial molar quantity.

### 7.3.3 Extent of reaction

The extent of reaction is a quantity used to describe the extent to which a reaction proceeds. It has the units of moles and is denoted  $\xi$ , following the relation :

$$d\xi = \frac{dn_i}{\nu_i} \quad (7.3.8)$$

Where  $n_i$  is the molar amount of specie  $i$  consumed/produced by the reaction, and  $\nu_i$  the stoichiometric coefficient associated with specie  $i$  for the reaction.

### 7.3.4 Conditions for a solution to be ideal

A solution is considered ideal when the enthalpy of mixing of its different components is null. In the case of proteins, this condition can be satisfied only in by considering the solute as being a combination of the protein and the waters of the hydration shells around that protein. In addition, the solute has to be very diluted such that the hydration shells do not overlap. A solution is considered ideal when the activity coefficient of a specie  $i$  is constant[68]

### 7.3.5 Standard state

The standard state of a chemical specie is arbitrary state serving as a reference point for the study of the reaction. It is always taken at 1bar, but the reference temperature is arbitrary. The standard state is defined for each phase of each chemical species.

### 7.3.6 Maxwell relations

General definition of Maxwell relations between thermodynamic parameters are expressed as follow :

$$\frac{\partial}{\partial x_j} \left( \frac{\partial \Phi}{\partial x_i} \right) = \frac{\partial}{\partial x_i} \left( \frac{\partial \Phi}{\partial x_j} \right) \quad (7.3.9)$$

From which are deduced the four most common relations :

$$\begin{aligned}\left(\frac{\partial T}{\partial V}\right)_S &= -\left(\frac{\partial P}{\partial S}\right)_V = \frac{\partial^2 U}{\partial S \partial V} \\ \left(\frac{\partial T}{\partial P}\right)_S &= \left(\frac{\partial V}{\partial S}\right)_P = \frac{\partial^2 H}{\partial S \partial P} \\ \left(\frac{\partial S}{\partial V}\right)_T &= \left(\frac{\partial P}{\partial T}\right)_V = -\frac{\partial^2 F}{\partial T \partial V} \\ -\left(\frac{\partial S}{\partial P}\right)_T &= \left(\frac{\partial V}{\partial T}\right)_P = \frac{\partial^2 G}{\partial T \partial P}\end{aligned}\tag{7.3.10}$$



## **7.4 Supplementary information**

### **7.4.1 High-Resolution Mapping of a Repeat Protein Folding Free Energy Landscape**

**Biophysical Journal, Volume 111**

**Supplemental Information**

**High-Resolution Mapping of a Repeat Protein Folding Free Energy Landscape**

**Martin J. Fossat, Thuy P. Dao, Kelly Jenkins, Mariano Dellarole, Yinshan Yang, Scott A. McCallum, Angel E. Garcia, Doug Barrick, Christian Roumestand, and Catherine A. Royer**

## Supplemental Information

### *Resonance Assignments*

Assignment of the  $^1\text{H}$  and  $^{15}\text{N}$  resonances were accomplished following the standard strategy based on [ $^1\text{H},^{15}\text{N}$ ] double-resonance experiments [ $^1\text{H},^{15}\text{N}$ ] NOESY- and TOCSY-HSQC) performed at 20°C on a Bruker AVANCE III 500 MHz equipped with a cryogenic 5 mm Z-gradient  $^1\text{H}$ - $^{13}\text{C}$ - $^{15}\text{N}$  probe head. Mixing times of 100 ms were used for the NOESY experiments, and isotropic mixings of 60 ms were used for the TOCSY experiments. Water suppression was achieved with the WATERGATE sequence.  $^1\text{H}$  chemical shifts were directly referenced to the methyl resonance of DSS, while  $^{15}\text{N}$  chemical shifts were referenced indirectly to the absolute frequency ratio  $^{15}\text{N}/^1\text{H} = 0.101329118$ .

Supplemental Table 1. Residue specific fit parameters for HSQC peak intensity loss vs pressure at 288 K

Residue Name	$\Delta G$ (cal/mol)	$\Delta G_{err}$	$\Delta V$ (ml/mol)	$\Delta V_{err}$	Cf (a.u.)	Cf_err	Cu (a.u.)	Cu_err	T (K)
2GluH	-3144.76143	539.5479	114.4241853	19.10376	9634.287	333.1059	-195.378	306.9667	288
3MetH	-2236.006518	314.1996	86.65427734	11.34857	10661.59	360.5459	41.09975	280.1842	288
4GlyH	-2986.529952	521.0207	110.6141743	18.6704	9791.034	339.1222	249.6151	301.7142	288
5ArgH	-3447.620323	562.3467	127.7583929	20.33198	12389.2	393.9561	-445.973	355.9119	288
6ArgH									
7IleH	-3634.184951	452.1502	131.5667116	16.19238	11561.8	250.4841	346.5684	269.7226	288
8HisH	-2814.737276	417.2485	108.847529	15.41178	12925.69	391.6243	258.6368	327.5661	288
9LeuH									
10GluH									
11LeuH	-2579.599468	178.6387	93.32665263	6.184052	14555.16	255.8361	-861.054	212.3354	288
12ArgH	-3014.298851	395.0976	106.594357	13.66785	16670.55	423.6278	21.94475	406.4667	288
13AsnH	-3020.108519	467.9621	121.8332946	18.12702	8533.114	269.9785	-13.8342	231.2717	288
14ArgH	-2251.489153	315.7222	85.01721802	11.20068	16911.62	597.3881	-859.164	489.1339	288
15ThrH	-2743.070119	255.6543	103.6580356	9.206871	16957.48	354.2754	766.0133	271.9727	288
17SerH	-3832.677891	340.3326	142.179822	12.37823	17529.36	270.5478	365.7706	246.3346	288
18AspH	-3711.607571	538.1001	137.2518544	19.48707	15343.68	381.8162	764.7278	348.2874	288
19ValH									
20LysH	-2804.126406	279.6215	102.9388988	10.91854	11342.69	190.937	358.2046	176.2734	288
21GluH	-2966.979264	195.9144	111.1051898	7.041126	17905.46	269.6465	-173.386	210.7751	288
22LeuH									
23ValH	-2652.727224	280.7183	103.6830028	10.41769	12607.97	289.4893	129.5936	230.9628	288
24LeuH	-2812.580673	679.9823	103.3558975	24.18598	6394.297	335.6761	-21.9827	291.3091	288
25AspH									
26AsnH									
27SerH	-2551.215232	357.8937	101.609246	13.43073	12764.74	397.002	165.9113	304.3832	288
28ArgH	-2134.084376	289.8729	89.37879391	11.50303	10575.36	350.9826	-259.099	235.8014	288
29SerH	-3290.988337	621.7835	109.3666115	21.30495	7913.425	284.4932	102.0608	435.2906	288
30AsnH									
31GluH	-2249.157569	331.5308	90.92280248	12.35755	7235.906	266.8259	-250.296	203.3382	288
32GlyH	-2143.196847	457.0447	81.45683372	16.19309	-9602.67	512.5606	134.0876	409.9023	288
33LysH	-3669.462441	275.8372	137.5360951	10.08998	15252.65	205.3378	130.6648	191.9945	288
34LeuH	-2616.017342	523.4725	99.24022842	18.47945	6130.005	294.3306	241.2752	212.6228	288
35GluH	-3099.612945	361.3385	111.0971511	12.62201	12570.94	308.6459	-756.99	291.4472	288
36GlyH	-2305.760053	291.896	90.6334627	10.69942	13916.49	416.1335	92.74161	317.5229	288
37LeuH	-2834.042884	325.3552	106.2525759	11.73411	22642.78	545.4844	29.63951	472.0268	288
38ThrH									
39AspH	-3295.475937	315.5931	124.9664241	11.61791	15809.7	289.4787	176.553	251.5206	288
40GluH	-3092.648226	499.0083	119.1361359	18.54629	14732.58	435.2559	1288.469	366.0769	288
41PheH									
42GluH	-2840.886337	389.3518	105.5720709	13.81317	13835.72	443.516	65.56564	369.0506	288
43GluH	-3696.22905	561.4197	122.5792581	18.41924	9307.388	249.1551	81.73276	287.143	288
44LeuH									
45GluH	-4032.819879	604.1856	143.491258	21.30194	8022.837	210.2011	-203.607	229.7126	288



96SerH	-4014.514613	790.7027	141.5724703	27.49517	10892.85	356.7136	387.7059	349.9903	288
97GlyH	-2351.551771	353.2163	87.31925396	12.4375	8387.63	301.6901	-225.368	257.581	288
98AsnH	-3187.415374	330.2269	112.0336326	11.36161	15825.21	326.229	-320.488	319.3955	288
99LysH									
100IleH	-3415.762163	497.0202	126.5222936	17.95432	13909.01	376.4242	230.9681	339.9851	288
101LysH	-3603.447459	330.3257	131.3302916	11.79062	10177.4	168.1453	174.1561	156.2912	288
102AspH									
103LeuH									
104SerH	-1343.708166	343.2572	57.24830888	14.67667	21000.86	1761.877	-1027.66	2060.139	288
105ThrH	-3648.240745	406.6217	142.2765536	15.44994	13845.23	274.9661	363.3165	233.0733	288
106IleH	-3510.772718	468.949	125.4604797	16.59421	12863.82	331.1643	-719.405	365.0598	288
107GluH	-3226.139015	383.3069	121.3288527	13.98856	17495.63	411.969	-239.191	361.3521	288
109LeuH									
110LysH									
111LysH	-3656.053417	438.7584	130.2021922	15.34791	17916.16	407.4302	-792.307	392.6009	288
112LeuH									
113GluH	-3772.246062	398.9215	131.2512651	13.68051	17183.47	321.6459	88.15171	321.8367	288
114AsnH	-3473.499544	600.6746	127.8850336	21.60008	11065.05	361.0708	-83.5584	329.5979	288
115LeuH	-2893.999334	382.965	107.481612	13.72507	14427.11	391.7125	155.4573	345.4583	288
116LysH	-3977.581884	824.2602	150.1585366	30.51706	7799.498	283.7292	-67.1076	253.1469	288
117SerH	-3108.116609	357.272	124.2156015	13.72584	11052.63	259.524	-200.967	207.4396	288
118LeuH									
119AspH	-2890.956909	676.7377	102.067258	23.26919	9019.222	451.4894	-351.485	436.1451	288
120LeuH	-4006.921957	734.5976	135.6676521	24.87545	7266.442	226.9812	-79.4405	306.9837	288
121PheH	-3253.151225	275.0134	110.3060201	9.189366	11721.71	192.9533	-192.601	202.0409	288
122AsnH	-3744.960566	675.287	128.8866281	22.93007	9699.782	300.0355	368.3929	305.606	288
123CysH	-4063.766367	393.9821	143.7983914	13.74758	17065.13	271.4232	725.7601	265.0568	288
124GluH									
125ValH									
126ThrH	-4158.049425	598.2383	156.1752123	22.0837	11648.18	276.9692	306.3831	249.6365	288
127AsnH	-2452.485941	317.4684	90.99827407	11.2168	16546.87	500.0016	-420.334	430.6083	288
128LeuH									
129AsnH	-1465.903436	296.2168	72.05406357	11.3685	11147.86	673.3457	236.2039	289.017	288
130AspH									
131TyrH									
132ArgH									
133GluH	-3310.713749	394.91	117.4575239	13.71611	18908.41	429.6908	83.30054	414.6524	288
134AsnH	-2929.942778	391.9502	106.4893203	13.8083	17621.54	499.9573	-578.738	458.3787	288
135ValH									
136PheH									
137LysH	-4123.611444	547.8017	149.6226925	19.57339	20043.9	442.9582	466.9055	417.7266	288
138LeuH	-4224.914114	325.4039	150.8925061	11.46519	17394.25	222.3178	207.6898	214.5046	288
139LeuH	-3429.698609	316.9974	124.0954162	11.2175	15542.85	271.1486	-42.7714	253.943	288
141GlnH	-2797.5808	284.2827	101.8283984	9.840497	18860.27	460.0191	-256.914	354.6854	288
142LeuH	-3838.898812	310.9937	143.2245766	11.48037	14110.41	188.4654	-22.9827	168.649	288
143ThrH	-3370.150317	954.197	119.0003192	34.20599	7877.398	373.4143	591.4523	600.5078	288
144TyrH	-2847.485266	600.4593	98.84885159	20.36049	9683.44	403.041	442.1523	402.2351	288
145LeuH	-3125.596473	837.2713	115.6091666	31.13145	8135.21	386.0608	471.7651	495.244	288

146AspH	-2996.123423	836.1916	110.0982722	30.31316	6963.714	387.8586	-38.9285	453.7197	288
147GlyH	-3547.08029	595.5872	125.2010071	20.85727	-7432.53	232.9487	225.4821	262.664	288
148TyrH	-3423.676759	406.6778	123.0147173	14.30146	11822.68	256.5647	348.0644	242.9379	288
149AspH	-3490.727724	529.8895	121.7545612	18.17005	10913.9	312.0286	-167.486	310.5269	288
151AspH	-3313.960856	407.1352	115.5848811	13.93652	15111.75	356.1316	-86.9268	353.9827	288
152AspH	-4346.53621	522.912	128.479482	15.5624	9734.46	192.4976	-287.609	259.1067	288
153LysH	-3137.633898	101.0423	114.9173941	3.45483	17398.89	141.1685	565.207	83.76088	288
154GluH									
155TrpH									
156LeuH									
157GluH									
158HisH									
	DG	DG_err	DV	DV_err	Cf	Cf_err	Cu	Cu_err	T
Average values	-3220.108506	449.4315	117.8826325	16.06304	12472.94	343.3919	51.01856	318.909	288

Supplemental Table 2. Residue specific fit parameters for HSQC peak intensity loss vs pressure at 293 K

Residue Name	$\Delta G$ (cal/mol)	$\Delta G_{err}$	$\Delta V$ (ml/mol)	$\Delta V_{err}$	Cf (a.u.)	Cf_err	Cu (a.u.)	Cu_err	T (K)
2GluH	-2197.130508	283.1276	81.23625435	9.855452	15070.13	486.421	-357.713	412.9389	293
3MetH	-3028.828907	404.711	107.3143802	14.00389	15020.51	426.2648	-34.968	392.0226	293
4GlyH	-2420.629473	304.2346	84.93568501	10.30722	13238.47	378.4183	95.59884	365.2795	293
5ArgH	-3297.059711	527.4138	113.9172602	17.65196	16683.91	598.2818	-659.659	512.9249	293
6ArgH									
7IleH	-3389.75777	323.5792	112.3059193	10.61347	16055.96	288.8647	-23.7317	316.382	293
8HisH	-2744.111572	321.5531	100.1858413	11.3076	17782.89	445.2549	123.5896	401.0394	293
9LeuH									
10GluH									
11LeuH	-3420.392232	383.7436	113.2856216	12.58505	19075.48	412.8001	-569.723	452.189	293
12ArgH	-2694.69944	225.3842	89.41958913	7.380593	21105.44	384.2736	-490.937	423.7439	293
13AsnH									
14ArgH	-3067.630543	408.6232	105.0767104	13.73866	19121.37	508.8257	15.06946	522.3626	293
15ThrH	-3572.798678	429.5927	123.1846936	14.5865	22114.2	484.9346	260.4665	492.3248	293
17SerH	-2689.494203	319.5414	96.44136651	11.10441	22667.92	594.1874	90.88679	539.904	293
18AspH	-2455.236039	222.5373	83.93162829	7.37309	20035.24	460.1929	-800.048	433.6883	293
19ValH									
20LysH	-2637.136971	296.4922	92.90726172	10.12137	14986.94	365.316	116.7036	351.7894	293
21GluH	-3080.862844	405.1542	111.4306051	14.24311	21402.94	557.6827	355.9486	517.3581	293
22LeuH	-2970.438835	246.1124	102.8670642	8.523233	19975.13	273.6153	0	0	293
23ValH	-3534.073086	392.9229	119.57017	13.12391	15705.72	320.3759	127.4012	336.4703	293
24LeuH									
25AspH									
26AsnH	-3360.699446	454.3702	108.1867309	14.56367	11760.21	306.0294	-303.934	356.3583	293
27SerH	-3649.208185	533.1113	130.5838331	18.71536	15849.79	413.2559	430.4633	394.8882	293
28ArgH	-2768.643927	332.3342	101.6631842	11.74765	13518.81	337.8674	384.5369	301.4538	293
29SerH	-2788.409783	435.4824	95.73218558	14.62203	12684.95	412.4533	147.8698	420.7849	293
30AsnH									
31GluH									
32GlyH	-2352.595952	573.9508	83.29553115	19.52783	-13342.8	786.316	474.7049	741.796	293
33LysH	-3053.700728	247.6076	108.0040365	8.54108	20654.46	337.9733	-9.88788	326.1545	293
34LeuH	-3042.219507	546.1116	104.2165321	18.35806	8864.189	314.6396	136.546	322.9219	293
35GluH	-2433.056889	403.5753	84.12931683	13.54425	17843.81	694.9593	-559.219	694.1968	293
36GlyH	-2538.616576	276.3844	91.6180864	9.580036	16530.55	408.2915	-187.69	372.5768	293
37LeuH	-2481.278729	368.4644	86.86681616	12.48113	29642.21	1023.524	-884.336	994.8728	293
38ThrH									
39AspH	-2672.343585	392.7121	92.46693214	13.23994	20554.53	657.533	-157.14	658.6498	293
40GluH	-3415.877493	334.0295	112.8514182	10.93193	18515.93	353.1746	-765.53	388.8712	293
41PheH									
42GluH	-3351.617064	740.1222	112.5099722	24.53409	18825.43	768.1583	505.5319	819.9946	293
43GluH	-3247.649666	516.3428	110.6787571	17.31703	13394.66	412.0266	16.66071	427.296	293
44LeuH									
45GluH	-3326.478182	502.9247	107.2635866	16.14141	10828.5	319.7811	-380.851	371.3878	293

46PheH									
47LeuH	-3118.53107	193.0358	102.8646454	6.303884	15563.97	196.1628	-573.844	217.8458	293
48SerH									
49ThrH	-3584.421959	398.9332	115.9515197	12.83631	14458.17	292.6919	-10.0933	335.2223	293
50IleH	-3493.779187	433.1942	114.329047	14.06978	10514.16	231.426	409.267	259.4183	293
51AsnH									
52ValH	-3147.078964	318.7124	101.1202762	10.20372	12709.36	251.7129	-104.029	296.3643	293
53GlyH	-3329.051862	383.6397	109.8083977	12.53617	12641.81	288.157	-535.2	318.7908	293
54LeuH	-3403.778129	409.1337	113.3312966	13.47496	17064.8	374.4156	496.6215	405.9508	293
55ThrH	-2829.589527	376.9827	95.67356934	12.52152	8468.696	240.1605	-189.517	253.3496	293
56SerH	-2843.6607	353.11	98.33097657	11.92878	15660.07	409.0934	-122.212	411.3004	293
57IleH									
58AlaH									
59AsnH									
60LeuH	-2992.977202	512.2887	98.40487891	16.68322	27569.76	990.2025	-1283.94	1110.289	293
62LysH									
63LeuH	-3349.784113	373.5397	111.3564777	12.284	16549.08	346.7793	139.8283	377.4573	293
64AsnH									
65LysH	-3067.484792	383.6971	102.7901346	12.6806	15009.43	370.7179	108.9166	398.2122	293
66LeuH									
67LysH									
68LysH									
69LeuH									
70GluH	-3062.749528	412.156	104.4814851	13.81206	13540.72	360.0116	159.2095	372.6436	293
71LeuH									
72SerH	-2473.420036	321.7842	91.03514744	11.29229	13809.94	390.9267	656.8146	341.6127	293
73AspH	-5642.237059	574.3924	187.0290869	18.93559	9590.505	141.3335	-49.8102	150.95	293
74AsnH									
75ArgH	-2902.778702	455.3839	98.4994349	15.17588	11495.1	377.5346	-214.581	395.1416	293
76ValH									
77SerH									
78GlyH	-2987.818977	275.3637	100.8688541	9.147685	23487	455.1358	-950.394	481.4783	293
79GlyH	-2585.180626	337.1362	91.5423814	11.53873	13865.19	409.5765	-333.534	389.5678	293
80LeuH	-2202.319473	338.0412	74.24917646	11.1345	16256.71	623.389	-696.486	660.9879	293
81GluH	-3106.400477	246.0573	107.2315347	8.327606	23384.6	367.4639	61.97658	371.5888	293
82ValH	-2863.683245	273.3036	99.55890997	9.273505	26829.7	546.3162	-664.623	543.3343	293
83LeuH	-2821.732175	441.7846	97.99830128	14.96912	21874.47	730.3282	-353.989	727.4632	293
84AlaH	-3825.477436	525.9508	133.8654695	18.1275	29436.6	708.9793	524.3668	702.9183	293
85GluH									
86LysH	-3783.995275	517.5257	121.6007066	16.56124	19395.95	472.7843	-187.232	545.9905	293
87CysH	-3140.262258	385.4448	106.6729441	12.88214	24994.01	616.0501	-474.03	643.103	293
89AsnH	-3196.311285	303.1297	108.6873948	10.14385	14955.81	275.2218	94.15775	286.6906	293
90LeuH	-3076.478197	402.1171	107.3953191	13.72828	18114.58	472.5252	75.23244	467.5105	293
91ThrH	-2273.670132	342.8723	76.21742618	11.26693	9103.003	334.1314	-333.268	359.87	293
92HisH	-3781.983243	414.0531	124.1660884	13.49265	13558.91	257.2758	315.7296	284.9783	293
93LeuH									
94AsnH	-3237.214605	415.7155	104.4524589	13.34993	13359.41	328.4548	-4.58611	381.934	293
95LeuH	-2800.154576	240.9863	86.49921608	7.530604	10686.98	199.2982	-695.188	263.8855	293

96SerH	-3355.047641	395.1318	113.0976672	13.14348	13482.59	290.2667	508.0815	307.3231	293
97GlyH	-2942.022722	403.9407	100.7042698	13.55642	10767.71	312.8298	-376.674	321.4702	293
98AsnH	-3247.251177	267.4445	110.1491871	8.935372	19339.02	303.1523	319.9039	317.2651	293
99LysH									
100IleH	-2964.213736	289.318	100.6040312	9.648468	19165.93	379.1107	51.73249	396.5992	293
101LysH	-2975.689003	364.2199	99.94776684	12.05145	12988.04	327.0002	-219.598	349.7635	293
102AspH	-4089.84435	369.994	129.3395371	11.67754	22330.78	334.6826	448.215	395.6947	293
103LeuH	-2372.436384	307.4844	82.00844503	10.30552	18336.81	550.9197	-70.8261	549.9809	293
104SerH									
105ThrH	-3588.892125	388.1262	123.8842382	13.19349	17633.56	353.3661	-110.458	358.0363	293
106IleH	-2817.894348	273.1125	95.5255716	9.087772	17728.08	367.3987	-426.882	385.3669	293
107GluH	-2811.636566	283.9098	93.16621039	9.29269	24174.65	525.9627	-892.819	581.0542	293
109LeuH									
110LysH									
111LysH	-3324.179363	290.2855	107.672891	9.349641	23969.01	406.1394	-638.149	467.0788	293
112LeuH									
113GluH	-2708.840193	287.4489	88.92595348	9.345307	23613.65	542.5546	-564.765	614.0739	293
114AsnH	-2931.938589	305.8328	97.73693056	10.05968	15019.79	325.6944	-324.358	354.347	293
115LeuH	-4222.220863	584.9724	139.2687411	19.15874	18449.52	408.3734	881.6883	445.6016	293
116LysH	-3224.339946	582.0655	110.0118488	19.53598	10794.31	387.038	-228.171	400.4904	293
117SerH	-3013.773172	307.8066	99.62355656	10.06492	15203.86	320.532	-482.881	354.9283	293
118LeuH	-2786.638403	270.1207	91.58074735	8.789878	11875.37	241.9042	-105.063	272.6408	293
119AspH	-3277.363003	393.9396	107.8983953	12.85142	12299.32	279.1052	157.0058	310.3004	293
120LeuH	-2873.63613	367.4991	94.27233859	11.9458	10646.53	289.8882	-423.135	327.5273	293
121PheH	-3655.943073	376.0356	115.3563178	11.78195	14427.41	237.9164	90.90752	292.0973	293
122AsnH	-3547.748645	470.3138	113.0253936	14.94669	12457.19	301.797	-53.2468	357.3764	293
123CysH	-4053.324749	376.7347	131.7587237	12.17696	20030.44	310.4585	482.5574	348.6048	293
124GluH									
125ValH									
126ThrH	-3757.289893	342.4935	122.1923783	11.07078	14670.78	236.3433	77.43711	266.7316	293
127AsnH	-3093.906691	314.471	106.8359565	10.64443	21814.29	444.9819	-146.098	449.6473	293
128LeuH									
129AsnH	-1915.579218	448.5149	84.46325112	16.56632	9785.172	721.3274	-169.731	327.3081	293
130AspH									
131TyrH									
132ArgH									
133GluH									
134AsnH	-3594.941901	206.6955	117.7515368	6.719819	22224.62	237.5852	-473.166	265.2694	293
135ValH									
136PheH									
137LysH	-3286.744676	305.3037	107.9784274	9.943376	25231.9	455.0454	-437.31	508.0845	293
138LeuH	-2699.246374	359.2671	91.95208315	11.98202	25475.52	736.214	-367.606	763.6403	293
139LeuH									
141GlnH	-2681.955111	216.3591	93.97557698	7.362932	23671.6	416.1082	-184.06	405.7747	293
142LeuH	-3279.622831	274.8321	110.5970885	9.140514	18494.94	308.7249	-666.068	326.7932	293
143ThrH	-2819.298139	360.9034	97.94773273	12.23144	10454.55	276.3201	173.3169	275.0268	293
144TyrH	-2910.485737	379.5285	98.22138143	12.59682	13227.39	346.3527	267.4519	366.8175	293
145LeuH	-4437.389022	813.3244	146.1196949	26.60892	10380.1	303.3325	254.9705	331.0327	293

146AspH	-3363.610357	401.5959	109.7044811	13.00599	9513.983	214.3619	-25.53	242.6501	293
147GlyH	-3735.415299	796.846	118.8483928	25.28944	-9740.12	381.8105	358.3618	450.653	293
148TyrH	-3943.638798	260.8381	128.5195434	8.448359	14872.92	166.8455	324.6134	186.8504	293
149AspH	-3089.649494	431.9925	101.5927681	14.07254	14928.16	409.8778	60.51952	458.5957	293
150ArgH									
151AspH	-3408.924113	414.1435	111.6212958	13.45575	19493.15	441.9399	50.64244	495.6464	293
152AspH									
153LysH	-3335.788801	436.9471	113.0805775	14.59998	22238.88	535.647	804.1076	561.0719	293
154GluH	-3632.40087	356.0629	117.6328499	11.36279	17488.3	275.4679	-75.1898	320.5846	293
155TrpH									
156LeuH									
157GluH									
158HisH									
	DG	DG_err	DV	DV_err	Cf	Cf_err	Cu	Cu_err	T
Average values	-3132.449929	382.0551	105.7295377	12.69881	16327.48	408.8547	-96.9119	419.0918	293

Supplemental Table 3. Residue specific fit parameters for HSQC peak intensity loss vs pressure at 298 K

Residue Name	$\Delta G$ (cal/mol)	$\Delta G_{err}$	$\Delta V$ (ml/mol)	$\Delta V_{err}$	Cf (a.u.)	Cf_err	Cu (a.u.)	Cu_err	T (K)
2GluH	-2197.130508	283.1276	81.23625435	9.855452	15070.13	486.421	-357.713	412.9389	293
3MetH	-3028.828907	404.711	107.3143802	14.00389	15020.51	426.2648	-34.968	392.0226	293
4GlyH	-2420.629473	304.2346	84.93568501	10.30722	13238.47	378.4183	95.59884	365.2795	293
5ArgH	-3297.059711	527.4138	113.9172602	17.65196	16683.91	598.2818	-659.659	512.9249	293
6ArgH									
7IleH	-3389.75777	323.5792	112.3059193	10.61347	16055.96	288.8647	-23.7317	316.382	293
8HisH	-2744.111572	321.5531	100.1858413	11.3076	17782.89	445.2549	123.5896	401.0394	293
9LeuH									
10GluH									
11LeuH	-3420.392232	383.7436	113.2856216	12.58505	19075.48	412.8001	-569.723	452.189	293
12ArgH	-2694.69944	225.3842	89.41958913	7.380593	21105.44	384.2736	-490.937	423.7439	293
13AsnH									
14ArgH	-3067.630543	408.6232	105.0767104	13.73866	19121.37	508.8257	15.06946	522.3626	293
15ThrH	-3572.798678	429.5927	123.1846936	14.5865	22114.2	484.9346	260.4665	492.3248	293
17SerH	-2689.494203	319.5414	96.44136651	11.10441	22667.92	594.1874	90.88679	539.904	293
18AspH	-2455.236039	222.5373	83.93162829	7.37309	20035.24	460.1929	-800.048	433.6883	293
19ValH									
20LysH	-2637.136971	296.4922	92.90726172	10.12137	14986.94	365.316	116.7036	351.7894	293
21GluH	-3080.862844	405.1542	111.4306051	14.24311	21402.94	557.6827	355.9486	517.3581	293
22LeuH	-2970.438835	246.1124	102.8670642	8.523233	19975.13	273.6153	0	0	293
23ValH	-3534.073086	392.9229	119.57017	13.12391	15705.72	320.3759	127.4012	336.4703	293
24LeuH									
25AspH									
26AsnH	-3360.699446	454.3702	108.1867309	14.56367	11760.21	306.0294	-303.934	356.3583	293
27SerH	-3649.208185	533.1113	130.5838331	18.71536	15849.79	413.2559	430.4633	394.8882	293
28ArgH	-2768.643927	332.3342	101.6631842	11.74765	13518.81	337.8674	384.5369	301.4538	293
29SerH	-2788.409783	435.4824	95.73218558	14.62203	12684.95	412.4533	147.8698	420.7849	293
30AsnH									
31GluH									
32GlyH	-2352.595952	573.9508	83.29553115	19.52783	-13342.8	786.316	474.7049	741.796	293
33LysH	-3053.700728	247.6076	108.0040365	8.54108	20654.46	337.9733	-9.88788	326.1545	293
34LeuH	-3042.219507	546.1116	104.2165321	18.35806	8864.189	314.6396	136.546	322.9219	293
35GluH	-2433.056889	403.5753	84.12931683	13.54425	17843.81	694.9593	-559.219	694.1968	293
36GlyH	-2538.616576	276.3844	91.6180864	9.580036	16530.55	408.2915	-187.69	372.5768	293
37LeuH	-2481.278729	368.4644	86.86681616	12.48113	29642.21	1023.524	-884.336	994.8728	293
38ThrH									
39AspH	-2672.343585	392.7121	92.46693214	13.23994	20554.53	657.533	-157.14	658.6498	293
40GluH	-3415.877493	334.0295	112.8514182	10.93193	18515.93	353.1746	-765.53	388.8712	293
41PheH									
42GluH	-3351.617064	740.1222	112.5099722	24.53409	18825.43	768.1583	505.5319	819.9946	293
43GluH	-3247.649666	516.3428	110.6787571	17.31703	13394.66	412.0266	16.66071	427.296	293
44LeuH									
45GluH	-3326.478182	502.9247	107.2635866	16.14141	10828.5	319.7811	-380.851	371.3878	293

46PheH									
47LeuH	-3118.53107	193.0358	102.8646454	6.303884	15563.97	196.1628	-573.844	217.8458	293
48SerH									
49ThrH	-3584.421959	398.9332	115.9515197	12.83631	14458.17	292.6919	-10.0933	335.2223	293
50IleH	-3493.779187	433.1942	114.329047	14.06978	10514.16	231.426	409.267	259.4183	293
51AsnH									
52ValH	-3147.078964	318.7124	101.1202762	10.20372	12709.36	251.7129	-104.029	296.3643	293
53GlyH	-3329.051862	383.6397	109.8083977	12.53617	12641.81	288.157	-535.2	318.7908	293
54LeuH	-3403.778129	409.1337	113.3312966	13.47496	17064.8	374.4156	496.6215	405.9508	293
55ThrH	-2829.589527	376.9827	95.67356934	12.52152	8468.696	240.1605	-189.517	253.3496	293
56SerH	-2843.6607	353.11	98.33097657	11.92878	15660.07	409.0934	-122.212	411.3004	293
57IleH									
58AlaH									
59AsnH									
60LeuH	-2992.977202	512.2887	98.40487891	16.68322	27569.76	990.2025	-1283.94	1110.289	293
62LysH									
63LeuH	-3349.784113	373.5397	111.3564777	12.284	16549.08	346.7793	139.8283	377.4573	293
64AsnH									
65LysH	-3067.484792	383.6971	102.7901346	12.6806	15009.43	370.7179	108.9166	398.2122	293
66LeuH									
67LysH									
68LysH									
69LeuH									
70GluH	-3062.749528	412.156	104.4814851	13.81206	13540.72	360.0116	159.2095	372.6436	293
71LeuH									
72SerH	-2473.420036	321.7842	91.03514744	11.29229	13809.94	390.9267	656.8146	341.6127	293
73AspH	-5642.237059	574.3924	187.0290869	18.93559	9590.505	141.3335	-49.8102	150.95	293
74AsnH									
75ArgH	-2902.778702	455.3839	98.4994349	15.17588	11495.1	377.5346	-214.581	395.1416	293
76ValH									
77SerH									
78GlyH	-2987.818977	275.3637	100.8688541	9.147685	23487	455.1358	-950.394	481.4783	293
79GlyH	-2585.180626	337.1362	91.5423814	11.53873	13865.19	409.5765	-333.534	389.5678	293
80LeuH	-2202.319473	338.0412	74.24917646	11.1345	16256.71	623.389	-696.486	660.9879	293
81GluH	-3106.400477	246.0573	107.2315347	8.327606	23384.6	367.4639	61.97658	371.5888	293
82ValH	-2863.683245	273.3036	99.55890997	9.273505	26829.7	546.3162	-664.623	543.3343	293
83LeuH	-2821.732175	441.7846	97.99830128	14.96912	21874.47	730.3282	-353.989	727.4632	293
84AlaH	-3825.477436	525.9508	133.8654695	18.1275	29436.6	708.9793	524.3668	702.9183	293
85GluH									
86LysH	-3783.995275	517.5257	121.6007066	16.56124	19395.95	472.7843	-187.232	545.9905	293
87CysH	-3140.262258	385.4448	106.6729441	12.88214	24994.01	616.0501	-474.03	643.103	293
89AsnH	-3196.311285	303.1297	108.6873948	10.14385	14955.81	275.2218	94.15775	286.6906	293
90LeuH	-3076.478197	402.1171	107.3953191	13.72828	18114.58	472.5252	75.23244	467.5105	293
91ThrH	-2273.670132	342.8723	76.21742618	11.26693	9103.003	334.1314	-333.268	359.87	293
92HisH	-3781.983243	414.0531	124.1660884	13.49265	13558.91	257.2758	315.7296	284.9783	293
93LeuH									
94AsnH	-3237.214605	415.7155	104.4524589	13.34993	13359.41	328.4548	-4.58611	381.934	293
95LeuH	-2800.154576	240.9863	86.49921608	7.530604	10686.98	199.2982	-695.188	263.8855	293

96SerH	-3355.047641	395.1318	113.0976672	13.14348	13482.59	290.2667	508.0815	307.3231	293
97GlyH	-2942.022722	403.9407	100.7042698	13.55642	10767.71	312.8298	-376.674	321.4702	293
98AsnH	-3247.251177	267.4445	110.1491871	8.935372	19339.02	303.1523	319.9039	317.2651	293
99LysH									
100IleH	-2964.213736	289.318	100.6040312	9.648468	19165.93	379.1107	51.73249	396.5992	293
101LysH	-2975.689003	364.2199	99.94776684	12.05145	12988.04	327.0002	-219.598	349.7635	293
102AspH	-4089.84435	369.994	129.3395371	11.67754	22330.78	334.6826	448.215	395.6947	293
103LeuH	-2372.436384	307.4844	82.00844503	10.30552	18336.81	550.9197	-70.8261	549.9809	293
104SerH									
105ThrH	-3588.892125	388.1262	123.8842382	13.19349	17633.56	353.3661	-110.458	358.0363	293
106IleH	-2817.894348	273.1125	95.5255716	9.087772	17728.08	367.3987	-426.882	385.3669	293
107GluH	-2811.636566	283.9098	93.16621039	9.29269	24174.65	525.9627	-892.819	581.0542	293
109LeuH									
110LysH									
111LysH	-3324.179363	290.2855	107.672891	9.349641	23969.01	406.1394	-638.149	467.0788	293
112LeuH									
113GluH	-2708.840193	287.4489	88.92595348	9.345307	23613.65	542.5546	-564.765	614.0739	293
114AsnH	-2931.938589	305.8328	97.73693056	10.05968	15019.79	325.6944	-324.358	354.347	293
115LeuH	-4222.220863	584.9724	139.2687411	19.15874	18449.52	408.3734	881.6883	445.6016	293
116LysH	-3224.339946	582.0655	110.0118488	19.53598	10794.31	387.038	-228.171	400.4904	293
117SerH	-3013.773172	307.8066	99.62355656	10.06492	15203.86	320.532	-482.881	354.9283	293
118LeuH	-2786.638403	270.1207	91.58074735	8.789878	11875.37	241.9042	-105.063	272.6408	293
119AspH	-3277.363003	393.9396	107.8983953	12.85142	12299.32	279.1052	157.0058	310.3004	293
120LeuH	-2873.63613	367.4991	94.27233859	11.9458	10646.53	289.8882	-423.135	327.5273	293
121PheH	-3655.943073	376.0356	115.3563178	11.78195	14427.41	237.9164	90.90752	292.0973	293
122AsnH	-3547.748645	470.3138	113.0253936	14.94669	12457.19	301.797	-53.2468	357.3764	293
123CysH	-4053.324749	376.7347	131.7587237	12.17696	20030.44	310.4585	482.5574	348.6048	293
124GluH									
125ValH									
126ThrH	-3757.289893	342.4935	122.1923783	11.07078	14670.78	236.3433	77.43711	266.7316	293
127AsnH	-3093.906691	314.471	106.8359565	10.64443	21814.29	444.9819	-146.098	449.6473	293
128LeuH									
129AsnH	-1915.579218	448.5149	84.46325112	16.56632	9785.172	721.3274	-169.731	327.3081	293
130AspH									
131TyrH									
132ArgH									
133GluH									
134AsnH	-3594.941901	206.6955	117.7515368	6.719819	22224.62	237.5852	-473.166	265.2694	293
135ValH									
136PheH									
137LysH	-3286.744676	305.3037	107.9784274	9.943376	25231.9	455.0454	-437.31	508.0845	293
138LeuH	-2699.246374	359.2671	91.95208315	11.98202	25475.52	736.214	-367.606	763.6403	293
139LeuH									
141GlnH	-2681.955111	216.3591	93.97557698	7.362932	23671.6	416.1082	-184.06	405.7747	293
142LeuH	-3279.622831	274.8321	110.5970885	9.140514	18494.94	308.7249	-666.068	326.7932	293
143ThrH	-2819.298139	360.9034	97.94773273	12.23144	10454.55	276.3201	173.3169	275.0268	293
144TyrH	-2910.485737	379.5285	98.22138143	12.59682	13227.39	346.3527	267.4519	366.8175	293
145LeuH	-4437.389022	813.3244	146.1196949	26.60892	10380.1	303.3325	254.9705	331.0327	293

146AspH	-3363.610357	401.5959	109.7044811	13.00599	9513.983	214.3619	-25.53	242.6501	293
147GlyH	-3735.415299	796.846	118.8483928	25.28944	-9740.12	381.8105	358.3618	450.653	293
148TyrH	-3943.638798	260.8381	128.5195434	8.448359	14872.92	166.8455	324.6134	186.8504	293
149AspH	-3089.649494	431.9925	101.5927681	14.07254	14928.16	409.8778	60.51952	458.5957	293
150ArgH									
151AspH	-3408.924113	414.1435	111.6212958	13.45575	19493.15	441.9399	50.64244	495.6464	293
152AspH									
153LysH	-3335.788801	436.9471	113.0805775	14.59998	22238.88	535.647	804.1076	561.0719	293
154GluH	-3632.40087	356.0629	117.6328499	11.36279	17488.3	275.4679	-75.1898	320.5846	293
155TrpH									
156LeuH									
157GluH									
158HisH									
	DG	DG_err	DV	DV_err	Cf	Cf_err	Cu	Cu_err	T
Average values	-3132.449929	382.0551	105.7295377	12.69881	16327.48	408.8547	-96.9119	419.0918	293

Supplemental Table 4. Residue specific fit parameters for HSQC peak intensity loss vs pressure at 303 K

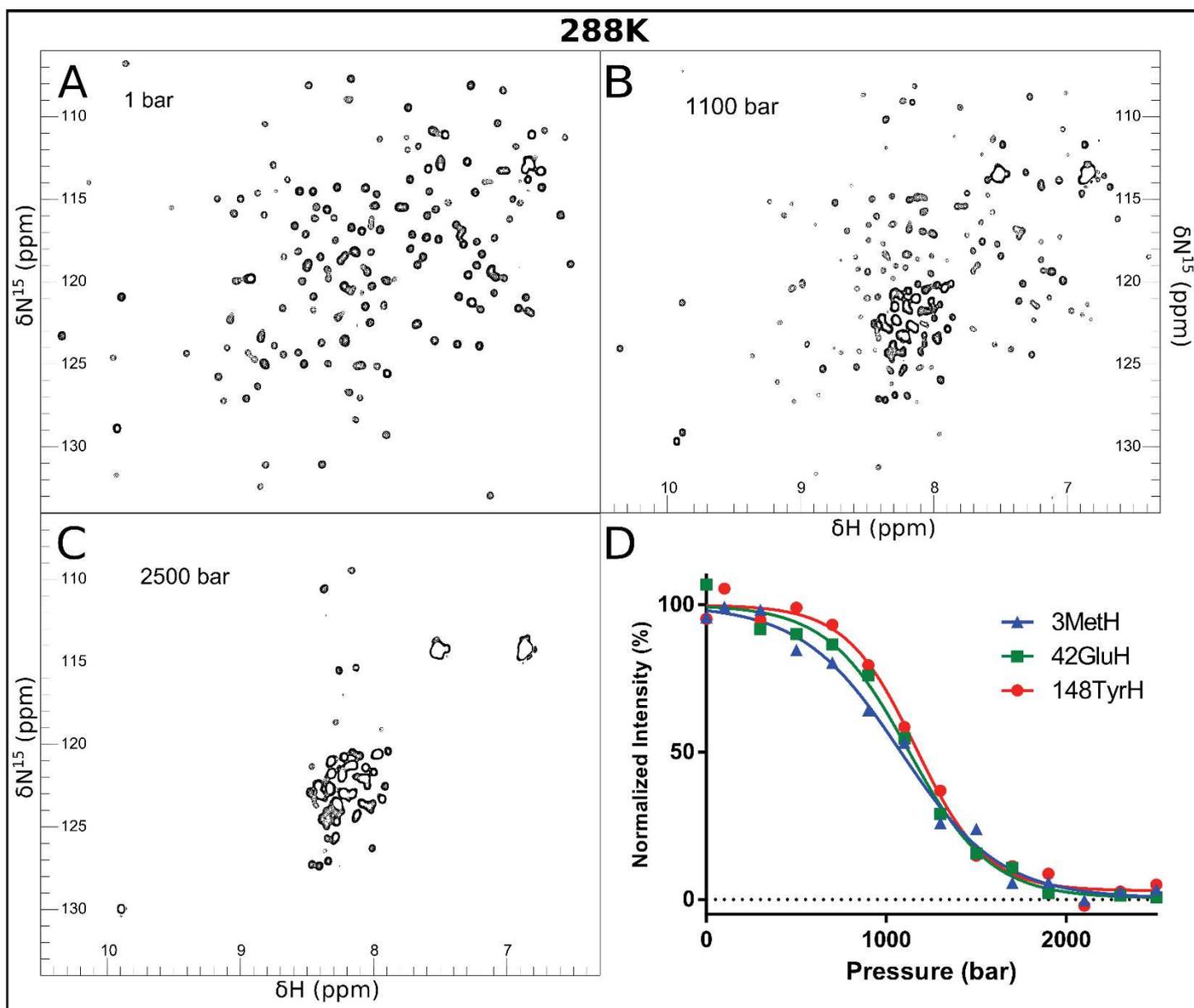
Residue Name	$\Delta G$ (cal/mol)	$\Delta G_{err}$	$\Delta V$ (ml/mol)	$\Delta V_{err}$	Cf (a.u.)	Cf_err	Cu (a.u.)	Cu_err	T (K)
2GluH	-2784.095208	856.7654	109.0440047	31.84204	9948.778	682.4199	-150.259	542.3309	303
3MetH	-1286.838444	164.7066	53.13142399	5.846773	15626.58	713.2111	-1336	529.8721	303
4GlyH	-1639.264509	456.2104	68.30021629	16.14746	10771.85	873.7399	199.8467	530.7688	303
5ArgH	-1863.426971	182.2478	73.38751992	8.22726	14692.3	0	-185.891	533.2302	303
6ArgH									
7IleH	-1402.283199	326.8355	58.7911087	11.16736	20088.54	1547.296	-736.756	899.2165	303
8HisH									
9LeuH									
10GluH									
11LeuH									
12ArgH									
13AsnH									
14ArgH									
15ThrH	-1782.407569	273.7474	70.31821531	9.572877	22577.94	988.9923	-174.231	691.7122	303
17SerH	-2507.651837	410.305	85.27373578	13.36235	18762.55	792.9476	-402.811	703.4694	303
18AspH	-2808.180757	485.8312	97.21311447	16.0612	18596.34	783.3298	-207.344	652.0025	303
19ValH									
20LysH	-1420.189253	300.1214	59.34814241	10.27316	15660.65	1090.864	-619.574	639.9402	303
21GluH									
22LeuH									
23ValH	-2300.941825	183.9569	76.68700392	5.853463	14792.71	321.0511	-321.926	289.4802	303
24LeuH									
25AspH									
26AsnH									
27SerH	-1927.540646	313.1581	73.97376429	10.92471	12811.31	602.0334	-901.416	455.8297	303
28ArgH	-1382.913365	297.566	60.2910927	10.30071	14064.41	983.4887	-178.017	521.3157	303
29SerH	-3529.983281	303.7102	123.1618237	10.38636	10502.7	139.4857	481.865	118.5945	303
30AsnH									
31GluH									
32GlyH	-2282.290583	386.5224	83.0340655	13.31957	-10942.3	443.2975	-184.76	390.2612	303
33LysH									
34LeuH	-3703.938658	908.8374	127.7027214	35.22829	10928.54	361.3143	854.5369	1213.32	303
35GluH	-1581.686587	330.4097	62.23086828	11.28688	18170.94	1188.247	-734.417	817.9097	303
36GlyH	-1908.704496	242.5094	72.49004135	8.398388	13136.72	471.5319	-540.757	364.963	303
37LeuH	-2349.171669	328.9426	83.97835624	11.23464	26719.79	895.9896	-110.157	823.8795	303
38ThrH									
39AspH	-2125.305537	334.4159	74.81866731	11.28047	17744.48	787.2471	-1215.78	704.4127	303
40GluH									
41PheH									
42GluH	-2846.667659	473.0731	91.89960736	14.91877	18472.7	732.0713	-444.574	726.0362	303
43GluH	-2350.804781	293.7205	73.47259888	9.321118	14407.5	444.8922	-1071.63	584.6891	303
44LeuH									
45GluH	-2493.345326	215.2705	85.50996143	7.18371	12411.21	245.0435	-16.8413	249.4962	303

46PheH	-2998.001172	219.8552	99.86444826	7.226385	15935.63	225.7814	893.8316	246.1528	303
47LeuH	-2616.40142	191	90.57854711	6.426579	15753.19	253.5708	159.5748	253.0791	303
48SerH									
49ThrH	-2744.611519	480.0866	94.46098356	16.09622	13898.28	492.6548	-14.7402	508.3446	303
50IleH	-1795.291411	349.1449	66.78793082	11.86613	12381.97	678.4563	-79.4129	546.3937	303
51AsnH									
52ValH	-2804.081113	161.4125	96.66986931	5.430603	11649.43	143.7991	-17.3556	145.2799	303
53GlyH	-1567.276056	203.0581	61.67147166	6.904607	13383.68	545.8601	-679.584	370.1104	303
54LeuH	-2232.029365	430.9381	77.63977293	14.42693	18936.5	876.7528	334.9752	909.4276	303
55ThrH	-2800.934738	395.4141	103.6920891	14.08972	7961.572	253.1454	100.3076	186.1658	303
56SerH	-1964.321034	234.8845	69.89988029	7.927977	16397.76	534.0667	-924.362	493.9242	303
57IleH									
58AlaH									
59AsnH									
60LeuH	-3454.094069	667.7062	118.7721517	22.95259	28412.17	962.588	1635.644	1170.428	303
62LysH									
63LeuH	-3175.117902	252.0186	99.15162937	7.90707	17117.2	272.3764	-515.013	345.0244	303
64AsnH									
65LysH									
66LeuH									
67LysH	-2335.774997	468.0342	76.63694702	15.20298	11494.74	562.4226	-482.562	641.9435	303
68LysH									
69LeuH	-2878.310638	331.4768	100.6959352	11.28722	14277.17	337.4041	441.5882	330.8352	303
70GluH	-2272.024204	380.8864	78.58907973	12.72279	13653.67	561.9705	-237.423	592.4361	303
71LeuH									
72SerH									
73AspH	-4352.601553	633.0791	135.626965	19.79364	8658.957	195.1332	555.2695	266.4928	303
74AsnH									
75ArgH	-3224.403695	521.377	111.6777216	17.6956	11408.09	355.7938	331.3492	357.5012	303
76ValH	-2383.370603	569.2388	81.85151114	18.97749	17675.45	1029.009	-703.799	1042.178	303
77SerH									
78GlyH	-2624.612287	234.1769	87.8295774	7.529311	21540.76	484.4859	-429.87	435.4069	303
79GlyH									
80LeuH	-2682.782469	316.4525	92.4095804	10.62187	17270.88	446.1031	32.49056	450.8448	303
81GluH	-2162.287788	219.5239	71.43154878	7.152197	22441.04	579.5056	-783.67	649.2938	303
82ValH	-2288.340385	314.0136	78.86620196	10.424	27261.68	970.9678	-400.119	907.5239	303
83LeuH	-2535.805928	393.6375	85.82978376	12.35097	23995.82	931.0539	-853.353	804.3582	303
84AlaH									
85GluH									
86LysH	-2221.544935	227.6135	75.9196008	7.545582	21444.07	554.7653	-758.146	567.1172	303
87CysH									
89AsnH	-2451.714148	228.6537	76.99014717	7.25914	15812.58	347.0461	-746.065	447.9244	303
90LeuH	-2703.346937	281.6007	84.84582492	8.841572	18228.24	453.5003	-683.752	527.6043	303
91ThrH	-2634.654791	368.5994	86.76745279	12.00437	9576.192	296.0609	-88.1713	333.3429	303
92HisH	-3102.429256	437.5497	96.88772117	13.73719	12903.54	353.5831	175.9192	449.4468	303
93LeuH	-2477.880655	197.2464	71.5492392	6.044327	13659.41	247.8767	-733.769	419.0695	303
94AsnH	-3942.393526	437.4223	129.02204	14.21635	13650.98	258.3479	465.9795	287.7382	303
95LeuH	-2255.602588	331.5711	70.86622147	10.56496	11164.89	389.8644	-59.5096	505.813	303

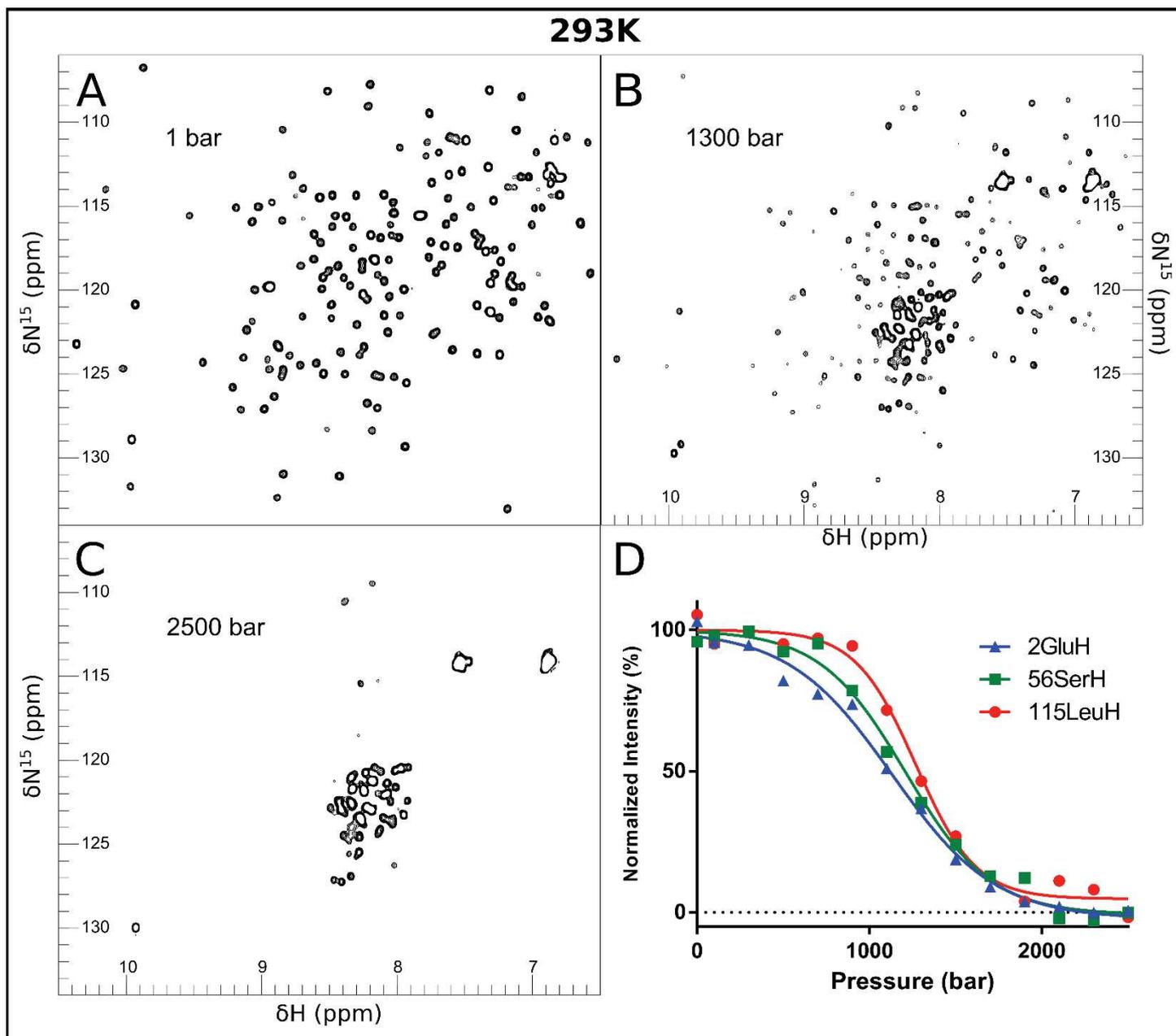
96SerH	-2159.275765	299.2505	75.83891539	10.05367	13819.65	478.8419	-30.5417	455.4865	303
97GlyH	-3351.608175	484.297	108.6755379	15.61113	9975.973	281.7093	-201.723	323.9665	303
98AsnH									
99LysH	-1682.746699	167.0564	68.89193859	5.849491	12377.29	355.7572	75.82555	227.7588	303
100IleH	-2556.859205	356.8371	85.27089632	11.7087	18209.02	582.3615	-462.113	635.539	303
101LysH	-2652.285555	389.6027	90.74015069	13.01046	10549	356.0915	-448.615	365.3868	303
102AspH	-2415.517705	326.0799	79.33452768	10.5984	23721.81	769.9338	-1089.88	875.6268	303
103LeuH									
104SerH	-1078.597668	370.8267	51.24167403	12.34836	17110.98	2075.696	-615.03	901.4541	303
105ThrH	-2368.991986	520.0576	79.64965129	17.12075	18991.87	1004.401	-563.267	1073.447	303
106IleH	-2889.658409	352.68	93.36136628	11.34377	17776.19	447.7963	44.19366	526.2277	303
107GluH	-1787.299871	293.0885	62.26105932	9.758184	30074.1	1459.891	-1792.1	1382.968	303
109LeuH									
110LysH									
111LysH									
112LeuH									
113GluH	-2763.611869	346.1069	90.78684712	11.25678	26912.72	716.8497	50.5682	810.517	303
114AsnH	-3072.126721	329.1682	94.85510957	10.25931	15763.65	332.3007	32.9224	436.1175	303
115LeuH	-3045.175106	486.5012	92.81767976	15.04298	18828.5	599.5749	-216.093	816.8867	303
116LysH	-2646.626092	303.0486	85.15514566	9.729816	12219.32	321.1873	-716.17	384.3366	303
117SerH	-2921.415751	424.3561	87.781056	12.84488	15734.29	497.4102	253.6403	616.911	303
118LeuH	-3399.023201	302.9623	112.7315636	9.99894	12276.21	192.3005	91.01187	214.6004	303
119AspH	-2184.791282	268.7436	67.24344677	8.741523	12718.75	393.8139	-780.086	570.3763	303
120LeuH	-3069.618821	432.6789	98.49721609	13.84474	9869.984	279.2601	-40.1497	331.8971	303
121PheH	-2939.009566	571.7687	92.38153063	18.05496	14307.47	564.8513	15.50597	711.4322	303
122AsnH	-3344.858681	552.4863	109.0688556	17.88695	12149.11	377.5696	247.0231	428.9214	303
123CysH	-3019.861304	361.4393	98.00061023	11.69506	17831.44	389.118	1207.121	492.9549	303
124GluH									
125ValH									
126ThrH	-2466.726342	425.9459	82.83177561	13.44835	15925.32	554.6573	-60.0648	584.3589	303
127AsnH	-2925.141943	242.4585	92.27412912	7.674849	18427.65	318.2421	-400.101	397.293	303
128LeuH									
129AsnH	-1287.680344	511.5288	68.9506399	20.45832	6517.528	930.8707	-286.55	430.181	303
130AspH									
131TyrH									
132ArgH									
133GluH	-2780.230138	391.928	89.59319265	12.49943	25132.59	798.7004	445.893	859.4969	303
134AsnH	-2772.042645	322.8214	84.41292022	9.829618	22825.99	621.8331	-375.18	732.0369	303
135ValH									
136PheH									
137LysH	-3128.735497	483.1386	95.29876473	14.76873	25054.32	841.4262	-91.8072	976.6266	303
138LeuH	-2641.024238	328.5178	82.90194593	10.42135	24918.12	682.5768	173.4715	850.856	303
139LeuH									
141GlnH	-2609.757141	291.9402	80.23297428	9.162286	21507.06	557.3625	-984.406	740.126	303
142LeuH	-2814.205626	214.278	86.8099061	6.698589	18858.11	304.1953	-487.577	406.3218	303
143ThrH									
144TyrH	-3573.011189	461.4217	104.9534673	13.8027	12866.9	293.4681	172.1154	427.0098	303
145LeuH	-3621.096495	468.2237	111.9680941	14.53746	10212.33	241.1957	-66.6482	306.2874	303

146AspH	-3151.886696	542.6465	93.76457986	16.49356	10015.86	339.248	-268.851	492.0029	303
147GlyH	-2748.647558	417.0517	84.92163468	12.86007	-10126.9	330.5486	330.846	437.7322	303
148TyrH	-3629.66696	433.984	108.6949743	13.14641	13910.69	293.7567	190.9306	436.6364	303
149AspH	-3012.525402	390.3038	93.79729819	12.66109	14731.63	349.4023	648.3137	578.0869	303
150ArgH									
151AspH	-4262.819465	590.6886	121.7018086	17.13713	17697.81	396.211	125.5589	591.7844	303
152AspH	-3442.643185	706.3655	102.4198597	21.361	12716.32	480.976	-71.9691	682.0756	303
153LysH	-2570.701318	189.6795	79.85934675	5.926175	21043.03	382.5649	-126.287	460.5018	303
154GluH	-3152.341376	439.0975	101.0698299	13.93638	20125.96	602.7067	92.86355	641.55	303
155TrpH									
156LeuH									
157GluH	-2255.969206	248.6619	74.99736831	8.129417	17366.37	482.8152	-882.064	531.6324	303
158HisH									
	DG	DG_err	DV	DV_err	Cf	Cf_err	Cu	Cu_err	T
Average values	-2607.603438	373.2347	87.267194	12.29289	15668.15	559.5554	-196.066	562.9099	303

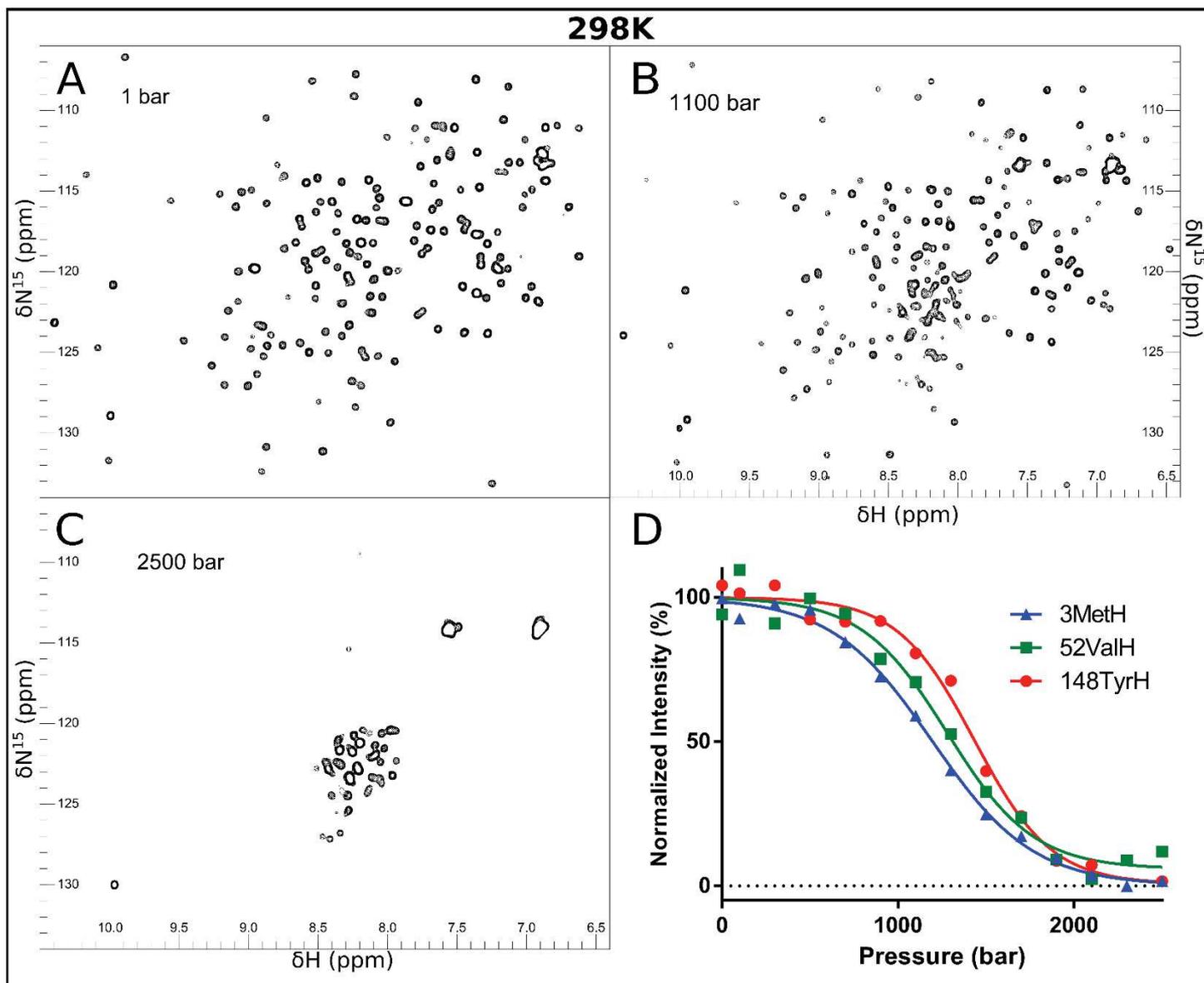




**Figure S2.** Pressure effect on the  $[^1\text{H}-^{15}\text{N}]$  HSQC spectra of pp32 at 288 K and 1.4 M urea. A-C Examples of  $[^1\text{H}-^{15}\text{N}]$  HSQC spectra at different pressures as indicated; D) Examples of 3 residues exhibiting distinct unfolding profiles.



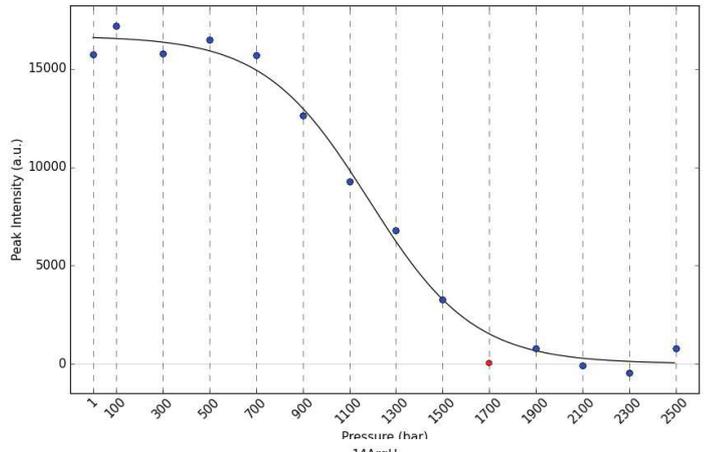
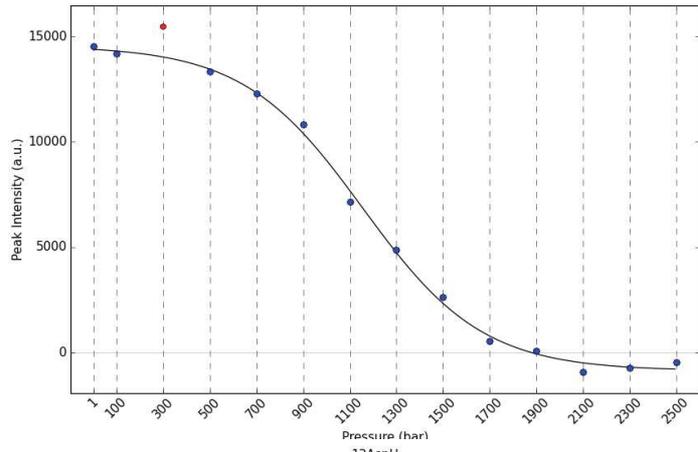
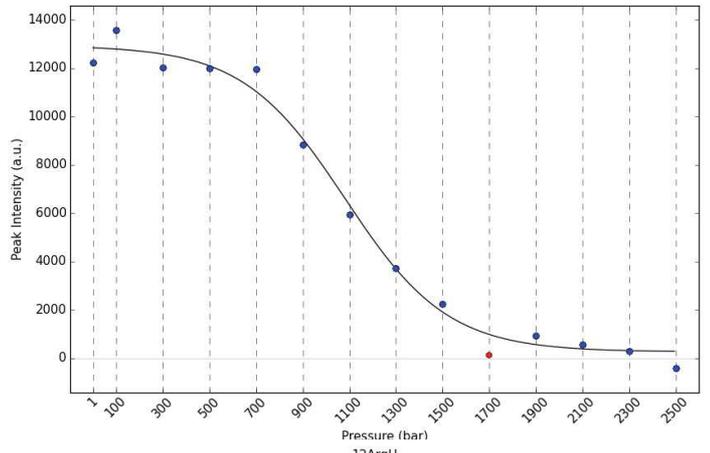
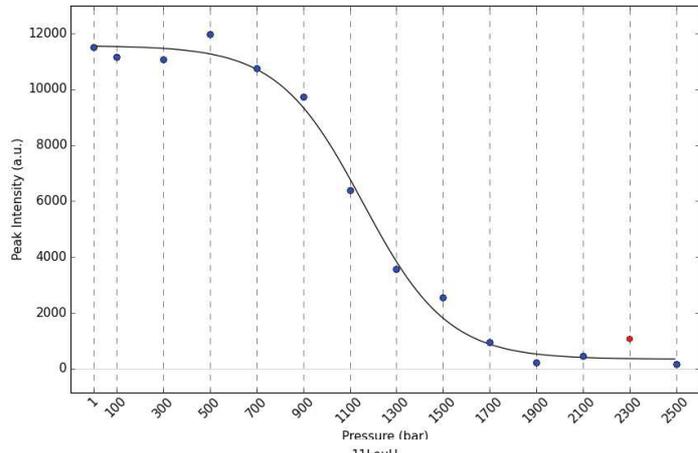
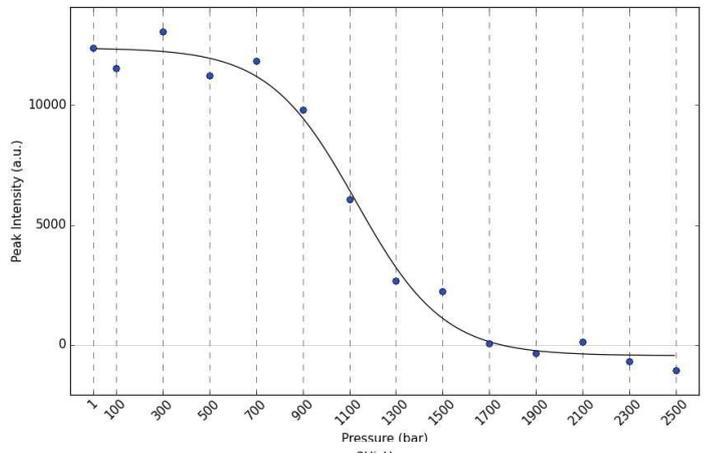
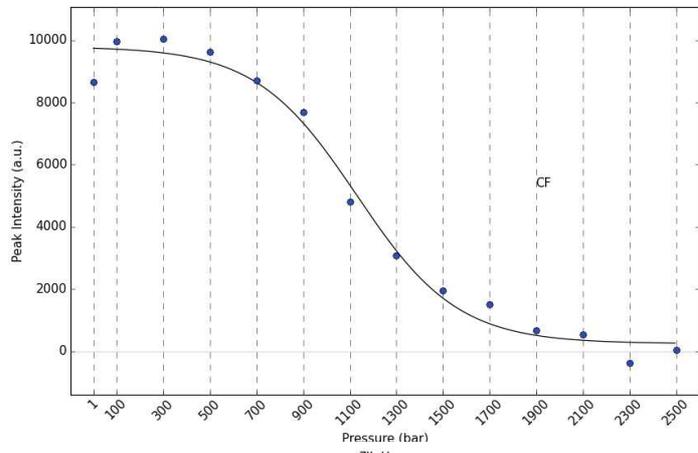
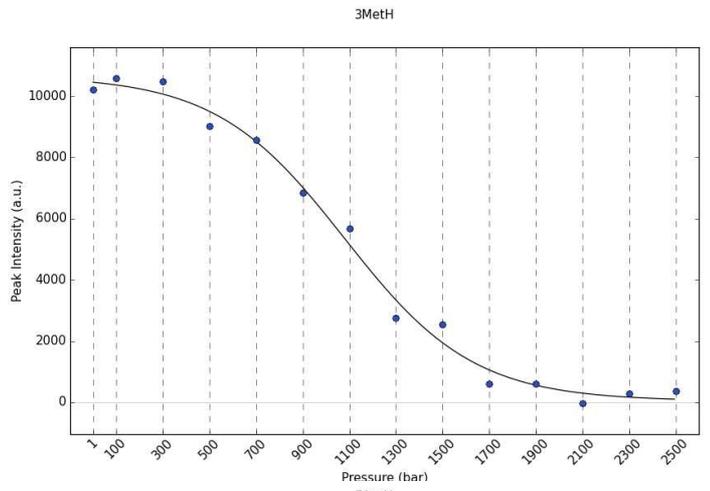
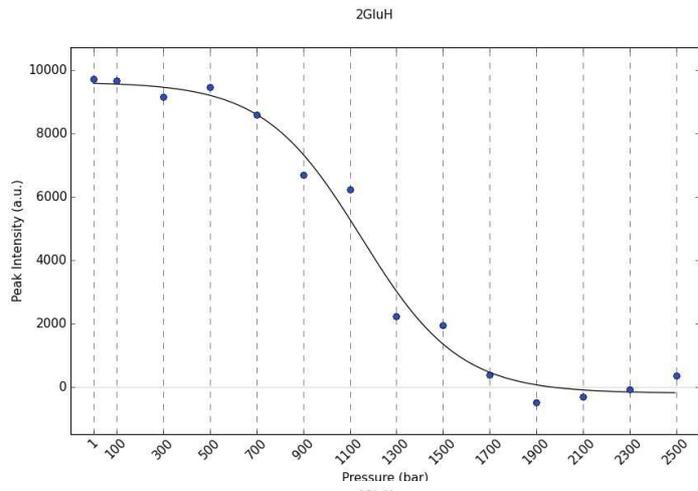
**Figure S3.** Pressure effect on the  $[^1\text{H}-^{15}\text{N}]$  HSQC spectra of pp32 at 293 K and 1.4 M urea. A-C Examples of  $[^1\text{H}-^{15}\text{N}]$  HSQC spectra at different pressures as indicated; D) Examples of 3 residues exhibiting distinct unfolding profiles.



**Figure S4.** Pressure effect on the  $[^1\text{H}-^{15}\text{N}]$  HSQC spectra of pp32 at 298 K and 1.4 M urea. A-C Examples of  $[^1\text{H}-^{15}\text{N}]$  HSQC spectra at different pressures as indicated; D) Examples of 3 residues exhibiting distinct unfolding profiles.

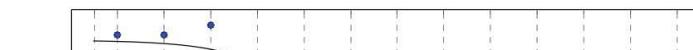
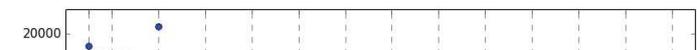
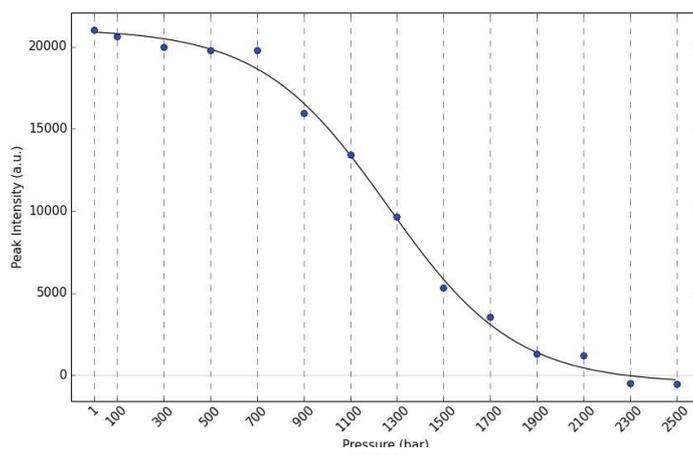
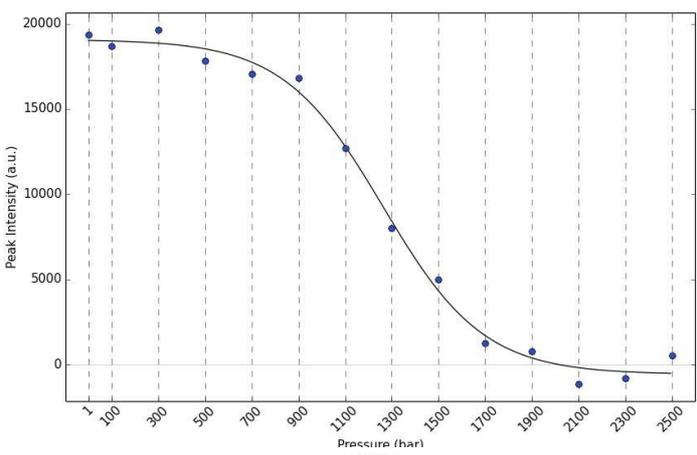
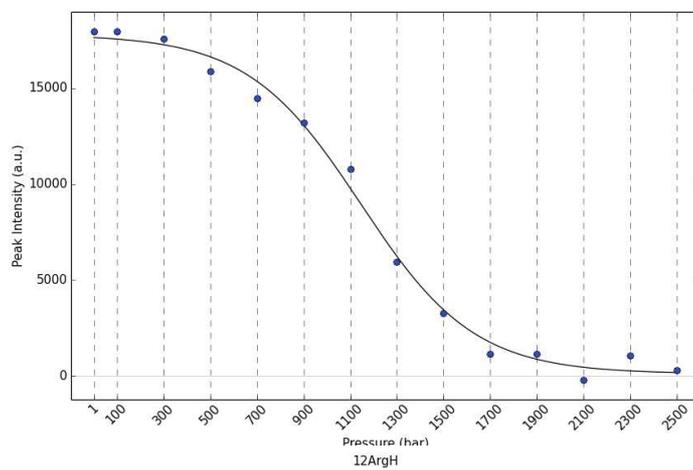
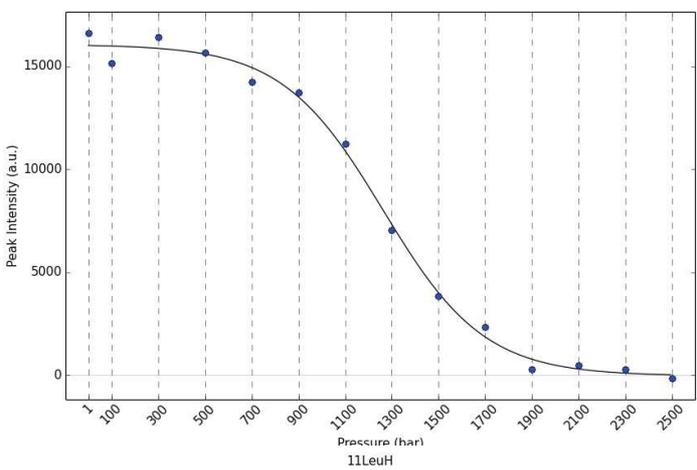
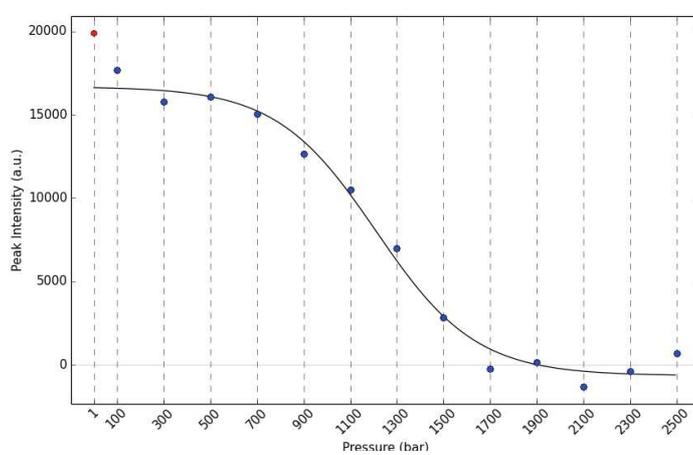
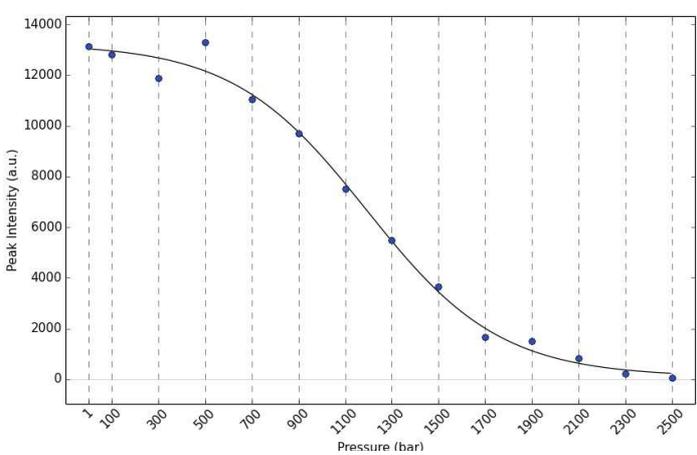
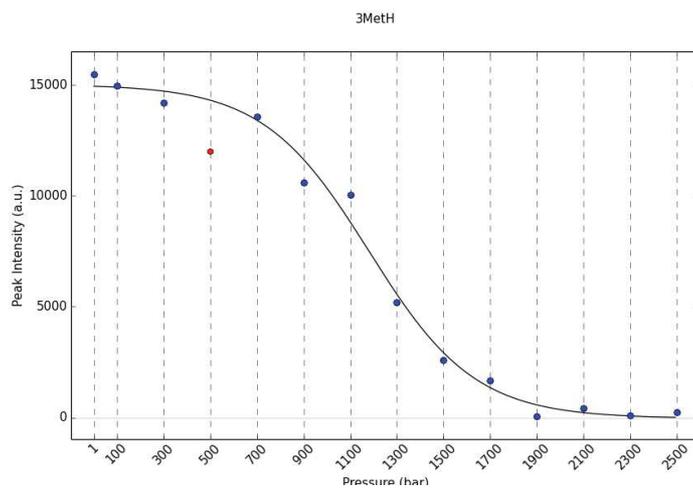
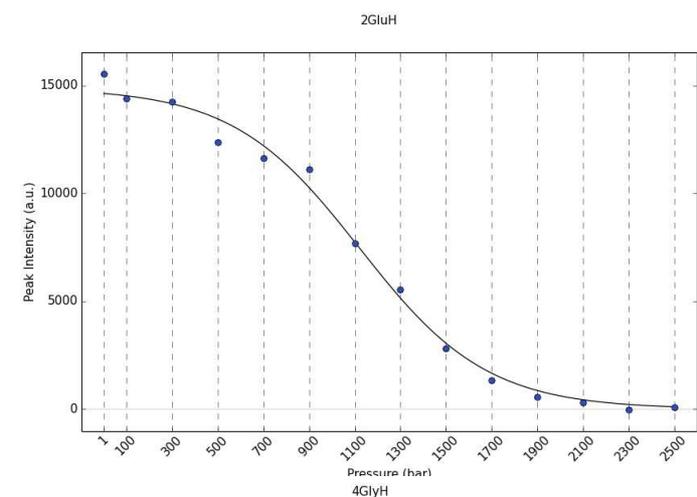
**Figure S5.** Individual fits of the HSQC peak intensity changes as a function of pressure at 288 K. Aliased peaks exhibit an increase in intensity. Data were analyzed as described in the Material and Methods section of the main text.

# Fits 288K



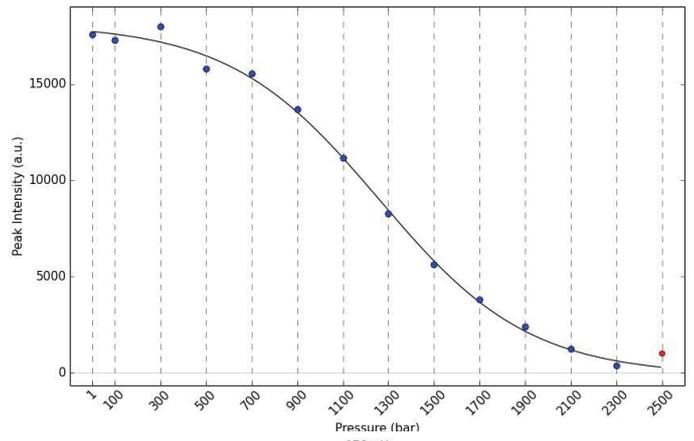
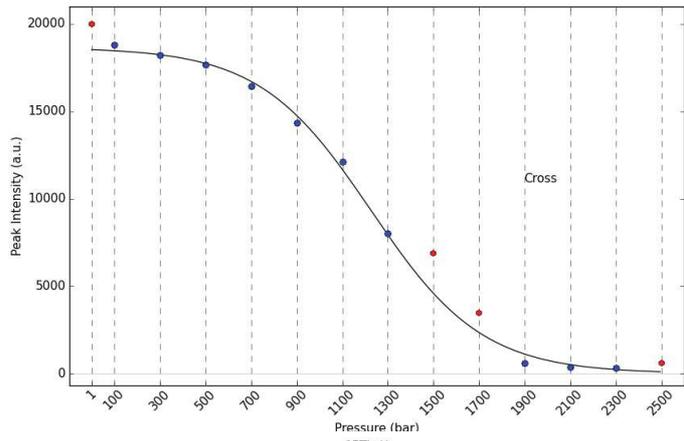
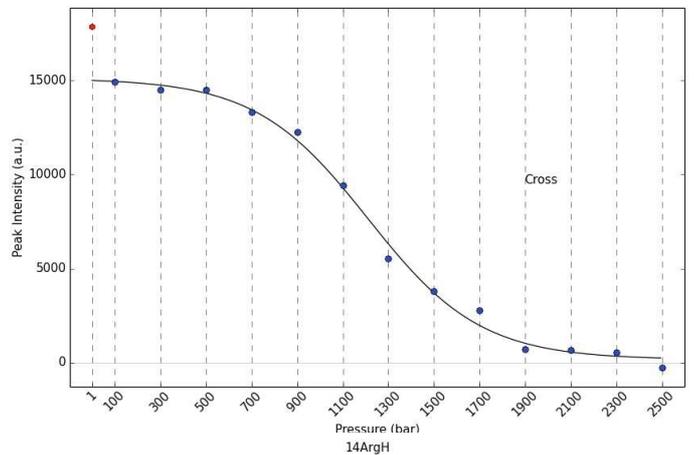
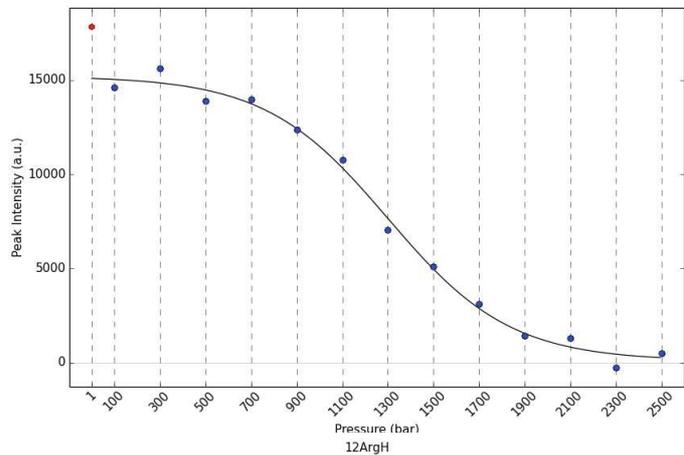
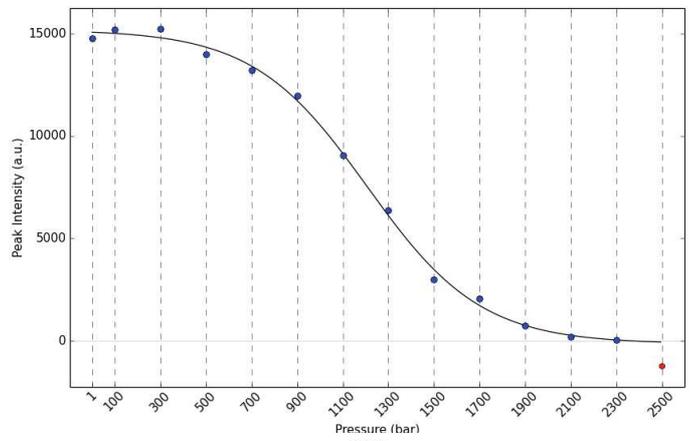
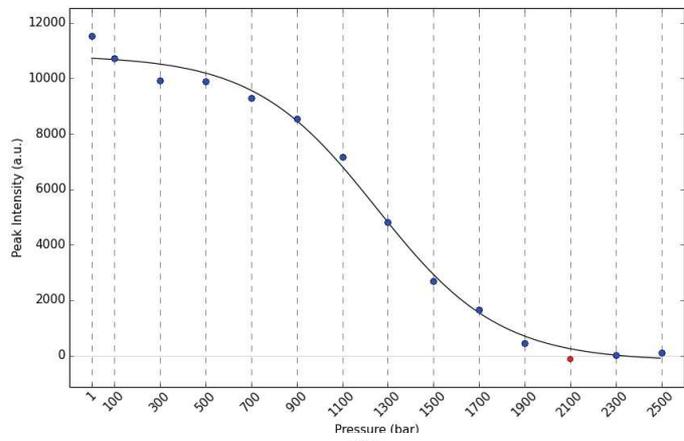
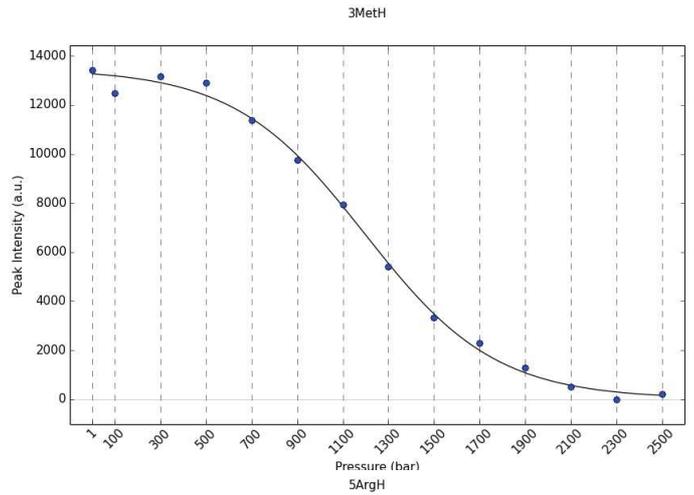
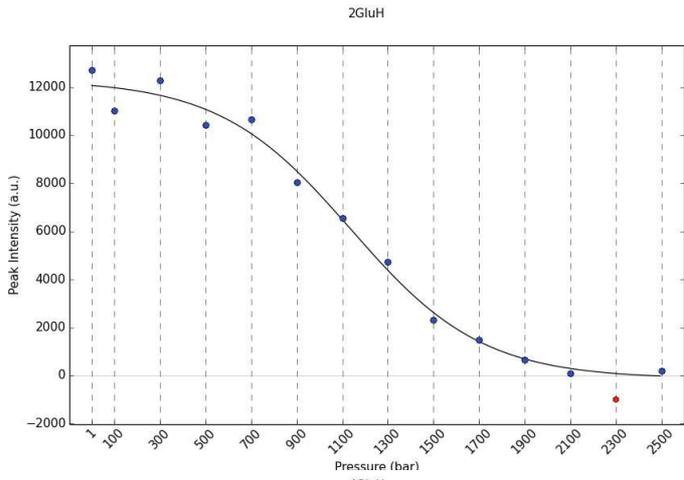
**Figure S6.** Individual fits of the HSQC peak intensity changes as a function of pressure at 293 K. Aliased peaks exhibit an increase in intensity. Data were analyzed as described in the Material and Methods section of the main text.

## Fits 293K



**Figure S7.** Individual fits of the HSQC peak intensity changes as a function of pressure at 288 K. Aliased peaks exhibit an increase in intensity. Data were analyzed as described in the Material and Methods section of the main text.

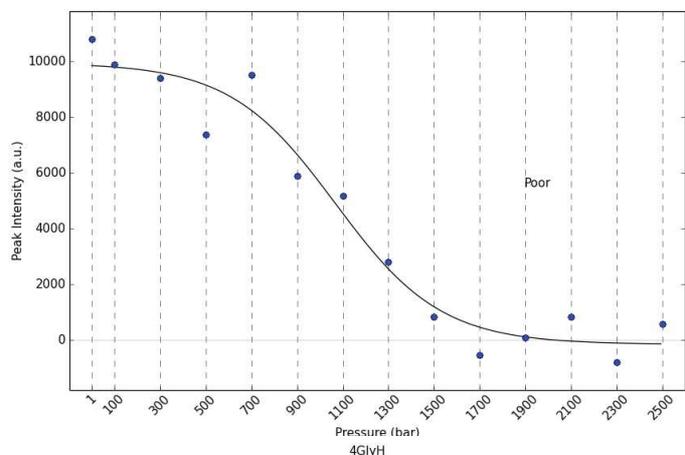
# Fits 298K



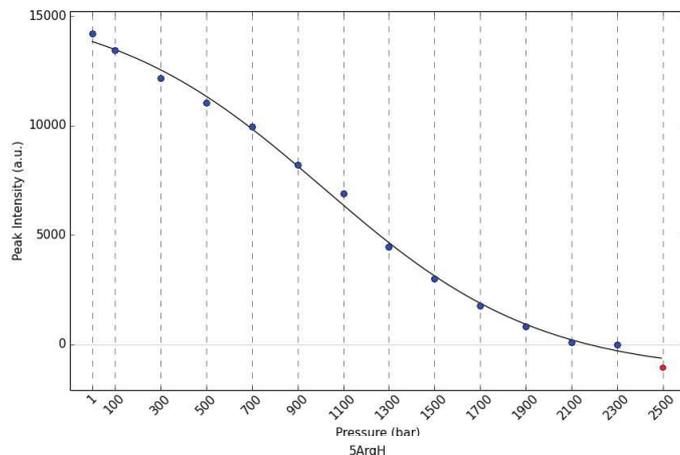
**Figure S8.** Individual fits of the HSQC peak intensity changes as a function of pressure at 288 K. Aliased peaks exhibit an increase in intensity. Data were analyzed as described in the Material and Methods section of the main text.

# Fits 303K

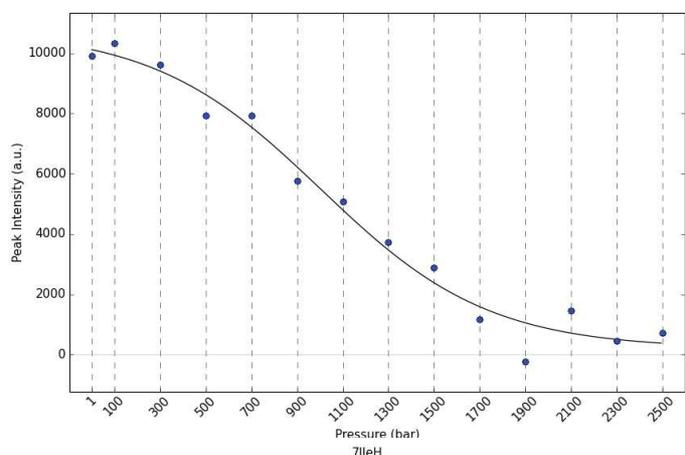
2GluH



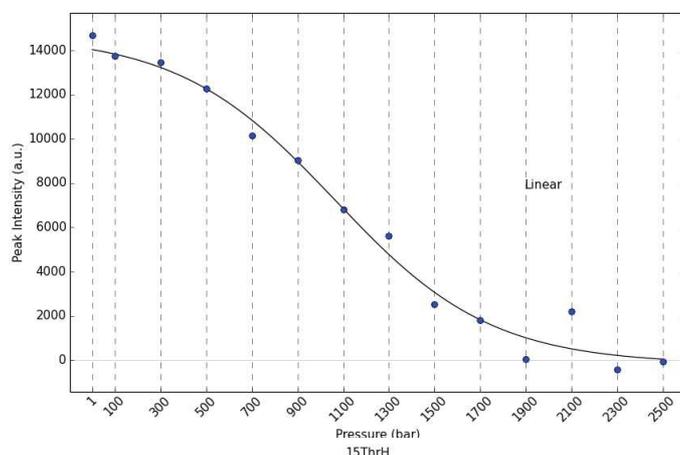
3Meth



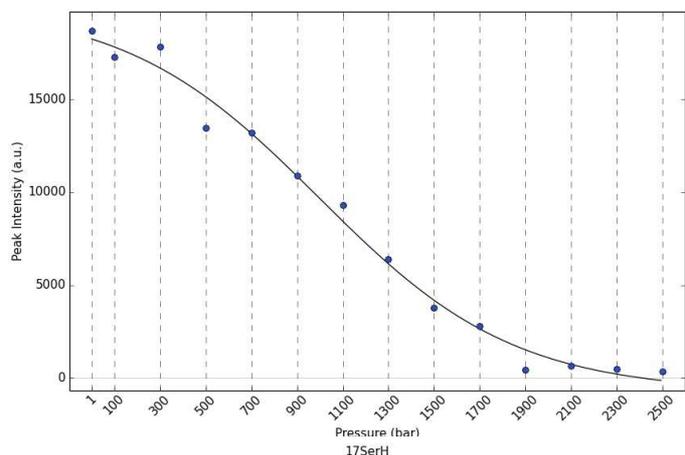
4GlyH



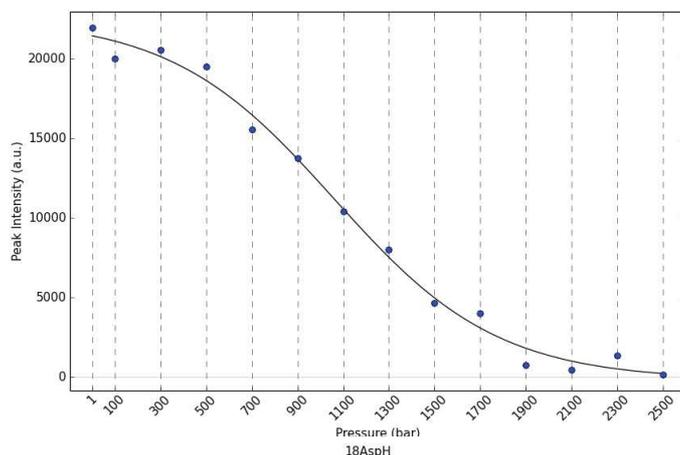
5ArgH



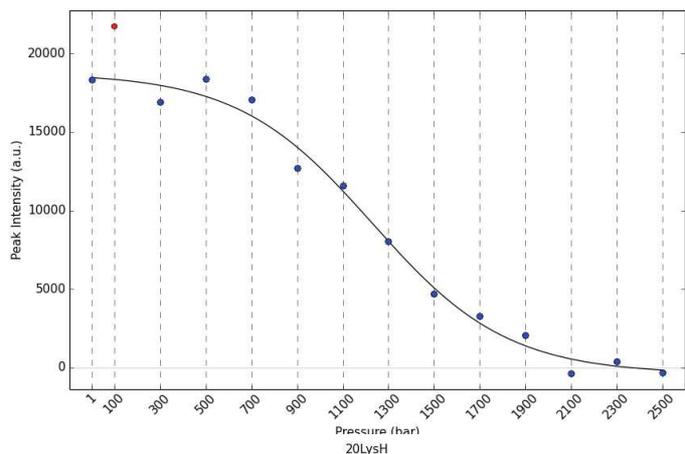
7IleH



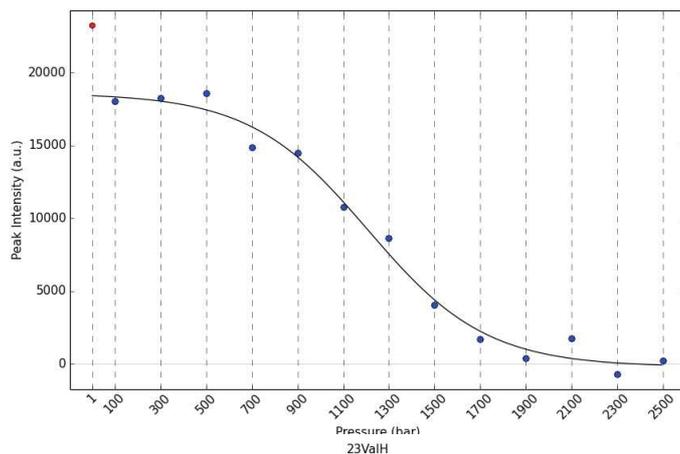
15ThrH



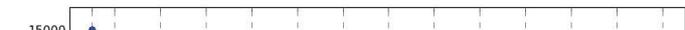
17SerH



18AspH

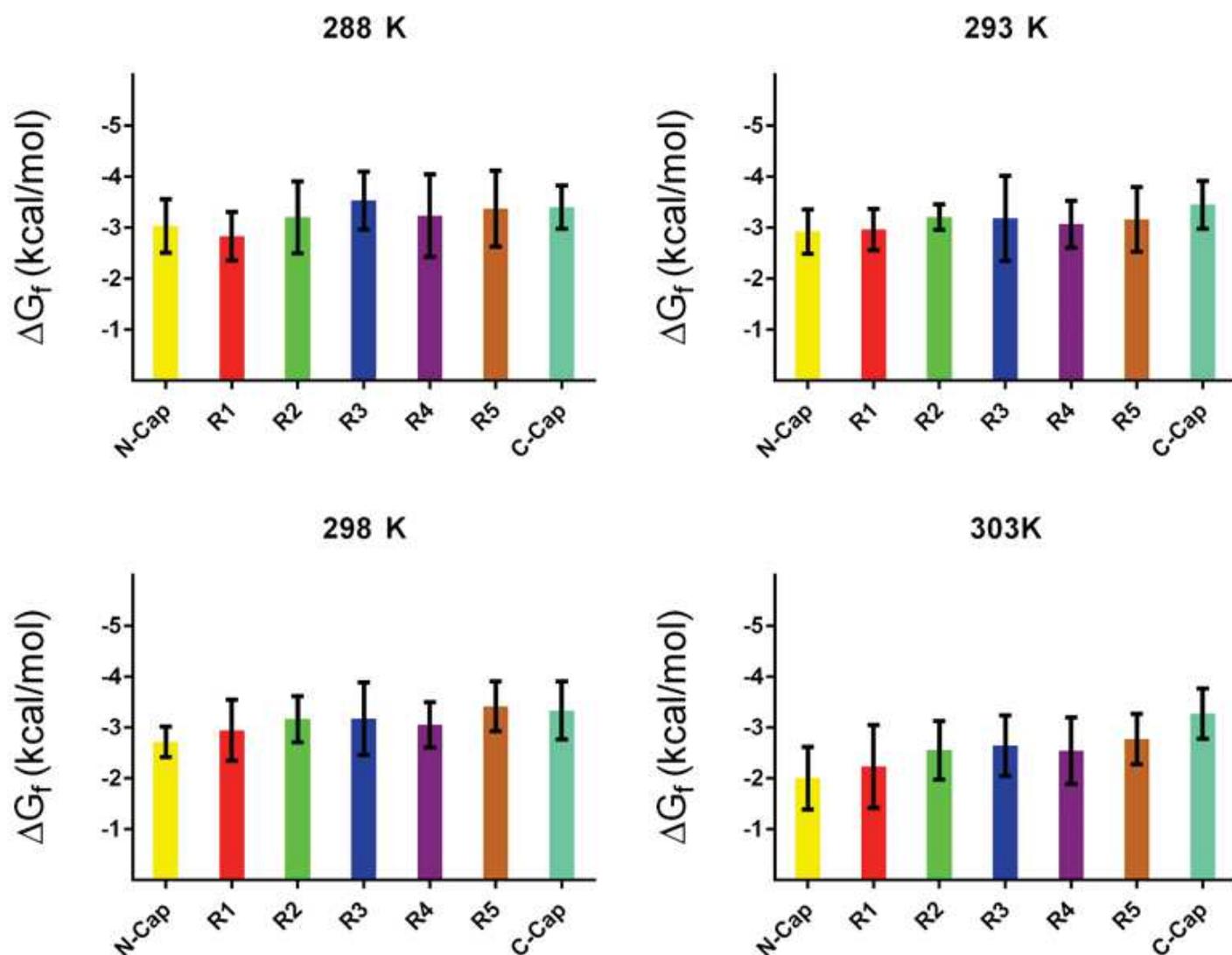


20LysH

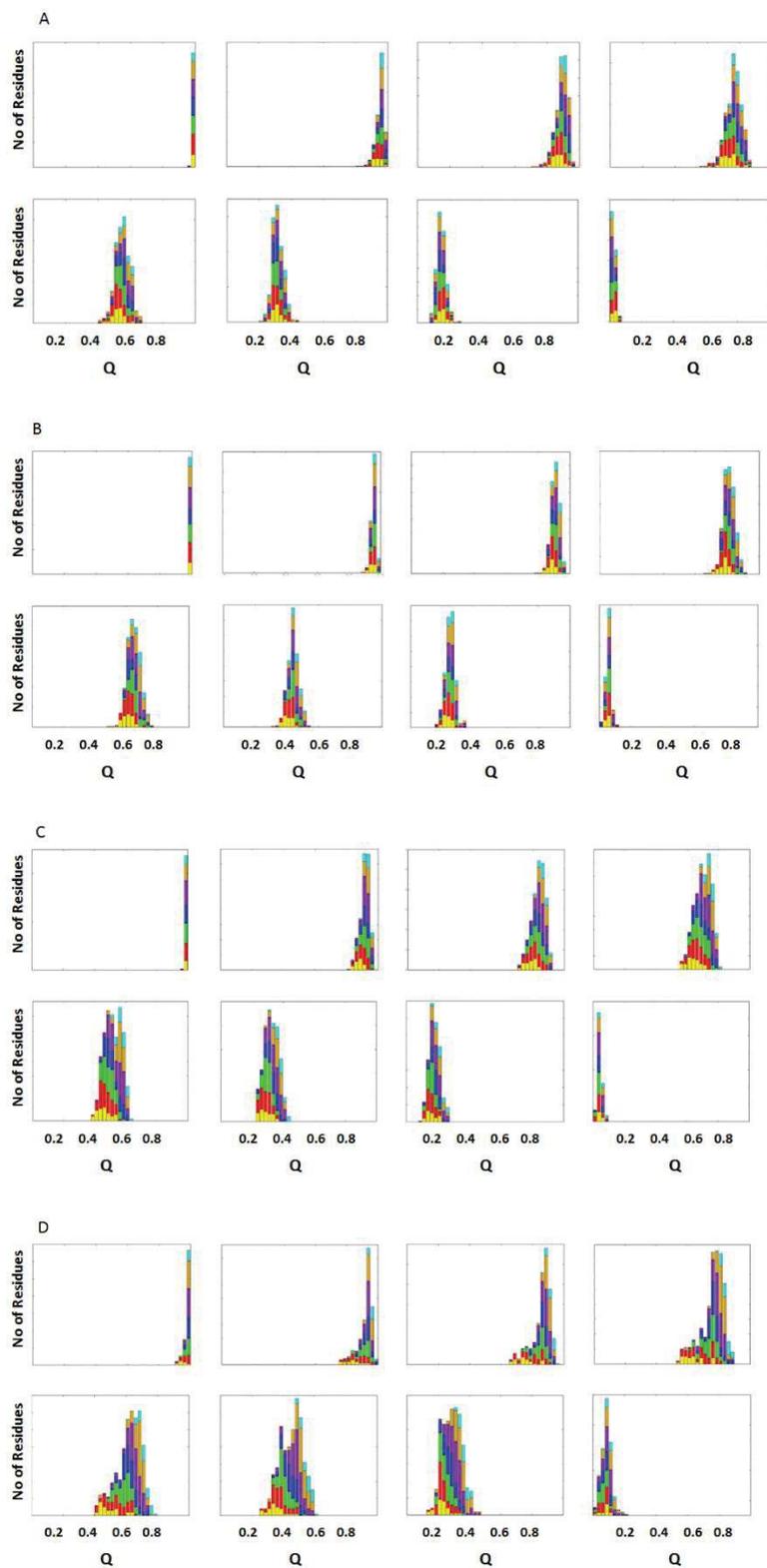


23ValH

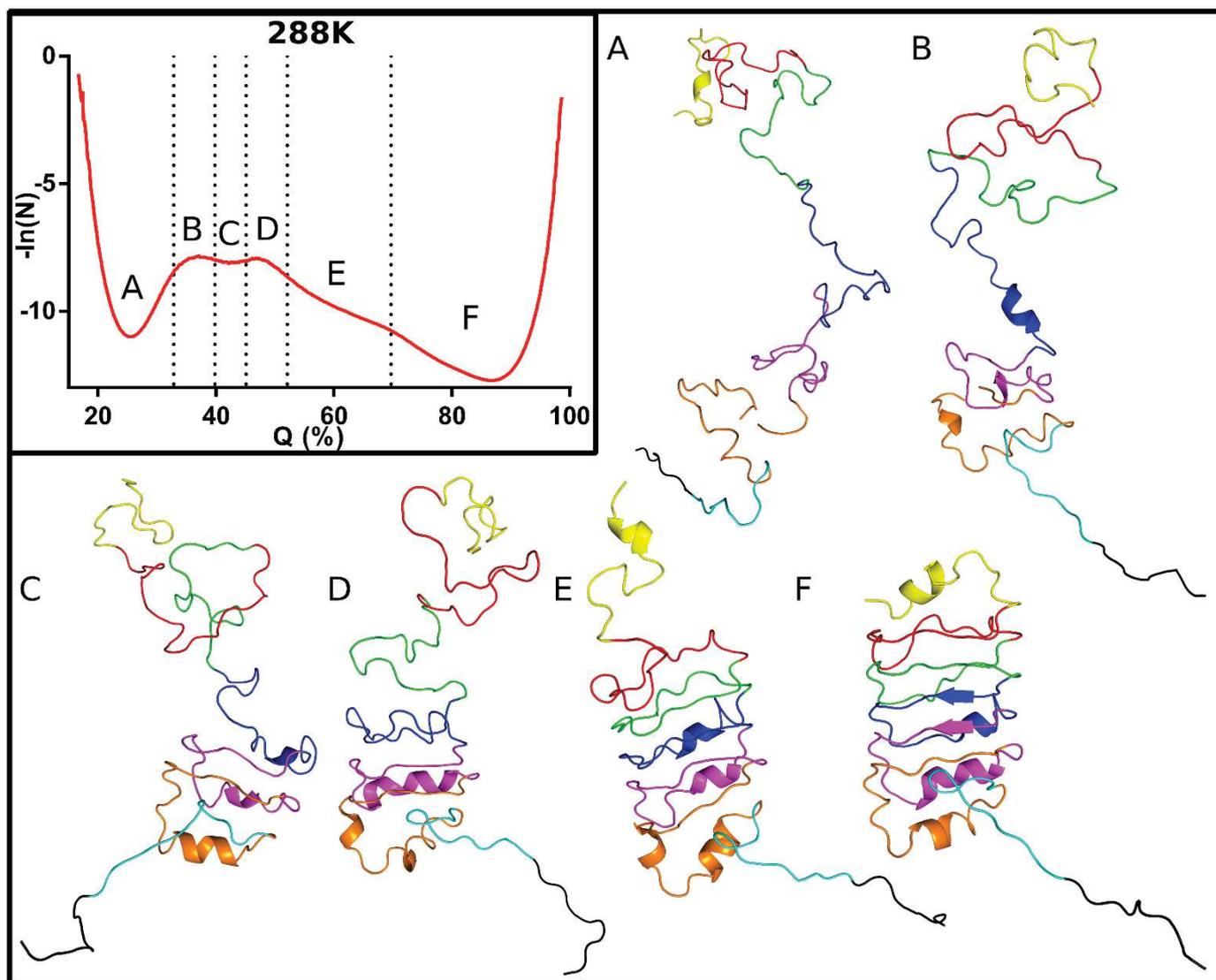




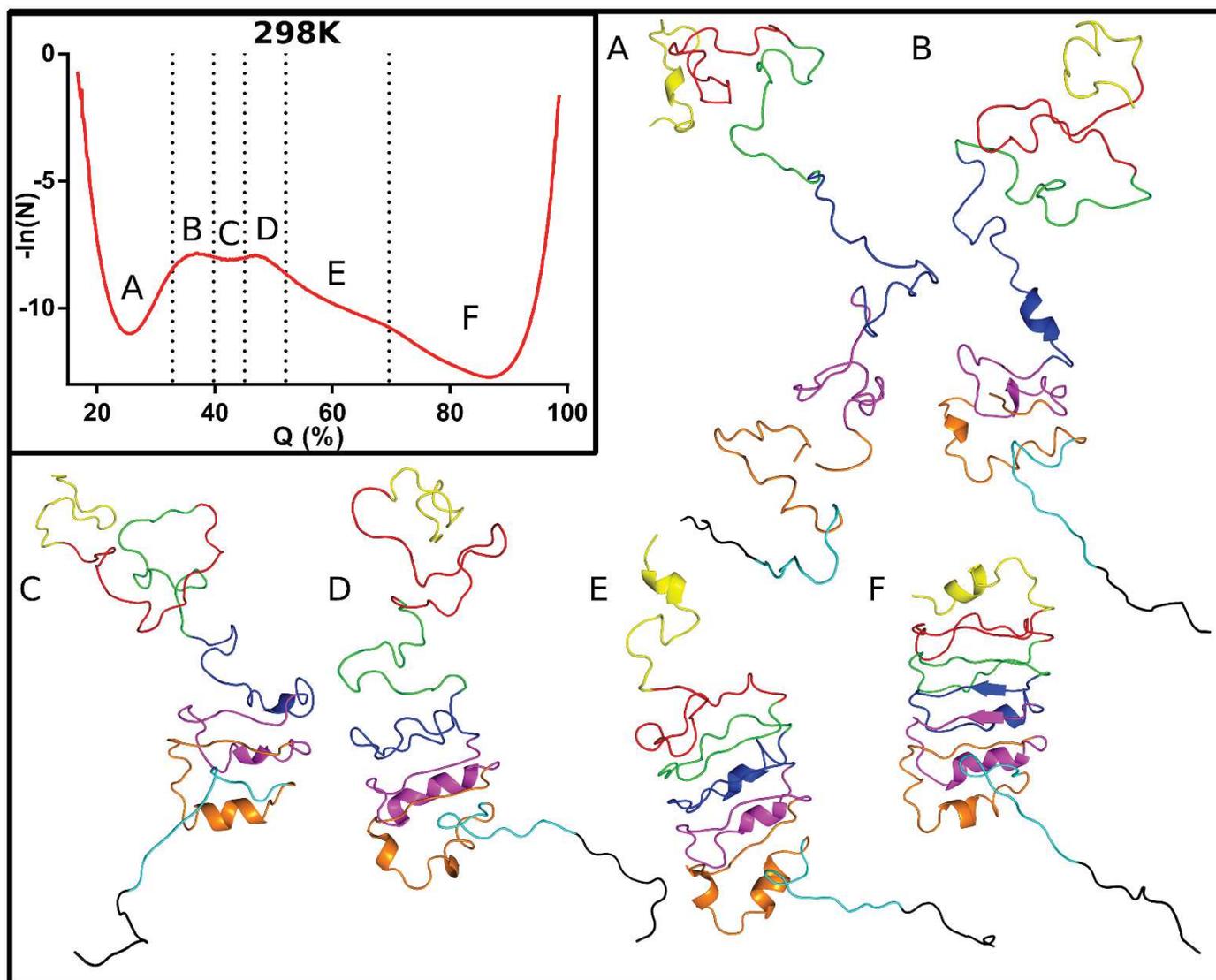
**Figure S9.** Average free energy of folding for the residues found in each repeat of pp32 noted on the x-axis and colored as per repeat as in Figure 1 in the main text at the four tested temperatures, as indicated. Free energy values for each residue were obtained from fits of the NMR HSQC peak intensity values vs pressure for each residue as described in the Material and Methods section. Error bars represent the standard deviation of the distributions of residue specific apparent free energy of folding values for each repeat.



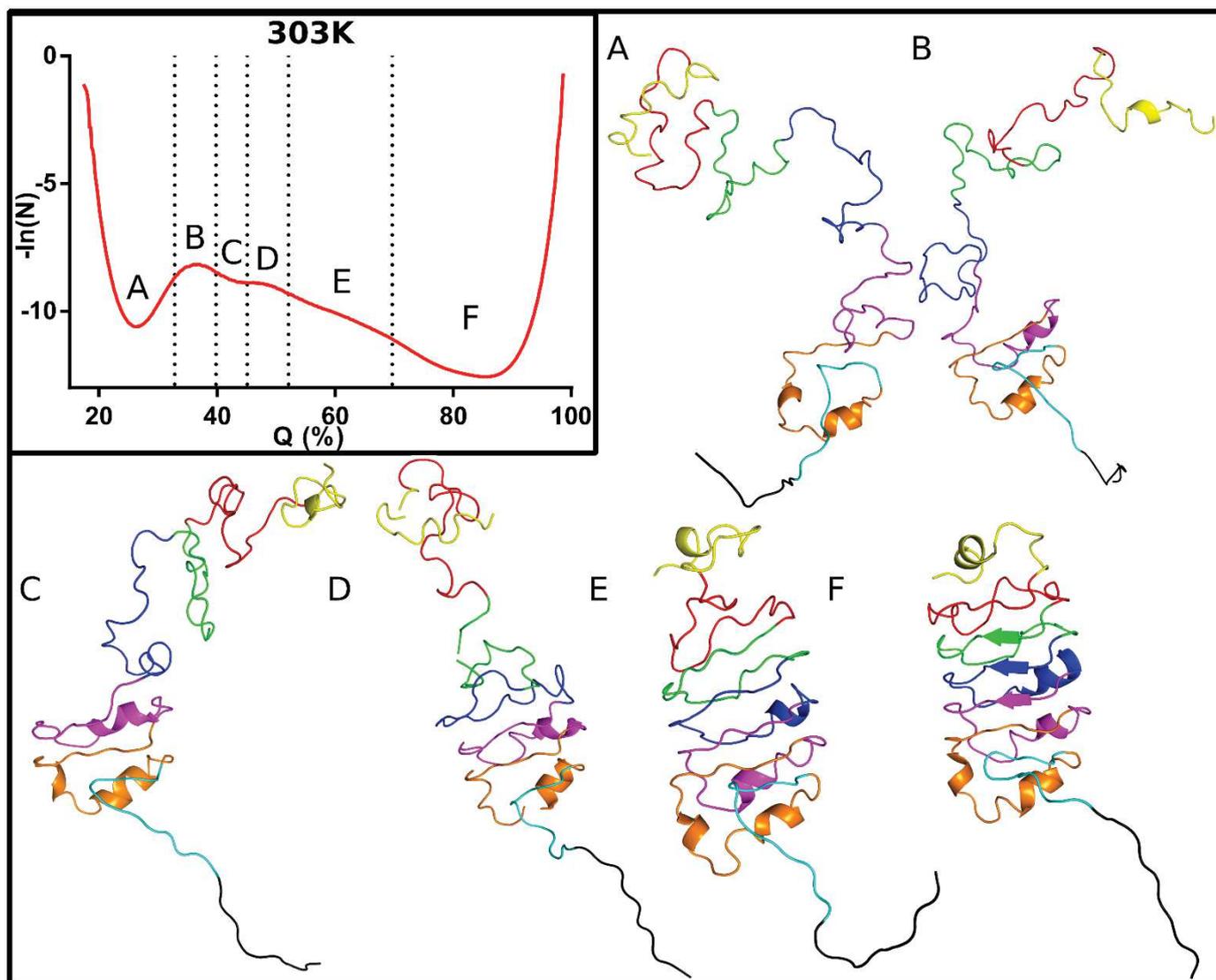
**Figure S10.** Histograms of the fraction of native contacts as a function of pressure at 288, 293, 298 and 303 K for A-D, respectively. Pressures were from left to right top) 1, 500, 700 and 900 bar, and left to right bottom) 1100, 1300, 1500 and 1900 bar for A-D. Histograms are colored for each repeat as in Figure 1 in the main text.



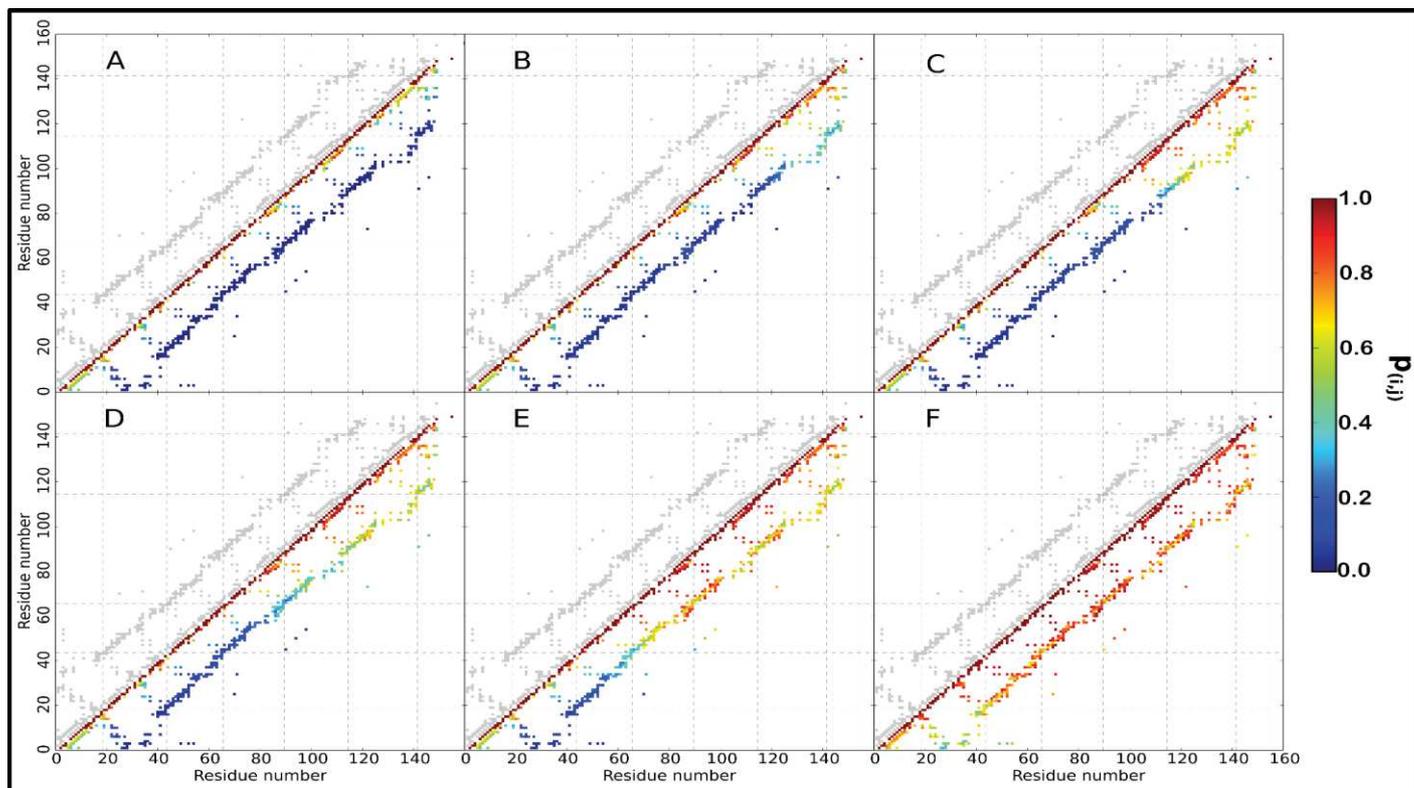
**Figure S11.** Pseudo energy landscape as extracted from the 288 K simulations. The displayed structures are reconstructed structures from the free energy regions indicated by the corresponding letters. Representative structures were obtained as described in the material and methods. Structures are colored as in **Figure 1** in the main text. Residues that do not appear in the crystal structure are displayed in black.



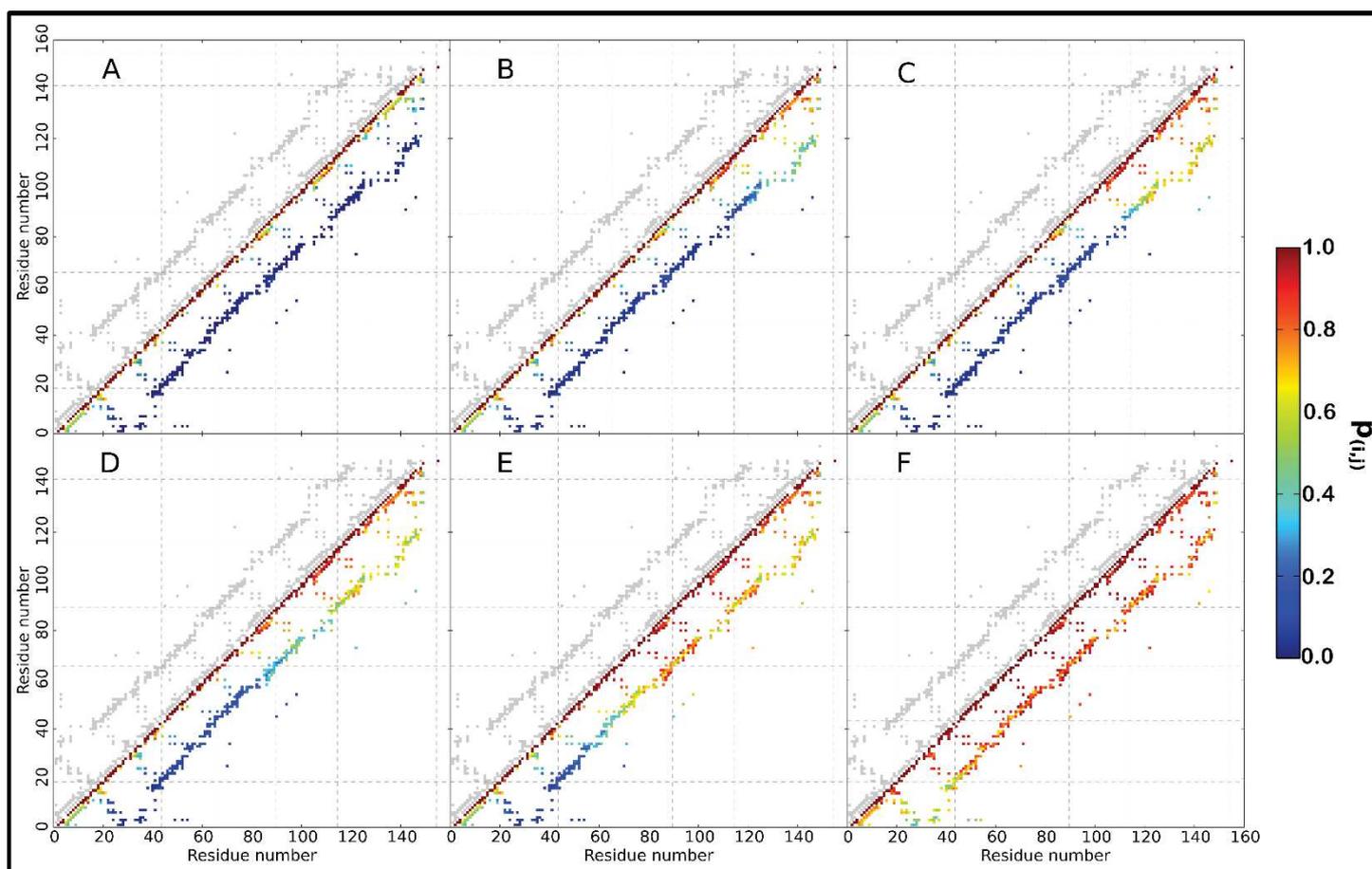
**Figure S12.** Pseudo energy landscape as extracted from the 298 K simulations. The displayed structures are reconstructed structures from the free energy regions indicated by the corresponding letters. Representative structures were obtained as described in the material and methods. Structures are colored as in Figure 1 in the main text. Residues that do not appear in the crystal structure are displayed in black.



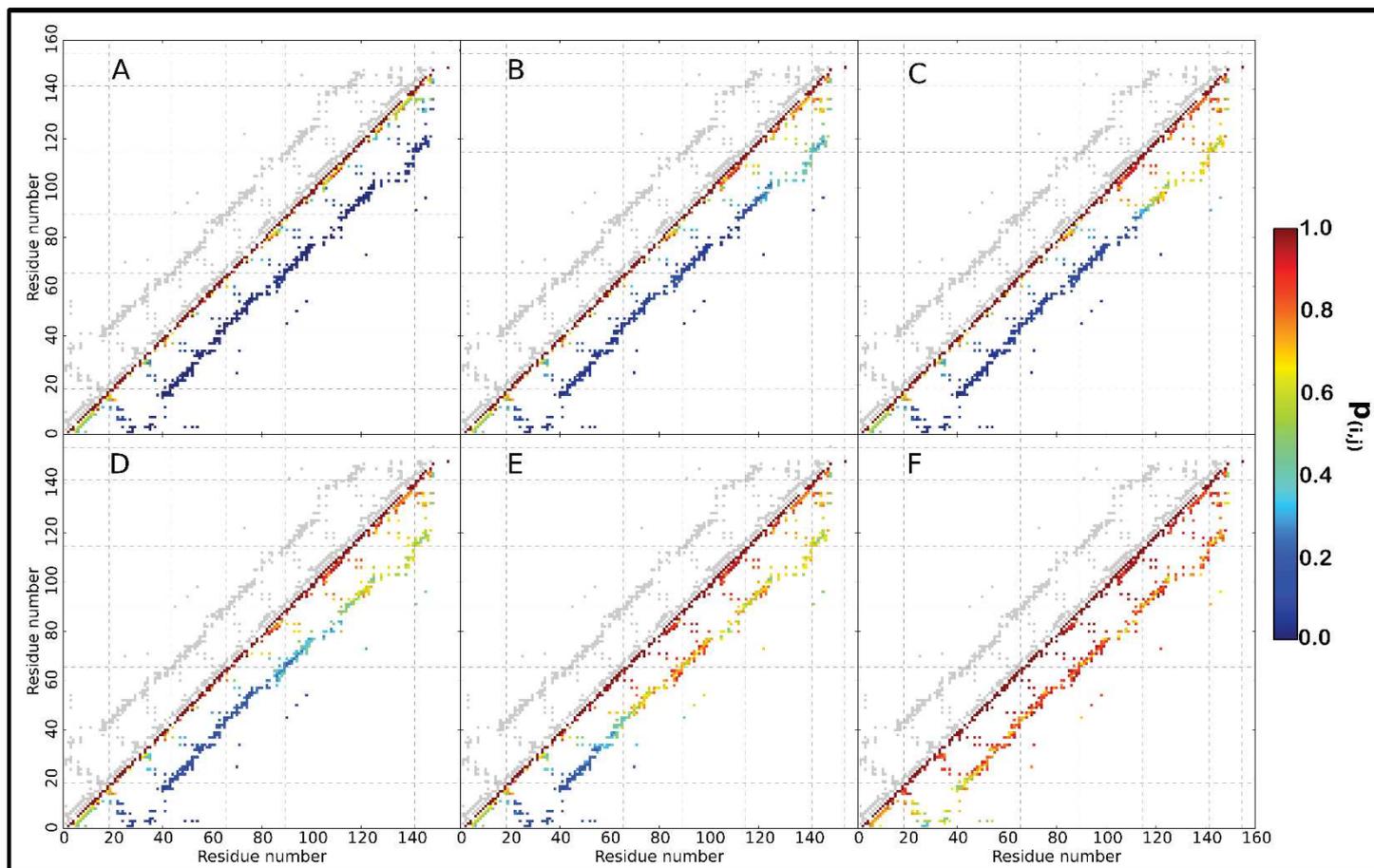
**Figure S13.** Pseudo energy landscape as extracted from the 303 K simulations. The displayed structures are reconstructed structures from the free energy regions indicated by the corresponding letters. Representative structures were obtained as described in the material and methods. Structures are colored as in Figure 1 in the main text. **Error! Reference source not found.** Residues that do not appear in the crystal structure are displayed in black.



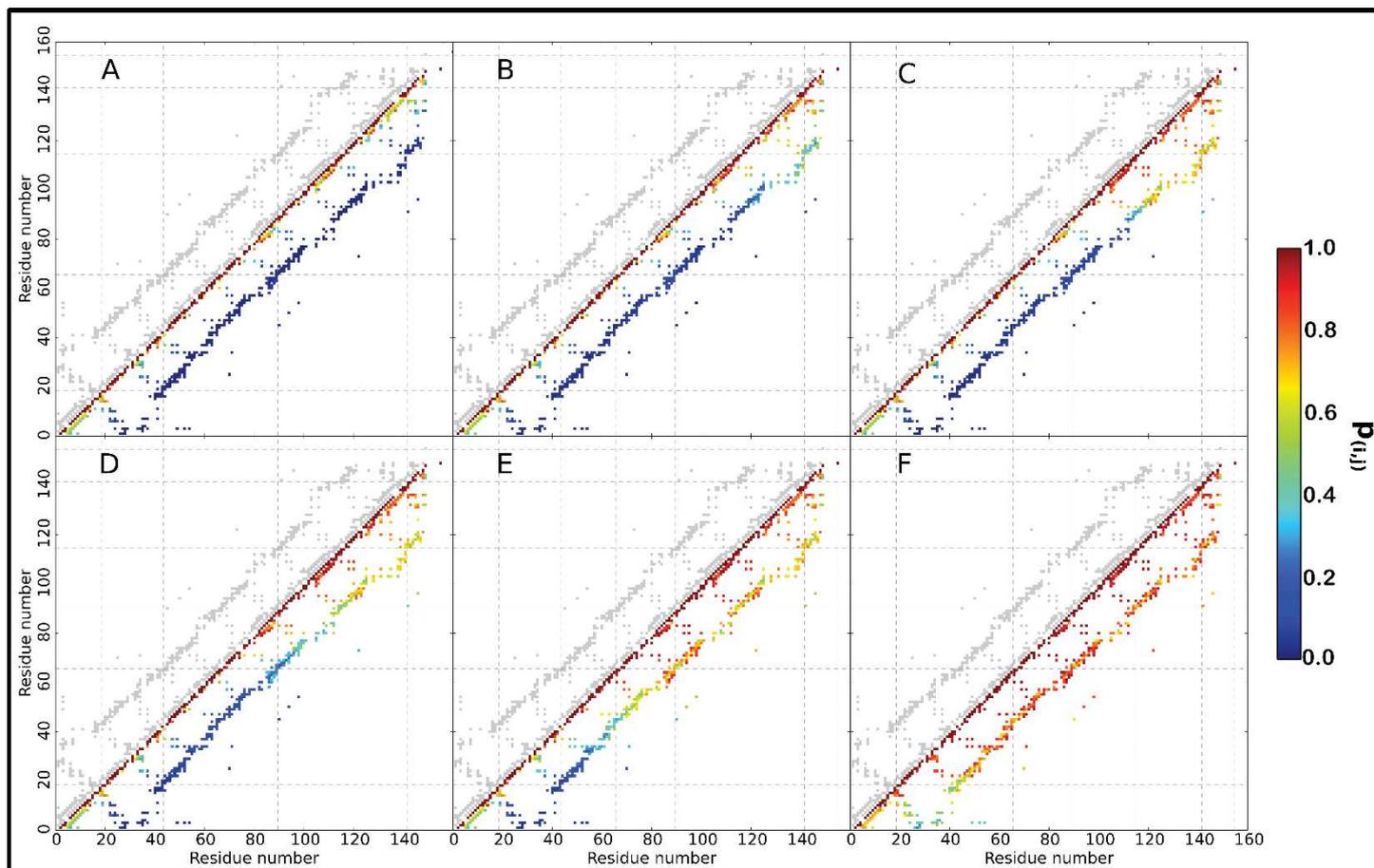
**Figure S14.** Fractional contact maps calculated from the pseudo free energy profile at 288 K. A-F correspond to the different Q value ranges defined in the Pseudo free energy profile in Figure S7. Grey dots correspond to the original contact map, colored dots correspond to the proportion of contact found in the given contact range, color code is given by the color scale of  $p(i,j)$ . Grey dashed lines represent the limits of each repeat in the protein primary sequence.



**Figure S15.** Fractional contact maps calculated from the pseudo free energy profile at 293 K. A-F correspond to the different  $Q$  value ranges defined in the Pseudo free energy profile in Figure 7 in the main text. Grey dots correspond to the original contact map, colored dots correspond to the proportion of contact found in the given contact range, color code is given by the color scale of  $p(i,j)$ . Grey dashed lines represent the limits of each repeat in the protein primary sequence.



**Figure S16.** Fractional contact maps calculated from the pseudo free energy profile at 298 K. A-F correspond to the different Q value ranges defined in the Pseudo free energy profile in Figure S8. Grey dots correspond to the original contact map, colored dots correspond to the proportion of contact found in the given contact range, color code is given by the color scale of  $p(i,j)$ . Grey dashed lines represent the limits of each repeat in the protein primary sequence.



**Figure S17.** Fractional contact maps calculated from the pseudo free energy profile at 303 K. A-F correspond to the different Q value ranges defined in the Pseudo free energy profile in Figure S9 in the main text. Grey dots correspond to the original contact map, colored dots correspond to the proportion of contact found in the given contact range, color code is given by the color scale of  $p(i,j)$ . Grey dashed lines represent the limits of each repeat in the protein primary sequence.



## **7.4.2 Evolutionarily Conserved Pattern of Interactions in a Protein Revealed by Local Thermal Expansion Properties**

### **Evolutionarily Conserved Pattern of Interactions in a Protein Revealed by Local Thermal Expansion Properties**

Mariano Dellarole<sup>1,a</sup>, Jose A. Caro<sup>2</sup>, Julien Roche<sup>1,b</sup>, Martin Fossat<sup>1,c</sup>, Philippe Barthe<sup>1</sup>, Bertrand Garcia-Moreno E.<sup>2</sup>, Catherine A. Royer<sup>1,c\*</sup> and Christian Roumestand<sup>1\*</sup>

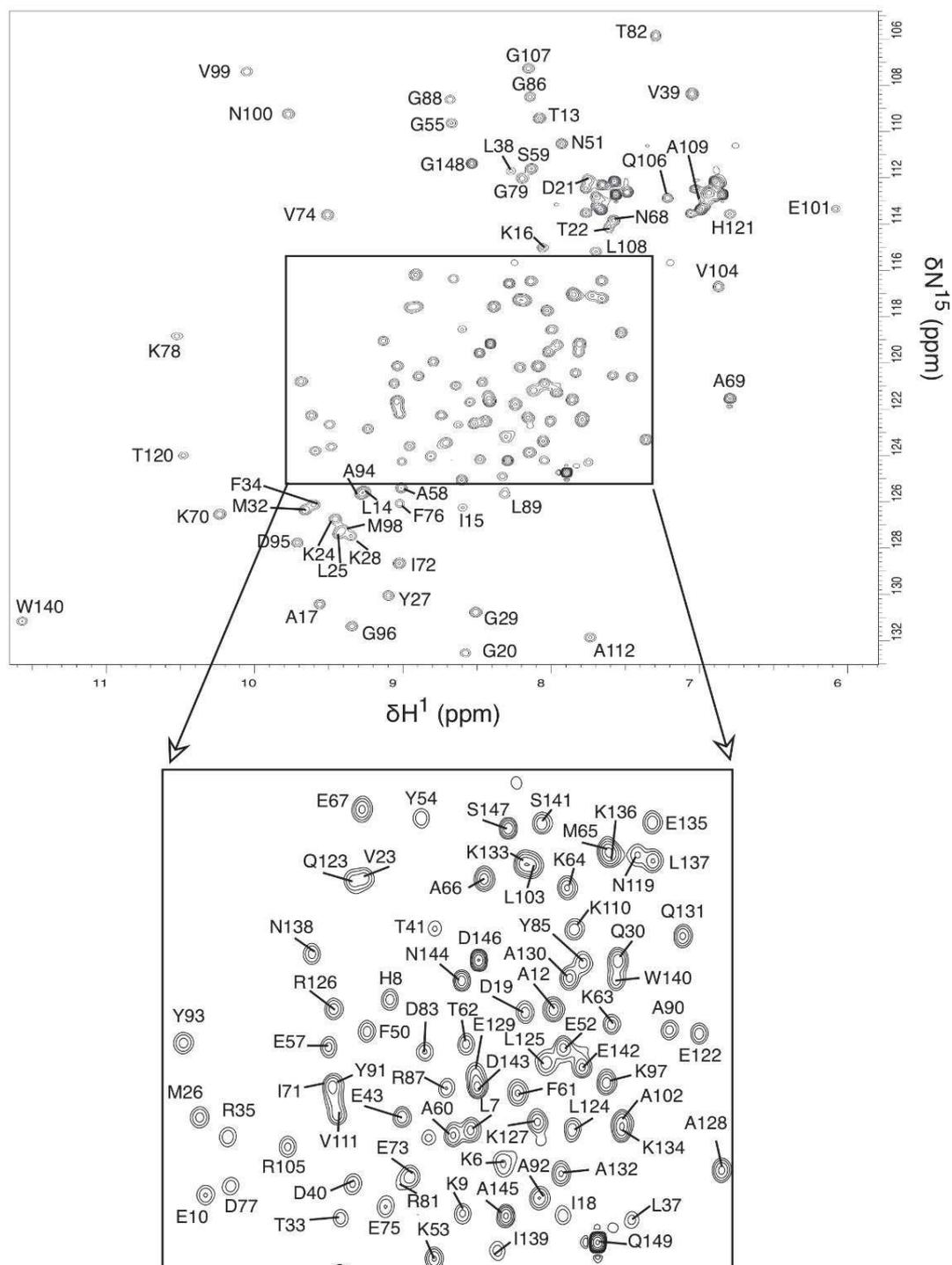
**Supplementary Material**

Supplementary table 1. Crystallographic parameters obtained for I92A/L125A.

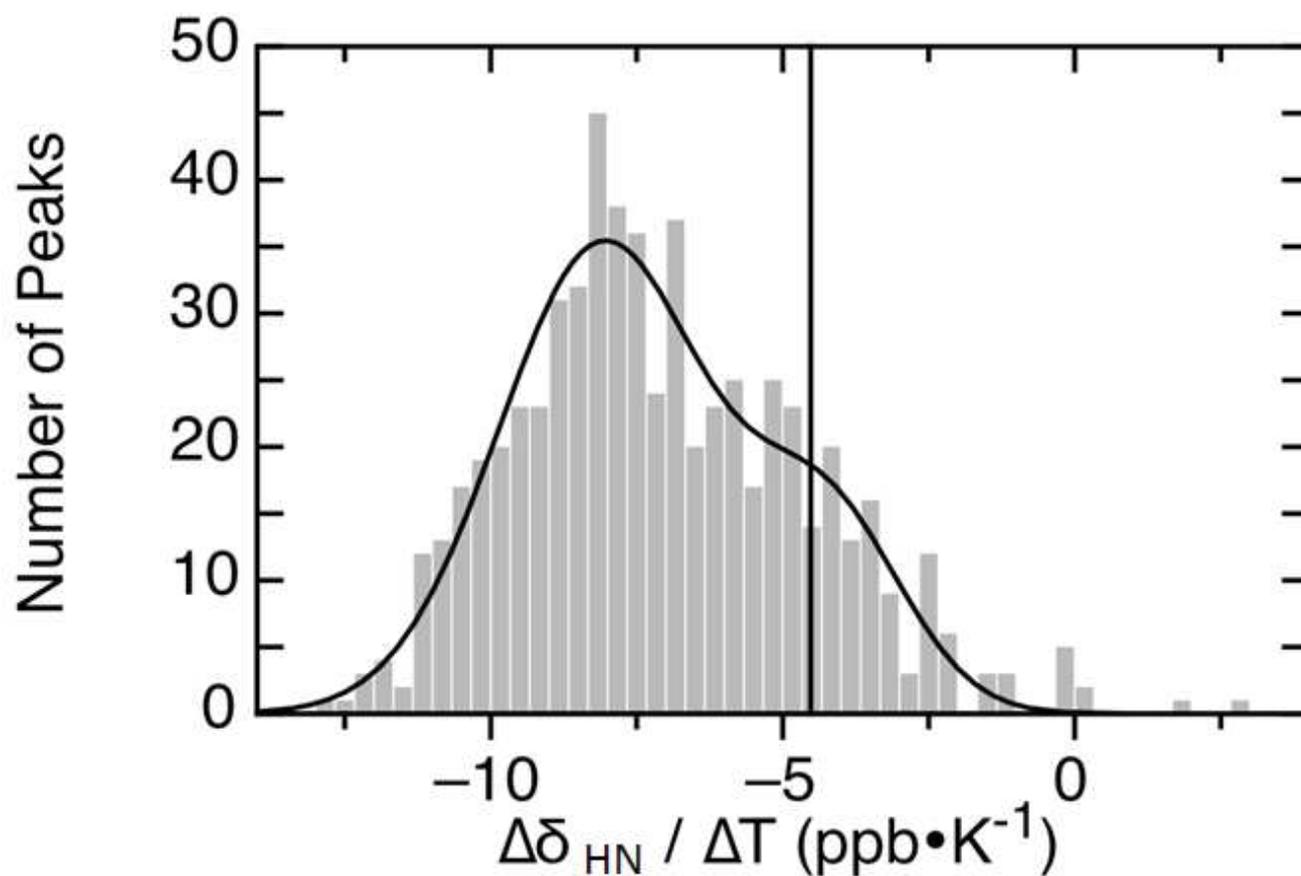
Protein	I92A/L125A
Crystallization conditions	
Buffer	25mM K phosphate
pH	7
Temperature (K)	277
Precipitant	20% (w/v) MPD
Additives	pdTp, CaCl <sub>2</sub>
Data collection	
Wavelength (Å)	1.1
Resolution (Å)	50.00-1.47 (1.50-1.47)
Unique reflections	23894 (1027)
Completeness	0.985 (0.862)
Redundancy	5.9 (3.4)
Average $I/\sigma(I)$	19.3 (8.3)
$R_{merge}$	0.054 (0.169)
Wilson B (Å <sup>2</sup> )	29.5
Space group	P2 <sub>1</sub>
Cell dimensions (Å ; °)	a = 31.07 ; α = 90.00 b = 60.40 ; β = 93.50 c = 38.48 ; γ = 90.00
Refinement	
Resolution (Å)	32.41-1.47 (1.51-1.47)
No. of non-hydrogen atoms	1225
No. of unique reflections	23838 (1449)
No. of reflections in test set	2397 (162)
$R_{work}$	0.174 (0.24)
$R_{free}$	0.213(0.27)
<i>RMS from ideal geometry</i>	0.018
Bonds (Å)	1.84
RMS angles (°)	24.0
<i>Average B-factors (Å<sup>2</sup>)</i>	31.3
Protein	22.4
Solvent	
Ion	98 (86.0)
<i>Ramachandran plot</i>	15 (13.2)
Most favored (%)	1 (0.9)
Additionally allowed (%)	

Disallowed (%)	114
No. of residues excluding Gly, Pro and termini	129
Total no. of residues	
PDB accession code	4DGZ
RMSD (Å) from $\Delta$ +PHS	0.35
Main chain only	0.14

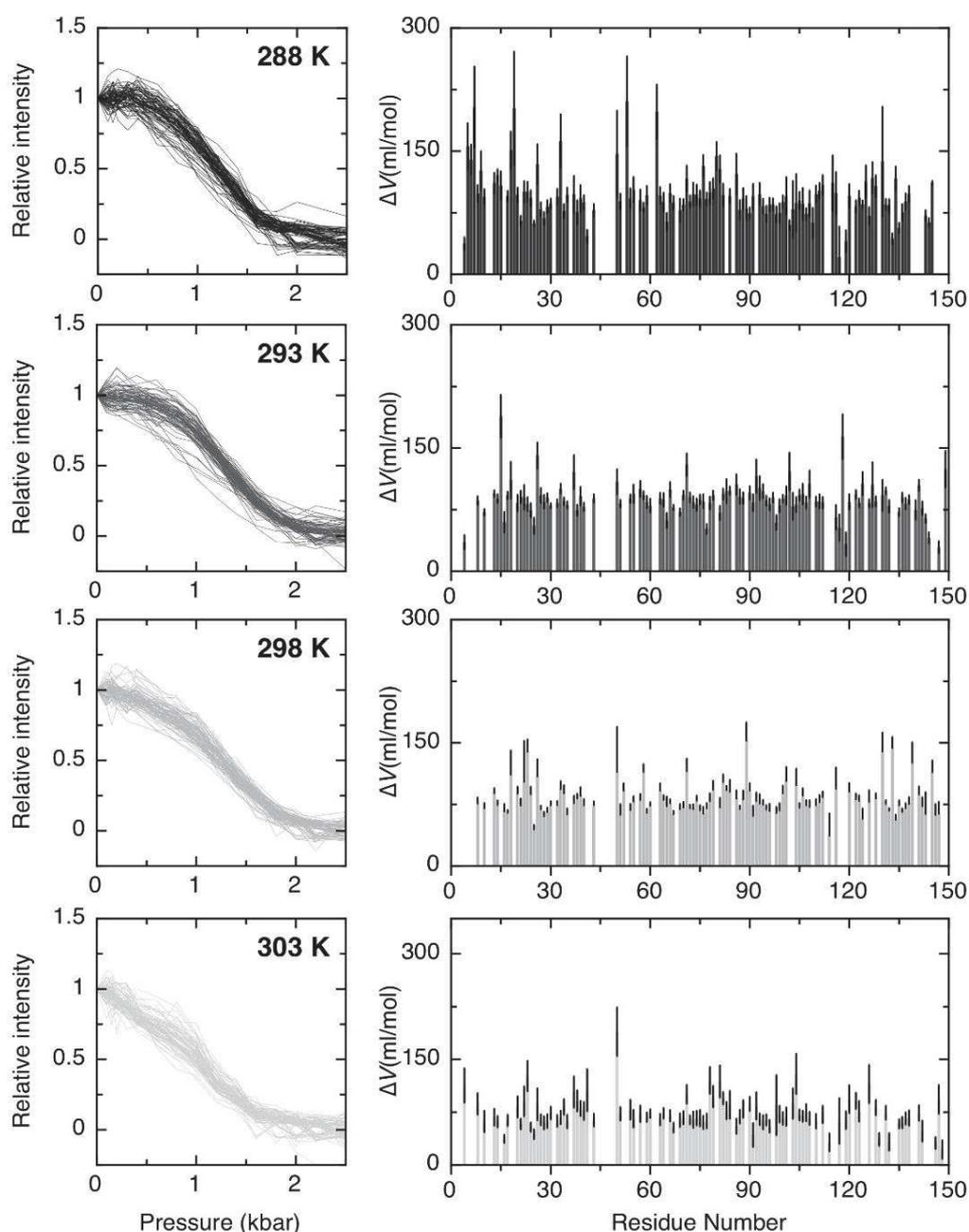
**Supplementary Figure 1.** Residue Specific assignments for backbone amide group of I92A/L125A double mutant. The assignments are given by the one letter amino acid code and the sequence positions in the corresponding 2D  $^1\text{H}^{15}\text{N}$  HSCQ spectra recorded at 298 K. Assignments of SNase,  $\Delta$ +PHS, I92A and L125A were defined earlier (1).



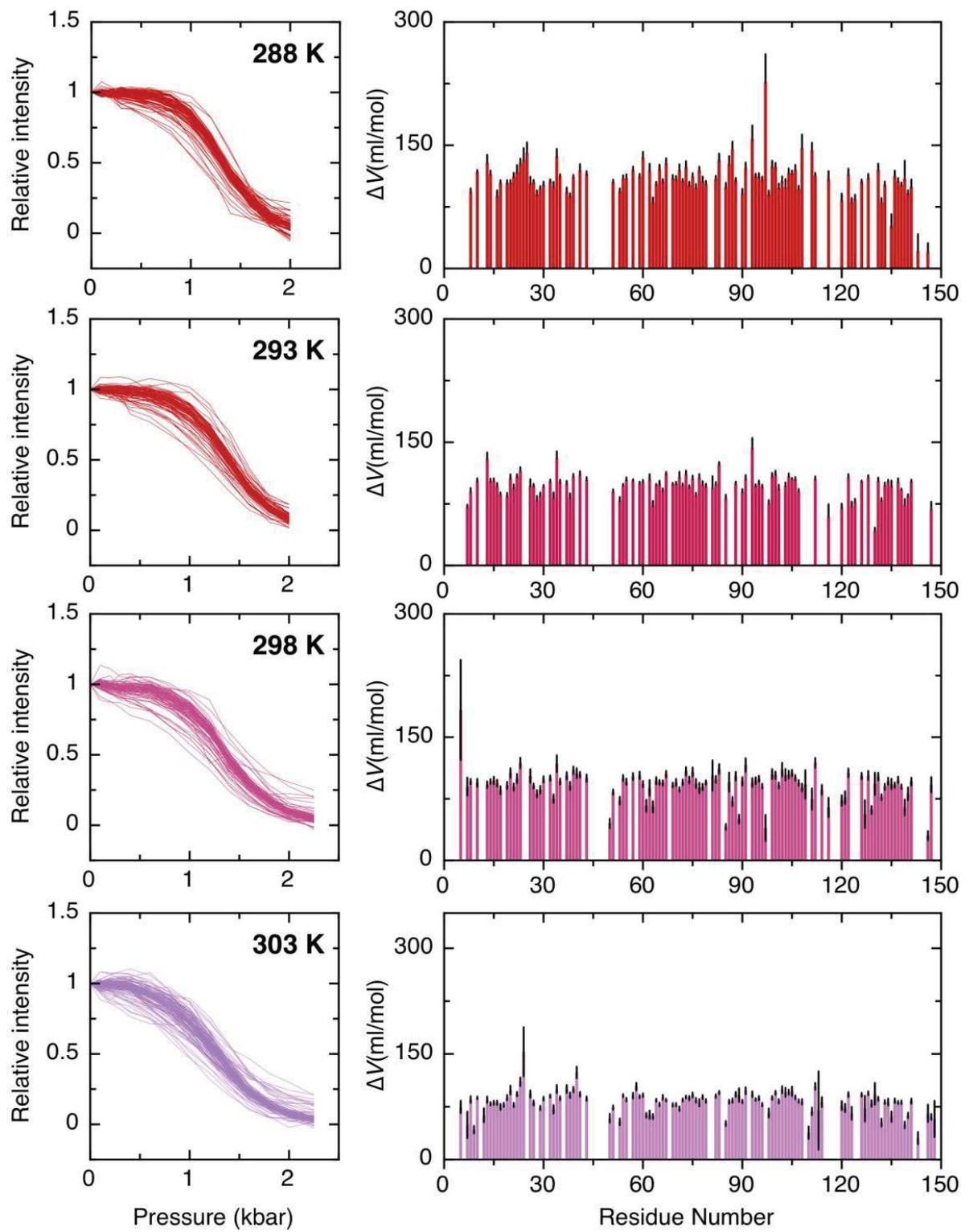
**Supplementary Figure 2.** Histogram of  $\Delta\delta_{\text{HN}}/\Delta T$  for all measured protein variants: SNase, the ultrastable  $\Delta$ +PHS and the corresponding cavity enlarging variants I92A, L125A and I92A/L125A. The line is a fit to a sum of two Gaussian functions centered at  $-8.1$  ppb/K (SD  $1.8$  ppb/K) and  $-4.5$  (SD  $1.3$  ppb/K). The vertical line is at  $-8.1 + 2 \times 1.8 = -4.5$  ppb/K. This value was used as the cut-off in Figure 1 to identify residues forming solid intramolecular H-bonds in solution (2).

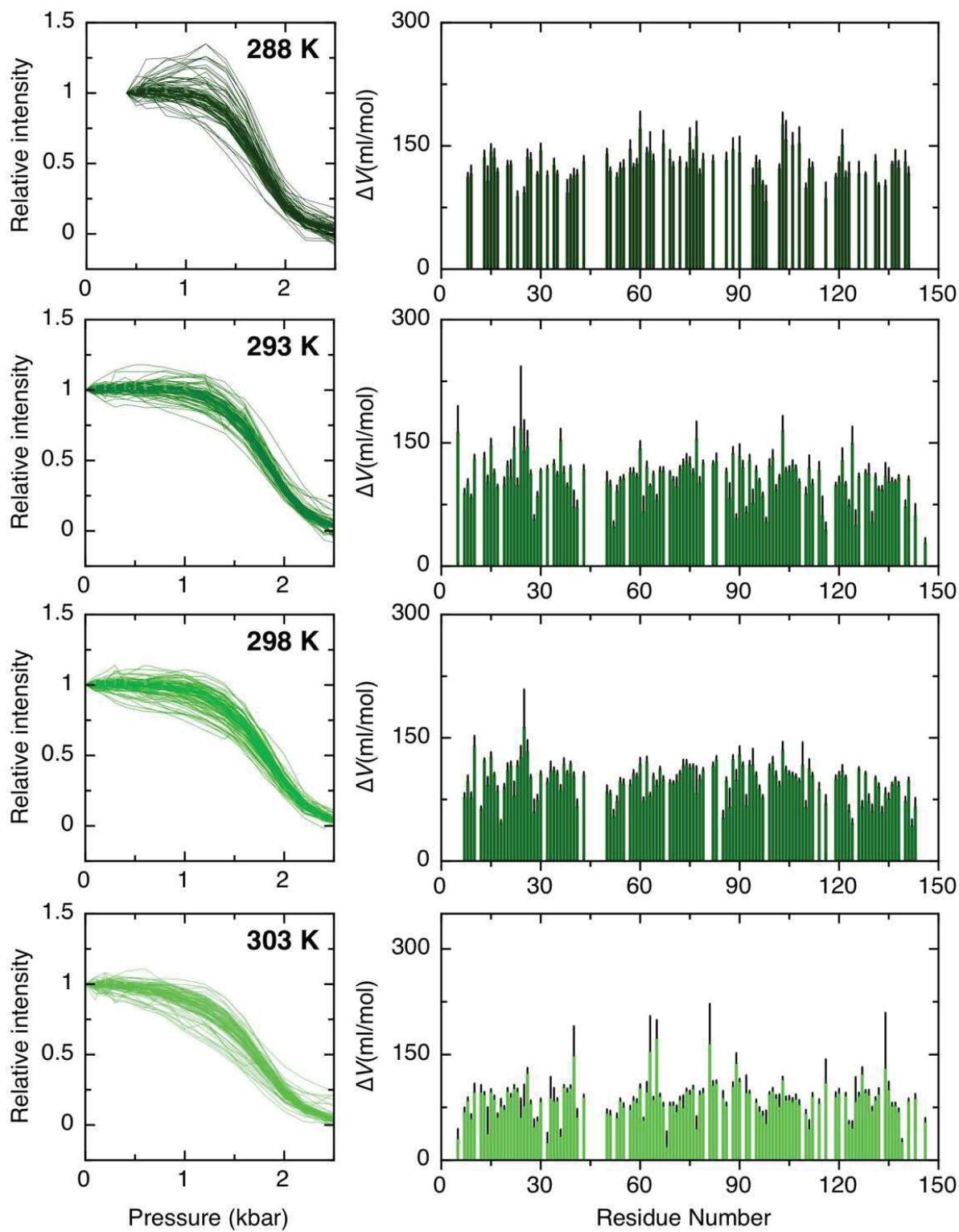


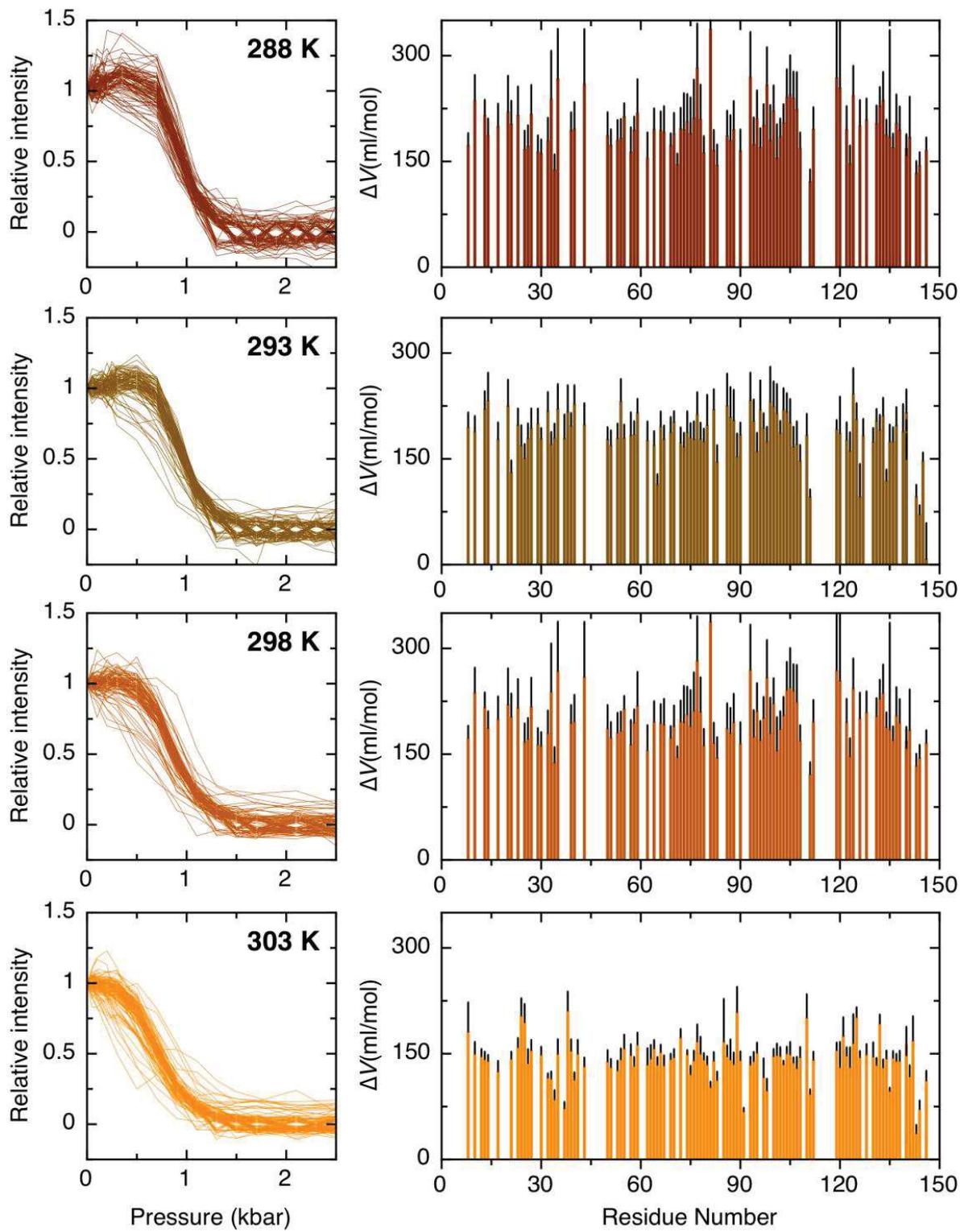
**Supplementary Figure 3** Thermal dependences of pressure induced unfolding of A) the ultrastable  $\Delta$ +PHS protein (grey) and the single mutants B) L125A (red) and C) I92A (green) and D) the corresponding double mutant L125/I92A (orange) as followed by 2D  $^{15}\text{N}$ - $^1\text{H}$  HSQC. First column: normalized intensity profiles of individual amides as a function of pressure. Second column: histogram of single residue  $\Delta V_f$  values derived from two-state unfolding model fit to intensity profiles. Under equilibrium conditions, native cross-peak intensities were integrated from the corresponding HSQC spectrum and the resulting intensity *versus* pressure data points were individually fitted for each resonance. The fitting procedure was equivalent to the one used for the high-pressure fluorescence experiments described in Figure S4, except no correction for quantum yield was applied. Experiments were recorded at 288 K, 293 K, 298 K and 303 K in 50 mM HCl-Tris buffer pH 7.0 and 1.8 M, 0.5 M, 0.75 M and 0 M GdmCl for  $\Delta$ PHS, L125A, I92A and L125/I92A respectively. Fitting  $\Delta V_f$  errors bars are shown as thin black lines.



**A.)  $\Delta$ +PHS**

**B.  $\Delta$ +PHS+L125A**

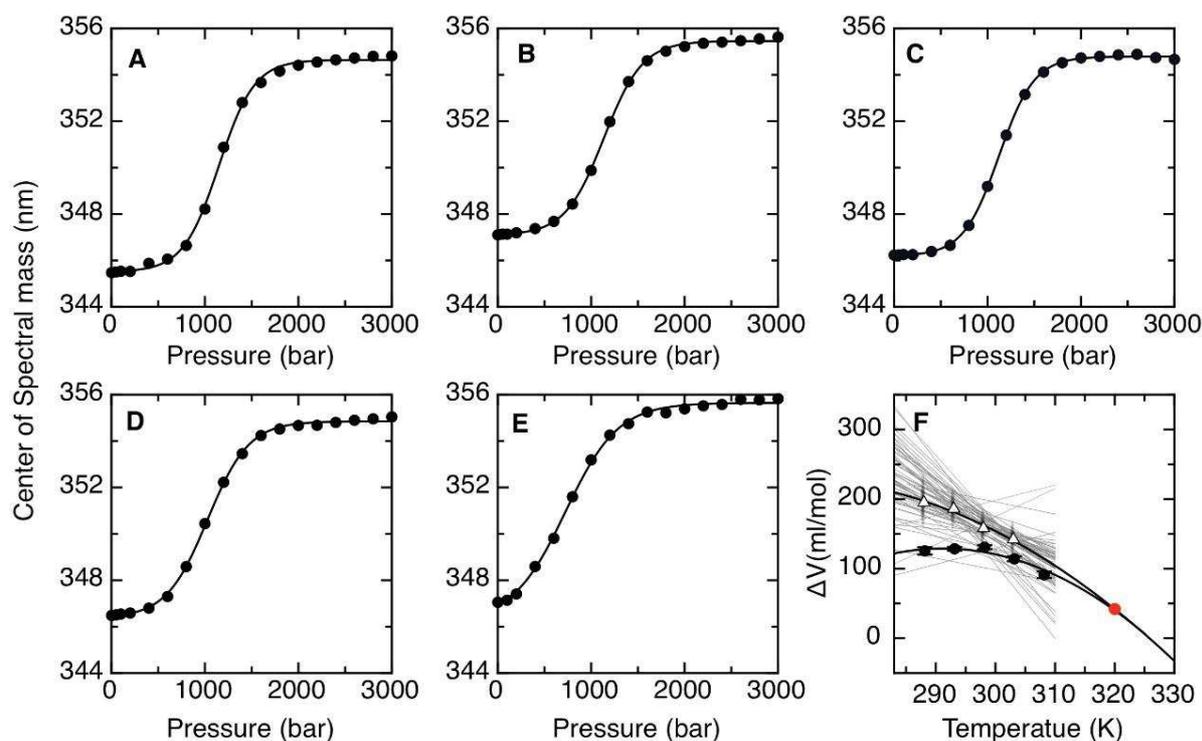
C.  $\Delta$ +PHS+I92A

D.  $\Delta$ +PHS+L125A+I92A

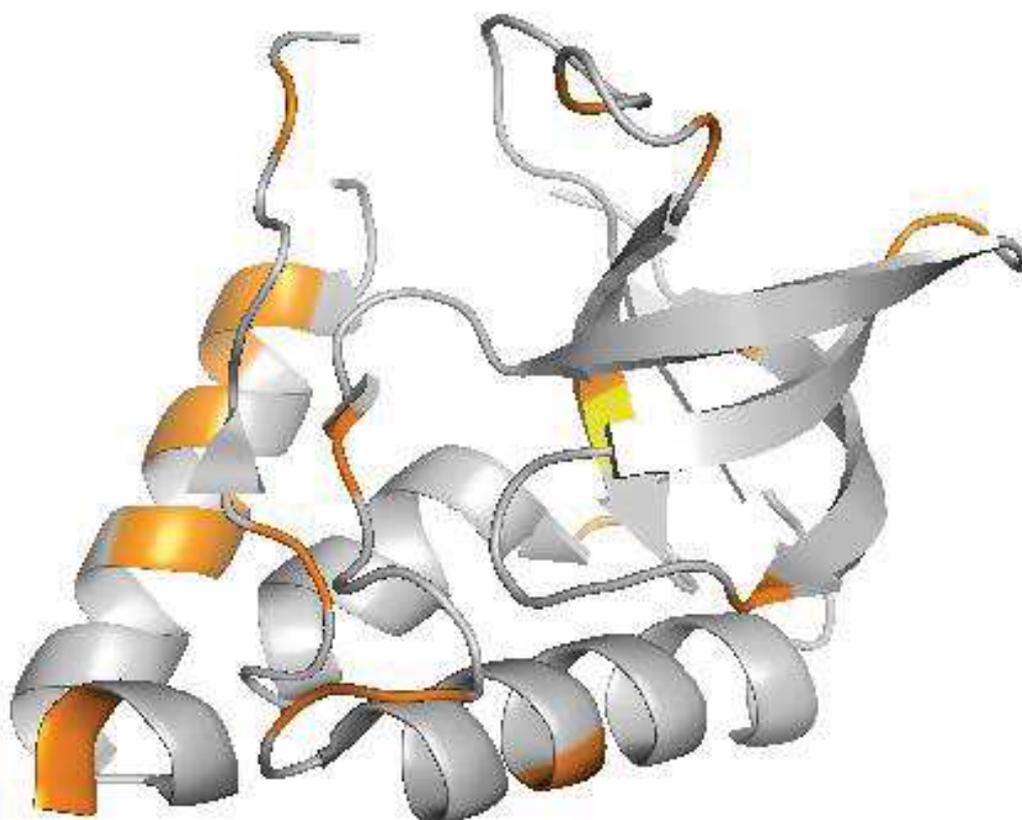
**Supplementary Figure 4. (A) to (E)**, pressure unfolding curves followed by the center of spectral mass of W fluorescence of I92A/L125A recorded at 288 K, 293 K, 298 K, 303 K and 308 K respectively. Lines are fits to the data to a two state unfolding model as described previously (3). For each experiment, a tryptophan emission spectrum was collected at equilibrium, from 320 to 450 nm, using an excitation wavelength of 290 nm. At equilibrium tryptophan emission spectrum was collected from 320 to 450 nm using an excitation wavelength of 290 nm. At each pressure, the intensity-weighted average wavelength  $\langle \lambda \rangle$  was calculated using the ISS software:  $\langle \lambda \rangle_j = \frac{\sum_j F_j \lambda_j}{\sum_j F_j}$  where  $j=320, 321...450$  nm.

Data were fitted to a two-state unfolding equilibrium as function of pressure for values of  $\Delta G_u^0$  and  $\Delta V_u^0$  using the BioEQS software, assuming a linear evolution of the free energy of unfolding with the pressure  $p$ :  $\Delta G_u(p) = \Delta G_u^0 + p\Delta V_u^0$  where  $\Delta G_u(p) = -RT \ln K_u(p)$  and  $K_u(p) = \frac{\langle \lambda \rangle_f - \langle \lambda \rangle_p}{\langle \lambda \rangle_p - \langle \lambda \rangle_u}$

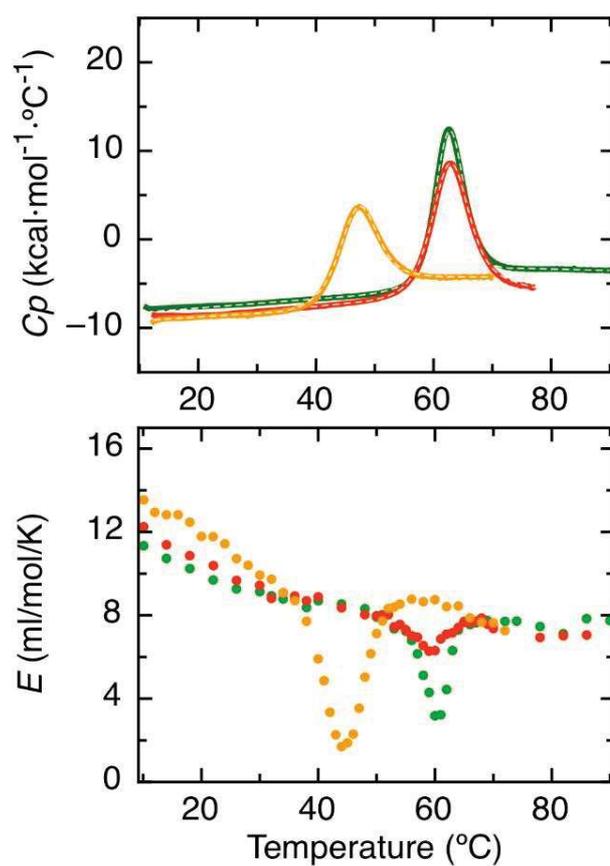
**(F)** Temperature dependence of  $\Delta V_f$  of I92A/L125A double mutant as measured by fluorescence (black circles), PPC (red circle) and NMR (open triangles). Black lines are quadratic fits to  $\Delta V_{\text{fluorescence}}$  and  $\Delta V_{\text{PPC}}$  or to  $\Delta V_{\text{NMR}}$  and  $\Delta V_{\text{PPC}}$  data. Grey lines are linear fits to microscopic  $\Delta V_{\text{NMR}}$  data. Fluorescence experiments were carried out in 50 mM HCl-tris buffer pH 7. PPC experiments were carried out in 50 mM sodium phosphate pH 7. NMR experiments were carried out in 50 mM HCl-Tris pH 7. Fluorescence detected pressure unfolding profiles for the double mutant yielded smaller values for  $\Delta V_f^0(T)$  on average than those obtained from NMR, as was previously the case for WT and  $\Delta$ +PHS and its variants with L125A and I92A (1) owing to the strong sensitivity of tryptophan fluorescence to small changes in local hydration.



**Supplementary Figure 5.** Structural cartoon of the I92A variant showing the residues exhibiting low values of  $\Delta V_f^p$ , which are indicative of deviation from two state behavior. Intensity is lost at lower pressures than for the average residue, indicating that prior to the main unfolding transition, there is partial unfolding involving the implicated residue. This tends to broaden the pressure dependent unfolding curve, and hence fitting two a simple two-state model results in a lower value of  $\Delta V_f^p$ . Positions on the backbone corresponding to residues with low  $\Delta V_f^p$  are colored in orange while the I92A position is in yellow.



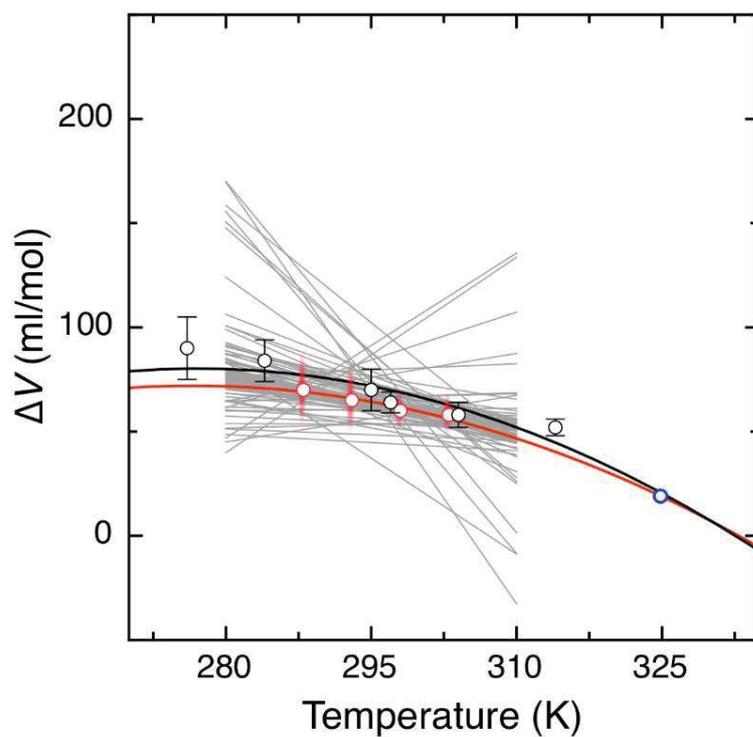
**Supplementary Figure 6.** Upper and lower panels show the DSC and PPC molar expansivity vs temperature calorimetric profiles respectively of  $\Delta$ +PHS single mutants I92A (green), L125A (red) and the double mutant I92A/L125A (orange). Thermodynamic parameters obtained for I92A/L125A variant are:  $T_{mDSC} = 320.4 \pm 10$  K;  $\Delta H_{DSC} = 73.3 \pm 0.3$  kcal/mol;  $\Delta C_p = 3.50 \pm 0.1$  kcal $\cdot$ C $^{-1}$ /mol;  $T_{mPPC} 317 \pm 5$  K and  $\Delta V_f(T_m)_{PPC} 42.1 \pm 1$  ml/mol. The corresponding parameters of I92A and L125A were previously reported (4). Experiments were carried out in 50 mM sodium phosphate pH 7.



**Supplementary Figure 7.** Coevolving network of the SNase protein family. **(A)** Circular representation of the multiple sequence alignment (MSA) and the mutual information (MI) of the SNase family. Squares indicate the MSA position and conservation (red: conserved, blue: variable). Lines connect pairs of residual positions with MI greater than 6.5. Connecting lines are colored grey and highlighted in red for high MI values or in black for intermediate MI values. MI interaction network is colored yellow (low MI) to violet (high MI) and represented as a web **(B)** or as straight sticks on the cartoon structure of SNase (PDB ID: 1SNC) **(C)**. Mutual information (MI) network was built using MISTIC server (<http://mistic.leloir.org.ar/>) (5) and SNase homologue family, Pfam ID: PF00565 (Finn *et al.* 2014). The multiple sequence alignment (MSA) derived from PF00565 has 1976 sequences in 838 clusters at 62% identity confirming a good MI predictive performance.



**Supplementary Figure 8.** Temperature dependence of SNase WT protein  $\Delta V$  as calculated by NMR (red circles), fluorescence (black circles) and PPC (blue circle). Red line is a quadratic fit to  $\Delta V_{\text{NMR}}$  and  $\Delta V_{\text{PPC}}$  data. Black line is a quadratic fit to  $\Delta V_{\text{fluorescence}}$  and  $\Delta V_{\text{PPC}}$  data. Grey lines are linear fits to single residue  $\Delta V_{\text{NMR}}$  data. PPC experiments were carried out in 10 mM Phosphate buffer pH 5.5. Fluorescence experiments were carried out in 10 mM Bis-tris pH 5.5 for 276 K, 284 K, 295 K, 304 K and 314 K and 50 mM Bis-Tris pH 6 for 297 K. NMR experiments were carried out in 10 mM Bis-tris pH 5.5.



## References

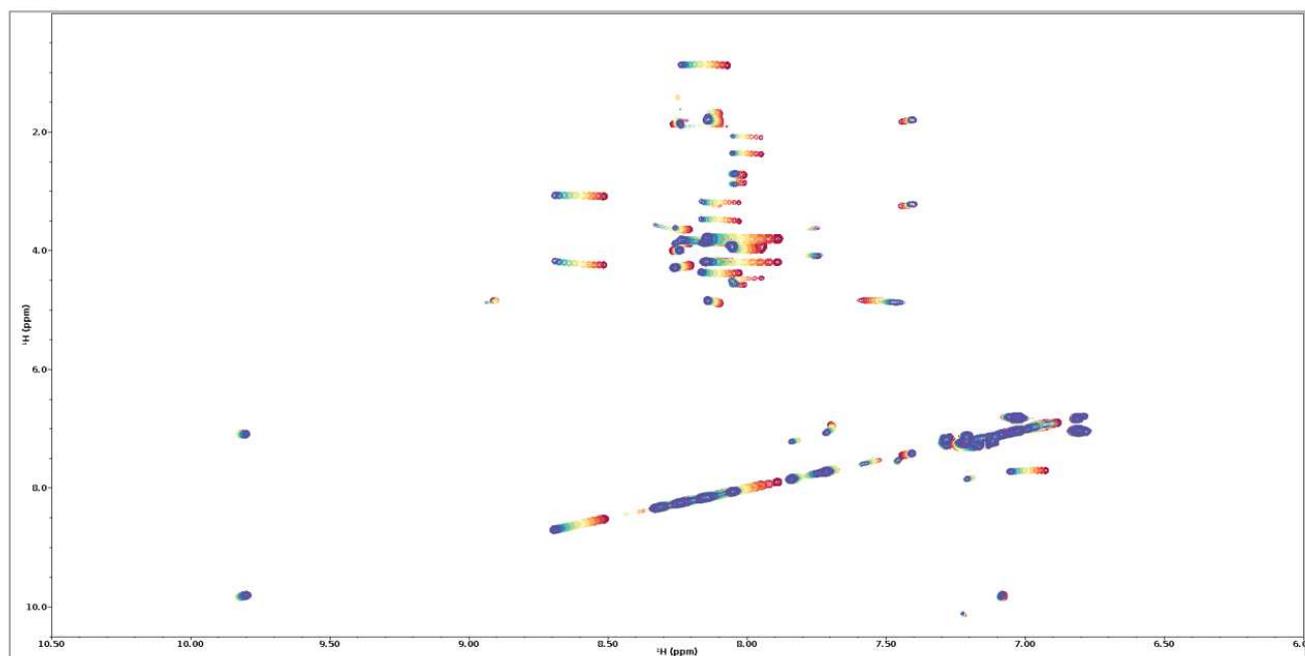
1. Roche, J., Caro, J. A., Norberto, D. R., Barthe, P., Roumestand, C., Schlessman, J. L., Garcia, A. E., Garcia-Moreno, B. E., and Royer, C. A. (2012) *Proc. Natl. Acad. Sci. U. S. A* **109**, 6945-6950.
2. Baxter, N. J. and Williamson, M. P. (1997) *J. Biomol. NMR* **9**, 359-369.
3. Dellarole, M. and Royer, C. A. (2014) *Methods Mol. Biol.* **1076**, 53-74.
4. Dellarole, M., Kobayashi, K., Rouget, J. B., Caro, J. A., Roche, J., Islam, M. M., Garcia-Moreno, E. B., Kuroda, Y., and Royer, C. A. (2013) *J. Phys. Chem. B* **117**, 12742-12749.
5. Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M., Marino, B. C. *Nucleic Acids Res* **2013**, **41**, W8.



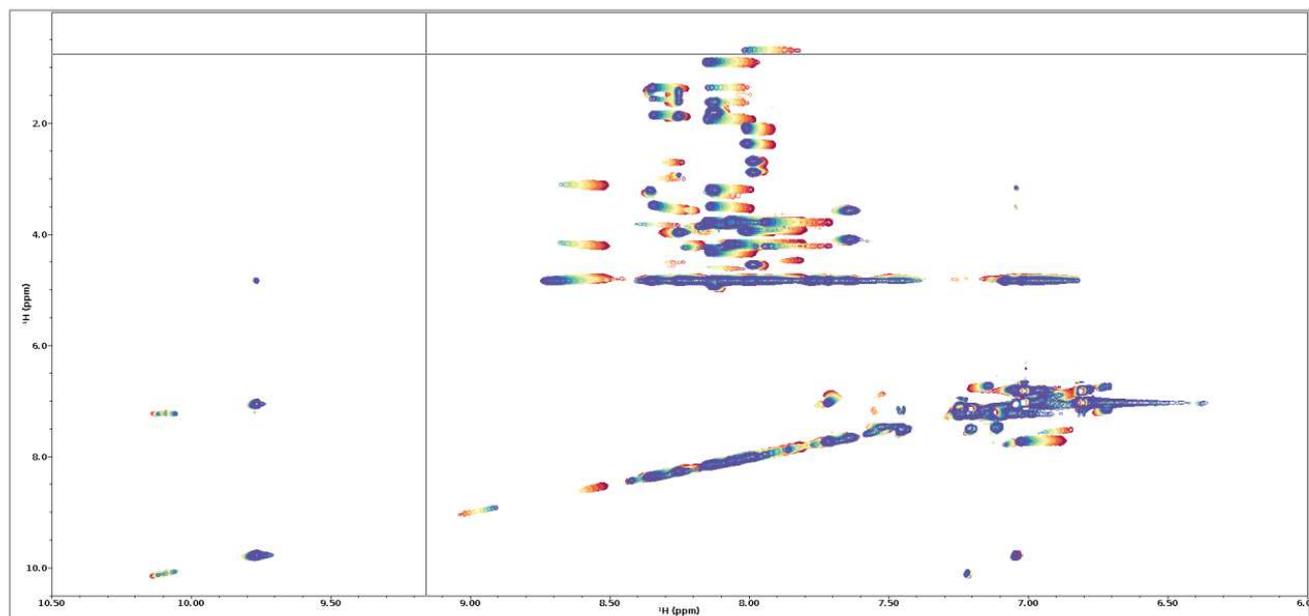
### 7.4.3 NMR and Computation Reveal a Pressure-Sensitive Folded Conformation of Trp Cage

**Supplementary Information**

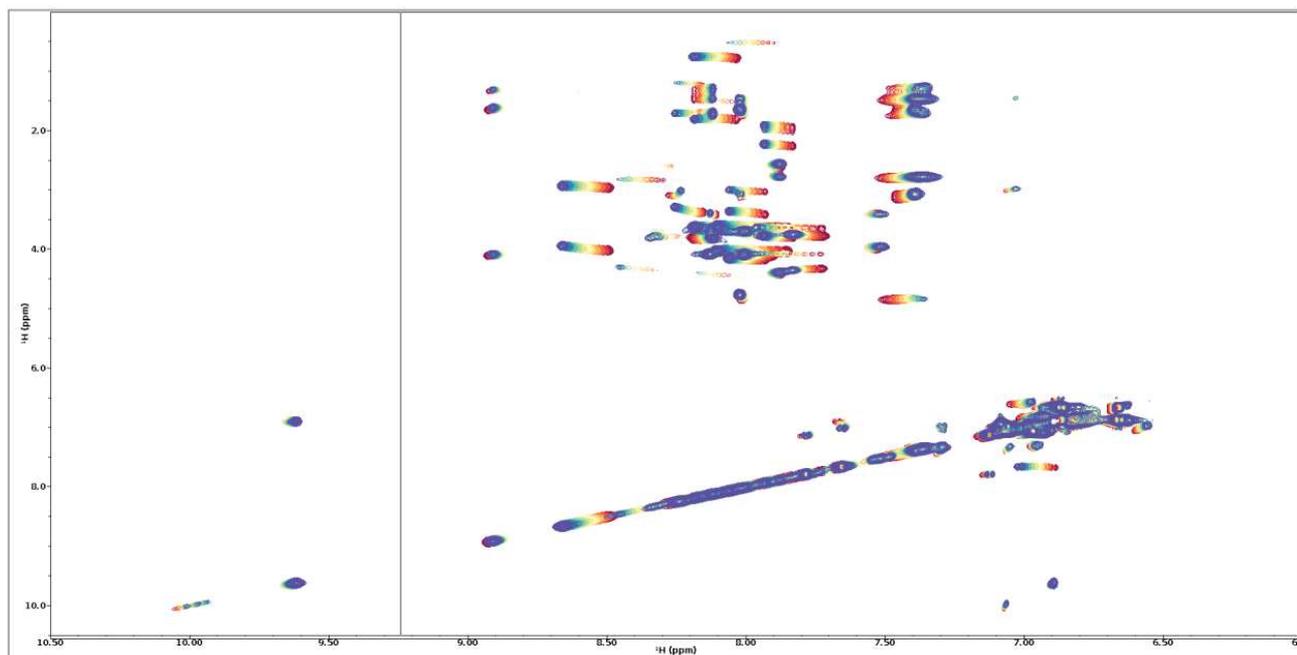
**Kitazawa et al.**



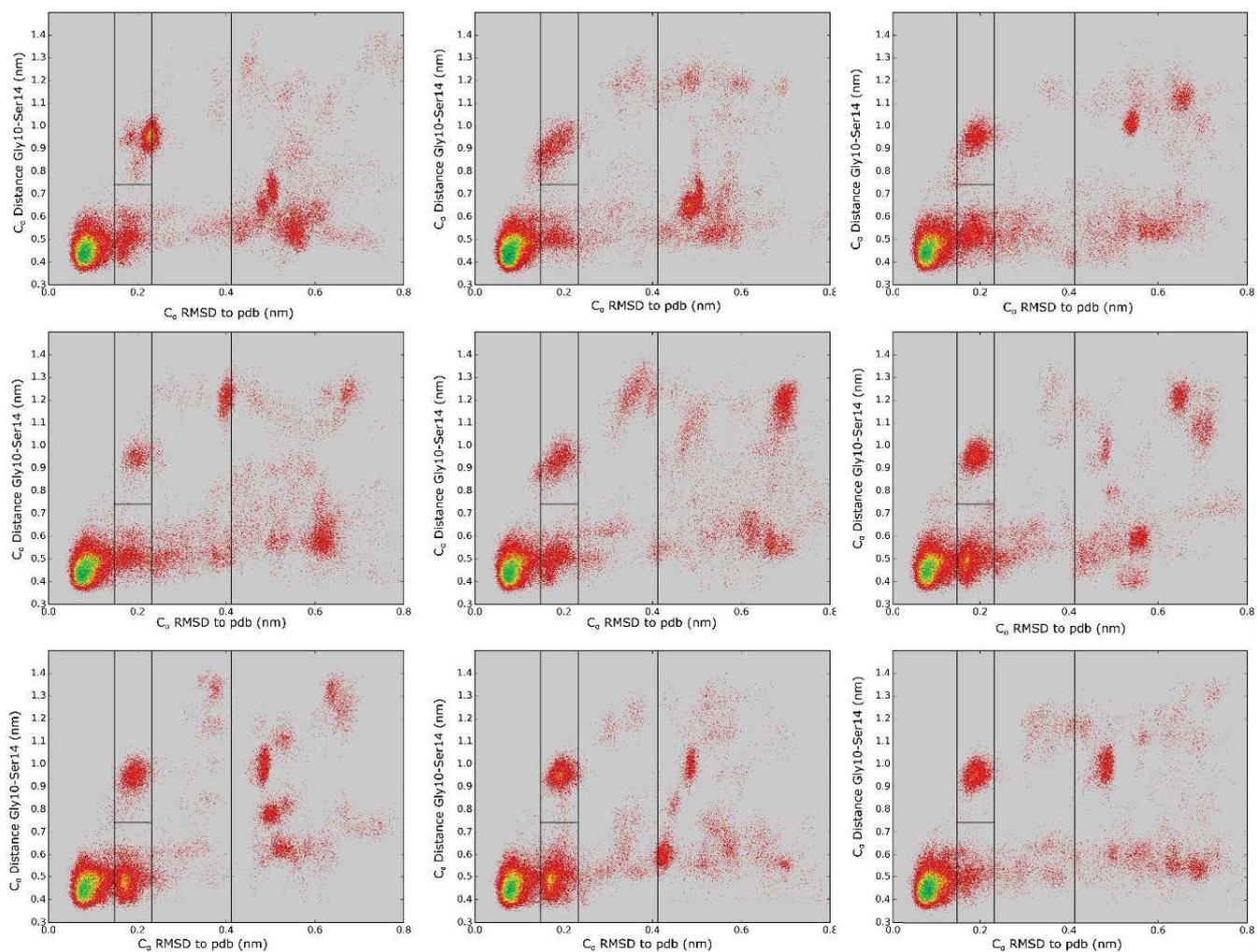
**Figure S1.** Pressure titration spectra for the 298K pH 7 conditions. Lowest pressure is displayed in Red. Highest pressure is displayed in blue.



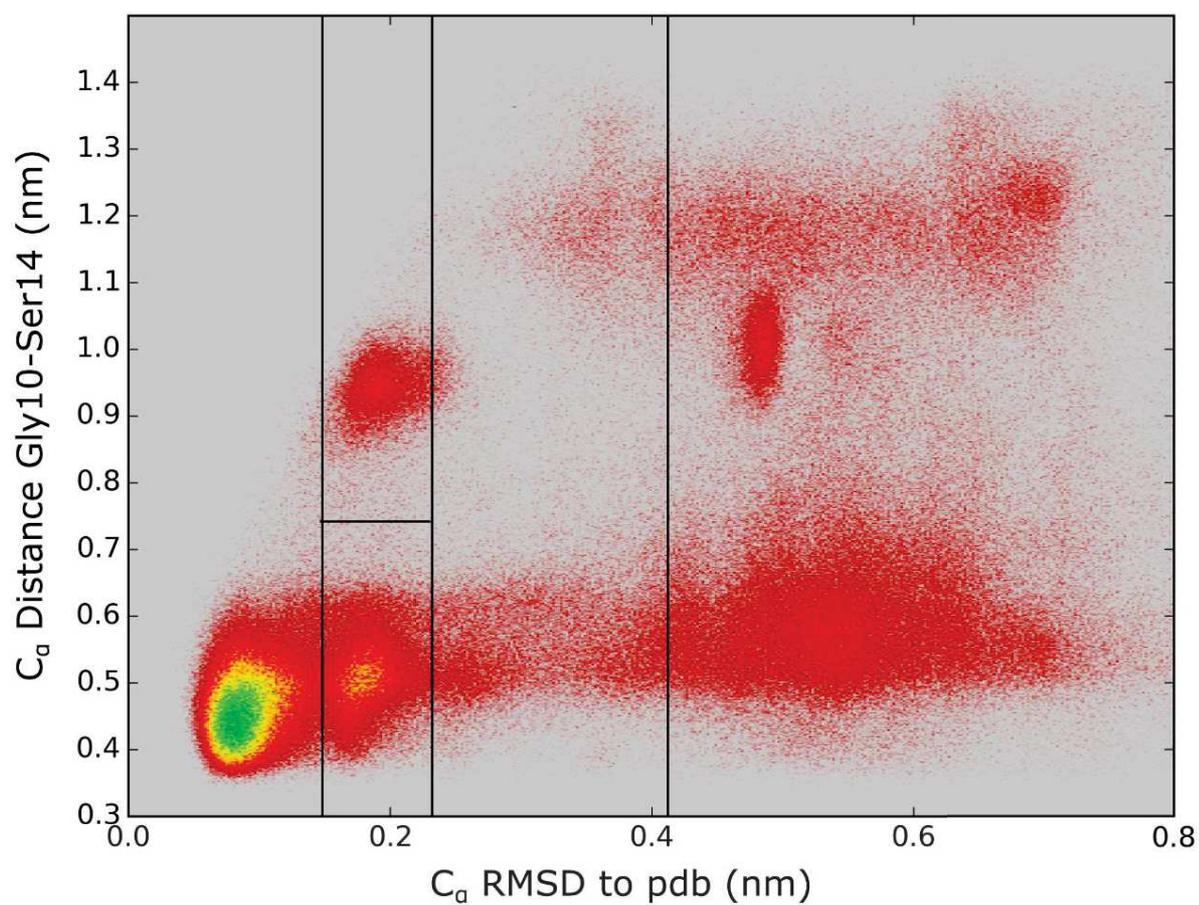
**Figure S2.** Pressure titration spectra for the 285K pH 5 conditions. Lowest pressure is displayed in Red. Highest pressure is displayed in blue .



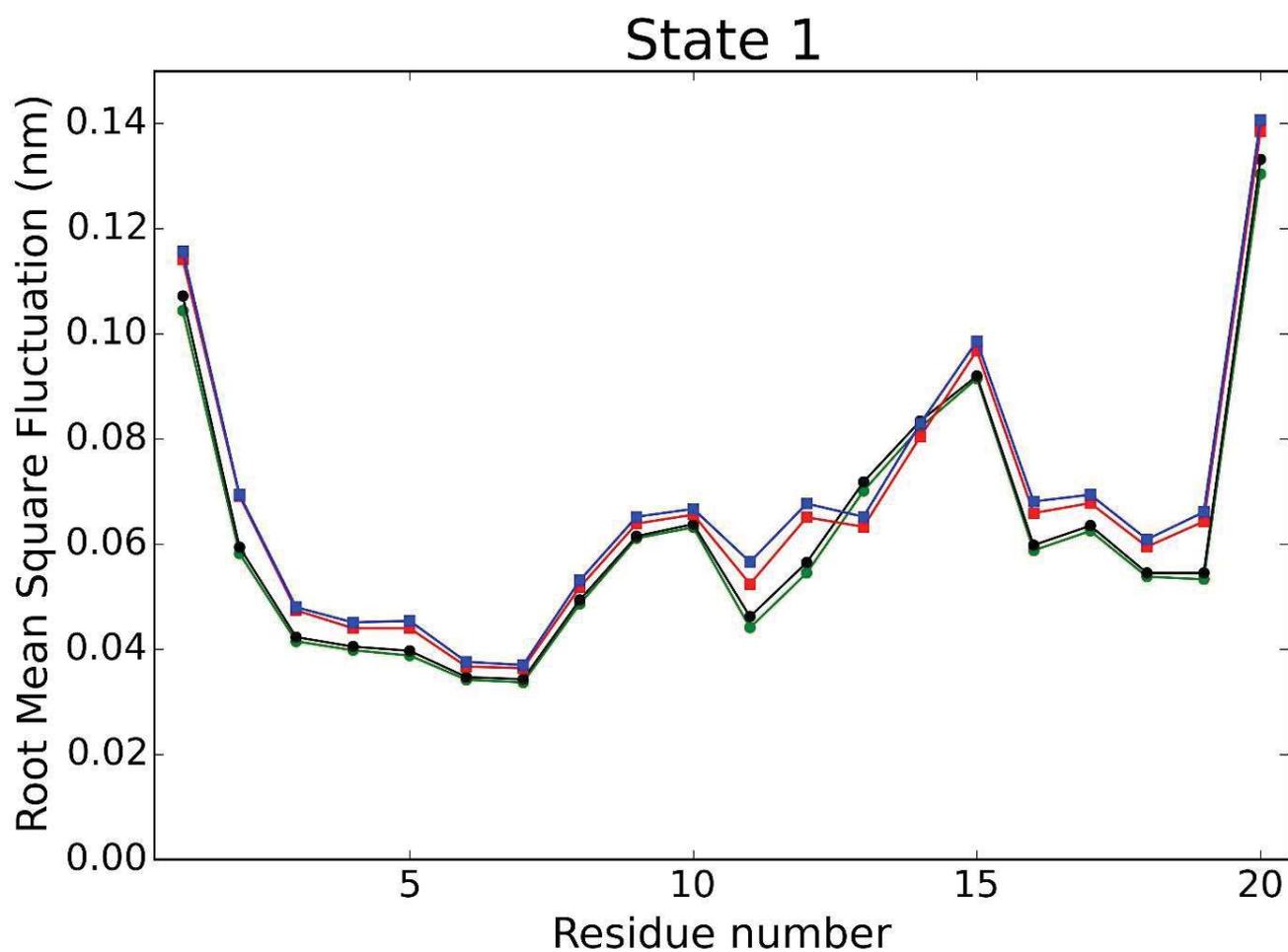
**Figure S3.** Pressure titration spectra for the 282K pH 5 conditions. Lowest pressure is displayed in Red. Highest pressure is displayed in blue.



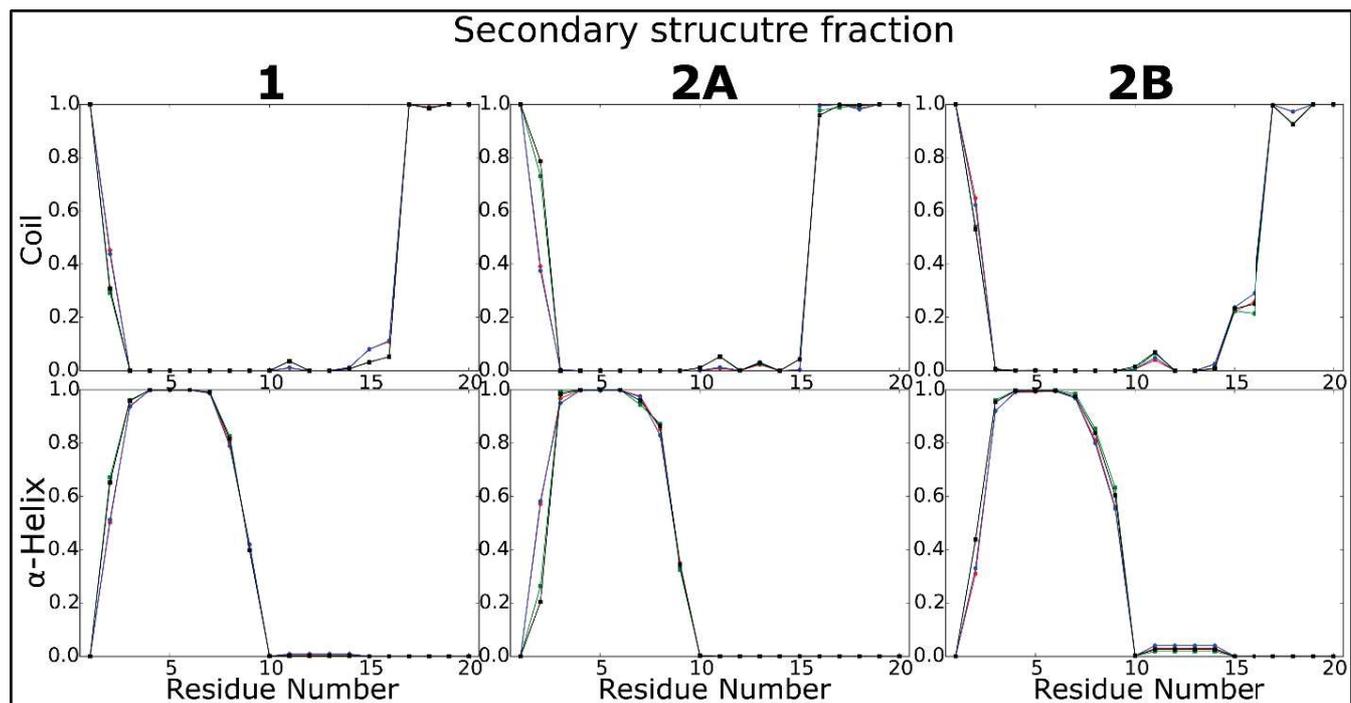
**Figure S4.** Block averaging of the Low density replica 1 simulation run. Each block is 50ns long.



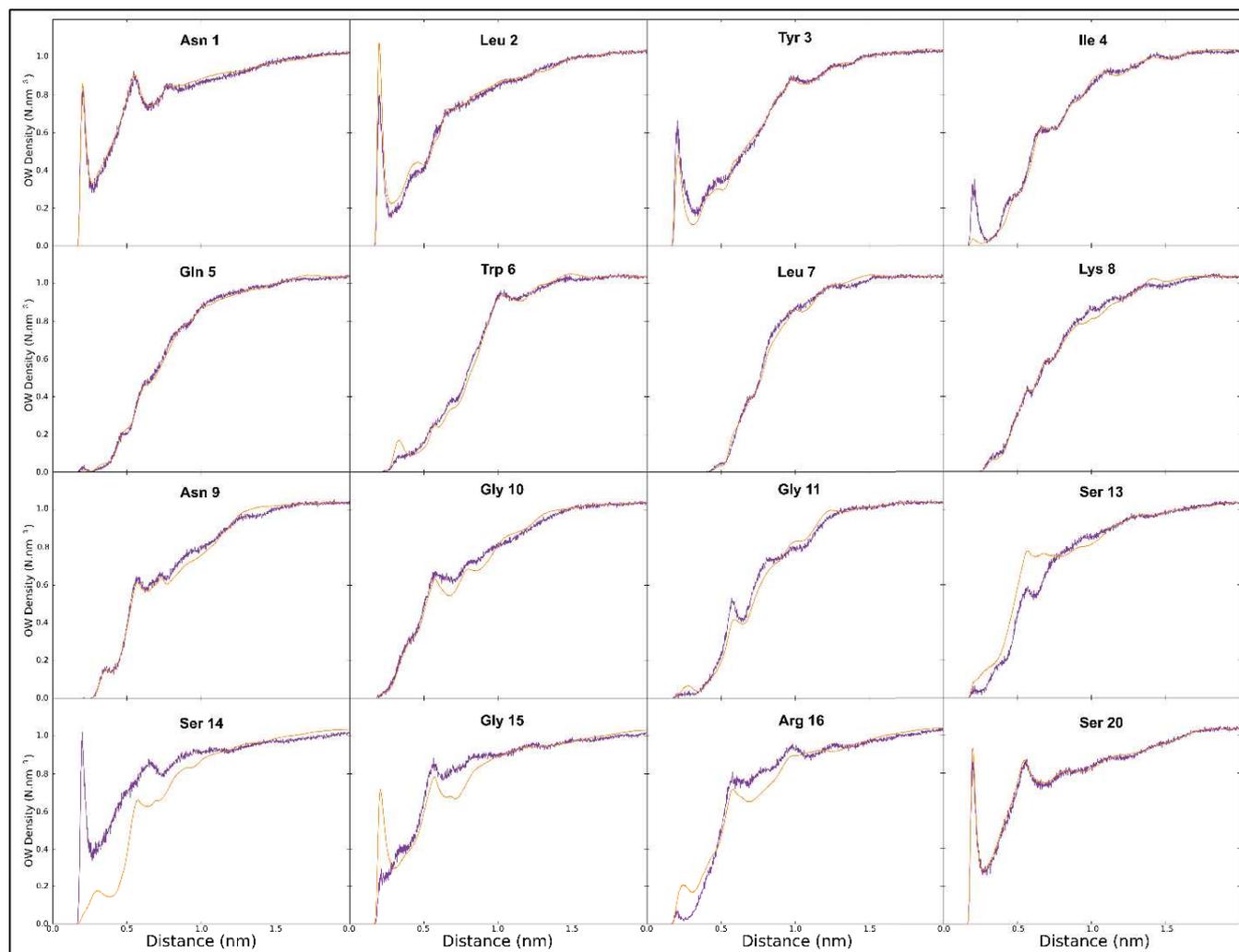
**Figure S5.** C<sub>α</sub> distance Gly10H<sub>N</sub>-Ser15H<sub>N</sub> vs C<sub>α</sub>RMSD to pdb structure (pdb 1l2y) for replica 11 yielding a pressure of 0 bar and a temperature of 330K. Black lines represent the boundaries chosen for RMSD based discrimination of structures at RMSDs of 0.147nm, 0.23 nm and 0.44 nm for the vertical lines and Gly10H<sub>N</sub>-Ser15H<sub>N</sub> distance of 0.75 nm for the horizontal lines.



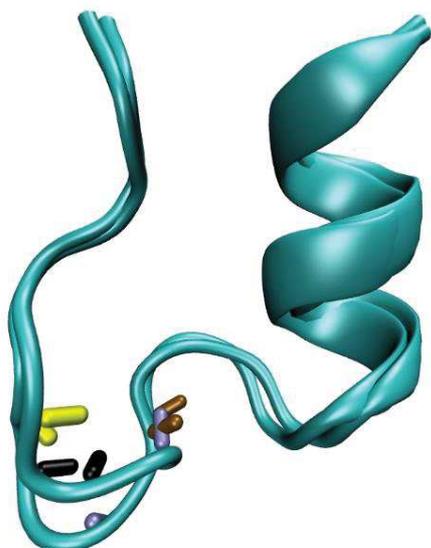
**Figure S6.** Root Mean Square Fluctuation for state 1. Green, and black correspond to high density runs at 285K and 297.6K, respectively, and red, and blue correspond to low density runs at 285K and 297.6K, respectively.



**Figure S7.** Fraction of secondary structure content by residue for each state as delimited in Figure 6, type of secondary structure is indicated on the left, points and line are colored by conditions, with blue, red, black, and green representing Low Density 297.6K, Low Density 285K, High Density 297.K and High Density 285K.



**Figure S8.** Radial Distribution Function of water oxygen around  $H_N$  of each residue averaged over all data sets for state 1 (orange) and state 2A (purple).



**Figure S9.** Difference in position of the amide hydrogen between state 1 and 2A for residue 13, 14, 15 and 16, displayed in brown, blue black and yellow, respectively.

## 7.5 Abbreviations

- Gdm : Guanidinium Hydrochloride
- Tris : tris(hydroxymethyl)aminomethane ((HOCH<sub>2</sub>)<sub>3</sub>CNH<sub>2</sub>)
- SNase : staphylococcal nuclease
- Hiapp : human islet amyloid polypeptide
- PP32 : Anp32
- LRR : Leucine Rich Repeat
- MD : Molecular Dynamics
- DNase :deoxyribonuclease
- DSS : 4,4-dimethyl-4-silapentane-1-sulfonic acid
- WT : Wild Type
- HP : High pressure
- NMR : Nuclear Magnetic Resonance
- SAXS : Small Angle X-ray Scattering
- DSC : differential scanning calorimetry
- PPC : Pressure perturbation calorimetry
- SBM : Structure Based Modeling
- DTT : Dithiothreitol



# Bibliography

- [1] Domenico Sanfelice and Piero Andrea Temussi. Cold denaturation as a tool to measure protein stability. *Biophysical Chemistry*, 208:4–8, 2016.
- [2] Katharina Wendler, Jens Thar, Stefan Zahn, and Barbara Kirchner. Estimating the hydrogen bond energy. *The journal of physical chemistry. A*, 114:9529–9536, 2010.
- [3] Tural Aksel, Ananya Majumdar, and Doug Barrick. The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding. *Structure*, 19(3):349–360, 2011.
- [4] Jonathan W. Neidigh, R. Matthew Fesinmeyer, and Niels H. Andersen. Designing a 20-residue protein. *Nature Structural Biology*, 9(6):425–430, jun 2002.
- [5] Bertrand E. García-Moreno, John J Dwyer, Apostolos G Gittis, Eaton E. Lattman, Daniel S Spencer, and Wesley E. Stites. Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophysical Chemistry*, 64(1-3):211–224, feb 1997.
- [6] Mariano Dellarole, Kei Kobayashi, Jean-Baptiste Rouget, José Alfredo Caro, Julien Roche, Mohammad M. Islam, Bertrand Garcia-Moreno E., Yutaka Kuroda, and Catherine A. Royer. Probing the Physical Determinants of Thermal Expansion of Folded Proteins. *The Journal of Physical Chemistry B*, 117(42):12742–12749, oct 2013.
- [7] Miguel A. De Los Rios, Mona Daneshi, and Kevin W. Plaxco. Experimental investigation of the frequency and substitution dependence of negative  $\Phi$ -values in two-state proteins. *Biochemistry*, 44(36):12160–12167, 2005.
- [8] Cyrus Levinthal. How to fold graciously. *Mössbauer Spectroscopy in Biological Systems Proceedings*, 24(41):22–24, 1969.
- [9] C Levinthal. Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65:44–45, 1968.
- [10] Ken a Dill, S Banu Ozkan, M Scott Shell, and Thomas R Weikl. The Protein Folding Problem. *Annual Review of Biophysics*, 37(1):289–316, jun 2008.
- [11] J D Bryngelson, J N Onuchic, N D Socci, and P G Wolynes. Funnels, Pathways, and the Energy Landscape of Protein-Folding - a Synthesis. *Proteins*, 21(3):167–195, 1995.
- [12] V L Arcus, S Vuilleumier, S M Freund, M Bycroft, and a R Fersht. A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *Journal of molecular biology*, 254(2):305–321, 1995.
- [13] Kirk C Aune, Ahmad Salahuddin, Mario H Zarlengo, and Charles Tanford. Evidence for residual structure in acid-and heat-denatured proteins. *Journal of Biological Chemistry*, 242(19):4486–4489, 1967.
- [14] Ken a Dill and David Shortle. Denatured States of Proteins. *Annual Review of Biochemistry*, 60(1):795–825, jun 1991.

- [15] Jean Baum, Christopher M. Dobson, Philip A. Evans, and Claire Hanley. Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry*, 28(1):7–13, jan 1989.
- [16] H S Chan, S Bromberg, and K a Dill. Models of cooperativity in protein folding. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 348(1323):61–70, 1995.
- [17] John F Brandts, Robert J Oliveira, and Chester Westort. Thermodynamics of protein denaturation. Effect of pressure on the denaturation on ribonuclease A. *Biochemistry*, 9(4):1038–1047, feb 1970.
- [18] Catherine A Royer. Revisiting volume changes in pressure-induced protein unfolding. *Biochimica et Biophysica Acta*, 1595(1-2):201–209, 2002.
- [19] Alexander Y. Yegorov and Sergey A. Potekhin. Moderate pressure has no distinct impact on hydrophobic hydration of proteins. *Thermochimica Acta*, 610:10–15, 2015.
- [20] Tuhin Ghosh, Angel E. Garcia, and Shekhar Garde. Enthalpy and entropy contributions to the pressure dependence of hydrophobic interactions. *The Journal of Chemical Physics*, 116(6):2480, 2002.
- [21] K P Murphy, P L Privalov, and S J Gill. Common features of protein unfolding and dissolution of hydrophobic compounds. *Science (New York, N.Y.)*, 247(4942):559–561, 1990.
- [22] Alan Cooper. Heat capacity of hydrogen-bonded networks: An alternative view of protein folding thermodynamics. *Biophysical Chemistry*, 85(1):25–39, 2000.
- [23] K A Sharp and B Madan. Hydrophobic Effect, Water Structure, and Heat Capacity Changes. *J. Phys. Chem. B*, 101(97):4343–4348, 1997.
- [24] Shekhar Garde, Gerhard Hummer, Angel E. García, Michael E Paulaitis, and Lawrence R Pratt. Origin of Entropy Convergence in Hydrophobic Hydration and Protein Folding. *Physical Review Letters*, 77(24):4966–4968, dec 1996.
- [25] Kim A. Sharp, Bhupinder Madan, Eric Manas, and Jane M. Vanderkooi. Water structure changes induced by hydrophobic and polar solutes revealed by simulations and infrared spectroscopy. *The Journal of Chemical Physics*, 114(4):1791, 2001.
- [26] Joseph D Bryngelson and Peter G Wolynes. Spin glasses and the statistical mechanics of protein folding. 84(November):7524–7528, 1987.
- [27] R a Goldstein, Z a Luthey-Schulten, and P G Wolynes. Optimal Protein-Folding Codes From Spin-Glass Theory. *Proc Natl Acad Sci USA*, 89(June):4918–4922, 1992.
- [28] Nataša Poklar, Jurij Lah, Mateja Salobir, Peter Maček, and Gorazd Vesnaver. pH and temperature-induced molten globule-like denatured states of equinatoxin II: A study by UV-melting, DSC, far- and near-UV CD spectroscopy, and ANS fluorescence. *Biochemistry*, 36(47):14345–14352, 1997.
- [29] C REDFIELD. Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods*, 34(1):121–132, sep 2004.
- [30] D.A. Dolgikh, R.I. Gilmanishin, E.V. Brazhnikov, V.E. Bychkova, G.V. Semisotnov, S.Yu. Venyaminov, and O.B. Ptitsyn.  $\alpha$ -lactalbumin: compact state with fluctuating tertiary structure? *FEBS Letters*, 136(2):311–315, dec 1981.
- [31] Robert L. Baldwin, Carl Frieden, and George D. Rose. Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins: Structure, Function and Bioinformatics*, 78(13):2725–2737, 2010.
- [32] K Kuwajima, E P Garvey, B E Finn, C R Matthews, and S Sugai. Transient intermediates in the folding of dihydrofolate reductase as detected by far-ultraviolet circular dichroism spectroscopy. *Biochemistry*, 30(31):7693–7703, 1991.

- [33] Mikio Ohgushi and Akiyoshi Wada. 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Letters*, 164(1):21–24, nov 1983.
- [34] Patricia A Jennings and Peter E Wright. Formation of a Molten Globule Intermediate Early in the Kinetic Folding Pathway of Apomyoglobin. *Science*, 262(November):892–896, 1993.
- [35] O.B. Ptitsyn, R.H. Pain, G.V. Semisotnov, E Zerovnik, and O.I. Razgulyaev. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Letters*, 262(1):20–24, mar 1990.
- [36] Santosh Kumar Jha and Jayant B Udgaonkar. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proceedings of the National Academy of Sciences of the United States of America*, 106(30):12289–94, 2009.
- [37] G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas', and R. I. Gilmanshin. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, 31(1):119–128, jan 1991.
- [38] Christina Redfield, Brenda A Schulman, M A Milhollen, Peter S Kim, Christopher M Dobson, A Michael, Peter S Kim, and Christopher M Dobson. Alpha-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat Struct Biol*, 6(10):948–952, 1999.
- [39] Gediminas J A Vidugiris, John L Markley, and Catherine A Royer. Evidence for a molten globule-like transition state in protein folding from determination of activation volumes. *Biochemistry*, 34(15):4909–4912, apr 1995.
- [40] Günter Pappenberger, Christophe Saudan, Michael Becker, André E Merbach, Thomas Kiefhaber, and E Merbach. Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements. *Proceedings of the National Academy of Sciences of the United States of America*, 97(1):17–22, 2000.
- [41] Jean Baptiste Rouget, Martin A. Schroer, Christoph Jeworrek, Matthias Pühse, Jean Louis Saldana, Yannick Bessin, Metin Tolan, Doug Barrick, Roland Winter, and Catherine A. Royer. Unique features of the folding landscape of a repeat protein revealed by pressure perturbation. *Biophysical Journal*, 98(11):2712–2721, 2010.
- [42] Ludovic Brun, Daniel G Isom, Priya Velu, Bertrand Garcia-Moreno, and Catherine A Royer. Hydration of the folding transition state ensemble of a protein. *Biochemistry*, 45(11):3473–3480, 2006.
- [43] T Kiefhaber and R L Baldwin. Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proceedings of the National Academy of Sciences of the United States of America*, 92(7):2657–61, 1995.
- [44] Patricia A Jennings and Peter E Wright. Formation of a Molten Globule Intermediate Early in the Kinetic Folding Pathway of Apomyoglobin. *Science*, 262(November):892–896, 1993.
- [45] Mourad Sadqi, Lisa J Lapidus, and Victor Muñoz. How fast is protein hydrophobic collapse? *Proceedings of the National Academy of Sciences of the United States of America*, 100(21):12117–12122, 2003.
- [46] T R Sosnick, L Mayne, and S W Englander. Molecular Collapse: the rate-limiting step in two state cytochrome c folding. *PROTEINS: Structure, Function, and Genetics*, 24:413–426, 1996.
- [47] Aaron K. Chamberlain and Susan Marqusee. Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. In *Advances in Protein Chemistry*, volume 53, pages 283–328. 2000.
- [48] J D Bryngelson, J N Onuchic, N D Socci, and P G Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, 21(3):167–95, mar 1995.

- [49] Thomas C. Taylor and Inger Andersson. Structural transitions during activation and ligand binding in hexadecameric Rubisco inferred from the crystal structure of the activated unliganded spinach enzyme. *Nature Structural Biology*, 3(1):95–101, jan 1996.
- [50] Anna R Panchenko, Zaida Luthey-Schulten, Ronald Cole, and Peter G Wolynes. The foldon universe: a survey of structural similarity and self-recognition of independently folding units. *Journal of Molecular Biology*, 272(1):95–105, sep 1997.
- [51] H. Maity, Mita Maity, Mallela M G Krishna, Leland Mayne, and S Walter Englander. Protein folding: The stepwise assembly of foldon units. *Proceedings of the National Academy of Sciences*, 102(13):4741–4746, mar 2005.
- [52] Mallela M G Krishna, Linh Hoang, Yan Lin, and S Walter Englander. Hydrogen exchange methods to study protein folding. *Methods (San Diego, Calif.)*, 34(1):51–64, sep 2004.
- [53] Marianne J. Rooman, Jean Pierre A. Kocher, and Shoshana J. Wodak. Extracting information on folding from the amino acid sequence: Accurate predictions for protein regions with preferred conformation in the absence of tertiary interactions. *Biochemistry*, 31(42):10226–10238, oct 1992.
- [54] Kevin W Plaxco, Kim T Simons, and David Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 277(4):985–994, apr 1998.
- [55] Igor B. Kuznetsov and Shalom Rackovsky. Class-specific correlations between protein folding rate, structure-derived, and sequence-derived descriptors. *Proteins: Structure, Function, and Bioinformatics*, 54(2):333–341, sep 2003.
- [56] M J Rooman and S J Wodak. Extracting information on folding from the amino acid sequence: consensus regions with preferred conformation in homologous proteins. *Biochemistry*, 31(42):10239–10249, 1992.
- [57] STEPHEN G. BRUSH. History of the Lenz-Ising Model. *Reviews of Modern Physics*, 39(4):883–893, oct 1967.
- [58] Theoretical Studies Of, Protein Folding, and N Go. Theoretical studies of protein folding. *Annual review of biophysics and bioengineering*, 12(1):183–210, 1983.
- [59] Tural Aksel and Doug Barrick. *Analysis of repeat-protein folding using nearest-neighbor statistical mechanical models.*, volume 455. Elsevier Inc., 1 edition, jan 2009.
- [60] Ellen Kloss, Naomi Courtemanche, and Doug Barrick. Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Archives of biochemistry and biophysics*, 469(1):83–99, jan 2008.
- [61] Cecilia C Mello and Doug Barrick. An experimentally determined protein folding energy landscape. 101(39), 2004.
- [62] Thuy Phuong Dao, Ananya Majumdar, and Doug Barrick. Highly polarized C-terminal transition state of the leucine-rich repeat domain of PP32 is governed by local stability. *Proceedings of the National Academy of Sciences*, 112(18):E2298–E2306, may 2015.
- [63] Martin Niss. History of the Lenz-Ising Model 1920-1950: From Ferromagnetic to Cooperative Phenomena. *Archive for History of Exact Sciences*, 59(3):267–318, mar 2005.
- [64] Tommi Kajander, Aitziber L. Cortajarena, Simon Mochrie, and Lynne Regan. Structure and stability of designed TPR protein superhelices: Unusual crystal packing and implications for natural TPR proteins. *Acta Crystallographica Section D: Biological Crystallography*, 63(7):800–811, 2007.
- [65] Tural Aksel and Doug Barrick. Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. *Biophysical journal*, 107(1):220–32, jul 2014.

- [66] Valerie Daggett and Alan R Fersht. Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences*, 28(1):18–25, jan 2003.
- [67] Victor Muñoz, Luis A Campos, and Mourad Sadqi. Limited cooperativity in protein folding. *Current Opinion in Structural Biology*, 36:58–66, feb 2016.
- [68] Pierre Perrot. *A to Z of Thermodynamics*. Oxford University Press, 1998.
- [69] Tamara N. Tsalkova and Peter L. Privalov. Stability of troponin C. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 624(1):196–204, jul 1980.
- [70] Robert L Baldwin. Temperature dependence of the hydrophobic interaction in protein folding. *Proceedings of the National Academy of Sciences*, 83(21):8069–8072, 1986.
- [71] George I Makhatadze and Peter L Privalov. On the entropy of protein folding. *Protein Science*, 5(3):507–510, mar 1996.
- [72] Litian Fu and Ernesto Freire. On the origin of the enthalpy and entropy convergence temperatures in protein folding. *Proceedings of the National Academy of Sciences*, 89(19):9335–9338, oct 1992.
- [73] K.P. Murphy and S.J. Gill. Group additivity thermodynamics for dissolution of solid cyclic dipeptides into water. *Thermochimica Acta*, 172:11–20, dec 1990.
- [74] Kenneth P. Murphy and Stanley J. Gill. Solid model compounds and the thermodynamics of protein unfolding. *Journal of Molecular Biology*, 222(3):699–709, dec 1991.
- [75] Peter L Privalov and George I. Makhatadze. Contribution of Hydration to Protein Folding Thermodynamics. *Journal of Molecular Biology*, 232(2):660–679, jul 1993.
- [76] George I Makhatadze and Peter L Privalov. Hydration effects in protein unfolding. *Biophysical Chemistry*, 51(2-3):291–309, aug 1994.
- [77] R V Pappu, R Srinivasan, and G D Rose. The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 97(23):12565–70, 2000.
- [78] a J Doig and M J Sternberg. Side-chain conformational entropy in protein folding. *Protein science : a publication of the Protein Society*, 4(11):2247–2251, 1995.
- [79] Michael C. Baxa, Esmael J. Haddadian, John M. Jumper, Karl F. Freed, and Tobin R. Sosnick. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proceedings of the National Academy of Sciences of the United States of America*, 111(43):15396–401, 2014.
- [80] Michael S. Marlow, Jakob Dogan, Kendra K. Frederick, Kathleen G. Valentine, and A. Joshua Wand. The role of conformational entropy in molecular recognition by calmodulin. *Nature chemical biology*, 6(5):352–358, 2010.
- [81] Vignesh Kasinath, Kim A. Sharp, and A. Joshua Wand. Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *Journal of the American Chemical Society*, 135(40):15092–15100, 2013.
- [82] Christine Peter, Chris Oostenbrink, Arthur Van Dorp, and Wilfred F. Van Gunsteren. Estimating entropies from molecular dynamics simulations. *Journal of Chemical Physics*, 120(6):2652–2661, 2004.
- [83] Jeffrey K Myers, C Nick Pace, and J Martin Scholtz. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein science : a publication of the Protein Society*, 4(10):2138–48, oct 1995.

- [84] Vakhtang V Loladze, Dmitri N Ermolenko, and George I Makhatadze. Heat capacity changes upon burial of polar and nonpolar groups in proteins. pages 1343–1352, 2001.
- [85] Alan Cooper. Protein heat capacity: An anomaly that maybe never was. *Journal of Physical Chemistry Letters*, 1(22):3298–3304, 2010.
- [86] Themis Lazaridis and Martin Karplus. Heat capacity and compactness of denatured proteins. *Biophysical Chemistry*, 78(1-2):207–217, 1999.
- [87] Silvina Matysiak, Pablo G. Debenedetti, and Peter J. Rossky. Role of hydrophobic hydration in protein stability: A 3D water-explicit protein model exhibiting cold and heat denaturation. *Journal of Physical Chemistry B*, 116(28):8095–8104, 2012.
- [88] Calvin R Chen and George I Makhatadze. ProteinVolume: calculating molecular van der Waals and void volumes in proteins. *BMC bioinformatics*, 16(1):101, 2015.
- [89] Heiko Seeman, Roland Winter, Catherine A Royer, and Heiko Seemann. Volume, Expansivity and Isothermal Compressibility Changes Associated with Temperature and Pressure Unfolding of Staphylococcal Nuclease. *J Mol Biol*, 307(4):1091–1102, apr 2001.
- [90] Stephen J. PERKINS. Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *European Journal of Biochemistry*, 157(1):169–180, may 1986.
- [91] David Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, sep 2005.
- [92] Jean-baptiste Rouget, Tural Aksel, Julien Roche, Jean-Louis Saldana, Angel E. Garcia, Doug Barrick, and Catherine A. Royer. Size and Sequence and the Volume Change of Protein Folding. *Journal of the American Chemical Society*, 133(15):6020–6027, apr 2011.
- [93] Pranav P. Pandharipande and George I. Makhatadze. Applications of pressure perturbation calorimetry to study factors contributing to the volume changes upon protein unfolding. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(5):1036–1042, 2016.
- [94] Julien Roche, Jose A Caro, Douglas R Norberto, Philippe Barthe, Christian Roumestand, Jamie L Schlessman, Angel E. Garcia, B. Garcia-Moreno E., and Catherine A. Royer. Cavities determine the pressure unfolding of proteins. *Proceedings of the National Academy of Sciences*, 109(18):6945–6950, may 2012.
- [95] Julien Roche, Mariano Dellarole, Jose A José a Caro, Douglas R Norberto, Angel E Garcia, Bertrand Garcia-Moreno, Christian Roumestand, Catherine Ann Royer, Bertrand Garcia-Moreno E, Bertrand Garcia-moreno E, Christian Roumestand, Catherine Ann Royer, and Bertrand Garcia-moreno E. Effect of Internal Cavities on Folding Rates and Routes Revealed by Real-Time Pressure-Jump NMR Spectroscopy. *Journal of the American Chemical Society*, 135(39):14610–14618, aug 2013.
- [96] Julien Roche, Jose A. Caro, Mariano Dellarole, Ewelina Guca, Catherine A. Royer, Bertrand García-Moreno E., Angel E. Garcia, and Christian Roumestand. Structural, energetic, and dynamic responses of the native state ensemble of staphylococcal nuclease to cavity-creating mutations. *Proteins: Structure, Function and Bioinformatics*, 81(6):1069–1080, 2013.
- [97] Vladimir P. Voloshin, Nikolai N. Medvedev, Nikolai Smolin, Alfons Geiger, and Roland Winter. Exploring volume, compressibility and hydration changes of folded proteins upon compression. *Phys. Chem. Chem. Phys.*, c, 2015.
- [98] Vladimir P. Voloshin, Alexandra V. Kim, Nikolai N. Medvedev, Roland Winter, and Alfons Geiger. Calculation of the volumetric characteristics of biomacromolecules in solution by the Voronoi-Delaunay technique. *Biophysical Chemistry*, 192:1–9, 2014.

- [99] Lung-nan Lin, John F Michael F Michael Brandts, Valerian Plotnikov, John F Michael F Michael Brandts, and Valerian Plotnikov. Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Analytical biochemistry*, 302(1):144–60, mar 2002.
- [100] Catherine Royer and Roland Winter. Protein hydration and volumetric properties. *Current Opinion in Colloid & Interface Science*, 16(6):568–571, dec 2011.
- [101] Gunda Panick, Gediminas J A Vidugiris, Ralf Malessa, G Rapp, Roland Winter, and Catherine A Royer. Exploring the Temperature-Pressure Phase Diagram of Staphylococcal Nuclease. *Biochemistry*, 38(13):4157–4164, mar 1999.
- [102] R. Ravindra and Roland Winter. Pressure Perturbation Calorimetric Studies of the Solvation Properties and the Thermal Unfolding of Proteins in Solution. *Zeitschrift für Physikalische Chemie*, 217(10-2003):1221–1244, oct 2003.
- [103] Lally Mitra, Jean-Baptiste Rouget, Bertrand Garcia-Moreno, Catherine A Royer, and Roland Winter. Towards a Quantitative Understanding of Protein Hydration and Volumetric Properties. *ChemPhysChem*, 9(18):2715–2721, dec 2008.
- [104] Vladimir P Voloshin, Nikolai N Medvedev, Nikolai Smolin, Alfons Geiger, and Roland Winter. Disentangling Volumetric and Hydrational Properties of Proteins. *The Journal of Physical Chemistry B*, 119(5):1881–1890, feb 2015.
- [105] Tigran V. Chalikian. On the Molecular Origins of Volumetric Data. *The Journal of Physical Chemistry B*, 112(3):911–917, jan 2008.
- [106] Mariano Dellarole, Jose Alfredo. Caro, Julien Roche, Martin Fossat, Philippe Barthe, Catherine Ann Royer, Christian Roumestand, Bertrand García-Moreno E., Catherine Ann Royer, and Christian Roumestand. Evolutionarily Conserved Pattern of Interactions in a Protein Revealed by Local Thermal Expansion Properties. *Journal of the American Chemical Society*, 137(29):9354–9362, jul 2015.
- [107] Alekos D Tsamaloukas, Neena K Pyzocha, and George I Makhatadze. Pressure Perturbation Calorimetry of Unfolded Proteins. *The Journal of Physical Chemistry B*, 114(49):16166–16170, dec 2010.
- [108] C.N Pace. [14]Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in enzymology*, 131:266–280, 1986.
- [109] G I Makhatadze. Thermodynamics of protein interactions with urea and guanidinium hydrochloride. *Journal of Physical Chemistry B*, 103(23):4781–4785, 1999.
- [110] C N Pace and K L Shaw. Linear extrapolation method of analyzing solvent denaturation curves. *Proteins*, Suppl 4(February):1–7, 2000.
- [111] A.K. Soper, E.W. Castner, and Alenka Luzar. Impact of urea on water structure: a clue to its properties as a denaturant? *Biophysical Chemistry*, 105(2-3):649–666, sep 2003.
- [112] Joseph D. Batchelor, Alina Olteanu, Ashutosh Tripathy, and Gary J. Pielak. Impact of Protein Denaturants and Stabilizers on Water Structure. *Journal of the American Chemical Society*, 126(7):1958–1961, feb 2004.
- [113] P.-O. Astrand, A. Wallqvist, G. Karlstrom, and P. Linse. Properties of urea-water solvation calculated from a new ab initio polarizable intermolecular potential. *The Journal of Chemical Physics*, 95(11):8419, 1991.
- [114] Y L A Rezus and H J Bakker. Effect of urea on the structural dynamics of water. *Proceedings of the National Academy of Sciences*, 103(49):18417–18420, dec 2006.

- [115] Hironori Kokubo and B. Montgomery Pettitt. Preferential Solvation in Urea Solutions at Different Concentrations: Properties from Simulation Studies. *The Journal of Physical Chemistry B*, 111(19):5233–5242, may 2007.
- [116] Lan Hua, Ruhong Zhou, D Thirumalai, and B J Berne. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proceedings of the National Academy of Sciences*, 105(44):16928–16933, nov 2008.
- [117] Peter J Rossky. Protein denaturation by urea: Slash and bond. *Proceedings of the National Academy of Sciences*, 105(44):16825–16826, nov 2008.
- [118] Deepak R Canchi, Dietmar Paschek, and Angel E García. Equilibrium study of protein denaturation by urea. *Journal of the American Chemical Society*, 132(7):2338–2344, 2010.
- [119] Deepak R. Canchi and Angel E. García. Backbone and side-chain contributions in protein denaturation by urea. *Biophysical Journal*, 100(6):1526–1533, 2011.
- [120] Michael Hoppert and Frank Mayer. Prokaryotes: Even without membrane-bounded compartments, prokaryotes display a high degree of subcellular organization. *American Scientist*, 87(6):518–525, 2016.
- [121] R John Ellis. Macromolecular crowding : obvious but underappreciated. *Trends in biochemical sciences*, 26(10):597–604, 2001.
- [122] Austin E. Smith, Larry Z. Zhou, Annelise H. Gorensek, Michael Senske, and Gary J. Pielak. In-cell thermodynamics and a new role for protein surfaces. *Proceedings of the National Academy of Sciences*, page 201518620, 2016.
- [123] Raj Kumar. Role of naturally occurring osmolytes in protein folding and stability. *Archives of Biochemistry and Biophysics*, 491(1-2):1–6, 2009.
- [124] Deepak R Canchi and Angel E García. Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem*, 64:273–93, 2013.
- [125] D Wayne Bolen and George D Rose. Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding. *Annual Review of Biochemistry*, 77(1):339–362, jun 2008.
- [126] Michael Senske, Lisa Törk, Benjamin Born, Martina Havenith, Christian Herrmann, and Simon Ebbinghaus. Protein stabilization by macromolecular crowding through enthalpy rather than entropy. *Journal of the American Chemical Society*, 136(25):9036–9041, 2014.
- [127] Huan-Xiang Zhou, Germán Rivas, and Allen P Minton. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences \*. *Annual Review of Biophysics*, 37(1):375–397, jun 2008.
- [128] Amir Levy, David Andelman, and Henri Orland. Dielectric constant of ionic solutions: A field-theory approach. *Physical review letters*, 108(June):227801, 2012.
- [129] A. E. Eriksson, W. A. Baase, X.-J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, and B. W. Matthews. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, 255(5041):178–183, 1992.
- [130] David Shortle, Wesley E. Stites, and Alan K. Meeker. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry*, 29(35):8033–8041, 1990.
- [131] Thuy P Dao, Ananya Majumdar, and Doug Barrick. Capping motifs stabilize the leucine-rich repeat protein PP32 and rigidify adjacent repeats. *Protein Science*, 23(6):801–811, jun 2014.
- [132] S. A. Hawley. Reversible pressure–temperature denaturation of chymotrypsinogen. *Biochemistry*, 10(13):2436–2442, jun 1971.

- [133] Gunda Panick, Ralf Malessa, Roland Winter, Gert Rapp, Kelly J Frye, and Catherine A Royer. Structural characterization of the pressure-denatured state and unfolding/refolding kinetics of staphylococcal nuclease by synchrotron small-angle X-ray scattering and Fourier-transform infrared spectroscopy. *Journal of Molecular Biology*, 275(2):389–402, jan 1998.
- [134] Yu.V. Griko, P.L. Privalov, S.Yu. Venyaminov, and V.P. Kutysenko. Thermodynamic study of the apomyoglobin structure. *Journal of Molecular Biology*, 202(1):127–138, jul 1988.
- [135] Revanur Ravindra and Roland Winter. On the Temperature - Pressure Free-Energy Landscape of Proteins. pages 359–365, 2003.
- [136] László Smeller. Pressure-temperature phase diagrams of biomolecules. *Biochimica et biophysica acta*, 1595(1-2):11–29, mar 2002.
- [137] Gordon S. Rule and T. Kevin Hitchens. *Fundamentals of Protein NMR Spectroscopy*, volume 5 of *Focus on Structural Biology*. Springer-Verlag, Berlin/Heidelberg, 2006.
- [138] Teodor Parella. NMR Periodic Table, 2000.
- [139] Gerhard Binsch. Unified theory of exchange effects on nuclear magnetic resonance line shapes. *Journal of the American Chemical Society*, 91(6):1304–1309, mar 1969.
- [140] Christopher A. Waudby, Andres Ramos, Lisa D. Cabrita, and John Christodoulou. Two-Dimensional NMR Lineshape Analysis. *Scientific Reports*, 6(April):24826, apr 2016.
- [141] J L Markley, a Bax, Y Arata, C W Hilbers, R Kaptein, B D Sykes, P E Wright, and K Wüthrich. Recommendations for the presentation of NMR structures of proteins and nucleic acids—IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *European journal of biochemistry / FEBS*, 256(1):1–15, 1998.
- [142] Hua Li, Hiroaki Yamada, and Kazuyuki Akasaka. Effect of Pressure on Individual Hydrogen Bonds in Proteins. Basic Pancreatic Trypsin Inhibitor. *Biochemistry*, 37(5):1167–1173, feb 1998.
- [143] Ryo Kitahara, Kazumi Hata, Hua Li, Mike P. Williamson, and Kazuyuki Akasaka. Pressure-induced chemical shifts as probes for conformational fluctuations in proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 71:35–58, may 2013.
- [144] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, sep 1997.
- [145] Erik Lindahl. Introduction to Molecular Dynamics with GROMACS Molecular Modeling Course 2007 Simulation of Lysozyme in Water. 2007.
- [146] Dirk Matthes and Bert L. De Groot. Secondary structure propensities in peptide folding simulations: A systematic comparison of molecular mechanics interaction schemes. *Biophysical Journal*, 97(2):599–608, 2009.
- [147] Hans W. Horn, William C. Swope, Jed W. Pitera, Jeffrey D. Madura, Thomas J. Dick, Greg L. Hura, and Teresa Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics*, 120(20):9665, may 2004.
- [148] E. R. Hernandez, Luis Manuel Montano Zetina, Gabino Torres Vega, Miguel Garcia Rocha, Luis F. Rojas Ochoa, and Ricardo Lopez Fernandez. Molecular Dynamics: from basic techniques to applications (A Molecular Dynamics Primer). In *AIP Conference Proceedings*, pages 95–123. AIP, 2008.
- [149] H J C Berendsen, J P M Postma, W. F. van Gunsteren, A. DiNola, and J R Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, oct 1984.

- [150] Shuichi Nose. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511, 1984.
- [151] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.
- [152] Denis J. Evans and Brad Lee Holian. The Nose-Hoover thermostat. *Journal of Chemical Physics*, 83(8):4069–4074, 1985.
- [153] Berk Hess, David van der Spoel, and Erik Lindahl. GROMACS User manual Version 4.6.3.
- [154] Berk Hess, David van der Spoel, and Erik Lindahl. Gromacs User Manual 4.6.7. In *SpringerReference*. Springer-Verlag, Berlin/Heidelberg, 2014.
- [155] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, nov 1999.
- [156] Ulrich H.E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281(1-3):140–150, dec 1997.
- [157] Ulrich H.E. Hansmann and Yuko Okamoto. New Monte Carlo algorithms for protein folding. *Current Opinion in Structural Biology*, 9(2):177–183, apr 1999.
- [158] Daniel Sindhikara, Yilin Meng, and Adrian E. Roitberg. Exchange frequency in replica exchange molecular dynamics. *Journal of Chemical Physics*, 128(2), 2008.
- [159] Tsuneyasu Okabe, Masaaki Kawata, Yuko Okamoto, and Masuhiro Mikami. Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. *Chemical Physics Letters*, 335(5-6):435–439, mar 2001.
- [160] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926, 1983.
- [161] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, nov 2006.
- [162] Ryan Day, Dietmar Paschek, and Angel E. Garcia. Microsecond simulations of the folding/ unfolding thermodynamics of the Trp-cage miniprotein. *Proteins: Structure, Function and Bioinformatics*, 78(8):1889–1899, jun 2010.
- [163] Bosco K Ho and Franz Gruswitz. HOLLOW: Generating Accurate Representations of Channel and Interior Surfaces in Molecular Structures. *BMC Structural Biology*, 8(1):49, 2008.
- [164] Jeffrey K. Noel, Paul C. Whitford, and José N. Onuchic. The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *Journal of Physical Chemistry B*, 116(29):8692–8702, 2012.
- [165] Jeffrey K. Noel, Mariana Levi, Mohit Raghunathan, Heiko Lammert, Ryan L. Hayes, José N. Onuchic, and Paul C. Whitford. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Computational Biology*, 12(3):1–14, 2016.
- [166] Piotr Rotkiewicz and Jeffrey Skolnick. Fast procedure for reconstruction of full-atom protein models from reduced representations. *Journal of Computational Chemistry*, 29(9):1460–1465, jul 2008.
- [167] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, dec 1983.
- [168] LLC MicroCal. VP-DSC MicroCalorimeter User’s Manual.

- [169] Lally Mitra, Nikolai Smolin, Revanur Ravindra, Roland Winter, and Catherine Royer. Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution - experiments and theoretical interpretation. *Physical chemistry chemical physics : PCCP*, 8(11):1249–65, mar 2006.
- [170] LLC MicroCal. Pressure Perturbation Calorimetry ( PPC ) DSC Application Note Solvation Effects.
- [171] Trevor Huyton and Cynthia Wolberger. The crystal structure of the tumor suppressor protein pp32 (Anp32a): Structural insights into Anp32 family of proteins. *Protein Science*, 16(7):1308–1315, jul 2007.
- [172] Nicola J Baxter and Michael P Williamson. Temperature dependence of  $^1\text{H}$  chemical shifts in proteins. *Journal of Biomolecular NMR*, 9(4):359–369, 1997.
- [173] Carlos Simmerling, Bentley Strockbine, and Adrian E. Roitberg. All-atom structure prediction and folding simulations of a stable protein. *Journal of the American Chemical Society*, 124(38):11258–11259, 2002.
- [174] Deepak R. Canchi, Pruthvi Jayasimha, Donald C. Rao, George I. Makhataдзе, and Angel E. Garcia. Molecular Mechanism for the Preferential Exclusion of Osmolytes from Protein Surfaces. *Biophysical Journal*, 104(2):189a, jan 2013.
- [175] Charles A. English and Angel E. García. Charged Termini on the Trp-Cage Roughen the Folding Energy Landscape. *The Journal of Physical Chemistry B*, 119(25):7874–7881, jun 2015.
- [176] D Paschek, S Hempel, and a E Garcia. Computing the stability diagram of the Trp-cage miniprotein. *Proceedings of the National Academy of Sciences*, 105(46):17754–17759, nov 2008.
- [177] Charles A English and Angel E García. Folding and unfolding thermodynamics of the TC10b Trp-cage miniprotein. *Physical Chemistry Chemical Physics*, 16(7):2748, feb 2014.
- [178] Kazuyuki Akasaka. Probing Conformational Fluctuation of Proteins by Pressure Perturbation. *Chemical Reviews*, 106(5):1814–1835, may 2006.
- [179] David A. Case. Chemical shifts in biomolecules. *Current Opinion in Structural Biology*, 23(2):172–179, 2013.
- [180] Gerhard Wagner, Arthur Pardi, and Kurt Wuethrich. Hydrogen bond length and proton NMR chemical shifts in proteins. *Journal of the American Chemical Society*, 105(18):5948–5949, 1983.
- [181] Gregorio Weber. *Protein interactions*. Chapman and Hall, New York, 1992.
- [182] William Robert Smith and Ronald William Missen. *Chemical reaction equilibrium analysis: theory and algorithms*. John Wiley and Sons, 1982.

## Résumé en français

Ce travail de thèse se concentre sur l'étude des protéines par l'usage de haute pression. Les articles présentés ici sont précédés d'une introduction présentant les différents modèles physiques utilisés pour décrire le repliement des protéines, une introduction posant les bases de la thermodynamique, ainsi que décrivant l'origine de la stabilité thermodynamique des protéines dans leur état plié. Il y a trois sujets principaux abordés dans ce mémoire. Le premier est l'étude de la coopérativité du repliement et du paysage de repliement de la protéine à répétition PP32 (Anp32a) à travers l'utilisation de la pression à différentes températures. La seconde étude concerne l'investigation de l'origine de l'expansivité thermique des protéines pliées grâce à l'utilisation de RMN haute pression et de la protéine très bien caractérisée Staphylococcal Nuclease (SNase) et certaines de ses mutantes. Finalement, un dernier article sur la stabilité sous pression de la variante TC5b de la mini protéine modèle tryptophan-cage grâce à une combinaison de RMN et de simulations moléculaires tout-atomes en « replica exchange ».

## Résumé en anglais

This thesis work focuses on the study of proteins through the use of high pressure. There are three main subjects that are being inquired here. The first is the study of folding cooperativity and folding landscape of a repeat protein (Anp32a) through the use of high pressure denaturation at different temperatures. The second concerns the investigation of the determinant of thermal expansivity in the folded state of protein using high pressure NMR, and the well characterized Staphylococcal Nuclease (SNase) and some of its mutants. Finally, a last article on the pressure stability of the model mini protein Tryptophan cage variant Tc5b by a combination of high pressure NMR and full atomic replica exchange simulations.