Study of protein folding within a chaperonin

Elisa Colas Debled

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

Elisa COLAS DEBLED

Thèse dirigée par **Jérôme BOISBOUVIER**

préparée au sein du **Laboratoire Institut de Biologie Structurale** dans **l'École Doctorale Chimie et Sciences du Vivant**

Etude du repliement des protéines au sein d'une chaperonine

—

Study of protein folding within a Chaperonin

Thèse soutenue le **2 avril 2019**,
devant le jury composé de :

Dr. Cécile Breyton
Directrice de Recherche, CNRS, Présidente du jury

Pr. Bruno Kieffer
Professeur, Université de Strasbourg, Rapporteur

Dr. Robert Schneider
Maître de Conférences, Université de Lille, Rapporteur

Dr. Wolfgang Hoyer
Professeur, Université de Düsseldorf, Examinateur

Dr. Emeric Miclet
Maître de Conférences, Université Pierre et Marie Curie, Examinateur

Dr. Jérôme Boisbouvier
Directeur de Recherche, CNRS, Directeur de thèse

Acknowledgements

Je souhaite avant tout remercier ma famille et plus particulièrement, mes parents Caroline et Eric, mon compagnon Benjamin, mes grands-parents Gisèle, Josette et Henry, et mes poilus Basile, Courgette, Kouros et Juliet pour leur soutien sans faille(s) durant ces années de thèse, et sans lesquels ce manuscrit n'existerait pas.

At the IBS, thanks to : Jérôme Boisbouvier for his supervision, patience with my manuscript and for sharing his NMR knowledge with me ; Tomas Hosek, Katharina Weinhaupl, Hugo Fraga and Vilius Karauskas for unforgettable moments in the lab and their very helpful english manuscript corrections ; Rida Awad for her kindness, her outstanding biology skills and her help with the "hardcore" Thermosome purifications ; Audrey Hessel for showing me round the lab and teaching me her great biochemistry tricks along with many other things, including best spots to pick chestnuts and spy on marmots ; Ricarda Toerner for her help with the last amyloid experiments ; Jia-Ying Guan for good scientific exchanges and her help in the lab with MSG ; Guillaume Mas for introducing me to the Thermosome wonders and mysteries ; Isabel Ayala for her great packing of the lab's homemade "sucrOse" column for Thermosome purifications and for teaching me the secrets of protein production and successful methyl labelling ; Paul Schanda for the great NMR advices ; Pierre Gans for his supervision at the beginning of my thesis ; Cedric Laguri for trying to make me appreciate Linux ; Adrien Favier for his assistance during my numerous spectrometer crashes ; Pavel Macek, Faustine Henot, Ricarda Toerner, Guillaume Mas and Carsten Krichel for very delightful office moments ; and the whole NMR Group, indeed.

In Düsseldorf, thanks to : Lothar Gremer and Wolfgang Hoyer, who purified amyloid proteins for my experiments, kindly hosted me for a month to share their amyloid knowledge and introduced me to "the Aggregates" ; Dieter Willbold for great advice and scientific talk during my thesis committees and Tanya Kupreichyk for very nice moments in the lab, AFM measurements and her precious tricks on fibrillation experiments.

At the ISBG platforms, thanks to : Daphna Fenel and Catherine Moriscot from the EM platform and their infinite patience with my numerous invisible fibrils samples ; Jean Pierre Andrieu for the amino-acid analysis, Caroline Mas for the help with the SEC-MALLS and Luca Signor for the MS.

I am grateful to many other people from the IBS (or not), included in this non-exhaustive list : Pacome Armagnat, Rida Awad, Alister Burt, Doriane Costa, Florian Deruet, Hugo Fraga, Benoit Gachet, Diego Gauto, Audrey Hessel, Tomas Hosek, Nicolas Jean, Vilius Karauskas, Carsten Krichel, Pavel Macek, Caroline Mas, Guillaume Mas, Roberto Maya, Kevin Pounot, Paul Schanda, Orso Subrini, Claire Verrier and last but definitely not least Katharina Weinhaupl, for the many great moments of laugh and friendship in and out of the institute, at the bar after long lab days, skiing and enjoying life in the great city of Grenoble.

Abstract

Chaperonins are molecular machineries involved in the prevention of protein misfolding. These large macromolecules (approximately 1 MDa) are present in all domains of life and globally organized in two stacked rings on top of one another, hosting a cavity in their respective centers. By hydrolyzing ATP within their cavities, these rings can switch between two major structural states, an open and a closed conformation, to trap and refold misfolded proteins. Among the different types of molecular chaperones, chaperonins are of particular interest because their mechanism of action is not yet totally understood.

This thesis focused on the study of **PhCPN**, the Chaperonin from Pyrococcus horikoshii, and its interaction with substrate proteins by various biochemical and biophysical techniques including NMR. In fact, NMR spectroscopy is a powerful tool to probe transient interactions in solution, at atomic resolution. Especially, specific isotope labeling of methyl groups is a technique of choice to study huge protein assemblies such as PhCPN chaperonin because they overcome the liquid-state NMR size limitation. To study the protein folding within the cavities of PhCPN, two different model substrate of various sizes and biological functions were selected. Particularly, one of these substrates (Malate Synthase G / **MSG**) forms amorphous aggregates when submitted to heat while the other (**Amylin**) is able to self-associate into amyloid fibrils. During this thesis, I have demonstrated that the Chaperonin PhCPN can prevent the aggregation of the chosen substrates.

In fact, the PhCPN Chaperonin is able to irreversibly bind thermally unfolded MSG in a 1/1 ratio. The MSG/PhCPN complex was isolated and characterized. Especially, the interaction surface between PhCPN and this large substrate protein was investigated using a combination of NMR and EM.

In addition, the inhibition of the Amylin fibrillation by the Chaperonin was investigated using NMR and ThT fluorescence assays. It was shown that the Chaperonin delays the

fibrils formation, no matter its oligomeric state. The role of the Chaperonin on the Amylin nucleation and fibril elongation mechanisms was investigated.

De plus, l'inhibition de la formation de fibres amyloïdes issues de l'Amyline par la Chaperonine a été étudiée par RMN et fluorescence de la ThT. Il a été notamment montré que la Chaperonine retarde l'apparition des fibres amyloïdes, quelque soit l'état oligomerique de PhCPN.

Keywords : Chaperonins, NMR, Amylin, IAPP, Amyloid fibrillation, MSG, Methyl labelling.

Abstract

Les chaperonines sont des machines moléculaires impliquées dans la protection des protéines contre le mauvais repliement et l'agrégation. Ces macromolécules de taille importante (environ 1 MDa) sont présentes dans tous les domaines du vivant et sont organisées en deux anneaux concentriques et empilés l'un sur l'autre, possèdent chacun une cavité en leur centre. Les chaperonines sont particulièrement intéressantes car peu caractérisées par rapport aux autres chaperones, notamment dû à leur grande taille et à leur complexité intrinsèque. Leur mécanisme d'action reste donc assez flou.

Ce travail de thèse est centré sur l'étude de **PhCPN**, la chaperonine de Pyrococcus horikoshii, et son interaction avec différentes protéines substrats, grâce à une combinaison d'outils biochimiques et biophysiques tels que la RMN. En effet, la spectroscopie RMN est un outil particulièrement adapté à l'étude des interactions moléculaires transitoires à l'échelle atomique. L'utilisation dans ce cadre du marquage isotopique spécifique des groupements méthyles permet d'étudier des ensembles protéiques de taille importante tels que PhCPN, tandis que la RMN plus classique reste limitée à des poids moléculaires inférieurs à 30 kDa. Afin d'étudier le repliement des protéines à l'intérieur des cavités de PhCPN, deux protéines substrats de taille hétérogène et d'activité différentes ont été sélectionnées. En particulier, l'un de ces deux substrats (la Malate Synthase G ou **MSG**), forme des agrégats amorphes lorsqu'elle elle chauffée tandis que la seconde (l'**Amyline**) est capable de s'auto associer de manière plus organisée, créant des fibres amyloïdes de haut poids moléculaire. J'ai observé lors de cette étude que PhCPN est capable d'empêcher l'agrégation de ces deux substrats.

En effet, la Chaperonine PhCPN est capable de se lier de manière irreversible à la proteine MSG, dépliée par une augmentation de la temperature, dans un ratio stoechiométrique 1/1. Le complex MSG/PhCPN a été isolé et characterisé. En particulier, la

surface d'interaction entre PhCPN et cette large protéine substrat a été déterminée grâce à la RMN et la mciroscopie électronique.

De plus, l'inhibition de la formation de fibres amyloïdes issues de l'Amyline par la Chaperonine a été étudiée par RMN et fluorescence de la ThT. Il a été notamment montré que la Chaperonine retarde l'apparition des fibres amyloïdes, quelque soit l'état oligomerique de PhCPN. Le rôle de la Chaperonine sur les méchanismes de nucléation et d'élongation des fibres amyloïdes de l'Amylin a également été étudié.

Mots clés : Chaperonines, RMN, Amyline, IAPP, Fibrillation amyloïde, MSG, Marquage méthyle.

Table of contents

List of figures **xiii**

1 Introduction **1**

1.1 Proteins : History and characteristics 2

1.1.1 Discovery 2

1.1.2 Insights on protein production within cells 2

1.1.3 Structure 3

1.2 Protein folding mechanisms 6

1.2.1 Historical perspective and Levinthal's paradox 6

1.2.2 Folding models 7

1.2.3 Limits of the proposed models 11

1.3 Protein misfolding 12

1.3.1 Causes 12

1.3.2 Protein misfolding can lead to aggregation 13

1.3.3 Compared aggregation mechanisms 16

1.4 Protein rescue systems in cells : The chaperones 17

1.4.1 History 17

1.4.2 Common features 19

1.4.3 Chaperoning mechanism 20

1.4.4 Chaperone networks in cells 21

2 NMR studies of chaperones in interaction with client proteins **25**

2.1 Nascent proteins 26

2.1.1 Context 26

2.1.2 Nascent chains 26

2.1.3 Co-translational folding of nascent chains . . . 29
2.1.4 Chaperones and nascent chains . . . 30
2.1.5 Ribosomal chaperones and newly synthesized proteins . . . 33
2.2 Ubiquitous cellular chaperones . . . 34
2.2.1 Foldase chaperones and their substrates . . . 34
2.2.2 Holdase chaperones and their substrates . . . 47
2.2.3 Chaperone-like proteins involved in transport . . . 55
2.2.4 Substrate aggregation and disaggregation . . . 57
2.3 Focus on chaperonins . . . 60
2.3.1 General refolding mechanism . . . 61
2.3.2 Type I Chaperonin . . . 64
2.3.3 Type II Chaperonin . . . 67
2.3.4 Thermosome from *Pyroccocus horikoshii* . . . 70
2.3.5 Type II Co-Chaperonin : Prefoldin . . . 73
2.3.6 NMR studies of Chaperonins and their substrates . . . 74

3 Oligomeric state of PhCPN 79

4 Amorphous aggregating protein model 85
4.1 Selection of the protein model . . . 86
4.1.1 Ubiquitin . . . 88
4.1.2 rsFolder . . . 89
4.1.3 Malate Synthase G . . . 90
4.1.4 MSG, the selected model protein . . . 91
4.2 Characterization of the interaction . . . 92
4.2.1 Chaperonin prevents aggregation and precipitation . . . 92
4.2.2 Identification and isolation of a stable complex . . . 102
4.2.3 Stoichiometry of the complex . . . 104
4.2.4 Substrate location . . . 109
4.2.5 Refolding . . . 114
4.3 Chapter summary and conclusion . . . 118

5 Amyloidogenic protein model 121
5.1 Suitable protein model . . . 122

5.1.1 α-Synuclein . . . 123
5.1.2 Amyloid-β 1-42 . . . 125
5.1.3 Amylin . . . 128
5.1.4 Amylin, the selected model substrate . . . 129
5.2 Characterization of the interaction . . . 131
5.2.1 Optimization of the experimental conditions . . . 132
5.2.2 Interaction of Amylin with PhCPN . . . 135
5.2.3 Aggregation kinetics . . . 143
5.2.4 Characterization of the fibrils . . . 153
5.3 Chapter summary and discussion . . . 158
5.4 Perspectives . . . 161
5.4.1 Remaining questions . . . 161
5.4.2 Future of the project . . . 163

6 Materials and Methods 165
6.1 Culture media and buffers . . . 166
6.2 Protein expression and purification . . . 167
6.3 NMR methods and application to chaperones and their substrates . . . 171
6.3.1 NMR to study proteins and their interactions . . . 172
6.3.2 Specific labelling to study large assemblies . . . 182
6.4 NMR experimental parameters . . . 185
6.5 Additional biophysical and biochemical techniques . . . 187
6.5.1 Monitoring amorphous and amyloid aggregation . . . 187
6.5.2 Imaging techniques . . . 190
6.5.3 Other techniques . . . 191
6.6 Images . . . 194
6.7 Physical and chemical theoretical parameters . . . 195

References 197

Supplementary figures 219
A.1 Protein parameters . . . 219
A.1.1 *Pyrococcus horikoshii* strain OT-3 Thermosome . . . 219
A.1.2 *Escherichia coli* Malate Synthase G . . . 220

A.1.3 *Homo sapiens* Amylin, or IAPP 222
A.2 Science Advances article and Supplementary figures 222

List of figures

1.1 Protein production and structural elements 4
1.2 Different folding models for unfolded proteins 9
1.3 Comparison of the amorphous aggregation and the fibrillation mechanisms 18
1.4 Chaperones networks in different organisms 23

2.1 Ribosome-associated nascent chains 27
2.2 TF binding models 32
2.3 The Chaperone HSP70 35
2.4 Interaction between HSP70 and various substrates 39
2.5 Functional cycle of HSP90 41
2.6 Interaction between HSP90, co-chaperones and substrates 44
2.7 Interaction between αBC and different substrates 50
2.8 Schematic illustrations of the two types of chaperonins 61
2.9 Schematic illustrations of the different models proposed to explain the refolding mechanisms occurring in the Anfinsen cage of chaperonins 63
2.10 The type I Chaperonin 65
2.11 The type II Chaperonin 68
2.12 The Thermosome from *Pyrococcus horikoshii*, PhCPN 71
2.13 Dimensions of PhCPN and its Methionine and Valine NMR assignments . . 72
2.14 Structure of the Prefoldin and its binding to type II Chaperonin 73

3.1 PhCPN dimensions analyzed by EM 80
3.2 Oligomeric state of PhCPN in different buffers 82

4.1 First interaction trials with Ubiquitin 88
4.2 Unfolding tests on rsFolder 89

4.3 NMR Spectrum, structure and aggregation of the *E. coli* MSG protein 91
4.4 Design of experiments to evaluate the effect of the buffer salt content on MSG aggregation 93
4.5 Measurement of the extinction coefficient of MSG and PhCPN 95
4.6 MSG aggregation in presence of PhCPN 96
4.7 Prevention of thermal aggregation of MSG by PhCPN monitored by NMR . . 98
4.8 MSG aggregation kinetics in presence of PhCPN monitored by NMR 100
4.9 Quantification of the amount of MSG in the precipitated pellet by resuspension in urea 101
4.10 Identification of an MSG/PhCPN complex by DOSY NMR 103
4.11 Isolation of a MSG/PhCPN complex by gel filtration 105
4.12 Coarse grain quantification of the MSG/PhCPN ratio in the MSG/PhCPN complex. 106
4.13 SEC-MALLS analysis of PhCPN, MSG and MSG/PhCPN complex 107
4.14 Amino-acid analysis of the MSG/PhCPN complex composition 108
4.15 EM pictures of the MSG/PhCPN complex 110
4.16 PhCPN 2D methyl SOFAST HMQC spectra in the presence or absence of MSG112
4.17 Spin labelling of MSG lysines by OXYL-1-NHS 113
4.18 PRE-NMR to map PhCPN binding site 115
4.19 Influence of the presence of 1 mM ATP on MSG after heating, in presence of PhCPN 116
4.20 Open and closed conformations of PhCPN during ATP hydrolysis, in the absence or presence of substrate proteins inserted within its cavities 117
4.21 Scheme describing the interaction between PhCPN and MSG at different temperatures 119

5.1 Structure and interaction of αSyn with GroEL and PhCPN Chaperonins . . . 124
5.2 Study of Aβ42 127
5.3 Diabetes mechanism and first Amylin interaction test 130
5.4 Elimination of Amylin degradation in presence of PhCPN 133
5.5 NMR spectra of Aymlin in MES and PhCPN buffers 134
5.6 NMR titration of Amylin in presence of different amounts of hexadecameric PhCPN at 10°C, in PhCPN buffer pH 6.5 136

5.7 NMR titration of Amylin in presence of different amounts of monomeric PhCPN at 10°C, in MES buffer . . . 137
5.8 Interaction tests with Amylin-CC . . . 139
5.9 Formulas to calculate the K_D of the Amylin/PhCPN interaction . . . 141
5.10 Amylin aggregation monitored by NMR . . . 143
5.11 Optimization of the fibrillation experiments . . . 146
5.12 Fibrillation curves processing . . . 149
5.13 Influence of different PhCPN ratios on Amylin fibrillation followed by ThT fluorescence . . . 151
5.14 AFM pictures of Amylin in MES buffer, in the absence or presence of PhCPN 154
5.15 Negative staining EM picture of Amylin fibrils in absence of PhCPN, in MES Buffer . . . 156
5.16 Negative staining EM pictures of Amylin in the presence or absence of PhCPN, in different buffers . . . 157
5.17 Preparation of the PRE experiments . . . 162
5.18 Monitoring the interaction of Amylin with PhPFD by NMR . . . 164

6.1 Purification of PhCPN . . . 168
6.2 Chemical shift perturbations . . . 173
6.3 Hydrogen-deuterium exchange . . . 175
6.4 The CMPG NMR experiment . . . 177
6.5 Saturation transfer difference principle . . . 178
6.6 PRE NMR . . . 180
6.7 The DOSY NMR experiment . . . 181
6.8 Specific methyl labelling for NMR . . . 184
6.9 The principle of light scattering . . . 187
6.10 Scheme of the ThT molecule and the ThT fluorescence assay . . . 189
6.11 SDS-PAGE gels composition . . . 191
6.12 The principle of amino acid analysis . . . 193
6.13 Hydrophobicity scale of Kovacs . . . 195
6.14 Homology model of PhCPN . . . 196

Chapter 1

Introduction

Les protéines sont des macromolécules essentielles à la vie, composées d'acides aminés. Chaque protéine assure une fonction particulière grâce à ses caractéristiques biochimiques et structurales. En effet, la séquence des protéines définit indirectement leur structure tridimensionnelle. Les protéines sont synthétisées linéairement par les Ribosomes et doivent se replier sur elles-mêmes afin d'acquérir leur conformation finale. Cependant, la structure ou la séquence d'une protéine peuvent être altérées par des mutations ou des perturbations extérieures. Ces modifications peuvent conduire au mauvais repliement des protéines et à leur agrégation. Ce phénomène d'agrégation peut prendre plusieurs formes. Les protéines mal repliées peuvent tout d'abord se regrouper entre elles de manière totalement désordonnée et former des amas amorphes, souvent toxiques pour les cellules. Elles peuvent également s'associer de façon beaucoup plus organisée et créer des oligomères puis des fibres amyloïdes. Certains de ces oligomères sont notamment impliqués dans des maladies neurodégénératives telles que la maladie d'Alzheimer et de Parkinson. Les processus de repliement et d'agrégation protéiques sont décrits dans le présent chapitre, ainsi que les éléments de théorie y étant rattachés.

1.1 Proteins : History and characteristics

1.1.1 Discovery

Proteins are macromolecules essential to life, composed of combinations of 20 different amino acids, encoded in the DNA sequence. Their existence was discovered in the early 19^{th} century owing to a series of studies on silk, serum and egg Albumin, and other natural compounds carried out by G. J. Mulder [Mulder, 1839]. They were termed by J. J. Berzelius in 1838, according to the Greek adjective $\pi\rho\omega\tau\epsilon\iota\sigma\sigma$ (proteios), meaning "primary", according to the belief that proteins were the main source of animal nutrition [Vickery, 1950]. Interestingly, the amino acids (often abbreviated **AA**), constituents of the proteins, were first isolated before the discovery of proteins. In 1806, Asparagine was extracted from asparagus, which makes it the first known amino acid. Later in 1902, Emil Fischer and Franz Hofmeister proved by acid hydrolysis that proteins can be broken down into individual AA.

1.1.2 Insights on protein production within cells

As explained before, proteins are macromolecules composed of a succession of amino acids. Protein size depends on the number of their constitutive amino acids, from a few AA (peptides) to massive globular proteins like the human Titin, the largest known protein (different isoforms, up to more than 33000 AA) [Freiburg et al., 2000]. The sequence of proteins is well defined and transcribed from the DNA sequence. The genetical information is organised in genes, coded with combinations of three nucleotides or codons, each combination coding for a particular AA [Jou et al., 1972]. Protein expression occurs in several steps (**figure 1.1, A**). The first phase of this mechanism is called transcription and takes place in the nucleus in eukaryotes and the cytosol of prokaryotic cells. The genetic information is read from DNA and transcribed into messenger RNA (**mRNA**) by the RNA polymerase and then moves to the cytoplasm through nuclear pores. In parallel, free amino acids are loaded on the amino acid specific transfer RNAs (**tRNA**), each one of them possessing an anticodon complementary with the codons of mRNA. Translation, the second step of proteins genesis, takes place in the cytoplasm. mRNAs are read by the Ribosomes, which are specialised molecular machineries composed of proteins and RNA. These Ribosomes assemble amino acid by amino acid, matching the codons from

mRNA with the corresponding anticodons of tRNA, loaded with corresponding amino acids. Once associated, a peptide bond is created between the amino acids, creating the peptide chain, leading step-by-step to the final protein. Once the protein synthesis is completed and the translation stopped by a stop codon encoded in the mRNA sequence, it is released from the Ribosome.

Post-translational modifications can occur anytime during the lifetime of proteins. They can consist of the modification or covalent addition of groups on specific AA of proteins (phosphorylation, glycosylation, etc.) and are essential for the regulation of the protein activity (activation, modification and inhibition of the activity). For eukaryotic cells, they mostly occur in the Endoplasmic reticulum [Ellgaard et al., 2016] or the Golgi apparatus after translation. Some of these modifications also occur in prokaryotes, even if less frequent and less characterised [Grangeasse et al., 2015].

1.1.3 Structure

As the protein exits from the Ribosome, the AA linked to each other by peptide bonds start to interact, depending on their hydrophobicity, charges and neighbouring AA. To reach a state of lowest energy, these amino acid chains fold. A consequent part of the proteome can spontaneously fold while other proteins, usually the larger ones [Finkelstein et al., 2013], need the assistance of chaperones to reach their native state (see **Section 1.4** and **Chap. 2**). The structure of the folded state depends on the sequence of the protein and its environment. Several types of native folding occur in nature and range from completely folders to disordered proteins [Gianni and Jemth, 2016]. These structures can be described at four different microscopic levels.

The primary structure is the protein sequence (**figure 1.1, B**). By convention, the first amino acid of the sequence is the one with a free amino terminus, the N-terminal often abbreviated **Nter**. Similarly, the last amino acid is the one exhibiting a free carboxyl terminus, the C-terminal also called **Cter**.

The secondary structure describes the local folding of the polypeptide backbone chain (**figure 1.1, C**). These structures emerge from a combination of steric constraints due to the side-chains of the AA and hydrogen bonds between the amino groups and carbonyls

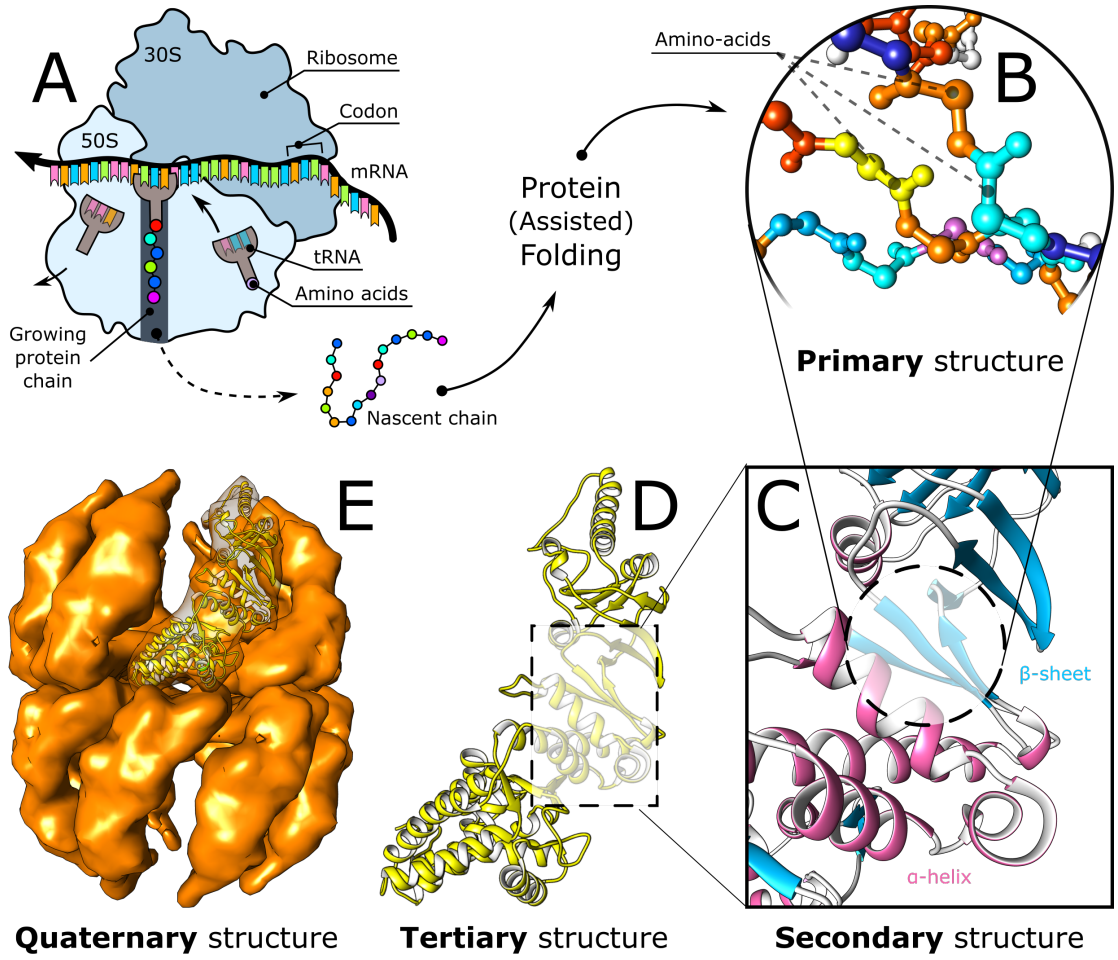


Fig. 1.1 : Protein production and structural elements. (**A**) Scheme of the protein production by the bacterial Ribosome. The protein chain is produced by successive addition of single amino acids to the native chain. The protein can start folding during translation or at the end of translation, with assistance from chaperones or not. (**B**) Scheme of the protein primary structure. (**C**) Scheme of the protein secondary structure. α-helices are depicted in pink, β-sheets in blue. (**D**) Scheme of the protein tertiary structure. A monomer of the Thermosome from Pyrococcus horikoshii, also named **PhCPN** in this thesis. (**E**) Quaternary structure of the full Thermosome 16-mer. (**B**), (**C**), (**D**) & (**E**) were modeled from the homology model of the Thermosome from Pyrococcus horikoshii, see **Chap. 6.7**.

from the backbone. The most frequently occurring secondary structure elements are β-sheets, α-helices and turns.

The tertiary structure corresponds to the global folding of the protein in space, its three-dimensional shape (**figure 1.1, D**). The correct tertiary folding is essential for the protein to ensure its role in the cells. This structure is maintained by a complex interaction network, such as disulphide bridges (covalent interactions), hydrogen bonds (electrostatic interactions), van der Walls interactions and interactions with the environment of the protein. The latest interactions play an essential role in the folding of proteins. Hydrophobic interactions are highly involved in the formation of hydrophobic patches in the proteins, to protect their hydrophobic residues from the solvent. It is particularly the case for membrane proteins, which need an amphipathic environment (lipid bilayer) similar to the cell membrane to fold correctly. Some proteins can be different in sequence but adopt the same tertiary structure and exhibit very similar functions. It is yet impossible to predict the tertiary structure of a protein directly from its AA sequence. The case of intrinsically disordered proteins (**IDP**) needs to be mentioned, as some proteins or regions of proteins actually do not exhibit a particular tertiary structure. For instance, approximately 75% of the mammals signaling proteins are predicted to contain disordered regions of more than 30 AA, and about 25% of the total protein pool is supposedly completely disordered [Dunker et al., 2008].

Finally, the quaternary structure gives insight into the oligomeric state of the protein and its eventual interacting partners if tightly assembled to form a complex (**figure 1.1, E**). Examples would be the Thermosome (an homo-hexadecamer chaperonin) and the Prefoldin (an heterohexameric molecular chaperone), which will be further described in the following chapters of this manuscript, **Chap. 2.3.3** and **Chap. 2.3.5** respectively.

A tight link exists between the structure, the sequence and the activity of proteins. The pioneering experiment of F. Richards in 1958 consisted in cutting an RNAse enzyme in two pieces [Richards, 1958]. Taken separately, the activity of these RNase pieces was found to be of around 5% whereas the original activity was fully recovered when both parts of the enzyme were mixed. From that discovery, it was possible to demonstrate that the activity of a protein is dependent on its amino acid sequence and structure. Another similar study was conducted by C.B. Anfinsen [Anfinsen et al., 1961, Haber and Anfinsen, 1962]. First,

the disulphide bonds of the Ribonuclease protein were reduced to denature the protein. A complete recovery of native properties was observed after progressive re-oxidation of the denatured protein, suggesting that the native protein folding was directly controlled by the amino acid sequence of the protein. In the second set of experiments, the arrangement of the disulphide bonds was chemically modified, and this perturbation changed the overall structure and activity of the protein. Full activity was also recovered after progressive re-oxidation of the protein to retrieve the native disulphide bonds location. These two sets of experiments proved that the native arrangement of the amino acid chain of Ribonuclease was the most thermodynamically stable one and that the activity of the protein was directly linked to its structure.

1.2 Protein folding mechanisms

Proteins start to fold during the latest steps of synthesis and after being released into the cells. Protein folding has been studied extensively and remains a matter of intense research. Up to date, there is still no consensus on the way proteins fold and especially on the folding mechanism. Thus, it is impossible to perfectly predict the structure of a protein solely from its amino acid sequence. In this chapter, some theories upon folding mechanisms will be considered.

1.2.1 Historical perspective and Levinthal's paradox

The first studies on protein folding mechanisms were carried out in the early 30s, 30 years before the emergence of the assumption that structure and protein sequence are linked, summarized in the citation "All the information necessary to achieve the native conformation of a protein in a given environment is contained in its amino acid sequence" [Anfinsen et al., 1961]. Interestingly, they were focused on denaturing the protein, before starting to speculate on the actual folding mechanisms. Some protein denaturations are reversible and denaturation corresponds to a transition from a native state to a more disorganized state of the proteins. These pioneering experiments were carried out and summarized by Wu [Wu, 1931], then gave rise to various theories on protein structure and folding that are still valid nowadays. His observations lead to the assumption that proteins are able to explore many conformations with similar structures, otherwise proteins could

not be specific in their activity. Their particular structures are caused by intramolecular interactions and the presence of water molecules, which stabilize their overall structure.

In 1959, Kauzmann added that denaturation seems to affect all proteins in a similar manner and that the denatured state could be characterized as a "random coil" state, where the protein is unfolded and can adopt a wide variety of random conformations [Kauzmann, 1959]. In his theory, denaturation arises from the rupture of intramolecular bonds and interactions, such as hydrogen bonds, disulphide bridges and hydrophobic forces. These hydrophobic interactions are of paramount importance for the proper folding of proteins. Their role was highlighted by the monitoring of large differences in hydrogen-deuterium exchange rates on the exchangeable protons of the amino acids of the same protein [Kauzmann, 1959]. It was then clear that these hydrophobic amino acids were inaccessible to water and buried in the protein's structure in a micelle-like manner, forming hydrophobic clusters inside the proteins.

From the kinetic point of view, the Levinthal paradox explicited in 1969 by Cyrus Levinthal [Zwanzig et al., 1992] is a good explanation of the problem raised by protein folding. To illustrate this paradox, one can consider a protein made of 101 amino acids which needs to fold from an unfolded state to its native state. The amino acids of this protein are linked together by 100 peptide bonds. If we consider that each bond is able to adopt a minimum of 3 different conformations then the unfolded protein could exhibit a minimum of 3^{100} different conformations. By calculating the time such a protein would take in order to properly fold, the paradox is raised. In fact, even if the protein would be able to adopt 10^{23} conformations per second, it would take an eternity (10^{27} years) to try them all. The protein folding would thus be impossible, whereas some unfolded protein can refold by themselves in less than a second. From this paradox came the conclusion that a protein does not adopt all random possible conformations prior to reaching the lowest energy conformation. This problem motivated the development of multiple protein folding theories, which will now be summarized, with focus on the most prevalent ones.

1.2.2 Folding models

Levinthal summed up the problem raised by protein folding in 1969 but he was not the first person to address this question. As mentioned earlier, folding theories started to emerge way before. In 1931, Wu already started to propose theories about protein structure and

folding, by making analogies with a simple organic carbon chain adopting many configurations [Wu, 1931]. He elaborated that such a comparison could not explain completely protein's behaviour since proteins have to explore a limited amount of conformations as some of them need to be specific. It was already a good start to define and explore protein folding.

The main problem in the study of protein folding is that protein folding is fast. Generally, one can only easily observe either the unfolded state of a protein or its folded state. It is thus difficult to gain information on the folding mechanism in these conditions, except when one can observe folding intermediates. It was of course not yet the case when the first folding theories emerged and started to flourish in the 70's. Herein will be described a variety of theories which try to reflect the processes involved in protein folding and overcome Levinthal's paradox.

Nucleation - Condensation

The nucleation-condensation model is one of the first "post Levinthal" models, conceived by Wetlaufer [Wetlaufer, 1973]. The protein is considered as an amino acid unfolded chain and local nucleation is needed to start protein folding (**figure 1.2, A**). This rapid self-assembly occurs in one or several small regions of about 8 amino acids, creating nuclei. Once formed, these structured regions of the protein serve as bases for rapid propagation of the structural organisation. This theory was later on adapted and reviewed by Fersht to fit with his experimental results [Fersht, 1997]. By using protein mutagenesis, he was able to trap chymotrypsin inhibitor-2 in a transient unfolded state containing folded regions the size of a 15 residues α-helix. He then assumed that these nucleation regions were much larger than previously thought and that they correspond to local folds in the transient states prior to native protein folding. This model evolved to a vision of protein folding quite close to the next presented protein folding model.

Hydrophobic collapse

The hydrophobic collapse model emerged from the work of Levitt & Warshel [Levitt and Warshel, 1975], as they tried to address how distant parts of the proteins can come together and achieve the final protein fold. They then conducted one of the first computer simulations of a protein, using coarse-grain molecular dynamics simulations because

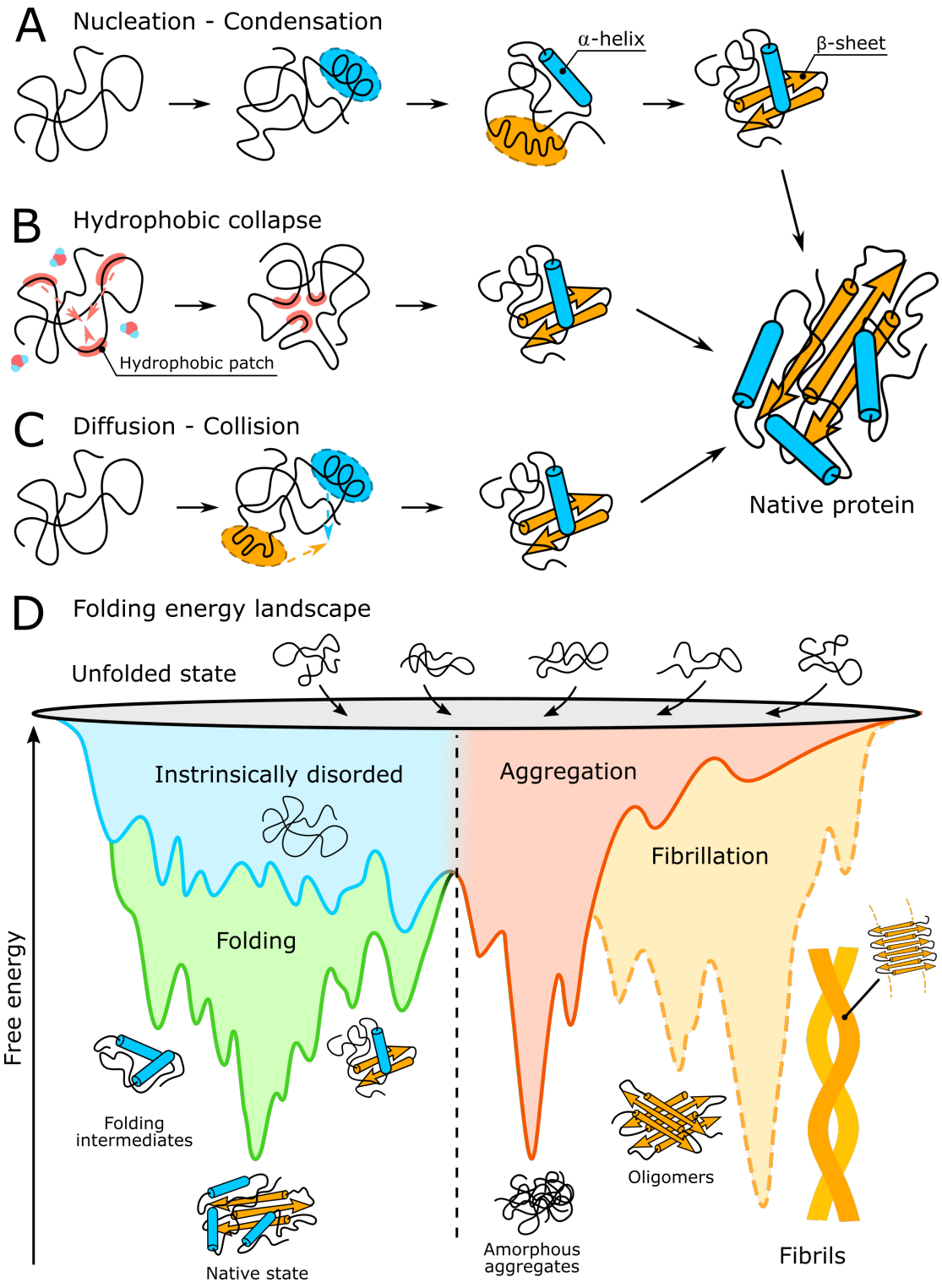


Fig. 1.2 : Different folding models for unfolded proteins. (**A**) Nucleation-condensation folding model. (**B**) Hydrophobic collapse folding model. (**C**) Diffusion-collision folding model. (**D**) Folding energy landscapes folding model, modified from [Jahn and Radford, 2005].

of the low computational power of early machines. By reducing degrees of freedom for the different amino acids, keeping the interactions with the solvent and by considering the different hydrophobicity of each amino acid, they managed to compute a relatively close model of a protein of known conformation (bovine pancreatic trypsin inhibitor). Their explanation of folding is that the collapse of the protein structure is driven by hydrophobic forces that lead to a compact protein shape with atoms already close to their native position. In a second step, short-range interactions form the final native protein conformation of lower entropy (**figure 1.2, B**).

Diffusion - Collision

The diffusion-collision model arises from the work of Karplus & Weaver. They elaborated their first model in 1976, which was already quite similar to the "modern" version of the nucleation-condensation model. The diffusion-collision model is different in the way that proteins are divided in small domains which are free to explore multiple sets of conformations rapidly [Karplus and Weaver, 1976], in order to overcome the Levinthal's paradox. Several of them need to collide, in order to coalesce and create higher order structures (**figure 1.2, C**). Both colliding entities need to be in their native state in order to lead to proper native protein folding. Incorrect folding has been observed and can be explained by the diffusion-collision model, caused by two entities colliding together with an initial wrong fold. This coalescence mechanism is believed to occur step-wise, otherwise the folding time calculated with their model would be too long to fit reality.

Folding energy landscapes

All these studies were complemented with experimental data and evolved over the years. Another approach to tackle the problem emerged in the early 90s owing to Bryngelson & Wolynes, who explained the general behaviour of protein through statistics [Bryngelson et al., 1995]. They described protein conformations and hydrophobic forces arising from contacts with the solvent in terms of Gibbs free energy (G), as they kept temperature and pressure constant in their model. G is low when hydrophobic residues form a cluster and high when they are apart and solvent exposed. They also specified that, to them, the Levinthal's paradox does not reflect reality. In fact, in their model, some conformations are more likely to occur than others, because of a favorable lower free energy. Once trapped

in these low energy conformations, it is more and more difficult to go backwards in terms of higher free energy as the energy needed to get out of these local minima would be too high. Proteins are then trapped in a "funnel-shaped energy landscape", where their free energy decreases as they go down the funnel, to finally reach their native state, of lowest energy (**figure 1.2, D**). However, in the case of disordered proteins, the local energy minima are close in terms of free energy. Thus, the proteins are able to switch from one conformation to another because the thermic energy given to the protein is in the same order of magnitude as the energy gap between the different conformations. Although this model gives a clear explanation of the folding pathways, it does not decipher the folding mechanism itself. It seems that through the years, it has been clear that the folding mechanisms are probably a mixture of different proposed folding models explained before, depending on the protein sequence, length, buffer conditions and many other variables.

1.2.3 Limits of the proposed models

It is difficult to conclude anything from the preceding theories, except maybe that protein folding is a really complex mechanism, which is still not yet quite understood, especially for large globular proteins [Gianni and Jemth, 2016]. Another problem with these models is that they consider a disordered, yet complete protein chain. The reality is somewhat different because the nascent proteins chains can start folding during their production by the Ribosome, as the Nter is already protruding out of the Ribosome pore and in contact with the solvent before the protein is completely expressed (see **Chap. 2.1.1**). The behaviour of these newly synthesized is of particular interest, as the protein sequence is not yet complete and can probably explore conformations impossible to reach with the full-length protein. In addition, the synthesized protein is also in contact with many other proteins before it starts to fold, such as molecular chaperones and particularly holdases to avoid instant folding of the nascent protein chains, which will be further described in the following chapters (see **Chap. 2.2.2**). In fact, *in vivo*, the concentration in Crystallins in the human eye and Hemoglobin of human red blood cells alone are respectively about 400 and 350 mg/mL [Delaye and Tardieu, 1983, Ellis, 2001], whereas *in vitro* studies rarely go above protein concentrations of 10 mg/mL (except for special biophysical techniques such as NMR). Some proteins also do not start to fold right after expression but they are transported in their unfolded state to other regions of the cell where they can finally

fold and achieve their role [Rapoport, 2007]. All these proteins at the same place and at the same moment can cause problems. A defective cell can start to produce mutated proteins, whose folding pathways are then modified, leading to misfolding. A misfolded protein present in a crowded environment has more chances to aggregate. The misfolding mechanisms are further developed in the next section.

1.3 Protein misfolding

Molecular chaperones are diverse and complementary, working in a complex protein network with a goal : to protect proteins from misfolding. In this section, the causes and mechanisms of misfolding, leading to various types of aggregation will be presented.

1.3.1 Causes

Folding, as previously explained, is based on a complex network of intramolecular interactions. Aggregation, on the other hand, relies on intermolecular interactions. The native state of a protein is not always corresponding to its free energy minimum. In fact, protein aggregates are extremely stable and can be even more stable than the corresponding native protein, as pictured in the previously described folding energy landscape (**figure 1.2, D**). For this reason, misfolded proteins can be kinetically trapped in these local low energy funnels, accumulate, start interacting with each other and finally aggregate. Protein homeostasis and aggregate clearance is ensured by the cellular protein quality control systems, such as chaperone networks including chaperone disaggregases and protein degradation machineries. When exposed to long-lasting stresses or aging, these systems can be overwhelmed by the presence of misfolded proteins, aggregates or mutations, leading to the accumulation of misfolded proteins. In addition, in human cells, about 30% of the newly synthesized proteins are directly processed by the proteasomal degradation pathway, which is due to a combination of protein misfolding and translational errors [Schubert et al., 2000]. Under such conditions, impairment of the degradation pathway or accumulation of additional misfolded proteins due to stresses can seriously perturb cellular homeostasis and lead to unwanted aggregation.

Aggregated and misfolded proteins cannot fulfill their cellular functions and moreover, the presence of the aggregates in cells is most of the time toxic [Gregersen et al., 2006].

The toxicity of these aggregates depends on the initial role of the misfolded protein, its cellular localization, the environmental conditions, and its native interaction partners. A low toxicity level results in an immediate response of the cell and elimination of the misfolded proteins through the degradation pathways. When the cell response is not fast enough to clear completely the presence of these aggregates, the remaining misfolded proteins can start to cause damages, and in most severe cases cause cell death.

1.3.2 Protein misfolding can lead to aggregation

Protein aggregation is a term which depicts two distinct mechanisms : the formation of disordered amorphous aggregates and the creation of ordered amyloid fibrils [Khan et al., 2018, Yoshimura et al., 2012]. Both types of aggregates are linked to various diseases, even if amyloid fibrils are more widely studied, as they are involved in several debilitating and deadly neurological diseases and other severe pathologies. Nevertheless, all proteins are able to form amorphous aggregates while a limited amount of them can also associate in amyloid fibrils [Chandel et al., 2018].

Amorphous aggregates

Amorphous aggregates are involved in various diseases associated with aging, such as cataract, Fanconi syndrome, light-chain deposition disease and primary systemic amyloidosis, reviewed in [Khan et al., 2018]. They also appear during in vitro protein expression and purification and are a major issue for industrial protein production, as aggregation is highly concentration dependent [Philo and Arakawa, 2009].

Structure - Amorphous aggregation is caused by non-specific and non-predictable agglomeration of misfolded proteins exposing their hydrophobic patches, which should normally be buried inside their native structure [Khan et al., 2018]. They differ in size and show no structural specificities.

Amyloid fibrils

Amyloid fibrils are protein fibers made of numerous amounts of amyloidogenic proteins monomers stacked together in an ordered fashion. They were first discovered in the mid-19th century in plants and mistaken for starch, *amylum* in Latin. Numerous autopsies

identified a similar substance in vital human organs such as spleen and liver and in 1959, the fibrils were identified as such [Kyle, 2001].

Pathogenicity - Amyloidogenic protein aggregation is linked to multiple debilitating and deadly diseases, referred to as amyloidosis. In amyloidosis cases, intra or extracellular deposits of aggregated amyloid proteins, referred to as amyloid plaques, accumulate in various organs [Ferreira et al., 2007]. In humans, more than 25 different amyloidogenic proteins are involved in an extended set of diseases including Alzheimer's disease or AD (Amyloid-β/**A**β protein, brain), type II diabetes mellitus or T2D (Amylin/**IAPP** protein, pancreas), Parkinson's disease or PD (α-Synuclein/α**Syn**, brain), infectious bovine spongiform encephalopathy (Prion protein/PrP, brain), Huntington's disease (Huntingtin, brain) and dialysis-related amyloidosis (β2-microglobulin, joint spaces), reviewed in [Eisenberg and Jucker, 2012].

However, the mechanism of these diseases and especially the mechanism of amyloid toxicity is not completely understood and still up to date subject for debates. In fact, amyloid fibrils were shown to be toxic for various types of cell cultures but the presence of the amyloid plaques was not directly linked to pathologies [Ferreira et al., 2007].

In the case of AD, for instance, memory loss is not caused by the presence of amyloid fibrils only and in T2D, amyloid plaques can be found in the pancreas of mice with normal glycemia. For these reasons, fibrils are not believed to be the most dangerous form of amyloid proteins and the interest has raised about soluble amyloid oligomers, which were proven to exhibit acute toxicity (reviewed in [Kayed and Lasagna-Reeves, 2012]). Actually, oligomers of Amyloid-β involved in AD induce oxidative stress, phosphorylation of the Tau protein also found in plaques, perturbs the communication between neurons and disrupt cellular calcium homeostasis by forming pores in the membranes, due to their detergent-like properties [Reiss et al., 2018]. Elevated levels of soluble oligomers are suspected to be responsible for the toxicity of Aβ, especially species ranging from dimers, trimers to 12-mers, 24-mers and 56-mers. Similar observations are made in the case of other amyloidoses. Another current direction of study is focused on investigating the transmission of the amyloidosis in a prion-like manner from cells to cells [del Río et al., 2018]. Such mechanism was only shown for the Prion protein but is highly suspected for other amyloids as they can act as "seeds", some can spread through blood or migrate in the brain when injected locally and induce disease [Eisenberg and Jucker, 2012]. A completely

opposite point of view on the subject also emerges, based on several observations of Aβ. It was found to be an antimicrobial peptide and to play a role in immune function, therefore raising doubts about its involvement in AD [Gosztyla et al., 2018]. An increase in infections during therapies reducing or cancelling cellular amounts of Aβ as well as an increase in Aβ expression during some bacterial or viral infections are also supporting that idea.

In every case, the mentioned amyloid proteins are found in healthy organs but in a monomeric form, where they ensure many different and yet not fully understood roles [Ferreira et al., 2007]. Functional amyloid aggregates or fibrils are also involved in the normal human physiology and can be found in red blood cells, or in teeth [Gosztyla et al., 2018]. They are also found in other eukaryotes and bacteria, where they ensure diverse functions such as the formation of biofilms or immune response [Eisenberg and Jucker, 2012].

Structure - The amyloidogenic monomers are often IDP, as for Aβ, Amylin and αSyn but can also show transient secondary structure, as in the case of the partly structured prion protein [Sabareesan et al., 2018]. The fibril propensity of these monomers seems to be mediated by the charge of the AA chains, their glycine and proline content and their length. Once associated into oligomers, they start to exhibit a characteristic structure made of beta-sheets perpendicular to the growing fibril axis. Each β-strand is tightly linked to the strand above and below, forming a compact and highly stable structure, the fibrils. Amyloid fibrils are of variable length, unbranched, insoluble and highly resistant to proteolytic digestion. They show an apple-green birefringence under polarized light, after staining with congo red dye and a cross-β structure based on its characteristic X-ray diffraction pattern forming an X. Fibrils usually contain two or more cross-β layers, attached and twisted together owing to side chain-side chain interactions [Tycko, 2014]. These layers are made of monomers organized to form parallel or antiparallel β-sheets perpendicular to the fibril axis, which amino acid identity can differ or not from one strand to the other [Toyama and Weissman, 2011]. Each fibril possesses a certain width and twist period, conserved along the fibril. Moreover, if short filaments of 20 nm to 50 nm are used as seeds to produce new fibrils, the characteristics of the seeds will be conserved in the new growing fibril. Cross-seeding between different amyloids has also been observed, as well as the interaction between oligomers and fibrillating proteins of other sorts [del Río et al., 2018].

Amyloid fibrils can exhibit different morphologies, called fibril polymorphs, even if made from the same original amyloidogenic protein [Eisenberg and Jucker, 2012]. These polymorph structures depend on growth conditions and the early organisation of the amyloid protein oligomers. They are more likely to occur in short amyloid peptides such as Aβ or Amylin[Tycko, 2014]. Different fibril morphotypes can be found in the brain of one AD patient to another.

1.3.3 Compared aggregation mechanisms

The formation of fibrils and amorphous aggregates are competing events, at least for proteins able to exhibit amyloidogenic properties. They exhibit very heterogeneous aggregations mechanisms, amorphous aggregation and fibrillation. Fibrillation can be compared to crystal formation, with fibrils formation following a nucleated-growth. Oppositely, the association of misfolded proteins forming the amorphous aggregates is completely disorganized and be compared to glass transition [Hall et al., 2015].

Amorphous aggregation - The formation rate of this type of aggregates shows short or nonexistent lag phase, not modified by agitation nor seed addition. This rate is high, as it depends on the available interaction surface with already formed aggregates and the available misfolded proteins. The association of misfolded proteins to the aggregates is thus completely random and leads towards final precipitation [Khan et al., 2018].

Fibrillation - As mentioned before, fibrillation is similar to the formation of crystals, organized in a compact and repetitive network of interactions. This aggregation mechanism can be broken down into two major steps. A first slow and unfavorable step consists of the formation of a fibrillation nucleus, as a result of interactions between misfolded proteins or IDP. Non-proteic particles can also serve as nuclei, such as impurities or roughness and adherence to the surfaces of the protein contenant [Philo and Arakawa, 2009]. Following a similar mechanism, this first step can be skipped by the addition of seeds. The second step is way faster and includes the growth of the fibrils from the fibrils nucleus upon monomer addition on the fibrils end, eventual breakage of the fibrils and other adverse events [Eisenberg and Jucker, 2012]. All these steps can be described by mathematical models [Hall et al., 2015]. Nevertheless, the concentration of amyloidogenic protein needs

to reach a critical concentration, otherwise no fibrillation can happen, in a similar fashion to the formation of detergent micelles above the critical micelle concentration.

From amorphous aggregation to fibrillation and *vice versa* - Some proteins can either form fibrils or amorphous aggregates, both associated with diseases. This is particularly the case for β2-microglobulin, which can form both types under controlled conditions and allows to compare the aggregation kinetics of amorphous aggregates and fibril formation [Yoshimura et al., 2012]. As expected, the rate of amorphous aggregates formation is higher since the available surface for the interaction with new misfolded proteins is larger, compared to the only two ends of each amyloid fibril [Khan et al., 2018]. Increasing salt concentrations at a given pH switch the aggregation from fibril formation to amorphous aggregates appearance.

1.4 Protein rescue systems in cells : The chaperones

Molecular chaperones are proteins which interact, stabilize or help other proteins to acquire or restore their native conformation. Most of these chaperones are heat shock proteins (**HSP**) and their expression is triggered or increased by cell-stress in order to prevent aggregation or help with protein transport. Chaperones are also believed to play a central role in evolution, as they intervene in the folding or stabilization of proteins exhibiting mutations, promoting genetic diversity [Hartl et al., 2012, Hirtreiter et al., 2009, Rutherford and Lindquist, 1998]. In this section, the different chaperones classes are introduced, as well as their common characteristics. Individual chaperone types are more precisely described in **Chap. 2**, with focus on chaperonins (**Chap. 2.3**).

1.4.1 History

Chaperones were first mentioned in 1976 and defined as proteins "sharpening the specificity and increasing the stability" in order to allow other proteins to ensure their proper function and prevent misfolding [Fohlman et al., 1976]. Fohlman's work on venom allowed him to investigate the role of Taipoxin, a protein ternary complex made of 3 different subunits. By monitoring the toxicity of these subunits taken individually, he noticed that two were not harmful, whereas the third was deadly. He then conjectured that these

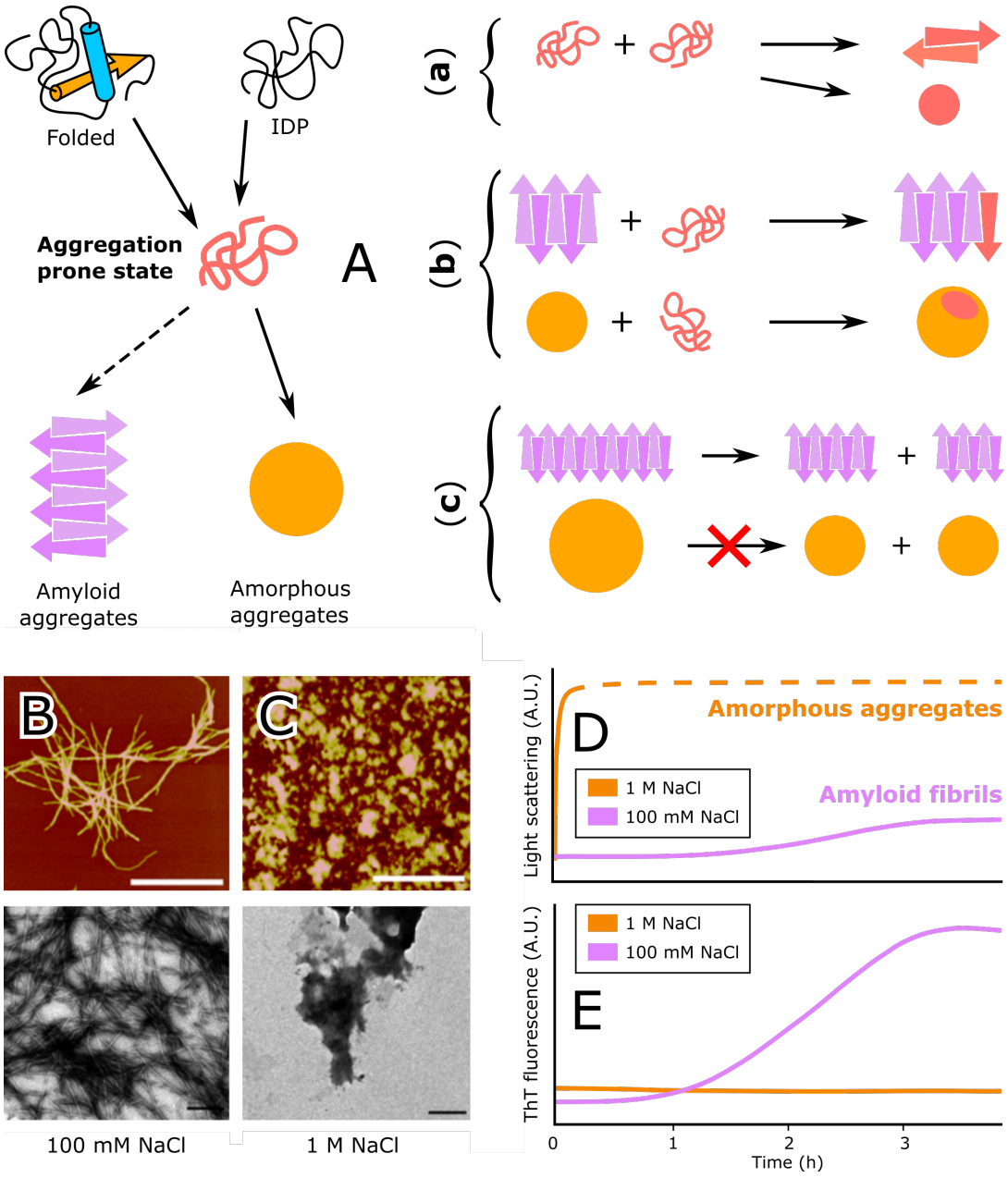


Fig. 1.3 : Comparison of the amorphous aggregation and the fibrillation mechanisms. (**A**) Stepwise progression of the fibrillation and the aggregation, starting from an aggregation-prone misfolded substrate. (**a**) Nucleation : the misfolded proteins interact together to form either amyloid seeds or small amorphous aggregates. (**b**) Elongation : the seeds grow into fibrils by stepwise addition of misfolded monomers, while the aggregates grow anarchically. (**c**) Secondary nucleation : the fibrils break to form small seeds and start a new fibril. Such a mechanism is not observed during amorphous aggregation. Adapted from [Hall et al., 2015]. (**B**) Fibrils of β_2-microglobulin observed by AFM (top) and EM (bottom), in low salt conditions. (**C**) Amorphous aggregates of β_2-microglobulin observed by AFM (top) and EM (bottom), in high salt conditions. Pictures from [Yoshimura et al., 2012]. (**D**) Schematic of the light scattering profile observed during aggregation of fibrillation. (**E**) Schematic of the ThT fluorescence scheme observed during fibrillation or aggregation.

proteins must have another role in the complex, thus the preliminary definition of chaperones. This role became clearer in 1978 with the studies of Laskey on histones, structural proteins which help the packing of DNA and thereby form nucleosomes [Laskey et al., 1978]. Interestingly, he noticed that when mixed together at physiological ionic strength, DNA and histones precipitate. Such an observation enlightened the fact that formation of nucleosomes needs a bit of help from an unknown assembly, which was later identified as a molecular chaperone protein. This chaperone was able to mask the charges of histones in order to prevent incorrect folding or non-specific interactions leading to precipitation of histones and DNA.

1.4.2 Common features

The main principle of chaperones action is to protect, host, unfold or refold misfolded or unfolded proteins with help from hydrophobic patches (rich in Met, Ala, Gly, Val, Leu, Thr & Ile) that can be found on the chaperones surface. In some cases, chaperones even co-associate in order to offer bigger hosting sites [Huang et al., 2016, Weinhäupl et al., 2018]. If the refolding is not completely successful, other chaperone networks can take over to either try other refolding rounds or directly address misfolded proteins to the degradation pathway, in order to get rid of a potential source of aggregation [Hartl et al., 2011]. Various chaperones are using ATP binding and hydrolysis as a primary source of energy to fuel vast structural rearrangements. The majority of such chaperones switch from an ATP bound to an ADP bound state and release a phosphate molecule [Mayer and Bukau, 2005]. The chaperones hold substrate proteins but also reduce their apparent concentration in the cells. This protection decreases the chances that misfolded proteins come in close contact and that their exposed hydrophobic regions associate together to lower their free energy. Chaperones are promiscuous and usually interact with a vast variety of substrates sharing common structural features that are recognized by the chaperones interaction regions [Fujiwara et al., 2010, Haslbeck and Vierling, 2015, Rudiger, 1997].

Many different types of chaperones, more than 200 different ones in the human organism, are necessary to ensure most physiological functions, with a great diversity of chaperoning mechanisms [Hartl et al., 2011]. They thus need a proper classification to differentiate them. The most well-studied chaperones are termed HSP. The modification of gene expression due to heat exposure was first uncovered in 1964, in Drosophila

melanogaster [Ritossa, 1964] and later associated with a change in protein production [Tissiéres et al., 1974]. The heat shock protein term is misleading, as HSPs are commonly expressed in healthy cells. Their cellular concentration can, however, increase from two to three times in response to various types of cellular stresses such as temperature variations, UV radiation, pH change, osmotic shock, infection, fever in eukaryotes [Pockley, 2003]. These perturbations yield translation errors, abnormal protein production or protein cleavage, leading to misfolding [Ananthan et al., 1986]. HSP chaperones act as a protection system for the cells, to refold or get rid of these undesirable proteins. The regulation mechanisms of the genes involved in the expression of HSP will not be described in this manuscript but are reviewed in [Beckmann et al., 1990]. HSP are named and classified according to their molecular weight (or the molecular weight of their monomers). Other chaperones are expressed in "normal" cells and participate in various cellular tasks such as nascent protein folding. They are termed constitutive chaperones or heat shock cognates (**HSC**). These constitutive chaperones are sometimes homologues of HSP, as explained in the next paragraph. In total, these chaperones correspond to 5 to 10% of the whole protein content of an organism submitted to normal conditions [Pockley, 2003].

1.4.3 Chaperoning mechanism

Chaperones protect most proteins from misfolding and aggregation in the cells. They also directly assist protein *de novo* folding during or after translation [Hoffmann et al., 2010]. Chaperones interact with unfolded or misfolded substrates through interactions with hydrophobic patches located on the chaperones surface. However, the whole chaperoning mechanisms vary with the chaperone classes and their ability to hydrolyse ATP or not [Saibil, 2013].

Holdases are chaperones that bind unfolded or misfolded substrates through their hydrophobic binding sites and protect them from aggregation (developed in **Chap. 2.2.2**). However, since their chaperoning mechanism is ATP-independent, they are unable to directly refold substrates and need assistance from other categories of chaperones.

Foldases chaperones are also protecting their protein substrates from aggregation. In addition, their ATPase activity provides them energy to switch between different conformations and actively promotes substrate folding or refolding (developed in **Chap. 2.2.1**).

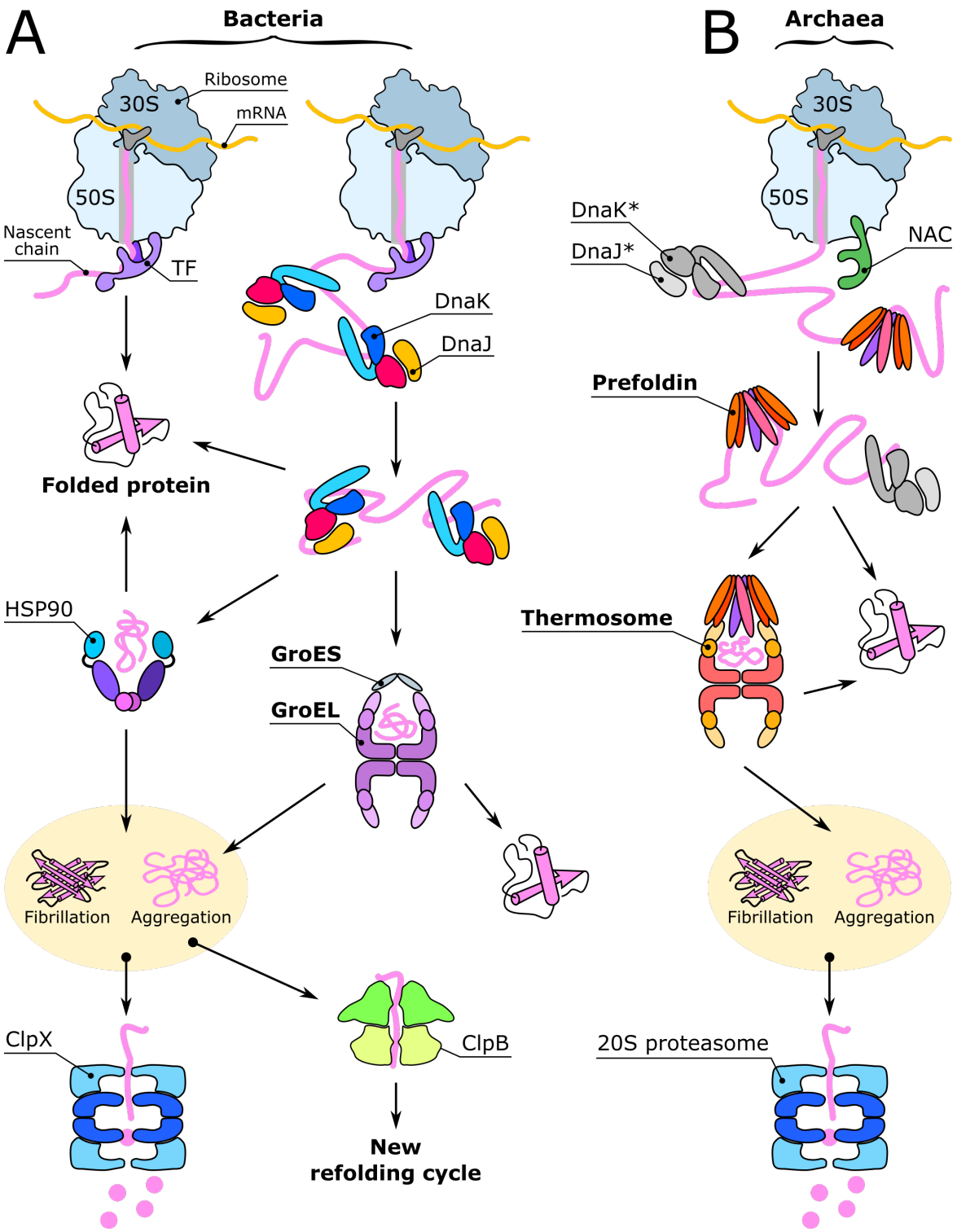
Some foldases also unfold their substrates, to destroy misfolded structures prior to refolding.

Disaggregases is a distinct class of chaperones, able to directly dismantle substrate aggregates into single unfolded proteins using ATP hydrolysis (developed in **Chap. 2.2.4**). Other chaperones can then take over and proceed to refolding. If chaperoning fails, the aggregated substrates can be directly addressed to the degradation machineries. Their role is also to disentangle protein aggregates prior to recycling.

1.4.4 Chaperone networks in cells

Some proteins do not need chaperones assistance for proper folding. Interestingly, chaperoning can also be insured by the nascent proteins themselves. Propeptides are intramolecular chaperones found in eukaryotes and prokaryotes. Their particularity is to be directly encoded in the sequence of the protein which needs chaperoning protection. Some proteins are expressed as preproproteins, the pre-region serving as a signal peptide to address the protein to the right cellular location, the propeptide catalysing the folding of their associated proteins [Shinde and Inouye, 2000]. Once proper folding is achieved, the propeptides are cleaved and degraded.

However, chaperone-assisted folding start during translation for many proteins . Ribosome-associated chaperones particularly help the initial co-translational folding steps of nascent proteins (**figure 1.4, A, B & C**). These chaperones can be localized near the exit tunnel of the Ribosome, ready to interact with synthesized proteins and prevent misfolding or aggregation. The best known and characterized Ribosome-associated chaperone is the trigger factor (**TF**) chaperone from the *E. coli* bacterium [Hoffmann et al., 2010] ; developed in **Chap. 2.1.4** and **Chap. 2.1.5**). The TF is the only Ribosome-associated chaperone found in bacteria, as well as chloroplasts, in excess compared to Ribosomes. Eukaryotes have two different, but similar Ribosome-associated machineries. The first system, found in yeast, is based on a Ribosome-associated complex (or RAC) composed of an HSP40 chaperone Zuo and an HSP70 chaperone Ssz, interacting with another HSP70 chaperone Ssb [Preissler and Deuerling, 2012]. A variant of this complex is also found in mammals but lacking an Ssb-like protein, which is replaced by cytosolic HSP70. The second system, highly conserved in the whole eukaryotic domain, is an heterodimeric nascent polypeptide-associated complex, or NAC.


A
Bacteria
Ribosome
30S
mRNA
50S
Nascent chain
TF
DnaK
DnaJ
Folded protein
HSP90
GroES
GroEL
Fibrillation
Aggregation
ClpX
ClpB
New refolding cycle
B
Archaea
30S
50S
DnaK*
DnaJ*
NAC
Prefoldin
Thermosome
Fibrillation
Aggregation
20S proteasome

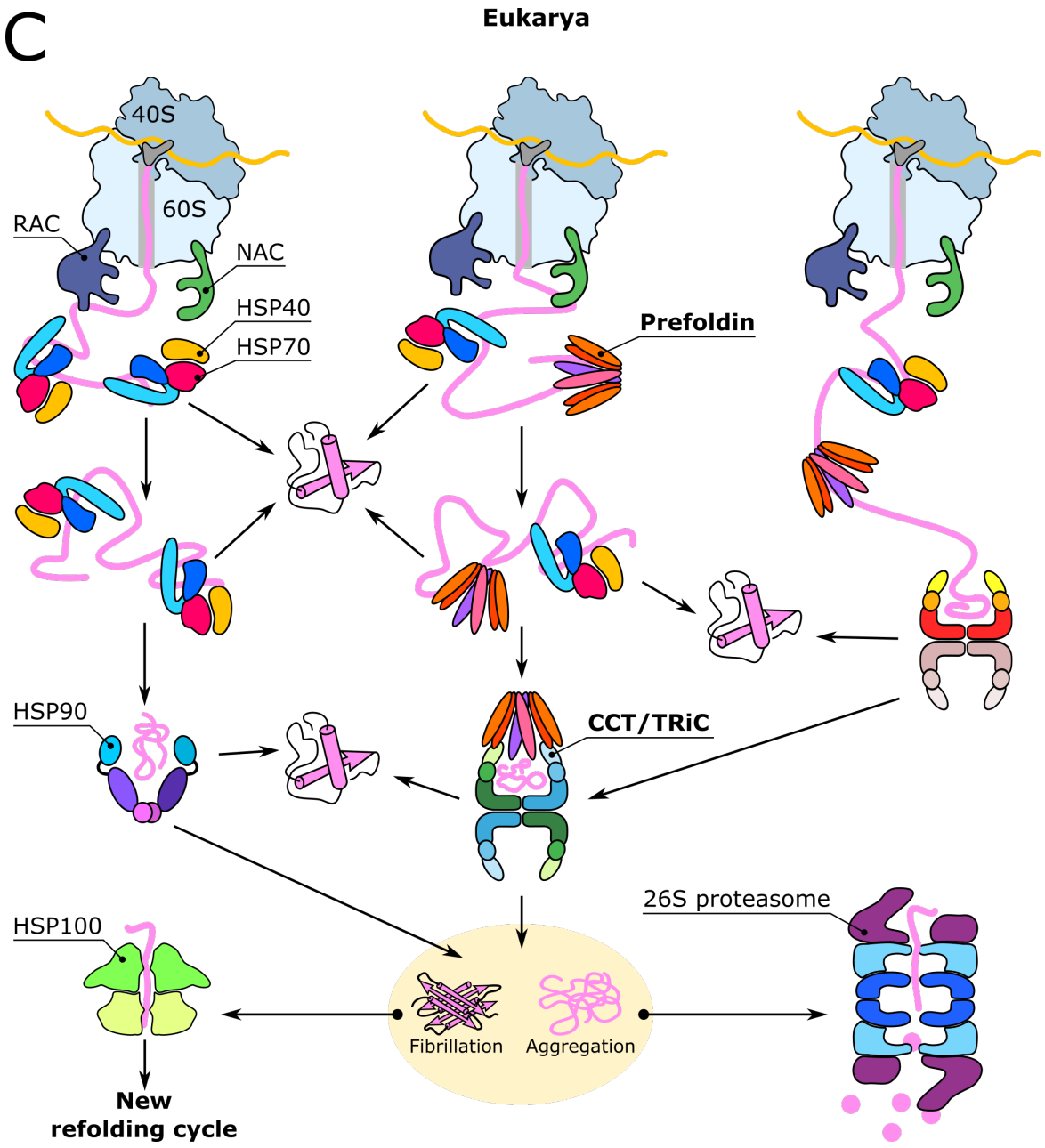


Fig. 1.4 : Chaperones networks in different organisms. (**A**) Main chaperones in bacteria. (**B**) Main chaperones in archaea. (*) The DnaK/DnaJ chaperone system does not exist in every archaeal organism. (**C**) Main chaperones in eukarya. The ATP/ADP and co-factors were not presented for clarity. They are precisely described in **Chap. 2**.

These Ribosome-associated chaperones also interplay directly with other chaperone machineries. Especially, they can deliver the substrate proteins for transport in targeted cellular compartments or assist protein folding. The main chaperone interacting with newly synthesized proteins is HSP70 and its co-chaperone HSP40. This ATPase chaperone system and its bacterial homologues DnaK/DnaJ are present in all the organisms (see **Chap. 2.2.1** for more developed characterization). Cellular chaperones are also specific to the cell organites or cellular compartments. For instance, the endoplasmic reticulum hosts a reticulum specific HSP70 chaperone named BiP. Other glucose-regulated chaperone proteins (Grp170/Grp94, homologs of HSP70) are also found in the endoplasmic reticulum under normal conditions and their presence is increased upon glucose starvation [Bracher and Verghese, 2015, Murphy, 2013]. These chaperones are involved in the folding and assembly of proteins or complexes in the ER, as well as the prevention of protein aggregation caused by stress. If HSP70-assisted protein folding fails, HSP60 chaperones, also entitled chaperonins, can take over. In fact, their chaperoning mechanism is very different from the one of HSP70, sequestering the substrates in an isolated cavity to provide an environment favorable to protein refolding (see **Chap. 2.3**). The chaperone HSP90 is also able to assist protein folding downstream to HSP70. In fact, this chaperone is ubiquitous and interacts with a very large set of substrates, through its modulable interaction site (see **Chap. 2.2.1**). As stated previously, chaperoning can be insufficient to prevent protein misfolding. Indesirable aggregates might then start to form within the cells, and need to be processed to prevent the development of diseases. Disaggregases and especially HSP100 (see **Chap. 2.2.4**) help disentangling these protein clumps into individual unfolded substrates, which can be later addressed to the previously cited chaperones for proper folding. Misfolded or aggregated proteins can also be directly brought to the proteolysis machineries for recycling and providing AA to synthesize new proteins.

Chaperones are ubiquitous and their mechanism of action very heterogeneous. A more detailed overview of the just mentioned chaperone types as well as their chaperoning mechanism is given in the next chapter.

Chapter 2

NMR studies of chaperones in interaction with client proteins

La plupart des protéines doivent se replier pour atteindre leur structure tridimensionnelle finale et pouvoir assurer leurs fonctions. Ce repliement peut être autonome ou assisté par d'autres machines protéiques, les chaperonnes. Les chaperonnes sont ubiquitaires et nécessaires au maintien des fonctions cellulaires dans tous les organismes. Elles permettent également de prévenir l'agrégation protéique en stabilisant les protéines mal repliées. Les grands groupes de chaperonnes et les chaperonnes les plus importantes sont décrits dans ce chapitre. Globalement, ce chapitre fait le point sur l'étude par résonance magnétique nucléaire (**RMN**) des chaperonnes et leur interactions avec des protéines clientes. Plus particulièrement, la chaperonne faisant l'objet de cette thèse est présentée, le Thermosome. Le Thermosome est une chaperonine dont les caractéristiques, le fonctionnement et les modes d'interactions avec ses substrats sont précisés ici.

2.1 Nascent proteins

In this chapter, a large overview of these different chaperones is given, and especially how they were studied by NMR to investigate their binding mechanism with unfolded proteins or IDP. This section will describe chaperones as they are encountered during a protein "lifecycle", from Ribosome-associated chaperones, to chaperones transporting substrate proteins and finally chaperones associated with degradation machineries.

2.1.1 Context

The Ribosome is a large cellular machinery found in both eukaryotes and prokaryotes, in charge of protein production. It is constituted of two distinct main subunits of different size, as well as a complex mixture of RNAs and additional ribosomal proteins. As a result of the tight cooperation between its two subunits, the Ribosome is able to read the genetic information encoded in mRNA, recruit tRNA loaded with amino acids needed for protein production in its peptidyl transferase center and finally create the peptide bonds in order to synthesize the proteins (**figure 2.1, A**). The nascent AA chain is then expelled through a 100 Å long exit tunnel, corresponding to the length of a growing polypeptide chain of about 30 amino acids [Eichmann et al., 2010]. This pore has a diameter of about 10 Å, enlarged at its end in order to provide a wider chamber for the nascent chains to explore different conformations before being released in the cellular environment (**figure 2.1, A**). The negatively charged surface of the Ribosome also acts as a protective environment [Deckert et al., 2016], in association with ribosomal chaperone proteins docked around, ready to take over and help the protein to properly fold and reach its active conformation [Wentink, 2014]. Different ways of modeling these nascent chains have been explored, yielding information on the different stages of protein production, from the nascent chains still trapped in the exit tunnel of the Ribosome, to protein chains in interaction with the Ribosome-associated chaperones, to be released in the cellular environment.

2.1.2 Nascent chains

A logical first step in the study of protein life cycle is to understand what happens to them during and just at the end of translation, when they start protruding from the exit tunnel of the Ribosome. At that time, the sequences of these polypeptide chains are not complete

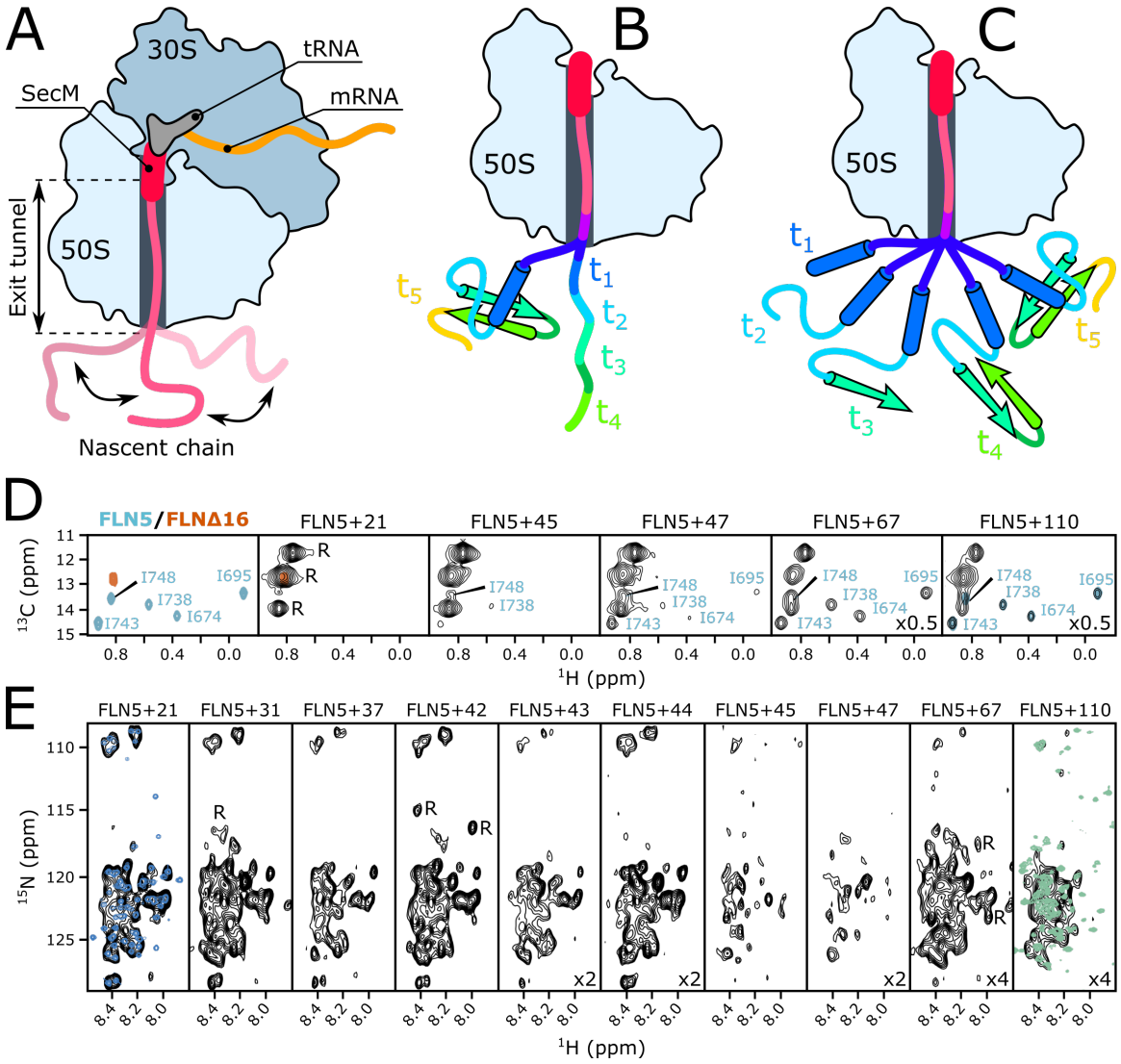


Fig. 2.1 : Ribosome-associated nascent chains. (**A**) Scheme of RNC arrested in an *E. coli* Ribosome by the SecM motif. (**B**) RNC folding after complete translation. (**C**) Progressive RNC folding during translation. (**D**) & (**E**) Nascent chains of FLN5 emerging from the Ribosome, as monitored by NMR spectroscopy. (**D**) ^{1}H-^{13}C correlation spectra of Ileδ1-[$^{13}CH_3$]-labelled FLN5 RNCs (black), isolated, natively folded FLN5 (cyan) and isolated unfolded FLN5Δ16 (orange). Resonances marked R arise from background labeling of 70S ribosomal proteins. (**E**) ^{1}H-^{15}N correlation spectra of U-^{15}N–labelled RNCs, compared with disordered controls, unfolded FLN5 in FLN5Δ12 (**blue**) and unfolded FLN6 in FLN5-6Δ18 (**green**). Figures and legends (**D**) and (**E**) were adapted from [Cabrita et al., 2016].

yet and thus the protein does not possess all the information contained in its sequence in order to properly fold [Anfinsen et al., 1961]. To simply simulate this situation, one can first mimic the nascent chains by using different lengths of the same protein. Such a protein needs to be cut at various point of its Cter region, the translation occurring from the Nter to the Cter linearly [Kurt et al., 2006].

Another elegant way to study these nascent chains is to trap them inside the Ribosome during translation, to form a Ribosome-bound nascent chain complex (**RNC**). With help from protein engineering, it is possible to suppress the stop codon in the mRNA protein sequence when producing the nascent chains in cell-free systems [Gilbert et al., 2004]. The protein production is stopped at the end of the protein sequence, trapping the mRNA and the nascent chain inside the Ribosome and thus forming RNC ready to be studied after proper purification. Such methods also allow isotopic labelling of the nascent chains, necessary for NMR characterization. In fact, even though the halflife of such a protein construction is quite short for NMR investigation (around 48 h), NMR was successfully applied to monitor its construction and dynamics. An RNC system modeled with the nascent chains of the Ig_2 protein construct from [Gilbert et al., 2004] was first explored by NMR in 2007, by the research group of Pr. John Christodoulou [Hsu et al., 2007]. Interestingly, even when bound to large Ribosomes, it is still possible to record acceptable quality backbone amides HMQC spectrum corresponding to the Nter region of Ig_2 sticking out of the Ribosome, due to its high mobility. At the opposite, residues inside the pore of the Ribosome are invisible because of their lack of motion because they are directly linked to a large protein machinery undetectable by classical liquid-state NMR without proper labelling. It was logically observed that cross-peaks corresponding to AA in the vicinity of the exit tunnel were broadened, similarly to those of the Nter region of the Ig_2, because of the electrostatic interactions between the tail of the protein and the negatively charged surface of the Ribosome. Most of the peaks show typical unfolded protein chemical shifts but some other exhibit way more dispersed resonances, proving that a part of Ig_2 is folded. Since it was impossible to prove the presence of the TF in the media, one can assume that the protein started folding during translation without the help of a chaperone. Liquid state NMR also provided information to confirm that the RNC were still attached to the Ribosome owing to diffusion experiments.

Similar studies were still carried out by the same group, this time by directly purifying the RNC from bacteria. The experimental protocol is described in [Cassaignau et al., 2016].

In fact, it is possible to stop polypeptide elongation in vivo, with help from the SecM arrest sequence [Cabrita et al., 2009] (**figure 2.1, A**). At the end of the translation, the SecM AA sequence interacts with the ribosomal exit tunnel, thus blocking expression at the last glycine and trapping the nascent chain and RNA. Such RNC were studied with Ig_2 as a nascent chain as well as other variants of the same protein, yielding similar results to previous observations. This RNC expression system was also recently used to study nascent chains of αSyn, an IDP involved in the Parkinson's neurodegenerative disease [Deckert et al., 2016]. As αSyn is intrinsically disordered, no secondary structure was observed for the NMR visible RNC residues, corresponding to most of the protein AA, excepted few Cter residues close to the Ribosome. The comparison of HMQC cross-peaks intensities and line broadening from RNC versus a mixture of separated Ribosomes and αSyn allows picturing that the mobility of the nascent chain is reduced in the vicinity of the exit tunnel, as well as at the Nter, as observed previously. Indeed, the resonances of IDP are highly dependant on media conditions and direct environment (flat energy landscape). Recently, a combination of MAS-NMR and CryoET was also used to monitor the binding of Ribosome arrested nascent chains to a membrane protein, YidC, in a native environment [Baker et al., 2018]. The C-Subunit of ATP synthase, a native substrate of YidC was fused with the SecM sequence and arrested in Ribosomes. The structure of native YidC was investigated by MAS-NMR while CryoET was used to picture the RNC interacting with the membrane protein.

2.1.3 Co-translational folding of nascent chains

The previously mentioned engineered RNC were also used to study the co-translational folding of nascent chains. Such work enabled to determine when the folding starts and if proteins really need chaperones for such while comparing folding *in vitro* to *in vivo*-like conditions. A first study of the co-translational folding of the SH3 domain of α-spectrin, aiming at providing snapshots of the structure of the nascent chain at different moments of the translation process was carried out [Eichmann et al., 2010]. Native SH3 exhibits β-sandwich conformation but requires its complete AA sequence to properly fold *in vitro*. A first folding-incompetent construct was used as a reference of unfolded nascent chain, arrested in the Ribosome by the SecM motif. The structures of the other constructs were different, with the wild type protein and successive fragments cut from the Cter, and

also bound to the Ribosome. By observing their backbone amides HMQC spectra, it was possible to determine the folding state of the protein and deduce that the sequence needs to be complete for folding, as the fragments lacking the Cter region of the protein were also folding incompetent (**figure 2.1, B**). Interestingly, another study showed a different result for a different substrate protein. The co-translational study of the folding of Ig_2 domains arrested in Ribosomes was carried out with two different labellings. Methyl selective labelling was performed on Ile to picture the folded state of the protein and identify less mobile residues showing less intense NMR peaks. A simple ^{15}N labelling was also used to follow the unfolded state of Ig_2 [Cabrita et al., 2016] (**figure 2.1, D & E**). It was then showed that the domain FLN5 of Ig_2 folds co-translationally and starts after the expression of the full domain plus 11 to 14 residues belonging to the FL5 domain. The interaction with the Ribosome is probably responsible of the folding inhibition until the protein is long enough to escape the Ribosome surface interactions and fold (**figure 2.1, C**). Dynamic nuclear polarization magic-angle spinning or DNP-MAS NMR is a powerful technique to investigate the conformational states of RNCs. It notably allows to obtain carbon correlation spectrum which gives good information on the secondary structure of the AA chain. In this case, the ^{13}C labelled DsbA signal peptide was fused with the SecM motif and arrested in Ribosomes to investigate its secondary structure [Lange et al., 2016]. The obtained spectrum was compared to one of empty Ribosomes, as DNP-NMR is a technique exhibiting really low detection limits, and can be sensitive to natural occurring isotopes of carbons in unlabelled assemblies. Results suggest that the RNCs Cα and Cβ chemical shifts correspond to coil-like secondary structure.

2.1.4 Chaperones and nascent chains

Even though some nascent proteins were shown to exhibit co-translational folding, help from chaperones can also be needed for other substrates. This assistance is particularly important in the case of multi-domain proteins since the different domains rely on the presence of the others to properly fold. As explained before, a simple way to study the interaction of chaperones with nascent polypeptide chains is to mimic them using different biologically engineered fragments of the same protein. With help from this technique, the interaction between artificial nascent Apomyoglobin (ApoMb) protein and the substrate binding domain of the DnaK chaperone protein DnaK-β from *E.coli* was investigated [Kurt

et al., 2006]. Nter fragments of ApoMb are known to misfold and self associate in absence of chaperones. A complex between the DnaK-β chaperone and ApoMb fragments was first isolated on an SDS-PAGE gel and NMR came as a proper tool to decipher the conformation of bound ApoMb. All the different lengths of ApoMb are unfolded at low pH, including the full-length protein. At physiological pH, however, full-length ApoMb exhibits residual secondary structure while all the fragments are unfolded and their NMR ^{1}H-^{15}N HSQC spectra exhibit low peak intensities. It is particularly the case for the shortest fragment of ApoMb, which resonances are broadened beyond detection, as the protein clearly aggregates. As confirmed by CPMG and diffusion experiments, the line broadening of AboMb peaks is due to the slow tumbling of the protein fragments associated together. A large part of the missing peaks is recovered in presence of DnaK-β as the fragments are bound to DnaK-β and stop interacting with each other. Nevertheless, the bound fragments still do not exhibit secondary structure. DnaK-β acts as a holdase chaperone and a folding enhancer, as it holds the small protein fragments and lets them spontaneously fold when long enough for spontaneous post-translational folding.

The *E. coli* trigger factor (**TF**) chaperone is directly anchored to the Ribosome through interaction with the L23 ribosomal proteins and located close to the exit tunnel, to be able to directly interact with nascent proteins as well as newly synthesized proteins. The structure of the TF is divided into three domains, a Ribosome binding domain at its Nter, a peptidyl prolyl isomerase domain and a Cter domain structurally located in between the two other domains, where substrates were reported to bind [Saio et al., 2014]. The Ribosome anchored TF is monomeric. When not interacting with protein client, the TF is dimeric and its structure and dynamics were also studied by NMR [Morgado et al., 2017]. The TF is a substrate independent chaperone and exhibits a high affinity for hydrophobic clusters and protein chains. The interaction between this chaperone and the nascent chains designed in the [Deckert et al., 2016] paper, monitored by NMR, showed only weak interaction resulting in small peaks shifts on the NMR HMQC spectrum of the labelled RNC (**figure 2.2, B**). Recent advances also allowed to record NMR spectrum of TF bound to the Ribosome, despite of its size, by ultrafast MAS-NMR and paved the way for further RNC studies by solid-state NMR techniques [Barbet-Massin et al., 2015].

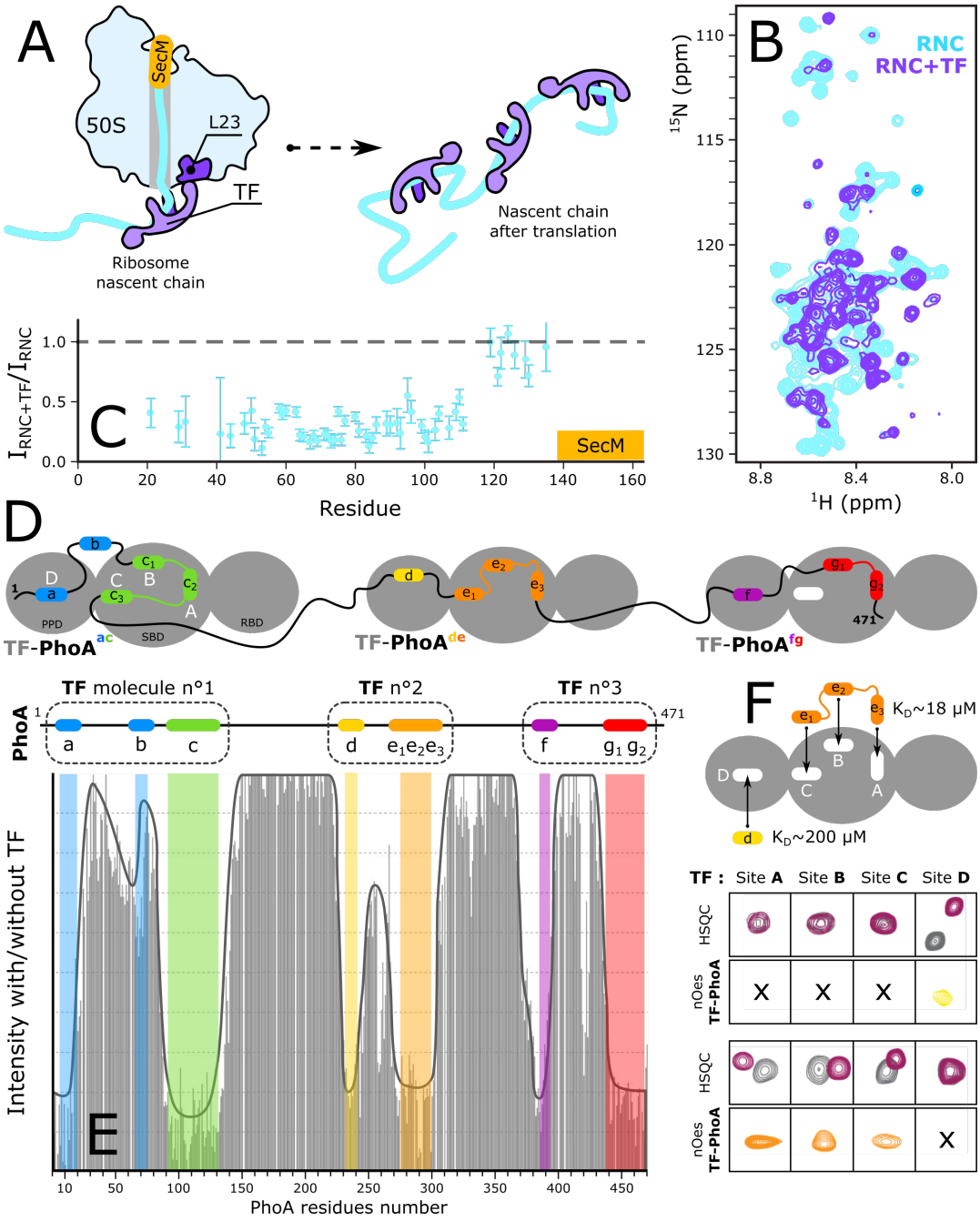


Fig. 2.2 : TF binding models. (**A**) TF anchored at the exit pore of the Ribosome interacts with RNCs. Several TF macromolecules are also able to stabilize a nascent chain completely expulsed from the Ribosome (**B**) Comparison of ^{1}H-^{15}N SOFAST-HMQC spectra of the αSyn RNC with and without 1 mol eq TF. (**C**) Relative cross-peak intensities of the αSyn RNC following the addition of TF. (**D**) Detailed scheme of the interaction between PhoA and three TF chaperones. (**E**) Plot of resonance intensity changes of PhoA residues as a result of complex formation with TF. (**F**) Binding of the second TF molecules to PhoA sites d through e. PhoA sites d and e are engaged by one TF molecule. The intermolecular nOes between TF and PhoA and the corresponding HSQC signals are also provided. Figures (**B**) and (**C**) were adapted from [Deckert et al., 2016], Data from figure (**D**), (**E**) & (**F**) were adapted from supplementary data of [Saio et al., 2014].

2.1.5 Ribosomal chaperones and newly synthesized proteins

It is also interesting to gain knowledge on post-translational folding. Since TF is usually bound to the Ribosome surface and interacts with nascent or newly synthesized proteins, its study in interaction with an unfolded protein is a good way to mimic post-translational folding helped by chaperones. To this extent, a study of the interaction between the TF and the PhoA protein, an *E. coli* alkaline phosphatase was carried out [Saio et al., 2014]. PhoA is unfolded in the reducing environment of the cytosol but still exhibit residual secondary structure. Interaction of this substrate with the TF is proven by line broadening analysis of the NMR resonances of PhoA in reducing conditions, while it disappears in an oxidizing environment. Regions of PhoA enriched in hydrophobic residues are interacting with monomeric TF in four distinct hydrophobic binding regions, three being located in the substrate binding domain of the TF (**figure 2.2, A, B, C & D**). Regions of PhoA with secondary structure do not interact with the trigger factor. One structure of the lowest energy state of the complex is proposed, obtained with a combination of intermolecular nOes between PhoA and the TF (**figure 2.2, D**). Three molecules of the TF are needed to stabilize a long protein like PhoA since one TF can apparently only bind about fifty residues. Rapid binding kinetics between the two proteins is observed by NMR relaxation dispersion experiments, probably to allow swift binding and release of the substrate. A mutant of MBP was also studied, with lower thermal stability compared to native MBP [Saio et al., 2014]. In fact, the unfolded state the MBP mutant is more populated at high temperatures. As a result, TF interacts more with the MBP mutant and is able to partly prevent its aggregation.

TF acts as a holdase and stabilizes the unfolded state of its substrates and displacing the protein equilibrium towards the unfolded state. The different substrates binding sites of the TF appear to have complementary affinities to stabilize a wide variety of substrates or nascent proteins in the cells and shield them from an aggressive and crowded environment.

2.2 Ubiquitous cellular chaperones

2.2.1 Foldase chaperones and their substrates

Cellular foldases are interacting with newly synthesized proteins as well as unfolded or misfolded substrates. Given the energy provided by ATP hydrolysis, they are able to fold substrates. The main types of foldases are presented in this section, HSP60 excepted. In fact, HSP60 or chaperonins, are the type of chaperones studied during this thesis and are described in a dedicated section at the very end of this chapter.

HSP70

Characteristics - The HSP70 class of proteins was first mentioned in 1982, with the discovery of the DnaK protein in *E. coli* [Georgopoulos et al., 1982]. HSP70 are ATPases, highly conserved in all domains of life [Mayer, 2013], which are able to bind, hold and refold unfolded or misfolded protein substrates, thus preventing aggregation. These HSPs are involved in the *de novo* folding of 10 to 20% of all bacterial proteins, and an even higher percentage in eukaryotes [Mayer and Bukau, 2005]. HSP70 are also implicated in membrane translocation mechanisms, interaction with regulatory proteins and are central actors of the chaperones cellular network, as they can also cooperate with other HSPs such as HSP100 and HSP90. Misregulation of HSP70 is linked to severe disease [Murphy, 2013].

For their function, they also require the assistance of diverse partner proteins, or co-chaperones, to ensure folding. The action of J proteins (also often referred to as HSP40, DnaJ in *E. coli*) is necessary, as they deliver the substrates for HSP70 and stimulate ATP hydrolysis. Nucleotide exchange factors, such as GrpE in *E. coli*, are also needed to catalyse the release of ADP, as ATP hydrolysis is extremely slow with HSP70 alone [Kampinga and Craig, 2010].

Structure - Different HSP70 can be found in each cell compartment in eukaryotes but their overall domain architecture is conserved [Bracher and Verghese, 2015]. HSP70 are monomeric and composed of two different domains : a 44 kDa Nter nucleotide binding domain (**NBD**) showing ATPase activity ; and a substrate binding domain (**SBD**) at its Cter, itself divided into a β-subdomain binding the substrates and an α-helical lid covering the β domain and encapsulating substrate proteins [Mayer, 2013], (**figure 2.3, B**). In fact,

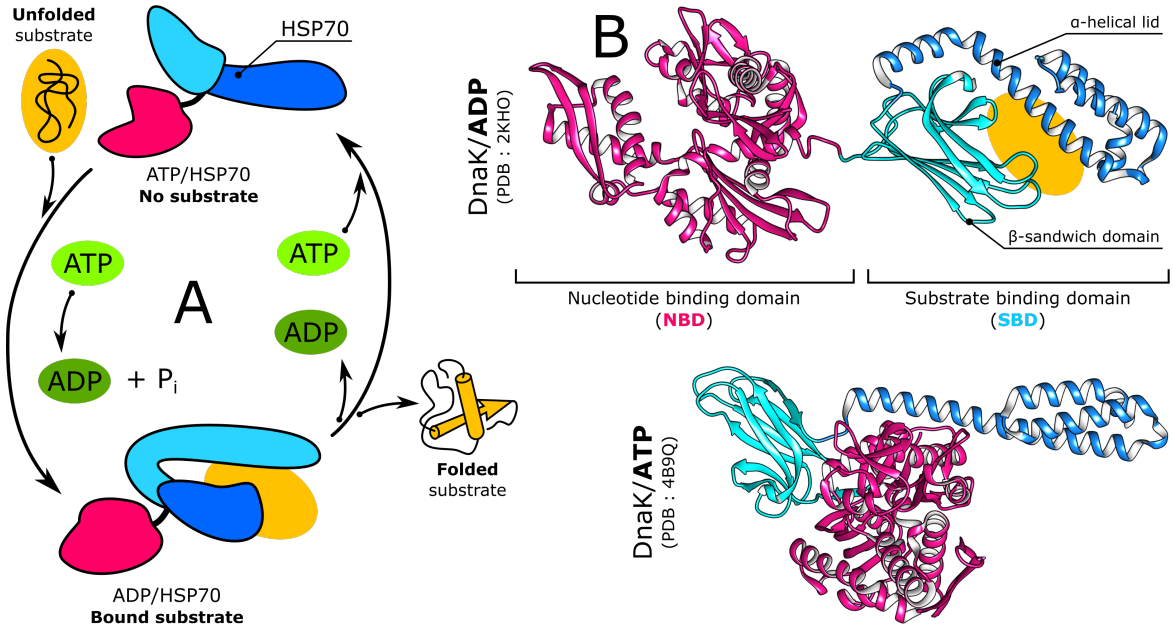


Fig. 2.3 : The chaperone HSP70. (**A**) Functional cycle of HSP70 in presence of a substrate. (**B**) Views of the ADP bound (PDB : 2KHO) and ATP bound (PDB : 4B9Q) states of DnaK, with different colors depending on the different structural domains.

HSP70 exhibits two radically different conformations when bound to ATP (open) or to ADP (closed), which changes its affinity for substrate proteins. When bound to ATP, the interaction with the substrate protein is weak, as the protein is not shielded by the helical α domain, while it is way stronger in the ADP-bound form [Mayer and Bukau, 2005], (**figure 2.3, A**). The substrates of these HSPs have a recognition motive composed of five hydrophobic residues framed by positively charged amino acids, found every 30 to 40 amino acids in proteins [Mayer, 2013, Rudiger, 1997]. These residues are usually buried in the hydrophobic core of native proteins but are exposed in the case of misfolded proteins, explaining why HSP70 does not bind native proteins.

Two modes of action are proposed to explain the folding activity of HSP70. First, the substrate unfolds locally within the substrate binding site due to interaction with its different hydrophobic patches. This unfolding then promotes proper folding and holding to lower the apparent substrate concentration in the cells and prevent aggregation. HSP70 can separate and refold small aggregates composed of 2 to 4 proteins but needs the assistance from the HSP100 chaperone network for greater aggregates. Several rounds of binding and release might be needed for proper substrate refolding. If refolding fails,

the substrate can be transferred to another refolding system or directly addressed to the degradation network [Hartl et al., 2011].

NMR structural studies - As just mentioned, the HSP70 chaperones are involved in many different cellular processes and are, to this extent, of paramount importance to cells. At the beginning of the NMR studies of HSP70, methyl-labelling was not available, thus the need to study the separated structural elements of HSP70. The secondary structure of a section of the SBD from the rat HSC70, believed to be a peptide binding domain, was first determined by NMR in the mi**d 90s (Morshauser et al., 1995)**. The calculation of the solution structures of the SBD of DnaK [Wang et al., 1998] and of the rat HSC70 [Morshauser et al., 1999] followed (complete assignment of the solution structure of its *E. coli* homolog, DnaK, bound to a peptide, only occurred in 2009 ; [Bertelsen et al., 2009]). Conformational changes of the NBD induced by ATP binding and hydrolysis into ADP were later investigated [Revington et al., 2004]. The NBD of *Thermus thermophilus* DnaK was showed to exist in a single conformation while binding ATP, whereas the ADP-bound state was in a slow exchange between two states, proving the central role of the NBD region in the ATP binding allosteric mechanism of HSP70. The identical NBD associated with a section of the SBD were suggested to be situated close with respect to each other, independently of ATP binding or its hydrolysis into ADP, with similar NMR signal line shapes suggesting similar rotational correlation, hence similar solution motions [Revington et al., 2005]. Substrate binding possibly caused conformational changes in the NBD, tightening the space between the two structural regions.

These findings were later questioned since another study proved that substrate binding did not affect the conformation of DnaK, confirmed by very similar hydrogen exchange patterns except for few residues situated in the binding pocket of the SBD [Swain et al., 2006]. In addition, previous studies where biased as the β domain of the SBD could self-associate, which is normally not the case *in vivo*. By adding mutations to prevent the self-association of the β domain, the interaction between the SBD and the NBD were shown to be reduced, as both spectra of isolated regions overlap quite perfectly with the spectrum of the two domains linked to each other. However, the communication between the two domains was proven central to the activity of *E. coli* DnaK [Swain et al., 2007]. In fact, the docking between the two domains occurs only in the ATP-bound state, while the two domains behave independently in presence of ADP. The presence of ATP

changes dramatically the stability of HSP70, the protein regions located near the SBD being stabilized while the substrate binding site is destabilized to trigger substrate release. The residues of the NBD responsible for the long-range allosteric changes between the ADP and ATP bound states of HSP70 were identified and proven to act as a "nucleotide-modulated switch" [Zhuravleva and Gierasch, 2011]. After ATP binding, the linker between the two domains of HSP70 interacts with the NBD, connecting themselves together, and changing the conformation of the substrate binding site. This rearrangement particularly increases the dynamics of the whole SBD, especially next to the substrate interaction pocket, providing plasticity to bind a wide range of substrates [Zhuravleva and Gierasch, 2015]. The plasticity of the ATP-bound NBD and SBD is especially increased for the BiP HSP70, the only HSP70 chaperone of the endoplasmic reticulum [Wieteska et al., 2017]. Nevertheless, eukaryotic cytoplasmic HSP70 can be significantly different from bacterial ones [Meng et al., 2018]. In the case of HspA1 and HSC70, the NBD and SBD are docked more closely together than in DnaK. In human HSP70, the contacts between the NBD and the SBD can be regulated by peptide binding, suggesting a different allosteric pathway for structural rearrangements [Umehara et al., 2018].

Interaction with co-chaperones and binding partners - *In vivo*, HSP70 are involved in complex chaperone networks. The interactions between HSP70 and these partners were also investigated by NMR. For instance, the cooperation between HSP40 (*E. coli* DnaJ) and HSP70 (*E. coli* DnaK) was investigated to determine the DnaK binding site on DnaJ [Greene et al., 1998]. It was shown by CSP and lineshape broadening that DnaJ interacts with both ATP and ADP-bound DnaK. The binding is located at the NBD and this binding mode is possibly dominated through electrostatic interactions. It was also suggested that DnaJ stimulates ATP hydrolysis of DnaK, when DnaJ binds DnaK while interacting with a peptide [Horne et al., 2010]. The two chaperones and the substrate then form a dynamic triple complex [Ahmad et al., 2011]. In this triple complex, DnaJ acts as a spring pulling on DnaK structure to change its conformation, stimulating ATP hydrolysis and substrate folding [Bascos et al., 2017]. ADP and substrate protein release in HSC70 can be accelerated by the members of the BAG family, upon interaction between BAG co-chaperones and the SBD [Rauch et al., 2016]. The NBD of DnaK also interacts with the ClpB disaggregase in *E. coli* to refold disaggregated proteins (see **Section. 2.2.4** [Rosenzweig et al., 2013]). Simulation of the lowest energy conformations was achieved

from the structural information obtained by CSP and PRE-NMR, using a docking program to generate structural models.

Substrates interaction monitored by NMR - A first interaction study between the vsvC peptide substrate and DnaK in the presence or absence of ATP was carried out [Landry et al., 1992]. The absence of long-range intramolecular nOe of the peptide showed that the bound peptide was in an extended conformation, with reduced mobility compared to the free peptide. As stated before, the conformation of DnaK does not change upon peptide substrate binding except for a few residues located in the binding pocket [Swain et al., 2006]. Binding studies of a protein substrate in two distinct structural states was also carried out [Kim et al., 2012]. The *E. coli* IscU protein is involved in the scaffold and the transport of Fe-S cluster assemblies and interconverts between two conformational states in solution : a structured state and an unfolded state which does not bind metal ions. HSP70 HscA binds predominantly with the disordered state of IscU and releases it upon ATP hydrolysis, while the DnaJ homolog HscB binds preferentially the structured conformation of IscU. These findings enable to conjecture a Fe-S cluster transfer mechanism, in which the cluster bound ordered IscU binds HSP40, which docks on HsP70, causing ATP hydrolysis and transfer of the substrate in its disordered form to HSP70, while HscB is released. A similar study on Human mitochondrial chaperones with identical IscU protein substrate showed that ATP hydrolysis of HSP70 homologue is stimulated by HSC20 (homologous to DnaJ) and increased again by addition of the IscUb substrate [Cai et al., 2013]. It was shown that the marginally stable SRC homology 3 domain SH3 substrate protein interacts with ADP-bound DnaK [Lee et al., 2015]. 2D DOSY measurements showed that detected SH3 is unstructured while interacting with DnaK (**figure 2.4, A & B**). However, this unstructured conformation is different from completely unfolded SH3, as proven by different HSQC spectra. In addition, the bound substrate is able to sample heterogenous globally disordered slowly exchanging conformations, which could explain the folding mechanism of DnaK. In fact, the bound substrate would be free to sample different dissimilar conformations in a protected environment, favoring refolding. This finding is consistent with the previously described plasticity of the binding domain [Zhuravleva and Gierasch, 2015].

A series of studies on the human telomere repeat binding factor 1 (**hTRF1**) as a substrate for *E. coli* Hsp70 (DnaK) was also carried out by Pr. Lewis Kay's group [Rosenzweig

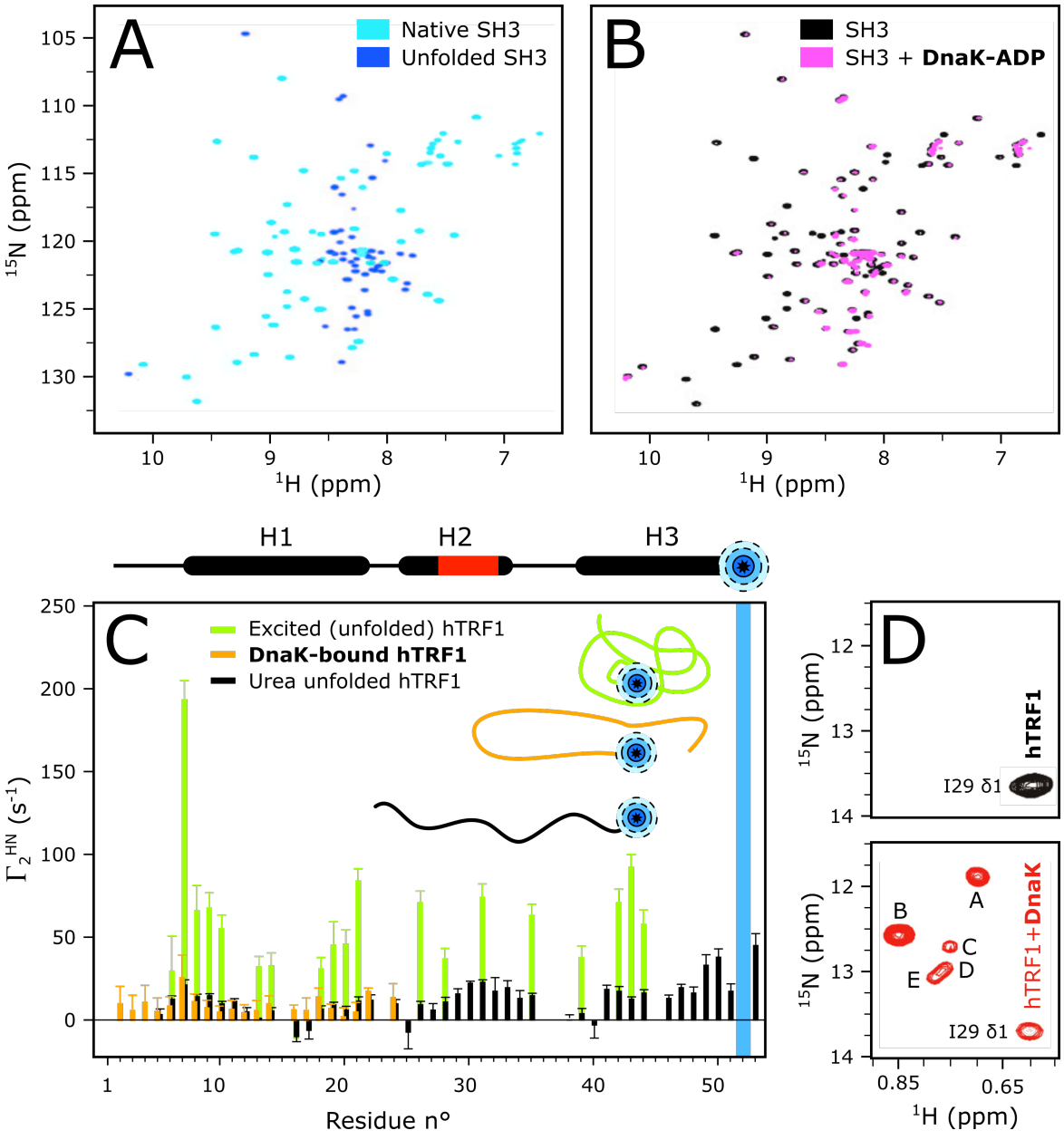


Fig. 2.4 : Interaction between HSP70 and various substrates. (**A**) 2D HSQC spectra of native and unfolded SH3. (**B**) 2D HSQC of SH3 in the presence or absence of DnaK. (**C**) PRE values of DnaK-bound hTRF1, unfolded hDRF1 and urea-unfolded hTRF1. (**D**) The Ile region of ^{1}H-^{13}C HMQC spectra of ILVM-$^{13}CH_3$ hTRF1 in the absence and in the presence of DnaK/ADP. (**A**) & (**B**) were adapted from [Lee et al., 2015] ; (**C**) was adapted from [Sekhar et al., 2016] and (**D**) was adapted from [Rosenzweig et al., 2017].

et al., 2017, Sekhar et al., 2015, 2016]. Similarly with SH3, hTRF1 is marginally stable and its native form is in equilibrium with an unfolded state, populated approximately 5% of the time at medium temperatures. hTRF1 is globally unfolded while interacting with DnaK [Sekhar et al., 2015]. CEST additionally shows residual secondary structure elements present in the non-interacting regions of the substrate chain identical to those of native hTRF1. Interestingly, its unfolded conformational states are very similar in the nucleotide-free DnaK, ATP-bound DnaK and ADP-bound DnaK forms of DnaK, even though DnaK bears large structural rearrangements between each state [Swain et al., 2007]. PRE data also proved that long-distance interaction between the secondary structure elements of hTRF1 are considerably weakened by DnaK binding, as shown by reduced PRE [Sekhar et al., 2016] (**figure 2.4, C**). The level of long-distance constraints is similar to what is observed for hTRF1 in urea, showing an extended unfolded conformation while small secondary structure elements remain, possibly the starting point of native folding in a diffusion collision mechanism (see **Chap. 1.2.2**). Methyl-labelling was also used to monitor the folding state of hTRF1 while binding DnaK. In fact, the backbone signal of the interacting regions of the substrate is broadened beyond detection due to the large size of the DnaK-hTRF1 complex of approximately 80 kDa [Rosenzweig et al., 2017]. The signals of the methyl bearing residues present in the expected binding sites of hTRF1 were multiplied upon DnaK binding, reporting the different conformation sampled by unfolded hTRF1, similarly to what was exposed for SH3 [Zhuravleva and Gierasch, 2015]. The translational diffusion of the species corresponding to these peaks are similar and globally slower than the one of free hTRF1, showing the existence of several conformations of hTRF1 in interaction with DnaK. a titration of hTRF1 by DnaK. DnaK can bind up to 4 different sites on hTFR1 with its non-selective substrate binding region. The interaction site is of minimal size compare to TF (see **Section 2.1.4** and SecB (see **Section 2.2.2**), the role of DnaK is thus not to hold but to help folding of substrates. This group carried out a last CEST NMR study, this time using the α-helical bundle domain R17* and the incomplete SH3 domain of the *Drosophila melanogaster* Enhancer, drkN SH3 [Sekhar et al., 2018]. They showed that bacterial and human HSP70 chaperones interact most likely only with the unfolded state of a substrate protein, holding and promoting refolding by the previously described structural conformation exchange of its bound substrate.

HSP90

Characteristics - HSP90 chaperonins are also ATPases, known as HtpG in **E. coli**, and are necessary for the survival of eukaryotes. Two different HSP90 are found in eukaryotes, one expressed upon heat shock, the other one being constitutive [Buchner and Li, 2013]. HSP90 bind, hold and regulate substrates involved in important cellular pathways, such as steroid receptors or protein kinases. They are involved in many physiological processes such as telomere maintenance, vesicle transport and targeted protein degradation. Eukaryotic HSP90 interacts with more than 20 co-chaperones, that help with the regulation and activation of HSP90 [Buchner and Li, 2013]. As for ClpB (HSP100), procaryotic HSP90 works cooperatively with DnaK for substrate transfer. HSP90 are believed to buffer the modifications of proteins and with this respect withstand evolution [Hartl et al., 2012]. The study of HSP90 is particularly relevant since mammalian HSP90 are involved in cancerous tumors development [Buchner and Li, 2013].

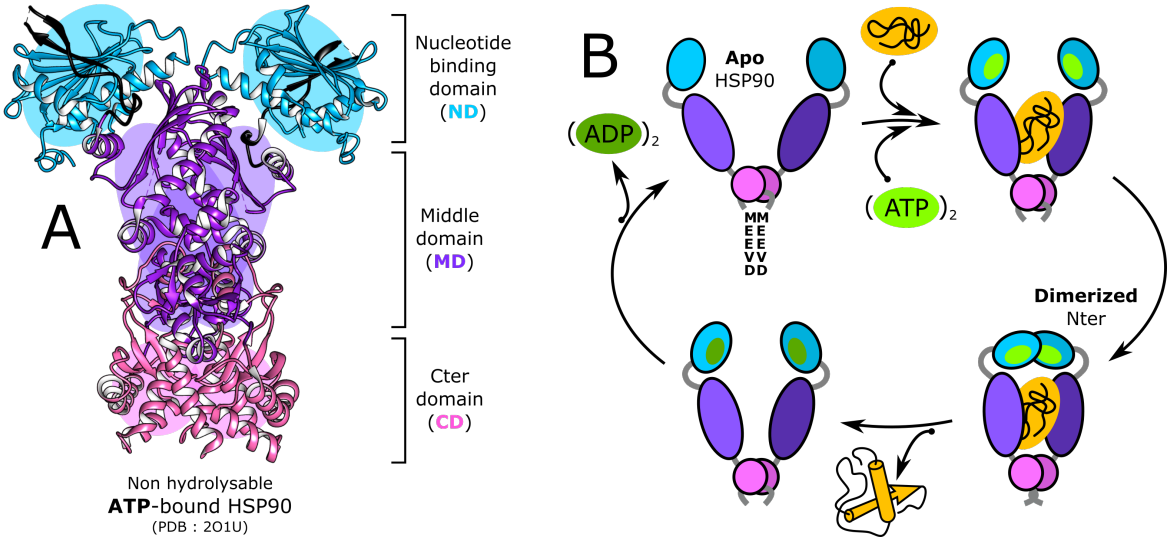


Fig. 2.5 : Functional cycle of HSP90. (**A**) Views of the AMP-PNP bound HSP90, a non-hydrolysable analog of ATP) (PDB : 2O1U), with different colors depending to the different structural domains. (**B**) Detailed functional cycle of HSP90 refolding a substrate protein.

Structure - HSP90 is an homodimer of approximately 180 kDa, with monomers divided into an ATP fixation and hydrolysis domain at its Nter, an intermediate domain and a dimerization domain at its Cter [Neckers et al., 2017], (**figure 2.5, A**). Substrates can be bound by the intermediate domain and the Nter dimerization domain as well [Karagöz

and Rüdiger, 2015]. The free dimer exhibits a clamp "V" shape with Cter domains bound together. This shape changes with ATP binding and hydrolysis, both Nter come close together and dimerize in order to close the clamp and hold substrates (**figure 2.5, B**).

NMR structural studies - NMR assignment of the isolated Nter ATPase domain of yeast [Salek et al., 2002] and human [Jacobs et al., 2006] HSP90 were performed due to the large size of the HSP90 dimer. Assignment of the Cter and the middle domain of yeast HSP90 are also accessible [Hagn et al., 2011]. Methyl resonances of Ala [Pederson et al., 2017] and Ile [Park et al., 2011b] of HSP90 from different organisms were also assigned to report on the full-length HSP90, since some substrates or interacting partners are binding several domains at the time and do not interact with the isolated domains [He et al., 2017].

Interaction with co-chaperones and binding partners - HSP90 works in concert with various and heterogeneous co-chaperones. The co-chaperone Sgt1 binding region was identified by CSP and linewidth changes upon binding of large HSP90, independently of the presence of ATP [Lee et al., 2004]. The identification of the binding site of the ATPase inhibition p23 co-chaperone on HSP90 was later performed by isolation of the different structural domains of both binding partners [Martinez-Yamout et al., 2006]. The isolated middle domain of HSP90 was shown to be responsible for p23 binding in absence of ATP. With the development of methyl labelling, the study of full-length HSP90 by NMR became possible and the previously exposed co-chaperones binding modes of p23 was revised. Especially, the p23 co-chaperone was found to bind exclusively to the Nter of HSP90 [Karagoz et al., 2011]. Binding of p23 in presence of ATP and after dimerization of the Nter domain of HSP90 induces conformational changes on the lid situated over the ATP binding pocket plus additional changes its middle domain leading to ATP hydrolysis perturbation. In fact, p23 possibly prevents the Nter domains from recovering their open conformation, blocking the ATPase activity. Similarly, precise mapping of the Nter of HSP90 and the middle domain of the Cdc37 co-chaperone (involved in the protection of protein kinases and ATP hydrolysis inhibition) interacting together was obtained with NMR provided restraints [Sreeramulu et al., 2009].

Aha1 is also a co-chaperone of HSP90, known to stimulate its ATP hydrolysis. The interaction site was determined by observing CSP, located at the middle domains of HSP90 [Retzlaff et al., 2010, Wolmarans et al., 2016]. The binding of this co-chaperone induces dimerization of the Nter of HSP90. Interestingly, only one Aha1 per HSP90 dimer is

sufficient to trigger the dimerization of the Nter. One Aha1 is also sufficient to stimulate ATP hydrolysis, through a cascade of conformational rearrangements from the middle of HSP90 to the Nter of the opposite HSP90 monomer not bound to Aha1 [Wolmarans et al., 2016]. HSP90 was also found to bind interactors such as Tah1 through MEEVD residues located at the very end of its Cter [Jiménez et al., 2012]. Precise mapping of the interaction was provided through structure calculations with intermolecular nOes in [Back et al., 2013].

The interaction between HSP90 and cis-trans peptidyl-prolyl isomerase co-chaperone, FKBP51 was also investigated [Oroz et al., 2018]. While other co-chaperones did not perturb the Nter dimerization of HSP90, FKBP51 binds and blocks HSP90 in its open conformation (**figure 2.6, A & B**) and decreases its ATP hydrolysis rate. Methyl labelling of Ile residues coupled with TROSY and PRE was used to map the interaction site between HSP90 and its co-chaperone. The interaction zone spans across the middle regions of HSP90 and both chaperones are in very close vicinity, as proven by the presence of intermolecular nOes. The presence of these nOes and additional SAXS measurements allowed to simulate the structure of the complex **figure 2.6, B**).

Substrates interaction monitored by NMR - The first study of HSP90 in interaction with a substrate was carried out by the group of Pr. Alan Fersht in 2002 [Rudiger et al., 2002]. The natural substrate of human HSP90 studied is p53, a tumour suppressor possessing different unstable domains. Especially, mutants of this protein are stabilized by HSP90, leading to abnormal oligomerization and eventually cancers. Aggregation of p53 is prevented by the presence of HSP90 below the melting temperature of HSP90. Interaction of HSP90 with p53 depends solely on the stability of the client protein since less stable mutants of p53 bind HSP90 at a lower temperature. The TROSY spectrum of p53 bound to HSP90 is characteristic of random coil secondary structure with restricted flexibility. Similar spectrum could be obtained with p53 alone, unfolded by urea. The formed complex was proven to be stable, HSP90 maintaining p53 in an unfolded conformation. Similar behaviour was reported by [Park et al., 2011a], with drastic signal reduction of p53 associated with HSP90 domains binding and increased hydrogen exchange pattern of p53 residues, suggesting a more solvent accessible conformation upon binding. However, a later study of the DNA binding domain of p53 (DBD-p53) and HSP90 questioned these findings. Using titration of the individual domains of HSP90 to determine the interaction site, it

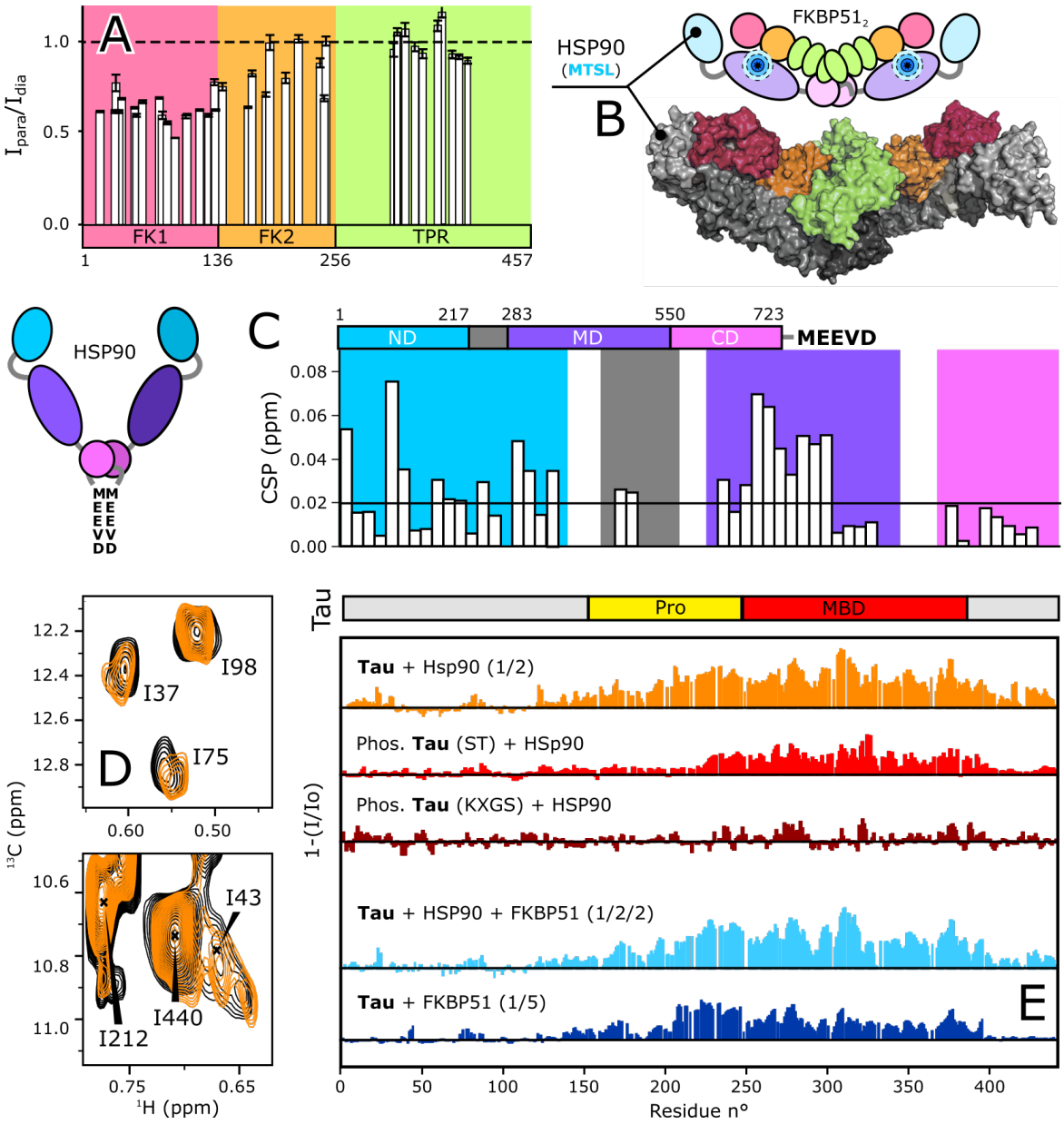


Fig. 2.6 : Interaction between HSP90, co-chaperones and substrates. (**A**) PRE mapping of the surface of FKPB51 interacting with HSP90. (**B**) Model of the interaction between HSP90 and FKPB51, based on PRE data, nOes and SAXS. (**C**) Regions of HSP90 affected by the interaction with M-TTR identified by chemical shift perturbation mapping. (**D**) Details of the 2D methyl TROSY spectrum of HSP90 labelled on Ile in the absence or presence of M-TTR. (**E**) NMR interaction profiles of Tau and phosphorylated Tau in presence of Hsp90 and the Hsp90/FKBP51/Tau complex. (**A**), (**B**) & (**E**) were adapted from [Oroz et al., 2018] ; (**C**) & (**D**) were adapted from [Oroz et al., 2017].

was showed that p53 is still in its native conformation when bound to HSP90, and that it remained able to bind DNA [Hagn et al., 2011]. The interaction was proven to be highly dynamic, with a strong affinity of DBD-p53 for the substrate binding domain (or Cter, 2 μM) and a weaker but non-negligible affinity for the other domains. In addition, DBD-p53 was protected from aggregation by full-length HSP90 only, showing the important site cooperation between the different domains of HSP90 to ensure its chaperoning.

This was consistent other studies proving that HSP90 binds structured regions of substrates [Linnert et al., 2013, Street et al., 2011]. The protein substrate Δ131Δ, a staphylococcal nuclease fragment, was proven to be globally unstructured while bound to HSP90 [Street et al., 2011]. However, some regions of the substrate maintained residual secondary structure. Interestingly, the region of Δ131Δ interacting with HSP90 corresponds to the most structured region of the substrate, confirmed by the presence of intramolecular nOes. This was confirmed in a second paper, showing that HSP90 binds a partially folded region of Δ131Δ, altering slightly its local and long-range structure [Street et al., 2014]. HSP90 also binds preferentially to the partially folded hormone-bound eukaryotic glucocorticoid receptor ligand binding domain (GR-LBD) [Lorenz et al., 2014]. Binding of GR-LBD on HSP90 particularly decelerates ATP hydrolysis and thus perturbs its conformational cycle. The interaction site was mapped by CSP and PRE, and is located at the middle domain of HSP90. By studying the binding of GR-LBD in presence of different mutants of HSP90, a key Tryptophan residue located in the middle domain at the position 300 could be identified [Rutz et al., 2018]. Mutation of this residues does not significantly change the overall conformation of HSP90 but residues located nearby are affected as proven by CSP, and these changes are propagated to distant residues. However, this residue is not directly involved with substrate binding.

Interaction with amylogenic substrates Transthyretin or TTR is an amyloidogenic protein involved in neuropathy and cardiomyopathy, interacting with the endoplasmic HSP90 GRP-94 *in vivo*. TTR is a 55 kDa homotetramer which structure is known but the monomeric misfolded structure had to be determined by NMR. In fact, it is really difficult to obtain a reliable structure of such protein, with a short lifetime and highly fibrillation prone. To this extent, a variant of TTR, proven monomeric by ITC and NMR relaxation thus called M-TTR was used for the assignment of the misfolded monomer structure [Oroz et al., 2017]. This misfolded protein, unlike native transthyretin (**TTR**), binds to

HSP90. This interaction between M-TTR and HSP90 was proven by peak broadening or total disappearance of the resonances during the addition of HSP90 (**figure 2.6, C & D**). The most flexible parts of M-TTR were less affected by the presence of HSP90 and the structure of HSP90 remained the same upon binding with client M-TTR. PRE was used to map the regions of M-TTR bound to HSP90. Some zones of the substrates are immobilized by the chaperone while some others become more flexible, but the detailed structure remains unknown, as methyl labelling only gives rough idea on the overall structure. The binding of the client amyloidogenic protein does not change and close the conformation of HSP90. The Nter of Huntingtin, which polyglutamine motif self-associates and causes Huntington's disease, also interacts with HSP90 independently of the presence of ATP [He et al., 2017]. Especially, the Nter exhibits an amphipathic α-helix important for the interaction with HSP90, since the interaction is dominated by hydrophobic interactions between the hydrophobic residues of the α-helix and the surface of HSP90. The middle and Cter domains of HSP90 are involved in the binding, while the two domains taken individually are not.

Quite a different study reported the interaction between HSP90 and the Tau protein, a natural substrate of HSP90 involved in Alzheimer's disease [Karagöz et al., 2014]. In fact, Tau is intrinsically disordered and was still reported to binds HSP90, while most of the previously reviewed studies showed that HSP90 was binding (partially) structured region of its substrates. The binding site of HSP90 on Tau coincides with a region bearing hydrophobic residues and positive charges, associated with Tau aggregation and microtubule binding. This region is different from the HSP70 interacting region of Tau (see **Section. 2.2.1**). Tau is bound with a low affinity to HSP90, around 5 µM, with or without ATP. By mapping CSP of attributed Ile residues of HSP90 and recording additional SAXS data, Tau was found to be bound to an extended substrate binding interface of HSP90, extended from the middle domain to the Nter [Karagöz et al., 2014] and its Cter [Oroz et al., 2018]. The binding of Tau additionally breaks the symmetry of the HSP90 dimer since signal splitting can be observed in the presence and absence of ATP [Karagöz et al., 2014]. Tau samples different positions in HSP90, as proven by the same PRE pattern in presence of MTSL labelled Tau, while being attached at different locations of its structure [Oroz et al., 2018]. The PRE is however perturbed in presence of the FKBP51 co-chaperone (see **previous paragraphs**), which reduces the mobility of Tau. Tau interaction with HSP90 is

also impaired by phosphorylation, phosphorylation being involved in the progression of Alzheimer's disease (**figure 2.6, E**).

2.2.2 Holdase chaperones and their substrates

sHSP and Crystallins

Characteristics - Small heat shock proteins (**sHSP**) are found in all domains of life, with various sizes ranging from 12 to 42 kDa [Haslbeck and Vierling, 2015]. Unlike most heat shock proteins, they do not rely on ATP hydrolysis to prevent the aggregation of substrates [Strauch and Haslbeck, 2016]. The best-studied members of the sHSP family are the two types of α-Crystallin (αAC and αBC, also named HSPB4 and HSPB5 respectively), found in vertebrates. Along with the other types of Crystallin proteins, they represent the main protein component of eye lenses in vertebrates. Their role is to prevent aggregation which would cause light diffraction and blurry vision. They are expressed in cells under normal conditions, where they bind client proteins involved in many cellular pathways, such as apoptosis regulation. Mutations of sHSP show clinical importance as they are linked to neuropathies, myopathies and many other diseases [Fu, 2014].

Structure - All sHSP share a common conserved structural domain called α-Crystallin domain (or HSP20 domain) of 94 amino acids, framed by a non-conserved Nter domain and a moderately conserved Cter shorter domain [Kriehuber et al., 2010]. The HSP20 domain of sHSP exhibits a β-sheet sandwich structure, which can dimerize. sHSPs are usually found as polydisperse oligomers of 12 to more than 30 subunits. The oligomers are formed from the dimers, associated through their Cter domains to form greater oligomers such as hexamers, that themselves associate by their Nter to form higher order oligomers. The oligomers are highly polydisperse, dynamic and exhibit different shapes such as barrel-like assemblies or spheres, commonly found in cells. When exposed to temperature variations, modified post-translationally (often phosphorylated) or in presence of substrate proteins, oligomers disassemble to form smaller structures. These small oligomers are able to strongly bind and hold misfolded proteins and to release them when the conditions are more favorable for proper refolding, with help from the HSP70/HSP40 chaperone machinery [Lee and Vierling, 2000, Strauch and Haslbeck, 2016]. When subjected to more drastic conditions or excesses of substrate proteins, substrates are protected and buried in

sHSP aggregates, and later resolubilized by disaggregases of the HSP100 chaperone family [Mogk et al., 2003, Strauch and Haslbeck, 2016].

NMR structural studies of αC - The stability of α-Crystallin (α**C**) fold was investigated by combined ^{1}H and ^{31}P NMR [Carver et al., 1993]. The protein complex is very stable and is resistant to temperatures up to 75°C. However, the αAC and αBC subunits taken individually show different susceptibility to pH conditions, with lower stability for αB, while both are properly folded when mixed together and forming a complex in identical conditions. This indicates that the interaction between the different monomers of α-Crystallin increases its overall stability. In denaturing conditions, the Cter of both αAC and αBC are unfolded first, possibly due to extended solvent exposition. The structures of the isolated Cter extensions of both αAC and αBC were investigated by ^{1}H-^{1}H correlation NMR and were shown to be unstructured, even in presence of membrane mimicking conditions [Le Breton and Carver, 1996]. The stability and role of the Cter domain of αAC was additionally explored in [Smulders et al., 1996]. By mutating the Cter, its hydrophobicity was either increased or decreased. Introduction of an hydrophobic Tryptophan residue at the Cter particularly perturbed its mobility and the thermal stability of the overall αAC subunit. The naturally occurring mutation R120G of the αBC Crystallin has a dramatic effect on chaperone activity and stability [Treweek et al., 2005]. This mutation is notably involved in various diseases including myopathy since αBC is also expressed in various non-lenticular tissues. Interestingly, the αC chaperone is also able to form amyloid fibrils in vitro. Especially, αAC, αBC and the mutant R120G-αBC are able to form amyloid fibrils in presence of 1 M Guanidine HCl and at high temperature [Meehan et al., 2007].

NMR structural studies of αBC - The Cter of αBC remains flexible in the fibrillar form of αBC and shows the same ^{1}H NMR trace as free αBC. The solution structure of the monomeric state of αBC isolated Crystallin domain (residues 64 to 162) was first determined by liquid state NMR, along with the structure of the polydisperse oligomers by MAS solid-state NMR [Jehle et al., 2009]. Similar structures were observed in both forms, except for the β8 and β4 strands or the Crystallin motif, which are probably corresponding to the contact zones between the monomers engaged in the oligomeric state. Particularly, αBC seems to be dimeric in solution and these dimers are engaged in higher order oligomeric structures. Soon after, MAS-NMR was applied to study full-length αBC [Jehle et al., 2010, Mainz et al., 2009]. The oligomers structure of 24 subunits engaged together could be

obtained by intersubunit interaction mapping and a combination of MAS solid-state NMR and SAXS [Jehle et al., 2010]. Additionally, this study provided a structural basis to explain the pH-dependant substrate binding self-inhibition. At physiological pH, the Cter motif of αBC binds the substrate binding groove to regulate substrate access while at lower pH the interaction is decreased, favoring substrate binding [Jehle et al., 2010, Rajagopal et al., 2015]. The variation of pH also influences the stability of the oligomeric state of αBC, as well as the size of its oligomers [Rajagopal et al., 2015]. The Cter regions of αBC was also identified to play a role in the assembly of αBC oligomers, as mapped by methyl labelling coupled with classic liquid-state NMR [Baldwin et al., 2011]. In fact, the dynamics of the Cter are useful to regulate the assembly/disassembly of the different oligomeric forms. The monomers engaged in the oligomers or in isolated dimers in solution exhibit very similar structure, as proven by almost identical Ile signals for both species. This shared structure is stable with temperature changes from 0 to 50°C, with its Cter remaining unfolded and highly dynamic, as proven by CPMG [Baldwin and Kay, 2012, Baldwin et al., 2012]. However, at temperatures below 0°C, the Cter adopts a different and more rigid conformation by interacting directly with the adjacent monomers, observed by both solid state and liquid state NMR. This conformation possibly regulates the access of the substrate to the substrate binding groove of αBC [Baldwin et al., 2012, Liu et al., 2018]. The size and the polydispersity of the oligomers are increased by the presence of Cu(II). The metal ion interacts strongly (K_D $11x10^{-12}$ M) with the Crystallin motif of αBC, at the interface involved in the monomer/dimer cohesion [Mainz et al., 2012]. The chaperone activity was also increased in presence of Cu as proven by decreased aggregation of β_L-Crystallin monitored by light scattering.

Substrates interaction monitored by NMR - Substrate recognition mechanisms of sHSP are still not well understood, even if they show a preference for smaller substrates as well as translation-related proteins [Haslbeck and Vierling, 2015]. Multiple binding sites are believed to interact with client proteins and the Nter region of sHSP seems to play a key role in substrate recognition mechanisms. Interaction with substrates studies of αC started soon after its first NMR characterizations. First, γ-Crystallin, another member of the Crystallin family involved in cataracts developement, was found to be immobilized and protected by αC in heating conditions [Carver et al., 1995]. The high molecular weight assemblies of αC chaperoning γ-Crystallin showed reduced or nonexistent residual

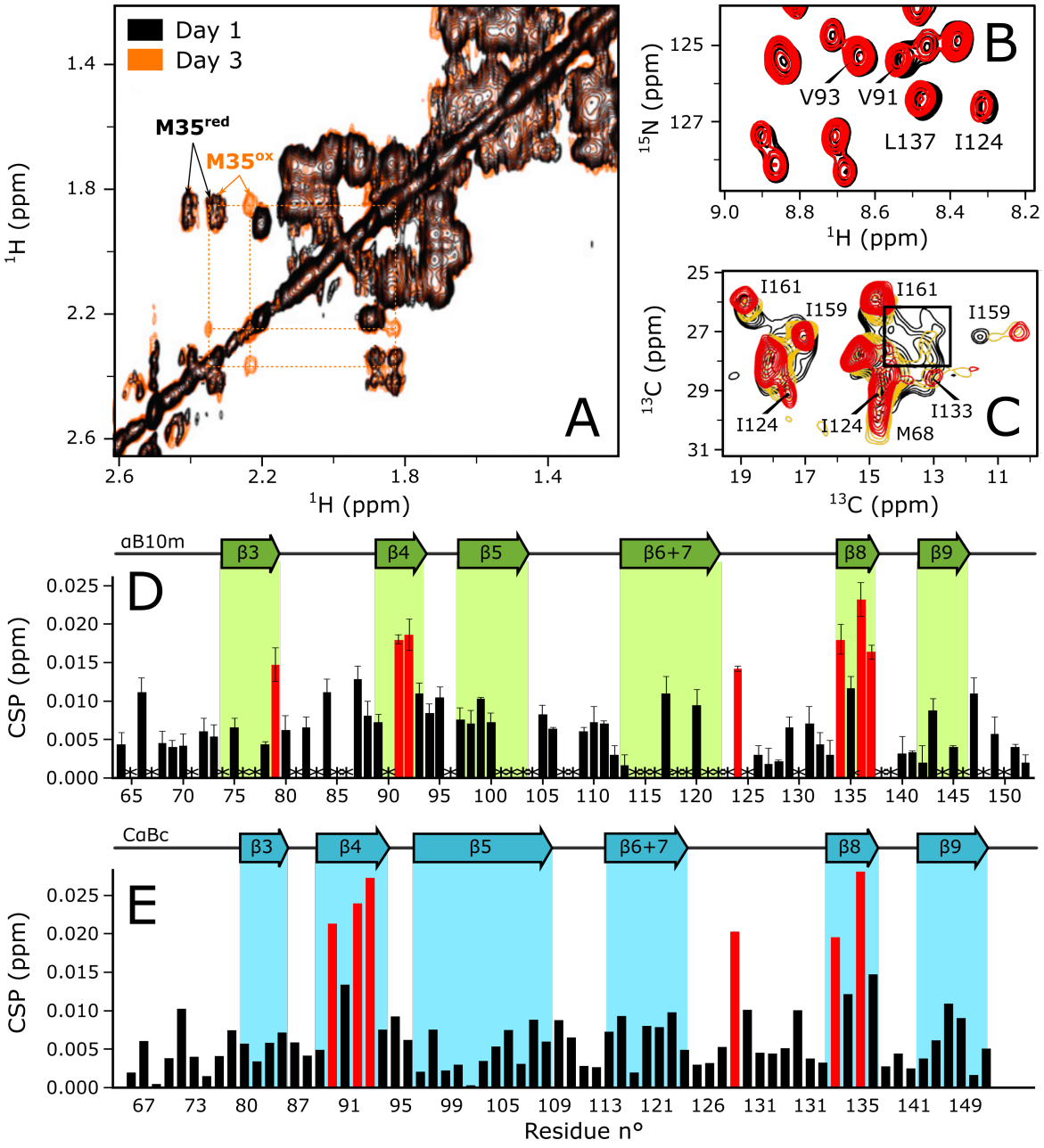


Fig. 2.7 : Interaction between αBC and different substrates. (**A**) TOCSY spectrum of Aβ40 in the presence of αBC as a function time. (**B**) Section of ^{1}H-^{15}N HSQC spectra of αB10 in presence and absence of Aβ40. (**C**) Isoleucine region of ^{13}C-^{13}C correlation MAS-NMR spectra of αB in the absence (red) or presence of unlabelled Lysozyme unfolded after addition of TCEP. (**D**) CSP profile of αBC upon addition of αSyn. (**E**) CSP profile of αBC upon addition of αSyn. (**A**) Adapted from [Narayanan et al., 2006] ; (**B**), (**C**) and (**D**) adapted from [Mainz et al., 2015] ; (**E**) adapted from [Liu et al., 2018].

movement of the Cter of αBC. Oppositely, αAC remained flexible, possibly due to positive charges of the Cter of αBC interacting with the substrate. The αC was also found to interact with α-Lactalbumin [Carver et al., 2002, Lindner et al., 1998, Treweek et al., 2010] and Insulin [Lindner et al., 1998], unfolded by DTT reduction. The substrates were also protected from aggregation and precipitation, even at high temperatures. Conversely, αC was not shown to interact with partially folded stable substrates. No major structural changes in αC upon substrate binding could be observed by NMR, except an increase in rigidity of the Cter [Lindner et al., 1998].

Interaction with amylogenic substrates - Interestingly, αBC exhibits two different binding modes for amyloidogenic proteins or amorphous aggregating client proteins [Mainz et al., 2015]. Lysozyme unfolds and aggregates under reducing conditions, which makes it a perfect client for αBC. A combination of MAS solid-state and liquid-state NMR using the TCEP reducing agent allowed to identify unfolded lysozyme binding site on αBC, involving the Cter and Nter ends of the chaperone. Especially, removal of these αBC regions inhibited its chaperone activity. Nevertheless, the interaction of the Aβ40 amyloidogenic protein is very different. Interestingly, Aβ40 and αBC can co-precipitate to form oligomers with increased neurotoxicity, which is counterintuitive for a chaperone and especially αBC [Narayanan et al., 2006]. To address this interrogation, the interaction between the two proteins was first proven by 1D and 2D-STD, since the Aβ40 STD signal dramatically increases in presence of αBC. By 2D STD, it was also possible to map the interaction site, located in the hydrophobic core of Aβ40. On the other hand, the binding of the Aβ40 amyloidogenic protein occurs at the βB4 and β8 of the Crystallin motif of αBC, as proven by CSP and additional PRE experiments, with a spin label attached to Aβ40 [Mainz et al., 2015]. The presence of αBC also influences the oxidation state of a key residue of Aβ40 involved in fibrillation, M35, which signal appears on the TOCSY spectrum of Aβ40 over time [Narayanan et al., 2006]. Oxidation of this Methionine increases polarity and solubility of Aβ40, which destabilizes monomer-monomer interaction and therefore fibrils formation. On the other hand, it also increases the toxicity of the formed oligomers by favorizing the propagation of free radicals. Accumulation of oligomers was actually reported by [Mainz et al., 2015] too. The β4/β8 region of the Crystallin domain is also responsible for the interaction of αBC with other amyloidogenic substrate proteins such as αSyn and Tau [Liu et al., 2018]. The binding can involves interactions with the Cter of

αBC, which disturbs its oligomeric state and weakens its chaperoning action. In fact, the isolated Cter of αBC binds the $\beta4/\beta8$ core more strongly than αSyn does, weakening the interaction with the substrate protein. The αBC chaperone was also shown to prevent fibrillation of β_2-microglobulin [Esposito et al., 2013]. The β-sheet core of the substrate interacts non specifically with αBC, as its NMR signal is broadened in presence of αBC. The intensity of the NMR spectra progressively decreases with time as the amyloid fibrils form and become indistinguishable with classical liquid-state NMR, while this process is slowed down in presence of the chaperone.

Skp

Skp is a homotrimeric periplasmic chaperone involved in the insertion and folding of membrane proteins. Its structure is similar to the prefoldin with a three-armed jellyfish-like architecture hosting a hydrophobic cavity in its middle. The three subunits of Skp exhibit very similar structure, as proved by a single set of NMR resonances reporting for the three subunits [Burmann et al., 2013]. Skp is a holdase chaperone which protects membrane proteins from aggregation in the periplasm, without the need for ATP hydrolysis.

Aggregation of OmpA is prevented by the formation of a 1/1 stable complex between the Skp trimer and the substrate protein, which can be isolated by gel filtration [Walton et al., 2009]. Interestingly, OmpA, 35 kDa, exceeds the size of the cavity of Skp and posses the particularity to be divided into a soluble periplasmic domain and β-barrel domain (tOmpA, 19 kDa), inserted in the outer membrane of bacteria. The amide backbone TROSY-HSQC spectrum of the β-barrel of OmpA bound to Skp is characteristic of an unfolded protein. Skp binds the β-barrel domain of OmpA in its unfolded state while the periplasmic domain can refold outside the cavity. Measurement of the rotational correlation of OmpA actually proved that the periplasmic domain protrudes outside of the cavity of Skp. Skp is also able to protect another membrane protein, OmpX, 18 kDa [Burmann and Hiller, 2012]. Its β-barrel architecture is very similar to the one of tOmpA. Skp binds unfolded OmpX and protects it from aggregation in the periplasm. The Nter region of OmpX, however, exhibits a reduced affinity for Skp. OmpX can be directly refolded after hosting in Skp, with a structural state very close to native OmpX, as shown by its ^{1}H-^{15}N TROSY spectrum. Interestingly, when unfolded in 8 M urea, two sections of

OmpX collapse into hydrophobic clusters. However, Skp does not interact preferentially with these.

Another study from the group of Pr. Sebastian Hiller addressed the structure of Skp while binding the two already mentioned substrates, tOmpA and OmpX [Burmann et al., 2013]. While the Skp chaperone is bound to a substrate, the symmetry of its three subunits is not broken, since they also exhibit identical resonances. No major structural rearrangement occurs upon substrate binding and the structure of Skp is similar with either OmpX or tOmpA inserted inside its cavity [Burmann et al., 2013, Callon et al., 2014]. Nevertheless, a reduction of the backbone motions of the helical "arms" of Skp upon substrate binding is observed, coupled with a global reduction of the dynamics of the chaperone. The unfolded substrate proteins bound within the hydrophobic cavity of Skp explore multiple and dynamic conformations [Burmann et al., 2013, Callon et al., 2014]. These conformations are random coil-like, even though the overall substrate is compacted inside the cavity, as proven by long-range intramolecular PRE, not observed when the substrate is unfolded in solution without the chaperone. Skp is also able to stabilize larger substrates such as FhuA (79 kDa), in an unfolded and highly dynamic conformation [Thoma et al., 2015]. Binding such a large substrate possibly requires several Skp at the time. Skp was also shown to bind selectively the locally frustrated segments of the Im7 and SH3 proteins substrates [He and Hiller, 2018, He et al., 2016]. This frustration arises from unfavorable contacts between neighbouring AA and causes instability of the protein chains.

Two other periplasmic chaperones, SurA and Spy were notably studied in [Burmann and Hiller, 2012, He and Hiller, 2018, He et al., 2016, Thoma et al., 2015] and exhibited similar substrates and chaperoning mechanism and will not be further developed.

SecB

The SecB chaperone acts as a holdase to maintain secretory proteins unfolded in order to be delivered to the SecA ATPase, presented separately in the next section (**Section 2.2.3**). Deletion of SecB is not lethal to cells but results in a dramatic increase of aggregation. Structurally, SecB is a tetramer composed of two dimers of a global size of 70 kDa. Cter thirteen amino acid residues of SecB were shown to be highly mobile by ^{1}H NMR, since their corresponding signal is very narrow compared to the rest of the spectrum, especially

for a protein of the size of SecB [Volkert et al., 1999]. Also, their role in the interaction with the co-chaperone SecA was proven central since their deletion dramatically perturbs the interaction and cooperativity between both proteins.

The unfolding of Barnase in presence of SecB was followed by H/D exchange [Zahn et al., 1996c]. In fact, Barnase possesses amide protons with different exchange dynamics, from slowly exchanging protons to protons which require complete Barnase unfolding to be solvent accessible. The last mentioned category of protons are buried within the central β-sheet of native Barnase and show faster exchange in presence of SecB. This shows that SecB is able to fully unfold and possibly hold Barnase for transport and later proper refolding of other chaperones. It was also shown that SecB possibly binds folded Barnase, as suggested by a decrease in linewidth of the folded Barnase NMR signal in presence of SecB [Zahn et al., 1996b]. This could also be explained by a displaced exchange equilibrium between folded and a folding intermediate of Barnase in presence of SecB, since SecB stabilized the unfolded state of Barnase.

A more recent study determined the structure of the SecB cytosolic chaperone in interaction with two different unfolded substrates, MBP and PhoA, using methyl isotope labelling coupled with state of the art liquid-state NMR [Huang et al., 2016]. Similarly to previously discussed chaperones, SecB posses different binding sites, enriched in hydrophobic residues, identified by their intermolecular nOe with fragments of MBP and PhoA, and confirmed by PRE experiments. These binding sites are keeping substrates unfolded, even if they are refolding competent when alone in solution. The whole bound substrates do not show a tertiary structure and the parts located directly in the binding sites do not show secondary structure either, as proved by their HSQC spectra. A combination of NMR, ITC and MALS show that the complex is formed in a 1/1 chaperone substrate ratio, and kinetics study using SPR and BLI showed that MBP associates with SecB ten times more than with the TF, explaining why is acts as a holdase. A final structure of that complex was calculated with nOe derived constraints and showed that SecB can accommodate approximately 250 interacting residues, 5 times more than the TF, which explains why SecB chaperones do not need to associate to bind a single molecule, unlike the TF.

2.2.3 Chaperone-like proteins involved in transport

SecA and SecYEG

A newly synthesized protein can exhibit a sequence encoded at its Nter of 15 to 30 AA, which gives information on whether or not a protein needs to be exported in a different cellular location. In fact, most proteins are synthesized in the cytoplasm but a third is exported. The Sec translocase is responsible for the secretory protein transport in bacteria. It works together with the SecB chaperone, which provides its substrates to SecA that are to be translocated across the membrane. With support from the energy provided by the ATP hydrolysis activity of SecA, the preprotein can then be exported through the SecYEG translocon and reach its final destination. SecA is a highly conserved homodimer of about 200 kDa which possesses ATPase activity providing energy for protein transport. SecA is made of different domains, including a nucleotide binding domain as well as a preprotein binding domain.

The mobile regions of SecA were identified by lineshape analysis of its ^{1}H-^{13}C 2D HSQC NMR resonances [Chou et al., 2002]. Sequential assignments of the protein allowed to conclude that the two most mobile regions of SecA correspond to its extreme Cter and NBD, which were proven essential for proper translocation activity of SecA. These regions are particularly affected by the presence of ATP and Mg^{2+}, as reported by large CSP, and possibly bear large structural rearrangements upon ATP hydrolysis. The Cter of SecA also possesses a zinc-binding domain, involved in the interaction between SecA and its co-chaperone SecB [Dempsey et al., 2004, Matousek and Alexandrescu, 2004]. The Cter zinc-binding domain was isolated by mutation and its fold was characterized in the absence or presence of Zn^{2+} [Dempsey et al., 2004]. Interestingly, the proper folding of this domain requires the presence of Zn^{2+} ions. The Cter also appears to shield the preprotein binding hydrophobic groove of SecA when it is not interacting with other proteins [Gelis et al., 2007]. It allows to restrict access to the groove to prevent nonspecific interactions with other proteins and SecB helps client proteins to bind SecA by canceling this auto-shielding process. The interconversion between these two SecA conformations could be monitored by exploiting intermolecular nOes and PRE, showing that the Cter protects the preprotein binding site approximately 10% of the time.

The effect of the ATP binding and hydrolysis on the conformation of SecA was also investigated deeper, and how it can provide enough energy for proper translocation of

preproteins. Especially, SecA posses different intramolecular regulator of ATP hydrolysis (or IRA) domains, that are involved in the ATP hydrolysis regulation, with an unclear mechanism of action [Keramisanou et al., 2006]. SecA with most of its Cter domain deleted was proven to be very dynamic for a protein of such a large size, as proven by the very good quality 2D HSQC spectrum, while full-length SecA is however less mobile. The IRA2 ATP hydrolysis regulation domain remains structured in full-length SecA. However, it exhibits significant spectral differences in the Cter deleted mutant, possibly due to interactions between IRA2 and the nucleotide binding domain. The IRA2 domain of this mutant is partly unstructured, as suggested by the low amount of intramolecular nOes. Binding of ADP to the NBD triggers large structural rearrangements which are propagated directly to the IRA2 domain. Especially, line broadening of the IRA2 signal shows a dynamic exchange between different conformations, a structured and an unstructured one. ADP binding also increases the order of SecA, and provides additional stability to all the domains of full-length SecA. The same stabilization is observed in presence of non-hydrolysable ATP.

SecA is able to "sort" secretory from nonsecretory proteins by interacting with the signal domain of the preprotein targeted for translocation. The only similarity between the secretory proteins is a stretch of hydrophobic residues preceded by a positively charged residue at their Nter, which questions on the recognition principle of SecA. This unknown selection mechanism was partly deciphered by studying the interaction of SecA with the signal peptide KRR-LamB, derived from the LamB porin, with help from NMR. The binding of the KRR-LamB signal peptide by SecA was first monitored by lineshape analysis of the ^{1}H-NMR spectrum occasioned by the presence of the chaperone in solution [Chou and Gierasch, 2005]. All the segments of the peptide were not equally interacting with SecA, as proven by increased line broadening of the signal corresponding to hydrophobic and positively charged residues of the peptide. Particularly, the charged residues are located at the Nter and some of the hydrophobic residues are involved in engaged in an α-helix conformation, as proven by the presence of intramolecular nOes [Chou and Gierasch, 2005, Gelis et al., 2007]. These regions and their organisation possibly constitute a motif which is selectively recognized by SecA. In fact, deletion of 4 residues located in the regions interacting with SecA completely cancel the interaction. To map more precisely the interaction with the KRR-LamB preprotein, abundant methyl groups bearing AA distributed along the whole SecA sequence were assigned by isolation of the different domains, recording of nOes and single point mutations [Gelis et al., 2007]. A single spin-

label was attached at different regions of the KRR-LamB protein and PRE NMR data were recorded to simulate the structure of the KRR-LamB / SecA complex. The peptide hydrophobic α-helix was shown to bind to a region of SecA located in a hydrophobic notch. This region is framed by polar and charged residues, forming salt bridges with the positively charged residues of the Nter of KRR-LamB while the Cter of the peptide remains unfolded.

WT LamB, alkaline phosphatase and M13 procoat were also proven to interact in the same region of SecA. To assess the role of the hydrophobic groove of SecA, its hydrophobicity was lowered by mutations, causing weaker substrate binding. Study of the complete Sec translocation machinery is more complicated since the SecYEG translocon is inserted in membranes. Nevertheless, DNP SS-NMR study of the LamB peptide interacting with SecYEG inserted into lipid bilayers could be carried out [Reggie et al., 2011]. By measuring the chemical shifts of the methyl labelled residues, the structure of the peptide could be predicted to be α-helical when interacting with SecA.

2.2.4 Substrate aggregation and disaggregation

Monitoring of fibrillation and amorphous aggregation by NMR

Monitoring of protein fibrillation or aggregation can be simply done by recording NMR spectra at different times, as used in [Cameron et al., 2014]. On classical HSQC / HMQC spectra, it is impossible to observe the formation of oligomers, which results in a decrease of the signal intensity as the size of the particle in solution increase, as mentioned previously.

Disaggregation machineries, HSP100

Structure - HSP100 (ClpB in *E. coli*, HSP104 in yeast) belong to the AAA+ protein family (or "ATPase associated with diverse cellular activities") and possess an AAA conserved domain where ATP binding and hydrolysis take place [Hanson and Whiteheart, 2005, Zolkiewski et al., 2012]. They were discovered in the early 90's and their main cellular activity is to disaggregate protein clusters and resolubilize individual proteins [Parsell et al., 1994]. The first structure of ClpB was obtained by X-ray diffraction in 2003 [Lee et al., 2003]. These assemblies organized in homohexameric rings possessing a central pore.

Misfolded or aggregated proteins are threaded through these pores and interact with inner loops composed of conserved aromatic residues [Hanson and Whiteheart, 2005, Mogk et al., 2018]. Once ATP hydrolysis takes place, the conformation of HSP100 and its loops changes and pulls on the substrate to disaggregate or unfold it mechanically.

Interactions with co-chaperones - The bacterial HSP100, ClpB, is working in association with other chaperones such as HSP70 (DnaJ, DnaK and GrpE) prior to disaggregation while HSP104 does not need such assistance [Zolkiewski et al., 2012]. The interaction between these large assemblies (580 kDa for ClpB) and their substrate is particularly challenging to study, since the monomers of the ClpB rings are very large (approximately 100 kDa) and in the case of *E. coli* systems, the disaggregation is ensured by the conjugated actions of several protein, ClpB, DnaK and GrpE [Zolkiewski et al., 2012]. The disaggregation mechanism of ClpB assisted by DnaK and a structural model of ClpB interacting with DnaK were determined by NMR, by [Rosenzweig et al., 2013]. The study of the complex was performed by double titration, selectively monitoring the DnaK binding site on ClpB or the opposite, owing to methyl-specific labelling. DnaK and ClpB bind through the nucleotide binding site of DnaK and the coiled-coil region of ClpB, in a 1/1 stoichiometry. Additional distance information provided by PRE allowed simulating one of the lowest energy conformations of the complex, using a docking program.

Substrates interaction monitored by NMR - The binding of ClpB with different type of substrates was also investigated by [Rosenzweig et al., 2015]. Specifically, the study was focused on the binding of the Nter region of ClpB, even though ClpB posses another Tyrosine-rich binding region which interacts specifically with polypeptides [Li et al., 2015]. NTD is particularly involved in disaggregation of strongly aggregated proteins [Barnett et al., 2005]. This work also notably relied on an elegant segmental labelling combined with specific methyl labelling strategy, to only label the Nter of ClpB in wtClpB [Rosenzweig et al., 2015]. The different parts of the protein were produced separately, with or without methyl labelling precursors and an artificial peptide bond was synthesized between the two protein segments to reconstitute integral ClpB.

The interaction region located at the Nter is a substrate binding groove that binds preferentially hydrophobic stretches and hydrophobic exposed residues, independently of ATP-hydrolysis [Rosenzweig et al., 2015]. Different aggregated substrates were tested, including firefly Luciferase, α-Casein and PhoA. It was proven that ClpB logically does not

bind folded proteins and that substrate binding is dependant on the presence of accessible hydrophobic residues on the aggregated substrates, from 3 to 6-AA long. The interaction is highly dynamic, the substrates are in fast exchange with the Nter of ClpB and the K_D ranges from 100 to 400 μM depending on the substrate. The isolated Nter or the Nter in the whole ClpB interacts with substrates in a 1/1 stoichiometry. Interestingly, the affinity for the substrate is increased in presence of integral ClpB, showing that the substrates possibly interacts with several monomers Nter at the same time. The substrates was shown to be destabilized and to unfold in the presence of ClpB. The Nter also block substrates from entering the pore prior to unfolding. Notably, the ATPase activity of full-length ClpB is enhanced by substrate binding at its Nter. In the case of substrate binding lacking its Nter however, ATPase activity is only weakly enhanced by substrate binding, through the Tyrosine-rich region of ClpB, showing the importance of the Nter in the interaction mechanism with substrates.

Interaction with amylogenic substrates - Investigation of the disaggregation mechanisms of HSP100 in the case of amyloidogenic substrates was also carried out. The prion-type protein Sup35 from Saccharomyces cerevisiae is amyloidogenic. The Gln and Asp residues situated at its Nter domain (residues 5 to 26) are especially critical for the fibrils formation, as deletion of this segments cancels fibrillation. The Sup355-26 peptide is highly aggregation-prone as proved by a rapid decrease of its NMR signal in solution due to its increasing size, and an increase of ThT fluorescence probably due to fibrillation [Narayanan et al., 2003]. Addition of HSP104, part of the Clp/HSP100 disaggregase chaperone family, dramatically slows down aggregation, even at catalytic concentrations. The population repartition of the oligomers of Sup355-26 varies greatly depending on the presence of HSP104. DOSY shows that, at 27°C, Sup355-26 exists in different oligomeric states, ranging from monomers to octamers, with a higher propensity to form octamers. When HSP104 is added, Sup355-26 is progressively disaggregated into lower molecular weight oligomers to finally monomers in a stepwise process. After disaggregation, HSP104 does not interact anymore with the monomers and seems to preferentially interact with certain oligomers (4 to 6 subunits) but does not interact with octamers, maybe because of symmetry difference [Narayanan et al., 2003]. Unexpectedly, since amyloid fibrils are usually made of a succession of β-sheets, combination of CD and NMR data analysis showed that low molecular weight oligomers of Sup355-26 have a high content of α-helical

structures [Narayanan et al., 2006]. As the signal disappears upon aggregation, the β-sheet content logically increases because of structural reorganization needed prior to fibrils formation. The structure of these oligomers is not perturbed by the presence of HSP104 as no chemical shifts nor signal decrease can be observed, but it probably lowers the energy required to disassemble low oligomeric states into smaller structures. The action of HSP104 is enhanced by ATP but ATP is not strictly needed, even though HSP104 is an ATPase [Narayanan et al., 2006].

2.3 Focus on chaperonins

Among the different types of molecular chaperones, chaperonins (often abbreviated **CPN**) are of particular interest because their mechanism of action is not yet totally understood and their large size makes their study challenging. Chaperonins are often referred to as HSP60, as explained previously. These protein assemblies are constituted of a combination of identical or different monomers of about 60 kDa (**figure 2.8, A**). Each monomer is divided into three different structural domains, namely the apical domain (possibly involved in the interaction with unfolded proteins through its hydrophobic patches [Hua et al., 2001, Kobayashi et al., 1999, Zahn et al., 1996a], the intermediate domain (hinge region between the two other domains, whose flexibility allows conformational changes between the open and closed states of the chaperonins) and an highly conserved equatorial domain (allows rings cohesion and ATP/ADP binding). These monomers are arranged to form two rings stacked on top of each other. Each ring embodies a cavity in its center often referred to as Anfinsen cage since 1991 [Ellis, 1996]. This cavity is of paramount importance for the function of chaperonins, as its goal is to encapsulate misfolded proteins and shield them from the crowded cellular environment.

This mechanism is only possible due to ATP binding on the equatorial domain and hydrolysis, which fuels a considerable structural change of the Chaperonin subunits. In fact, the central cavity of the HSP60 is believed to be generally hydrophobic and ATP binding switches the assembly from an open to a closed state. This rearrangement seems to be the key of protein refolding, as it exposes more hydrophilic residues and perturbs the global hydrophobicity of the chaperonin cave, in order to allow proteins to refold properly in a more confined environment. This refolding process is efficient but, in some

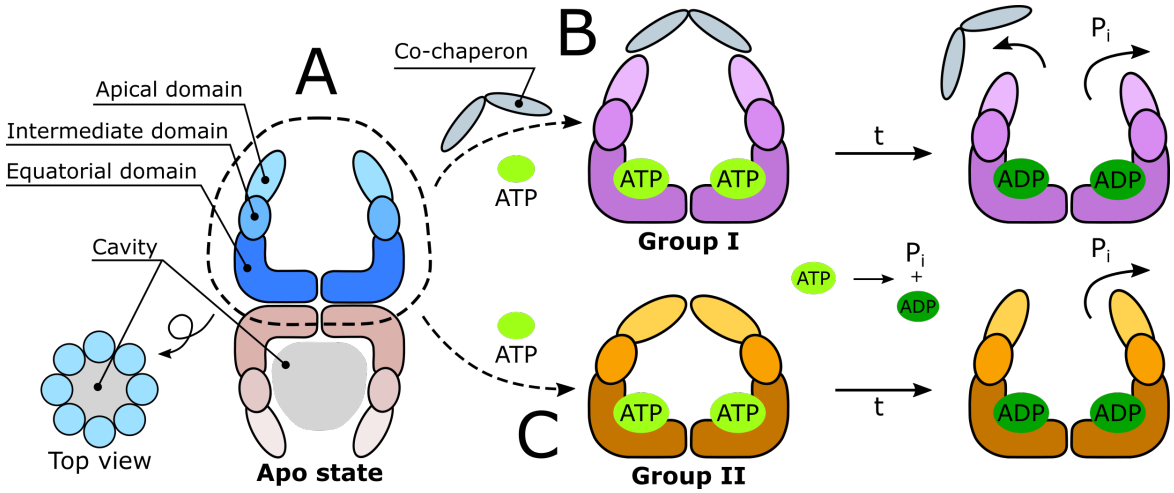


Fig. 2.8 : Schematic illustrations of the two types of chaperonins. (**A**) Top and side view of a schematic chaperonin and its different sub-domains. (**B**) Group I chaperonin and its co-chaperone during ATP hydrolysis. (**C**) Group II chaperonin during ATP hydrolysis. In (**B**) and (**C**), only one ring is represented for clarity.

cases, several rounds of binding and release into the cellular environment need to occur before achieving the substrates native state [Ditzel et al., 1998]. Chaperonins are separated into two distinct categories, type I and type II [Hemmingsen, 1992], depending on their structural features and refolding mechanism (**figure 2.8, B**). Even if classified differently, they both share a common ancestor [Bross, 2015]. A third classification type was also proposed, based on phylogenetic analysis, which does not show major differences in structure and mechanism compared to type II chaperonins [Techtmann and Robb, 2010]. It will thus not be mentioned further. In the next section, type I & II chaperonins will be described in details and their differences and similarities will be presented.

2.3.1 General refolding mechanism

Several mechanisms are proposed to explain the refolding abilities of the chaperonins. As for the mechanisms of folding, the biological truth probably lies in a combination of different hypothesis, described [Bross, 2015] and reviewed in [Motojima, 2015].

Confinement - A first proposed mechanism would be confinement [Baumketner et al., 2003, Brinker et al., 2001]. Since the substrate proteins are isolated from the cellular environment during the refolding cycle of the chaperonin, they would be shielded from

the other proteins which could start interacting and aggregating with the substrate protein (**figure 2.9, A**).

Hydrophobicity change - During the conformational cycle of ATP hydrolysis, the general hydrophobicity of the chaperonin cavity changes. In fact, in its open state, the cavity exposes hydrophobic residues of the monomers to provide a favorable binding site to horst unfolded substrates. The closing mechanism of chaperonins is followed by a change in the global hydrophobicity the cavity, due to individual motions of the monomers forming the cavity [Ditzel et al., 1998]. They switch from a completely hydrophobic to a more hydrophilic environment by rotation and structural rearrangements of the monomers upon ATP binding (**figure 2.9, B**). This could allow proteins to refold in a more favourable environment and not to stick to the chaperonin core, as unfolded or misfolded proteins tend to expose their hydrophobic patches. Once the cavity opens again after complete ATP hydrolysis, the substrate can be released from the cavity.

Iterative annealing - In this alternative model [Motojima, 2015, Todd et al., 1996], chaperonins trap unfolded proteins and change their wrong fold slightly, and by processing to several refolding cycles, allow misfolded proteins to partially fold and eventually allow them to leave a local energy minimum and reach their native conformation (**figure 2.9, C**).

Mechanical stretching - Since each chaperonin ring can be made of multiple identical subunits with different hydrophobic patches, an unfolded protein could interact with different subunits at the same time. The closing mechanism of the cavity could then stretch the amino acid chain and possibly prevent non-native interactions to help the substrate reach its native state (**figure 2.9, D**) [Shtilerman et al., 1991].

Chaperone-percolator model - This last hypothesis is also based on the stretching eventually caused by the interaction between the substrate and the different chaperonin monomers (**figure 2.9, E**). The difference with respect to the previous model is that this stretching could allow water molecules to penetrate the collapsed structure and hydrophobic core of the unfolded substrate protein and to force it to refold in a more native way [Csermely, 1999].

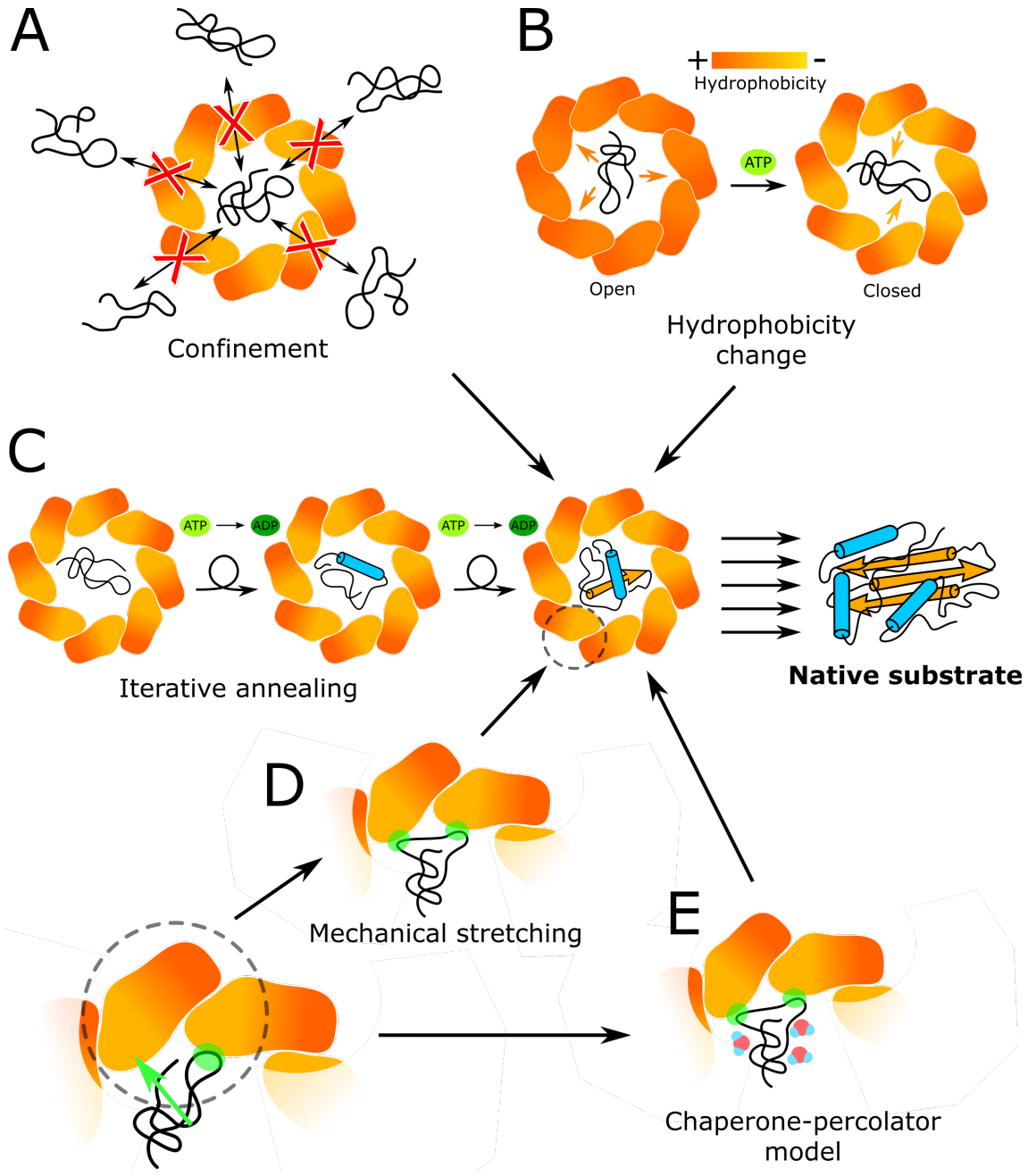


Fig. 2.9 : Schematic illustrations of the different models proposed to explain the refolding mechanisms occurring in the Anfinsen cage of chaperonins. (**A**) Confinement : the substrate is confined in the cavity. (**B**) Hydrophobicity change : the substrate is attracted by the hydrophobic environment of the open chaperonin ; becoming more hydrophilic upon ATP binding owing to structural rearrangement. (**C**) Iterative annealing : the substrate is exposed to several rounds of cavity binding and release, and progressively refolded. (**D**) Mechanical stretching : the substrate is stretched by interacting with different hydrophobic patches inside the cavity. (**E**) Chaperone-percolator model : the substrate is stretched and water molecules enter its core. (**D**) and (**E**) are zooms of (**C**).

2.3.2 Type I Chaperonin

History - The best-characterized member of type I Chaperonin family is GroEL/ES from *E. coli*. Type I chaperonins are found in the bacterial cytosol and chloroplasts, as well as in the mitochondrial matrix. Their existence was first speculated by Ellis in 1977, as a result of his observations on RuBisCO, an enzyme present in cell leaves, responsible for photosynthesis [Ellis, 1977]. RuBisCO was isolated and shown to interact with an unknown large protein complex, identified as GroEL/ES (named Gro because their absence inhibits the growth of the bacteriophage T4 head assembly). Roughly ten years later, GroEL/ES chaperonins from plants as well as from *E. coli* were sequenced and were already shown to play an important role in the assembly of RuBisCO [Hemmingsen et al., 1988]. A first elaborated model of the functional cycle of GroEL/ES was proposed in 1993 [Martin et al., 1993], and paved the way for chaperonin mechanism studies, nowadays still under investigation.

Structure - Structure-wise, Type I chaperonins consist of two back-to-back heptameric and asymmetric rings (GroEL, L stands for Large, (**figure 2.10, A**)). Its ATP-bound form is capped by a dome-shaped smaller ring assembly, also composed of seven monomers (GroES, S for Small, also named co-chaperonin, (**figure 2.10, B & C**)). The role of the GroES ring is to close properly the GroEL cavity, while the substrate protein is being refolded. In fact, the apical domain of the GroEL subunit is too short to close the cavity. The binding of the GroES ring to GroEL results in a major conformational change, pictured in (**figure 2.10, B & C**). GroEL then switches from a high-affinity state with an exposed hydrophobic core, to a lower affinity state exposing more hydrophilic residues, as previously explained. The volume of its cavity is doubled as a consequence of this rearrangement, allowing refolding [Ditzel et al., 1998]. First studies on GroEL/ES refolding activity showed that only one GroES ring was fixed per GroEL particle during the refolding cycle (bullet-shaped GroEL/ES, (**figure 2.10, B**)). Another assembly shape could be found, known as football-shaped GroEL/ES, with two GroES rings for one GroEL, but this form was initially believed not to be active (**figure 2.10, C**). More recent studies reviewed in [Iizuka and Funatsu, 2016] showed that the latest assumption was wrong and that the football-shaped ($GroEL/ES_2$) symmetric complex was formed and active in the case of high local concentrations of unfolded substrates.

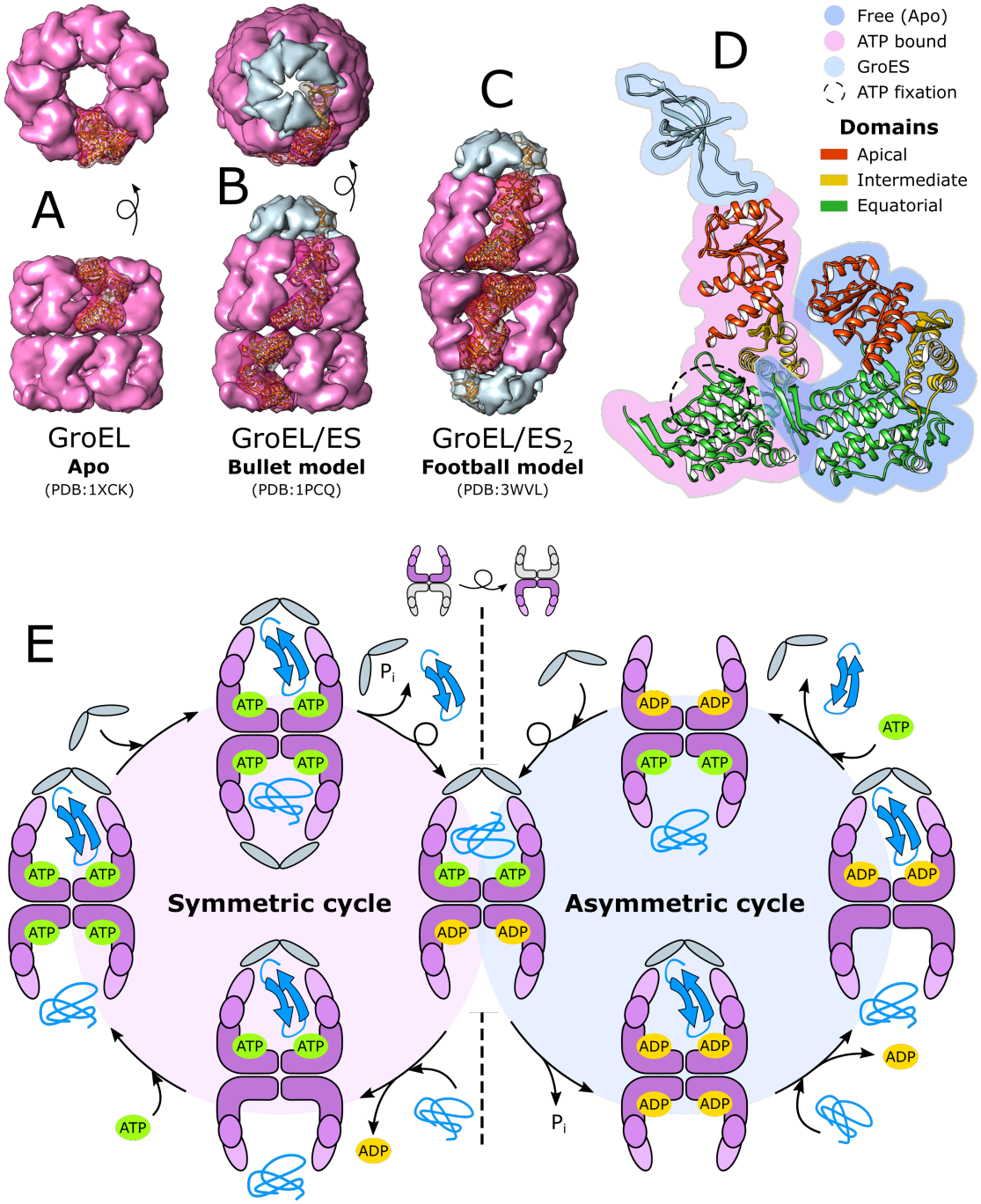


Fig. 2.10 : The type I Chaperonin. (**A**) Top and side views of Apo GroEL. (**B**) Top and side views of GroEL bound to GroES, in a bullet-shaped complex. (**C**) Side view of GroEL bound to GroES, in a football-shaped complex. (**D**) Side view of the monomers in the open (PDB : 1XCK) and closed (PDB : 1PCQ) conformations, with different colors depending on the different structural domains. (**E**) Schematic representation of the two models proposed to explain the mechanism of action of type I chaperonins, adapted from [Sameshima et al., 2010].

Refolding cycle(s) - Considering the latest discoveries on the refolding activity of GroEL/$GroES_2$, two refolding cycles are proposed (**figure 2.10, D**). The first refolding cycle is the classical one, with the formation of the asymmetric complex GroEL/GroES. This cycle occurs in the case of absence or low unfolded substrates concentrations in the vicinity of GroEL/GroES. The unfolded protein first binds the apical domain of the chaperonin due to hydrophobic interactions. GroES then binds to the ring as ATP is fixed on the equatorial domains of the seven monomers, pushing the unfolded protein inside the cavity. The substrate protein is now trapped, as it is impossible to transit from one ring to another, due to the inner shape of GroEL. In this ring, the structure of the monomers has changed due to ATP binding, exposing more hydrophilic residues (rotation of the apical domain by 90°, called hinge motion). Meanwhile, allosteric changes occur in between the two heptameric rings in order to prevent the binding of a second GroES on the other cavity. The refolding cycle takes about 20 seconds and is timed by ATP hydrolysis. Once ATP is hydrolyzed into ADP, GroES is ejected and the substrate protein can leave the cavity, refolded or not. Each cycle consumes 7 molecules of ATP [Tyagi et al., 2010]. The second refolding cycle includes the formation of the symmetric GroEL/$GroES_2$ complex, in the case of high unfolded substrate concentrations. The high concentrations in unfolded substrates seem to promote the dissociation of ADP in the ring opposed to the one where the refolding occurs, changing its conformation. The timing of the cycle is then modified, leaving the opportunity to a second GroES molecule to be bound to the opposite ring, while the first ring is still closed by a GroES molecule, leading to football-shaped GroEL/$GroES_2$.

Selectivity - Type I chaperonins interact with a broad spectrum of proteins. In the case of bacterial GroEL/ES, the cavity is 85 Å high and 80 Å wide, and could theoretically host proteins of a size up to 142 kDa. The reality is somehow different since GroEL can bind proteins of more than 70 kDa but cannot encapsulate them with GroES, the limit being around 60 kDa [Sakikawa et al., 1999]. Around 10 to 15% of the cytoplasmic proteins in *E. coli* are interacting with GroEL, and this amount increases to 30% under stress conditions [Ewalt et al., 1997]. This interactome is composed of approximately 250 proteins, of which 60 are obligate substrates of GroEL [Fujiwara et al., 2010]. Most of these proteins are aggregation-prone enzymes.

2.3.3 Type II Chaperonin

History - The Type II Chaperonins can be found in the cytosol of eukaryotes (**CCT/TRiC chaperonins**) and archaea (**Thermosomes**). Eukaryotic chaperonins are evolutionarily derived from archaea and show similar genetic origins with GroEL/ES [Bross, 2015]. The first type II chaperonin was discovered and purified in 1991 from the hyperthermophilic archaeon *Pyrodictium occultum* and called Thermosome because of its resistance to extremes temperatures [Phipps et al., 1991]. The same type of assemblies was also observed in other archaea and exhibited different symmetry compared to GroEL, as well as heterogeneous monomers (subunits α and β). Their expression was increased upon heat shock, like GroEL/ES in **E. coli**. Only one year later, the eukaryotic cytosolic equivalent of the Thermosomes was identified in human HEp-2 cells and named t-complex polypeptide-1 or TCP-1 [Lewis et al., 1992]. The complex was already believed to play a role in protein folding mechanisms. It was later renamed TRiC (**TCP-1 ring complex**) or CCT (**chaperonin containing TCP-1**), since TCP1 is only one of the 8 different monomers of CCT.

Structure - Type II chaperonin rings are more heterogeneous compared to Type I Chaperonins. Each ring can be composed of either 8 or 9 subunits, identical or not. The complexity of these rings is such that eukaryotic CCT chaperonins are composed of 8 different subunits per rings, with a defined order. Even in archaea, the subunits can be different, as proven by the a, b and c types of subunit found in the Thermosomes from the Sulfolobus genus, forming “Rosettasome” with 9-fold symmetry or chaperonin rings of 8 subunits, depending on the local abundance of monomers [Wang et al., 2010]. The structure of their apical domain is so different compared to type I chaperonins that, owing to an α-helical protrusion working as a built-in lid, they do not need the assistance of a co-chaperonin in order to close their cavities [Klumpp et al., 1997]. This apical domain seems to be specific to the type of chaperonin and does not exhibit similarities in sequence. It also seems that the cavity of type I chaperonin is bigger and is able to host bigger proteins compared to type II [Ditzel et al., 1998]. Even though the two types of chaperonins vary in their structures and folding mechanisms, there is a high degree of sequence similarities in between GroEL and Thermosomes (about 45%), especially between the different equatorial and intermediate domains.

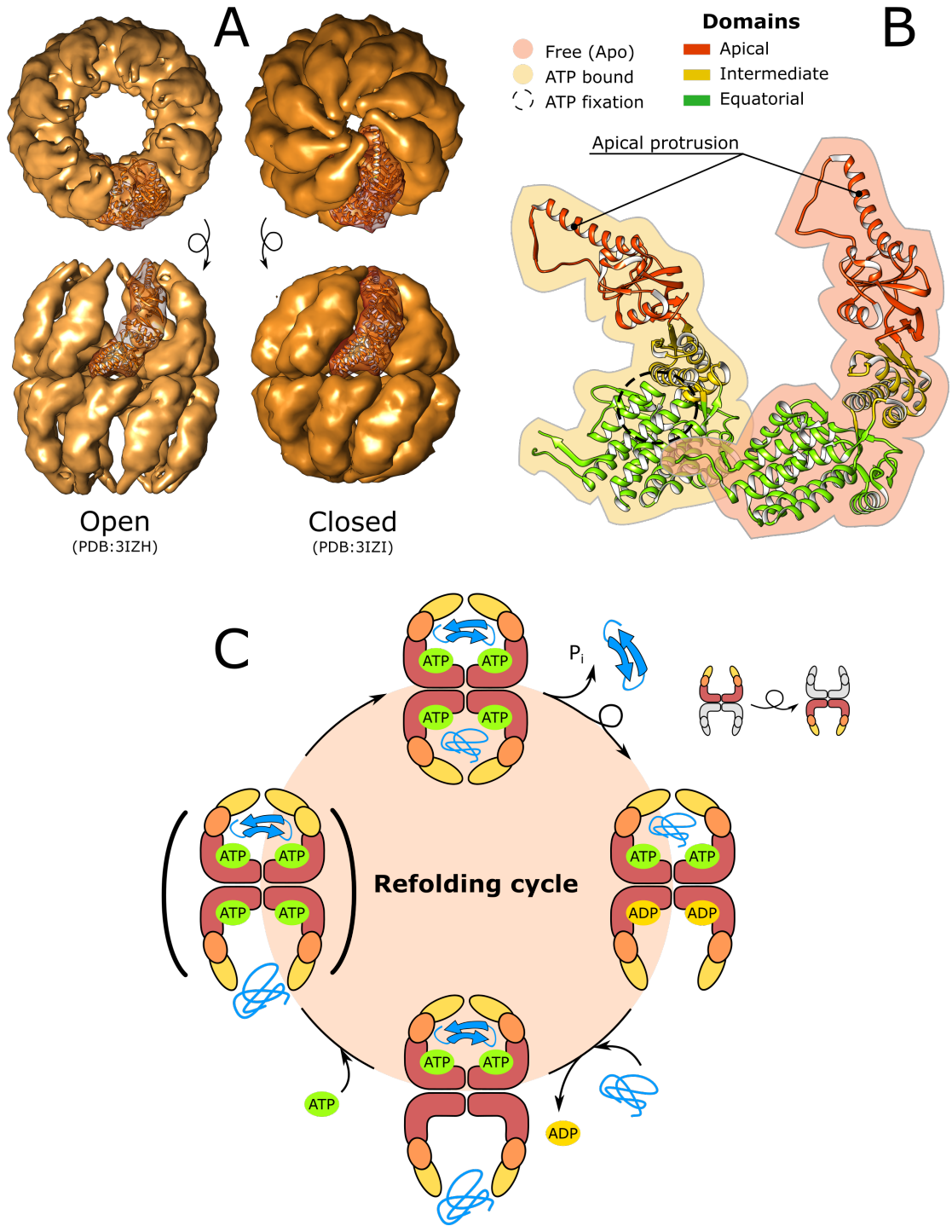


Fig. 2.11 : The type II Chaperonin. (**A**) Top and side views of the open and closed conformations of *Methanococcus maripaludis* Thermosome. (**B**) Side view of the monomers in the open and closed conformations, with different colors depending on the different structural domains. (**C**) Schematic representation of the refolding models proposed to explain the mechanism of action of type II chaperonins, adapted from [Mas et al., 2018].

Refolding cycle - The refolding cycle is similar to the one of GroEL/ES. The unfolded or misfolded protein enters the cavity, ATP binding closes the cavity, with no escape for the misfolded protein, since there is no way to transit from one cavity to the other. This structural rearrangement is possible due to a clockwise rotation of the subunits of approximately 15° and bending of the monomers to close the cavity. Once closed, the cavity exposes more hydrophilic residues, providing a more favorable environment for the substrate to refold. One ATP is hydrolyzed to ADP, the cavity opens and releases the substrate, folded or not. The refolding cycle lasts for approximately 20 seconds, depending on the chaperonin origin and environmental conditions [Mas et al., 2018]. Some Thermosomes also exhibit poor refolding properties, which seems to depend on the metabolic rate of certain archaea strains. In the case of Thermosomes, it appears that the presence of Mg^{2+} and K^{+} is of great importance for the ATP-hydrolysis activity [Wang et al., 2010].

Selectivity - Not much is known about the substrates and the selectivity of archaeal CPNs. In fact, research is mostly focused on eukaryotic CCT, as it is a more promising target to understand the mechanism of human chaperonin. Nevertheless, an interesting study was carried out to compare the substrates of type I CPN versus type II archaeal CPN [Hirtreiter et al., 2009]. Usually, type I and II CPN do not coexist but they both have been found in an archaeon of the genus *Methanosarcina*, *Methanosarcina mazei*. In this organism, the GroEL chaperonin structure is classical and the Thermosome is composed of three different subunits, in a ratio 2/1/1 and organized in octameric rings. The study shows that 13% (333) of the proteins found in the cytosol interact with the chaperonins, of which 24% (81) are GroEL specific, 45% (150) are Thermosome specific and 31% (102) interact with both CPNs. Since most of the substrates are chaperonin specific, one can conclude that type I and II chaperonins exhibit different interaction and refolding pathways. By comparing the properties of the various kinds of substrates, it seems that GroEL interacts preferentially with conserved and more hydrophobic proteins while the Thermosome is less specific and interacts with fast evolving proteins, of different and more complex folds. The average size of the substrates is of 29 kDa and 38 kDa for the Thermosome and GroEL, respectively. It was also shown that both chaperonins can interact with larger substrates of 80 to 210 kDa, with a clear preference for the Thermosome, even though such substrates cannot completely enter the chaperonin cavities. In the case of CCT-like

chaperonins, the selectivity is low with approximately 10% of the cytosolic proteome being client [Lopez et al., 2015], these substrates are often highly aggregation prone. In particular, CCT is known to fold cytoskeletal proteins such as actin and tubulin as well as cell cycle regulators. Even if its cavity is smaller compared to GroEL, it can interact with proteins bigger than 70 kDa and is involved in the association of multi-protein complexes. Human CCT mutations on subunits 4 and 5 are associated with disease and proteins linked to cancers are part of the interactome of CCT. In the *C. elegans* worm model, genes involved in the production of CCT were predicted to be linked to Alzheimer's disease [Khabirova et al., 2014]. Human CCT also interacts with the Nter of a protein associated to Huntington's disease and prevents its aggregation. Interestingly, the apical domain of its subunit 1 alone also exhibits the same behaviour, whereas the subunit 3 does not [Tam et al., 2009]. Another reason for this observed effect is that CCT deficiency causes a lack of autophagy which could also explain why the presence of CCT inhibits the aggregation of Huntingtin [Pavel et al., 2016].

2.3.4 Thermosome from *Pyroccocus horikoshii*

The work of this thesis focuses on the type II chaperonin from the archaea bacteria *Pyrococcus horikoshii*, that will now be described in more details The variety of molecular chaperones which can be found in hyperthermophilic archaea is reduced, comprising only chaperonins and small HSPs. *Pyrococcus horikoshii* is a hyperthermophilic anaerobic archaeon first discovered in 1998 [González et al., 1998], in samples collected on hydro-thermal vent chimneys in the Okinawa Trough in the Pacific Ocean (**figure 2.12, A & B**). This hyperthermophilic archaeon proliferates at 1395 m depth and its optimal growing temperature is 98°C, even though the cells can survive at 105°C for a few hours. This hyperthermophilic strain is of particular interest because of the exceptional stability of its constituting proteins, even used in the textile industry to biopolish cotton at high temperature [Ando et al., 2002]. It shares the same cocci morphology with its related other *Pyrococcus* archaea, *P. furiosus* and *P. abyssi* (**figure 2.12, C**). The genome of *P. horikoshii* was also sequenced and encodes for about 2000 genes [Kawarabayasi et al., 1998], including a type II Chaperonin, the Thermosome (**figure 2.12, D & E**) which will be abbreviated **PhCPN** in the rest of this manuscript.

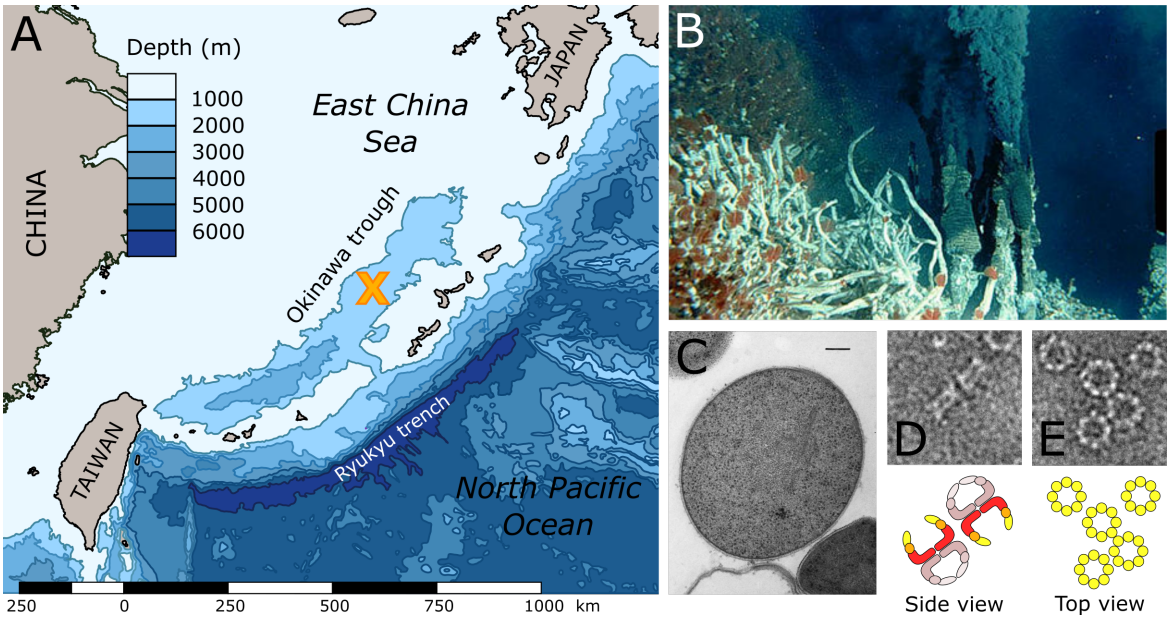


Fig. 2.12 : The Thermosome from *Pyrococcus horikoshii*, PhCPN. (**A**) Location of the Okinawa Trough and the *P. horikoshii* discovery site marked by an orange cross. (**B**) A hydrothermal vent. (**C**) Transmission electron micrograph of a thin section of *P. horikoshii*. Length of bar: 0.1 μm, from [González et al., 1998]. (**D**) & (**E**) PhCPN negative staining electron microscopy pictures of 50 x 50 nm, (**D**) picturing a side view and (**E**) a top view of the PhCPNs.

PhCPN exhibits the classical double-ring structure of type II chaperonins, with identical monomers of 59,7 kDa. *P. horikoshii* is an obligatory heterotrophic organism and needs to be complemented with Tryptophan, which maybe explains its absence in the PhCPNs sequence (see **Chap. A.1.1**). This protein, once purified, possesses a great thermal resistance, which is an undeniable quality for NMR studies. Coupled with deuteration and methyl labelling, one can easily observe an NMR spectrum of good quality, even if its size can first appear as a limiting factor. In fact, the Thermosome particle as a whole is roughly of 1 MDa. Optimisation of the purification of the protein, assignment of methyl groups (Val, Met, (**figure 2.13, C & D**)) as well as refolding cycle studies of this particular protein were carried out in our group prior to this thesis [Mas, 2015] and will not be further discussed in this section.

The exact structure of PhCPN itself is unknown. Nevertheless, another analog type II chaperonins were studied by crystallography, which allowed the determination of their structure. Homology models allow to calculate approximately the structure of PhCPN, and

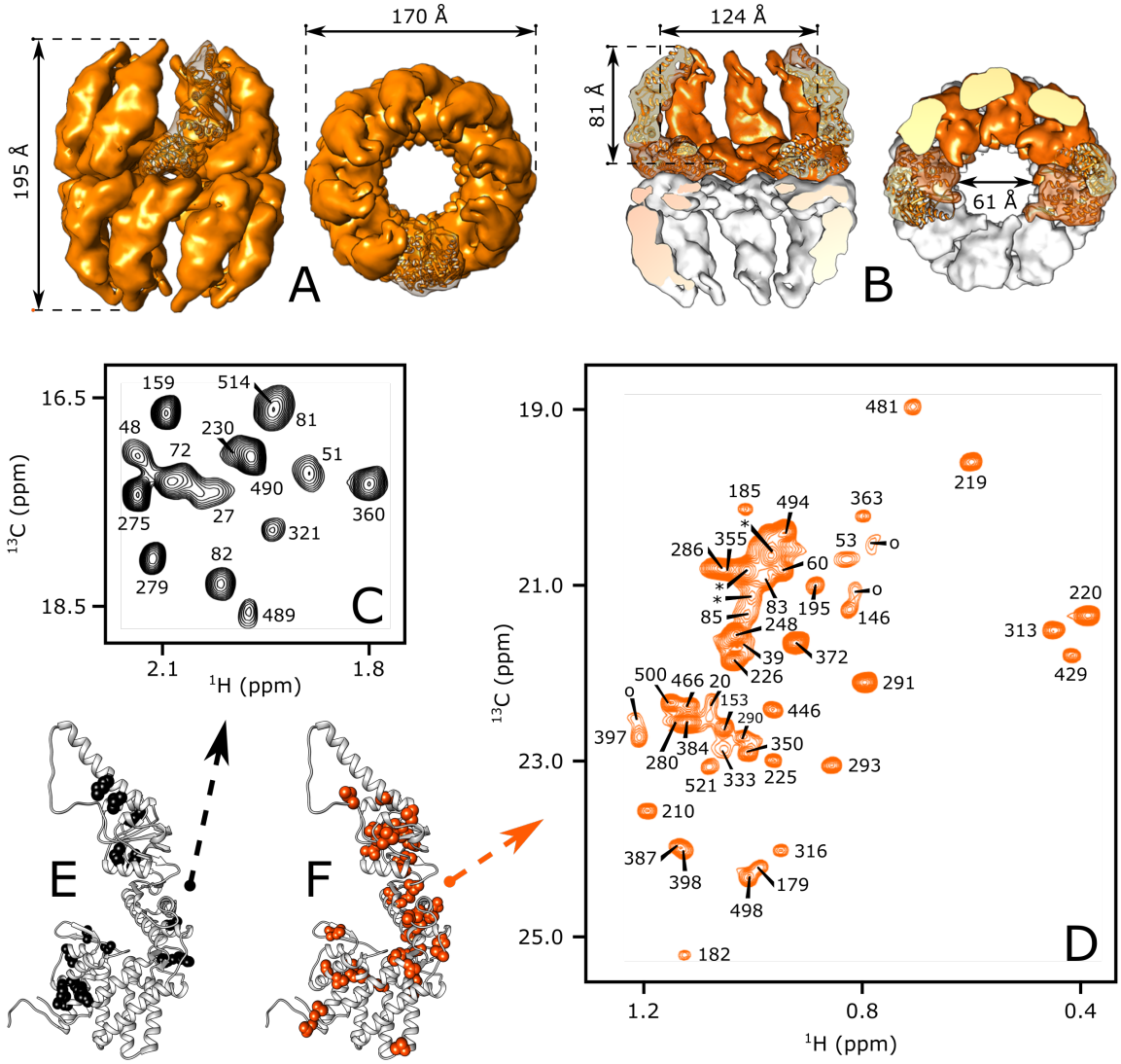


Fig. 2.13 : Dimensions of PhCPN and its Methionine and Valine NMR assignments. (**A**) Global view of the homology model of the open conformation of PhCPN, simulated with the SwissProt structure calculation tool. (**B**) Section of the homology model of PhCPN to show its internal structure and dimensions. (**C**) Attribution of the Methionine methyl groups of PhCPN. HMQC (^{1}H, ^{13}C) spectrum of the Methionine region of PhCPN recorded at 75°C, in D_2O PhCPN buffer, on an NMR spectrometer operating at a proton frequency of 850 MHz. (**D**) Attribution of the Valines methyl groups of PhCPN. HMQC (^{1}H, ^{13}C) spectrum of the Methionine region of PhCPN recorded at 75°C, in D_2O PhCPN buffer, on an NMR spectrometer operating at a proton frequency of 850 MHz. (**E**) Modelisation of the Methionines of PhCPN on its monomer structure, obtained by homology. (**F**) Modelisation of the Valines of PhCPN on its monomer structure, obtained by homology. Spectra (**A**) & (**B**) are adapted from [Mas, 2015].

the SwissProt structure calculation website in particular (**figure 2.13, A & B**). The detailed analysis of the structure of the open conformation of PhCPN is described in **Chap. 6.7**.

2.3.5 Type II Co-Chaperonin : Prefoldin

Even though type II CPN do not require a co-chaperone to close their cavity, they still need the assistance of the heterohexameric chaperone prefoldin (**PFD**), in order to recruit unfolded substrates in vivo. Prefoldin was first identified in eukaryotes in 1998, as a necessary protein which binds unfolded Actin [Vainberg et al., 1998], and which is involved in Tubulin biogenesis [Geissler et al., 1998].

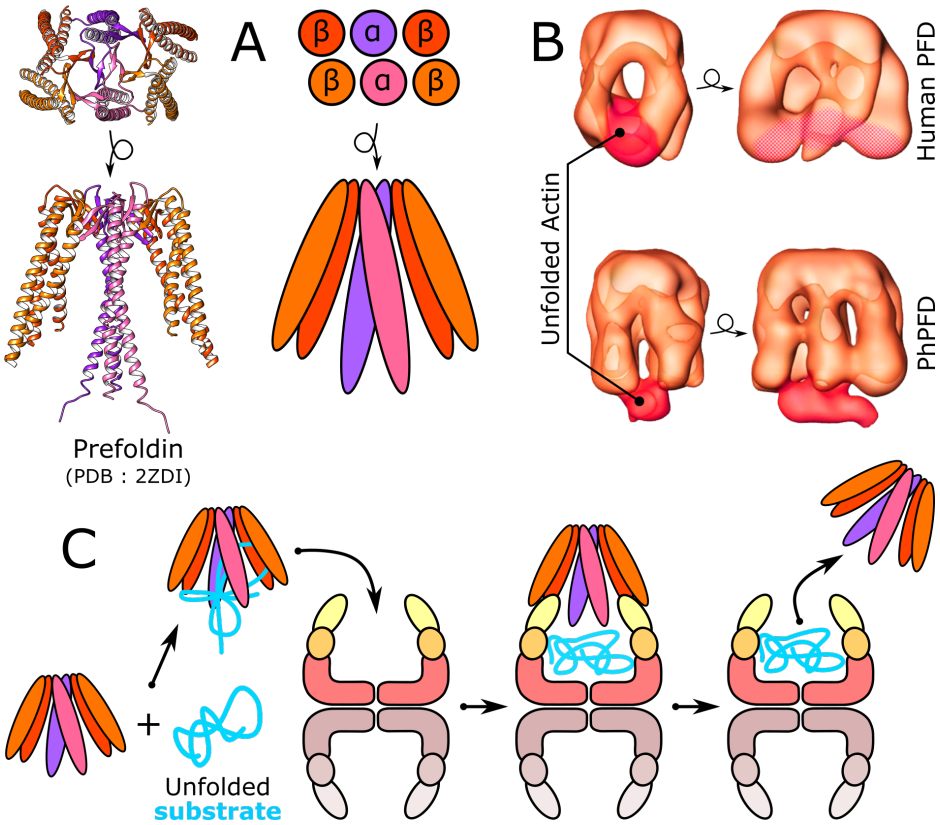


Fig. 2.14 : Structure of the Prefoldin and its binding to type II Chaperonin. (**A**) Structure and scheme of the Prefoldin from *Pyrococcus horikoshii*, PhPFD. (**B**) Binding of unfolded Actin by human PFD and PhPFD, adapted from [Martín-Benito et al., 2007]. (**C**) Scheme of the binding of unfolded substrates by PFD and its transfer to type II chaperonin.

As for chaperonins, eukaryotic PFD is made of 6 different subunits noted 1-6, whereas archaeal PFD exhibits a lower complexity, with two α and four β subunits only [Cowan

and Lewis, 2001] ; (**figure 2.14, A**)), with a global molecular weight of about 80 kDa. Its shape can be compared to a jellyfish, the bell consisting of a double β-barrel and the tentacles of six coiled coils sticking out of it. These coiled coils are exhibiting hydrophobic grooves oriented towards the inside of the protein, especially on the β subunits [Sahlan et al., 2018], where the unfolded substrates are believed to interact (**figure 2.14, B**). In fact, the Nter and Cter regions of the β subunit are proven critical for interaction with the substrates [Zako et al., 2016]. PFD exhibits interesting plasticity as it is able to change its conformation when bound to substrates, owing to a conformational change of the tentacles [Martín-Benito et al., 2007]. Unlike for chaperonins, ATP hydrolysis is not needed to provide sufficient energy to fuel these conformational changes and the transport of unfolded substrates. Prefoldin is able to transfer unfolded proteins to the chaperonin without releasing the substrate protein into the solution [Vainberg et al., 1998]. Once the substrate is in close contact with the chaperonin, it is irreversibly transferred to the CPN [Sahlan et al., 2018] ; (**figure 2.14, C**)). The role of PFD within cells is to protect unfolded or nascent proteins from aggregation. The substrate binding mechanism differs from archaeal PFD (bound at the end of the tentacles tips) to eukaryotic PFD (bound and encapsulated inside the cavity), independent of the substrate size ([Martín-Benito et al., 2007] ; (**figure 2.14, B**)). Archaeal PFD has been shown to protect citrate synthase from thermal aggregation and to slow down the spontaneous refolding of denatured GFP, as it traps its unfolded state [Okochi et al., 2002]. GFP can then be properly refolded after the addition of archaeal CPN and ATP. The β subunits alone are proven to decrease the risk of aggregation too but are still less efficient than the whole PFD. Human PFD is also able to reduce the fibrillation of Amyloid-β 1-42 (**A**β), involved in Alzheimer's disease [Sörgjerd et al., 2013]. Interestingly, the soluble Aβ aggregates found in the presence of human PFD or archaeal PFD differ in structure and in toxicity, probably due to a difference in mechanism for both PFDs. More generally, dysfunction of PFD is a cause of disease, leading to increased protein aggregation within cells, and thus cell death [Takano et al., 2014].

2.3.6 NMR studies of Chaperonins and their substrates

While chaperonins are ubiquitous proteins in the three domains of life, the NMR interaction studies of chaperonins and their substrates are mainly focussed on the type I

chaperonin GroEL/ES from *E. coli*. The first study of GroEL interacting with a substrate by NMR was carried out with unfolded RuBiSco [Shtilerman et al., 1991]. The interaction was demonstrated by comparing the hydrogen/tritium exchange rate (similar to H/D exchange, see **Chap. 6.3.1**) of unfolded RuBisCo protons by NMR in presence or absence of GroEL, GroES and ATP. Most of the protons were in fast exchange, except for a core of 12 protected hydrogens in slow exchange, possibly due to residual secondary structure. No difference of exchange in presence of GroEL and ATP without GroES could be monitored. However, after the addition of GroES, the slowly exchangeable protons became more accessible to solvent, showing an unfolding event caused by the whole GroEL/ES+ATP machinery. In addition, in certain conditions, some unfolded RuBisCo could reach its native state, while it was impossible without GroEL/ES. This study paved the way for further investigations of the unfoldase and foldase activity of chaperonins. Later, GroEL was also shown by H/D exchange NMR and lineshape analysis to promote unfolding with residual secondary structure elements of the Cyclophilin [Zahn et al., 1994], Barnase [Zahn et al., 1996c], DHFR [Goldberg et al., 1997] and human β2GPI [Gozu et al., 2002] protein substrates. Especially in the case of Cyclophilin, the interaction with GroEL was shown to be transient and highly dynamic, with the protein equilibrium shifted towards the unfolded state while protected from aggregation [Nieba-Axmann et al., 1997]. In addition, the unfolding pattern of Cyclophilin in solution or when bound to GroEL/ES was proved different, with a final unfolded protein exposing more hydrophobic residues in presence of the chaperonin. The refolding rate of Cyclophilin was accelerated by the presence of GroEL/ES, even without ATP. The global unfolding performed by GroEL/ES could be useful for efficient refolding especially in the case of substrates trapped in low energy minima. Similar studies were carried out with a single ring variant of GroEL, SR1. This mutant was proven to bind human DHFR in a random coil conformation [Horst et al., 2005]. Particularly, the bound substrate was highly mobile, sampling different unfolded conformations, possibly inside the cavity. Similar unfolding was observed with SR1 or GroEL/ES, with higher peak intensities with SR1 probably due to different relaxation times, confirming the validity of the model. The refolding of DHFR was also monitored after denaturation, in presence or absence of SR1, GroES and ATP [Horst et al., 2007]. The same folding trajectories were observed inside and outside of the chaperonin cavity by H/D exchange, contradicting previous observations. Refolding speeds were also similar but the refolding was two times higher in presence of the chaperonin, due to reduced

misfolding and aggregation. In that case, the chaperonin was proven to act as a foldase only, preventing misfolding and aggregation with shielded refolding in its cavity, without interfering in the refolding pathways of the substrate protein. A similar observation was carried out by [Koculi et al., 2011] on Rhodanese.

Interestingly, the interaction between GroEL and peptides was also investigated and showed strikingly different results. While completely disordered in solution, the vsv-C peptide was shown to retained an α-helical secondary structure in presence of GroEL, with mobile sidechains [Landry et al., 1992]. Additional investigation with derivatives from the Rho peptide by nOe and lineshape analysis showed that α-helices as well as β-strands secondary structure elements can be formed or retained upon GroEL binding [Wang et al., 1999]. Especially, binding of the peptides correlates quantitatively with their hydrophobicity/exposed hydrophobic surfaces. Such an interaction was believed to be driven by hydrophobic interactions between the apical domains of the GroEL monomers and the exposed hydrophobic residues of the peptide. A more recent study of the Aβ40 peptide bound to GroEL was carried out, this time aiming at investigating the so-called "dark state" of the substrate directly bound to GroEL [Libich et al., 2013]. In fact, since GroEL is very large and tumbles slowly, its magnetization relaxes faster and the recorded signal is broadened, most of the time beyond detection, "hiding" the substrate in its bound state by artificially increasing its apparent size. Using relaxation based NMR experiments including lineshape analysis, CPMG and DEST (see **Chap. 6.3.1** and **Chap. 6.3.1** respectively), the transient nature of the interaction could be observed again and the lifetime of the complex (< 1 ms) computed. The K_D of the interaction between one Aβ40 and one GroEL cavity was also determined (70 µM). In addition, the peptide was bound to GroEL through interaction with its hydrophobic regions, in an extended unfolded state, contrary to what was observed before with other peptides. It is very likely that Aβ40 exchanges between different subunits on a very fast time scale compared to the complete lifetime of the bound state. A study based on the same NMR experiments, also carried out by the group of Dr. Marius Clore, this time focusing on a metastable protein bound to GroEL [Libich et al., 2015]. With help from methyl labelling and multinuclear relaxation-based NMR, the interconversion rate between the native state and a folding intermediate of a mutant of the SH3 domain was explored in presence of GroEL. This exchange rate was shown to be accelerated more than 20 times by the presence of GroEL, while less than 1% of the total substrate quantity was bound to GroEL. The substrate was most likely bound

inside the cavity, with a K_D of 60 μM for its partially folded intermediate and 7 mM for the completely folded substrate.

Most of the studies focussed on the substrate location but little was known to where the binding really occurred since GroEL is a very large protein assembly and tumbles slowly (see **Chap. 6.3.2**). Nevertheless, mutants of GroEL were first engineered to overcome this problem (in absence of methyl labelling at the time) and gain information on the substrate binding region of GroEL. The isolated apical domain of GroEL was expressed and purified, and its interaction with the Rho protein was investigated since the apical domain was believed to be the main interaction site of chaperonins [Kobayashi et al., 1999]. Similarly as in presence of intact GroEL, the peptide maintained secondary structure while interacting with the apical GroEL, as showed by intramolecular nOes. This time, the apical binding site could be identified by mapping CSP on the sequence and was shown to be located next to two helices, H8 and H9. The apical domain of a GroEL archaeal homologue was also shown to interact with insulin [Hua et al., 2001]. However such studies were incomplete and with the development of methyl labelling, studying the binding site on the whole chaperonin particles became possible. Unlike what was reported for previous substrates and especially peptides, the HEWL substrate protein was proven to be hosted within the very core of the PhCPN [Mas et al., 2018]. In fact, a combination of DOSY and PRE data showed successively that both proteins were exhibiting the similar translational diffusion properties when interacting, and that signals corresponding to methyls located inside the cavity were affected by the presence of spin-labelled unfolded HEWL. Methyl labelling of the substrate Ile and Val gave information on the folded state of the HEWL substrate, even when bound to large PhCPN. The presence of PhCPN displaces the equilibrium between the folded and unfolded state of the substrate at high temperature towards the unfolded state.

Recently, an extensive study by fast liquid-state NMR of the closure mechanism of the cavity of the type II chaperonin from *Pyrococcus horikoshii* PhCPN shed light on the mechanisms of these less studied (at least by NMR) chaperonins [Mas et al., 2018]. Especially, there is still to date a debate whether the closing of the cavity is caused by ATP hydrolysis [Douglas et al., 2011] or by ATP binding only. Using methyl-labelling techniques, it is possible to differentiate the populations of open and closed states of PhCPN cavities, since some methyl groups are located in the lid of the apical domain. Additional methyl groups located near the ATP binding site also give information on the nucleotide binding

state. With help from these structural reporters, it was shown that closing is caused by the binding of ATP, especially because binding of an non-hydrolysable ATP-analog is enough to close the chaperonin cavities. In addition, by following the populations of the open and closed states of PhCPN as a function of time, it was possible to determine a precise ATP hydrolysis cycle of the chaperonin.

This cycle is actually accelerated in presence of a substrate protein. Another study monitored the degrees of freedom of the Cter of PhCPN and the Nter of the tentacles of its co-chaperone, the Prefoldin of *Pyrococcus horikoshii* (**PhPFD**) [Kurimoto et al., 2007]. The two different subunits α and β of PhPFD were labelled on Methionines, since Met residues are located at the Nter of the subunits only and thus report their conformational changes. The Methionines were assigned by deconstruction and reconstruction of the complex with labelled or unlabelled subunits. The evolution of the methyl signal upon addition of PhCPN was monitored and showed a decrease in peaks intensity. Especially, the M11 residue of the β subunits was dramatically affected, which was possibly caused by restrained dynamics caused by the interaction with PhCPN. To confirm this hypothesis, the 12 first residues of the β subunits were removed by mutagenesis, resulting in impaired interaction with PhCPN. The deletion of this region also perturbs substrate binding, possibly indicating a central role of this Nter in the substrate transfer from PhPFD to PhCPN. Inversely, the Cter Met residues of PhCPN are very mobile as proven by their sharp NMR signal and remain highly mobile upon PhPFD binding.

Chapter 3

Oligomeric state of PhCPN

Le Thermosome de l'archée Pyrococcus horikoshii, PhCPN, est un homo-hexadécamère. Provenant d'une archée hyperthermophile, PhPCN est de nature très stable. Cependant, la présence d'ions magnésium à l'intérieur de ses cavités est nécessaire pour maintenir son état oligomérique. Ce court chapitre prouve qu'il est possible de dé-oligomériser PhCPN en utilisant différents tampons pauvres en sels et en magnésium. Il montre également qu'il est nécessaire de choisir des conditions de tampons compatibles avec l'état oligomérique de PhCPN souhaité.

PhCPN is a type II homo-hexadecameric chaperonin, its overall structure forming characteristic two stacked rings with a cavity in their respective centers (**figure 3.1, A**). Even though PhCPN is stable at high temperatures due to its thermophilic archaeal origin, the lack of salts in the working buffer can lead to destabilization of its oligomeric structure. This short introductory chapter is meant to draw the reader's attention on the importance of the buffer composition in the results which will be described afterwards in the following results chapters. Most of the experiments described in this manuscript were performed in PhCPN buffer (**figure 3.1, A**), a buffer containing 50 mM NaCl, 25 mM $MgCl_2$, 100 mM KCl and 50 mM MES pH 6.5, which was optimized to increase the ATP hydrolysis of PhCPN [Mas, 2015].

In this buffer, PhCPN is hexadecameric and its dimensions measured by Electron Microscopy or EM (**figure 3.1, A**), are in agreement with the simulated structure of PhCPN presented in **Chap. 6.7** (**figure 3.1, B & C**). Even if highly resistant to heat, PhCPN is sensitive to buffering conditions, to maintain the cohesion of its hexadecameric structure. As an example, if EDTA is used as an antiprotease during the purification process in PhCPN buffer, it chelates magnesium ions and the yield in final hexadecameric PhCPN decreases dramatically, while seeing the number of PhCPN aggregates increase (personal observation, **data not shown**).

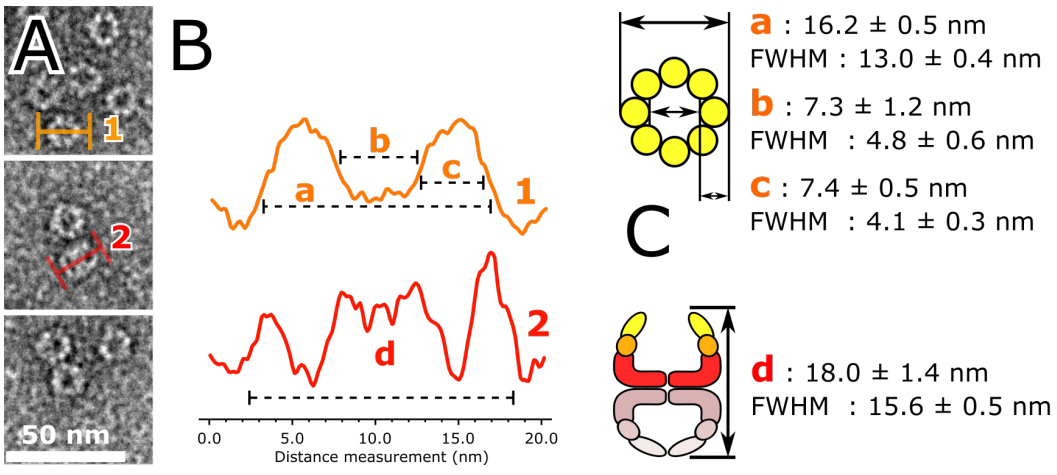


Fig. 3.1 : PhCPN dimensions analyzed by EM. (**A**) Negative staining EM pictures of PhCPN profile and top views. (**B**) Gwyddion software measurements of PhCPN dimensions. (**C**) Precise measurements of PhCPN in PhCPN buffer.

Inversely, the amount of NaCl in PhCPN buffer does seem to influence much the ATPase activity of PhCPN, as proven by ATPase improvement assays [Mas, 2015]. For this reason, the stability, des-oligomerization and re-oligomerization of PhCPN were tested in diverse buffer conditions. Especially, the influence of Mg^{2+} ions on the oligomeric state was investigated.

PhCPN is in its normal hexadecameric state in PhCPN buffer (**figure 3.2, A**). In absence of salts (the sodium ions brought by pH equilibration of the MES are neglected), the amount of visible hexadecameric PhCPN is drastically decreased while small particles can be distinguished at the background of the images (**figure 3.2, B**). This results would suggest that PhCPN is mainly monomeric in such buffer, due to the lack of salts (**figure 3.2, C**). To further investigate this hypothesis and precisely characterize the oligomeric state of PhCPN in each buffer, samples were analyzed by SEC-MALLS. PhCPN was des-oligomerized in pH 6.5 MES buffer without salts, in which PhCPN is believed to be mainly monomeric as suggested by the EM images. The buffer was re-exchanged with the buffer of choice, to attempt re-oligomerization of PhCPN (**figure 3.2, E**). In the case of PhCPN buffer, the homo-hexadecameric PhCPN particle elution volume is of 1.5 mL and exhibits a molecular weight of approximately 1 MDa, which was to be expected. In the case of MES buffer, a 50 mM pH 6.5 MES buffer without salts, the main peak is composed of monomers and is eluted at 1.8 mL. Three other samples containing separately each salt composing the PhCPN buffer were also tested, in order to determine which salt was the most important to maintain the hexadecameric state of PhCPN. The presence of NaCl and KCl did not promote the formation of the hexadecamer and PhCPN remained mainly monomeric, with an average molecular weight of approximately 70 kDa. In the case of $MgCl_2$ however, its presence promoted greatly the formation of the hexadecamer, with very few monomers left in solution. A similar effect of $MgCl_2$ was previously reported for type I Chaperonin GroEL des-oligomerization and re-oligomerization [Wälti et al., 2017].

The influence of this salt was more precisely tested to determine the minimum amount of $MgCl_2$ necessary to form the hexadecamer (**figure 3.2, E**). The influence of $MgCl_2$ even in low concentrations is significant and can be detected on the SEC profile, with drastic decrease in the monomer peak intensity and a progressive increase of the whole PhCPN particule amounts. In the presence of 25 mM $MgCl_2$, the sample is composed almost only of the hexadecamer and the SEC and MALLS profiles are similar to the one of PhCPN in PhCPN buffer.

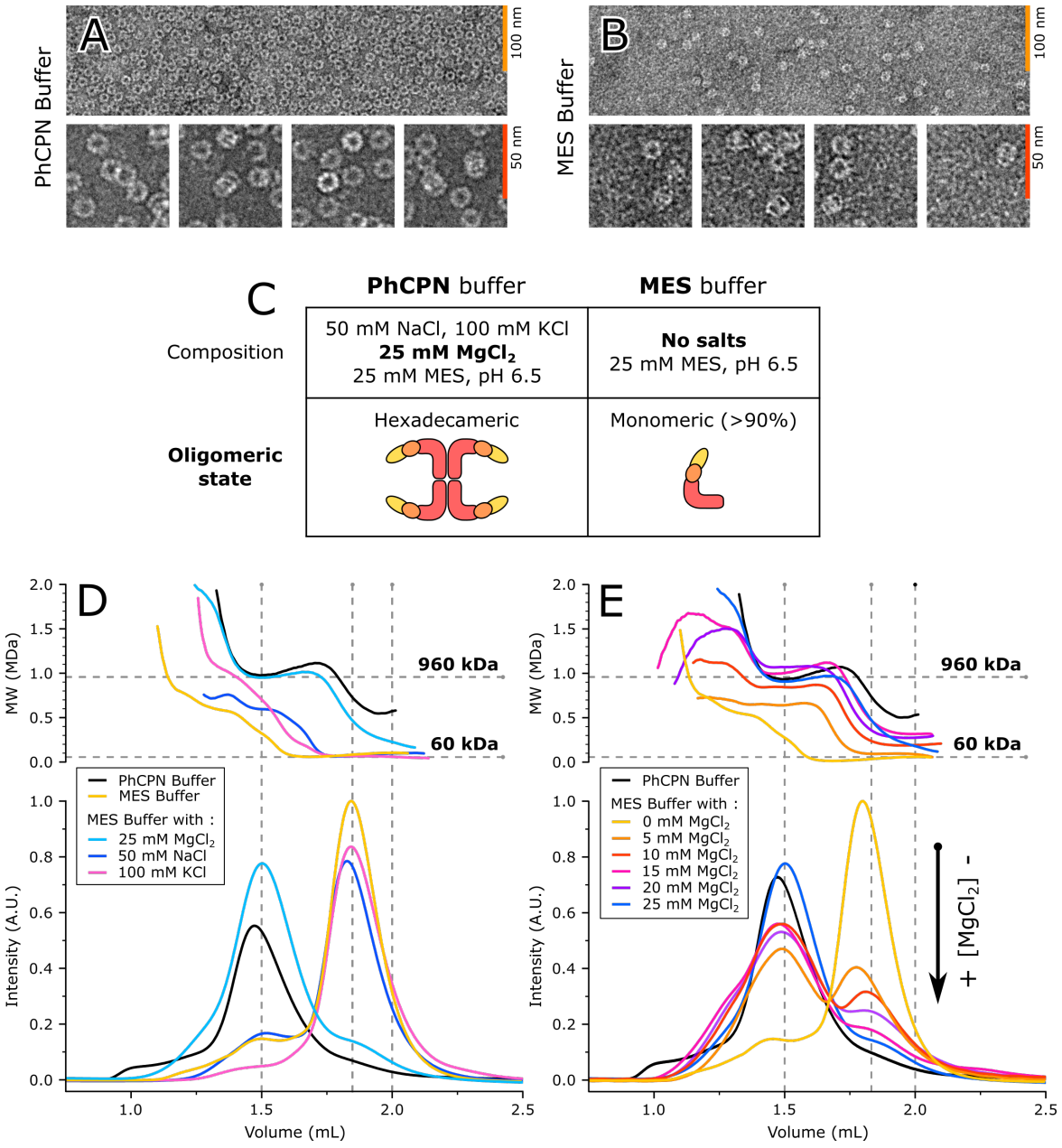


Fig. 3.2 : Oligomeric state of PhCPN in different buffers. (**A**) & (**B**) Ammonium molybdate negative staining EM pictures of (**A**) 0.2 µM PhCPN in PhCPN Buffer and (**B**) 0.2 µM PhCPN in MES Buffer. (**C**) Composition of the MES and PhCPN buffers and the corresponding PhCPN oligomeric state. (**E**) Sec-MALLS profiles of 10 µM PhCPN in different buffers. (**D**) Sec-MALLS profiles of 10 µM PhCPN in buffers containing different amounts of $MgCl_2$. (**F**) Oligomeric state of PhCPN in PhCPN and MES buffer.

From this experiment, it is obvious that $MgCl_2$ is of paramount importance for the formation of the normal hexadecamer quaternary structure of PhCPN. It is possible to play with the $MgCl_2$ concentration in order to modify the oligomeric state of PhCPN. In fact, knowing interaction patterns with monomers or the whole particle in very similar buffers could allow monitoring the effect of the presence of the cavities and the eventual cooperation in between the monomers. Nevertheless, when possible, PhCPN buffer remained the buffer of choice, PhCPN buffer. In **Chap. 5**, both MES and PhCPN buffers are used (**figure 3.2, E**), and concentrations are always expressed as 16-mer concentration, even in the case of monomeric PhCPN in MES buffer, for clarity purposes. These results were obtained at the end of the thesis lab work and the results which are presented in the next chapters were processed and interpreted accordingly to consider the oligomeric state of PhCPN.

Chapter 4

Amorphous aggregating protein model

Le Thermosome de l'archée *Pyrococcus horikoshii*, PhCPN, est capable de protéger ses protéines substrats de l'agrégation. Ce chapitre résume l'étude de PhCPN en interaction avec la Malate Synthase G (**MSG**), une protéine de taille importante sujette à l'agrégation amorphe lorsqu'elle est soumise à des températures importantes. Tout d'abord, le rôle protecteur de PhCPN est mis lumière en prouvant qu'il protège MSG de l'agrégation, même à des températures très élevées. La RMN combinée à d'autres techniques biochimiques et biophysiques permettent de montrer que PhCPN maintient MSG dans un état déplié afin de la protéger de l'agrégation. Il est également montré que MSG est maintenue au centre de la cavité de PhCPN, dans un environnement la séquestrant du reste des protéines mal repliées. PhCPN ne peut par ailleurs protéger qu'une macromolécule de MSG à la fois, malgré la présence de ses deux cavités distinctes. Il est proposé que cet effet est dû à la taille importante de MSG, perturbant le fonctionnement normal de PhCPN.

PhCPN (*Pyrococcus horikoshii* Chaperonin) is a chaperonin which interacts with unfolded proteins (see **Chap. 2.3**). Here is reported a study on one group of substrates : aggregating substrates. Since PhCPN is highly resistant to heat, it is possible to study its interaction with an aggregating substrate, at high temperatures. Previous work in the IBS NMR group used as a model substrate Hen egg white lysozyme (**HEWL**) [Mas, 2015, Mas et al., 2018]. It was proven that PhCPN interacts only with unfolded HEWL, with a stoichiometry of one substrate protein per PhCPN cavity. This study also showed that the melting temperature of HEWL is decreased by the presence of PhCPN as the chaperonin stabilizes the unfolded substrates ; and that the unfolded substrate is hosted at the bottom of PhCPN cavity [Mas et al., 2018]. This chapter describes the study of PhCPN interaction with a very large protein substrate compared to previously studied small globular proteins, including :

- The selection of the aggregating protein substrate
- The investigation of the substrate aggregation process by NMR
- The isolation of the Substrate/PhCPN complex and the determination of its stoichiometry
- The mapping of the substrate interaction site with PhCPN
- The eventual refolding effect of PhCPN on the substrate in presence of ATP

As PhCPN and the studied substrate are two very large assemblies and tumble slowly, NMR combined with methyl-specific labelling had to be employed (see **Chap. 6.3.2**), as well as the combination of other biophysical techniques, precisely detailed in **Chap. 6**.

4.1 Selection of the protein model

Chaperonins like PhCPN are ubiquitous and are able to interact with a various set of substrates (see **Chap. 2.3** and **Chap. 2.3.6**). However, not all of these substrates are suitable for studies, particularly involving liquid-state NMR. The following sections describe the substrate requirements, the substrate selection process and provide details about the model substrate selected for further investigations. The first part of this project tested

the feasibility of studying the interaction between PhCPN and an aggregating protein by liquid state NMR coupled to a set of various biophysical techniques. In order to carry out such study, several requirements must be fulfilled by the model protein substrate :

- Easy to unfold using elevated temperatures or low amounts of denaturants
- Unfolding and aggregation at temperatures reachable by NMR instruments (below 75°C). This is particularly important as some NMR probes components are sensitive to heat and the cryoprobes of the IBS institute are unfit to heat above 75°C
- Suitable for expression in E. coli, with acceptable protein yield in M9-D_2O medium allowing labelling of specific methyl groups. Nowadays, methyl labelling is also possible in insect cells or cell-free but, at the time of the study, these techniques were still under development in the NMR group.
- Previously studied by NMR, with a preferable available assignment of the (methyl) peaks
- Fits in PhCPN cavity, even though there is no consensus on the accommodation size limit of PhCPN [Hirtreiter et al., 2009]

The screening was principally focussed on large size substrates to complement the work and observations obtained prior to this thesis with HEWL (14.5 kDa, “medium” size, [Mas et al., 2018] and with Amylin (4 kDa, “small” size) , presented in the next chapter (see **Chap. 5**). Several proteins met the requirements specified above. In fact, literature is rich with examples of protein models used for interaction studies with chaperonins. Particularly, prokaryotic GroEL/ES was extensively studied in interaction with various substrates, since its structure is simpler compared to archaeal and eukaryotic type II chaperonins (see **Chap. 2.3.3**). Among them, three substrates were already expressed, labelled and routinely studied in the NMR group : Ubiquitin [Brutscher et al., 1997], Green Fluorescent Protein [El Khatib et al., 2016, Makino et al., 1997] ; rsFolder, a variant, is studied in the NMR group ; Dr. Bernard Brutscher & Isabel Ayala, personal communication) and Malate Synthase G [Ayala et al., 2009].

4.1.1 Ubiquitin

The first protein to be tested was *Homo sapiens* Ubiquitin with an uncleaved His-tag (**figure 4.1, A**). Such a protein model is convenient since it is extensively studied by NMR and yields excellent quality spectrum when alone and (^{15}N, ^{13}C)-labelled or methyl labelled in solution (8.5 kDa). Its aggregation temperature was first monitored by light scattering in a fluorimeter, using a protocol adapted from [Dahiya and Chaudhuri, 2014] and described in **Chap. 6.5.1**. The protein was first exposed to a temperature ramp ranging from 20 to 90°C at low protein concentration (0.9 mg/mL) and did not show aggregation. At a higher concentration however (2.0 mg/mL or 78 µM), the protein started to aggregate above 80 °C (**figure 4.1, B**). Even though temperatures above 80°C are not fit for the NMR available equipment for the reason presented above, the impact of PhCPN on Ubiquitin aggregation was tested.

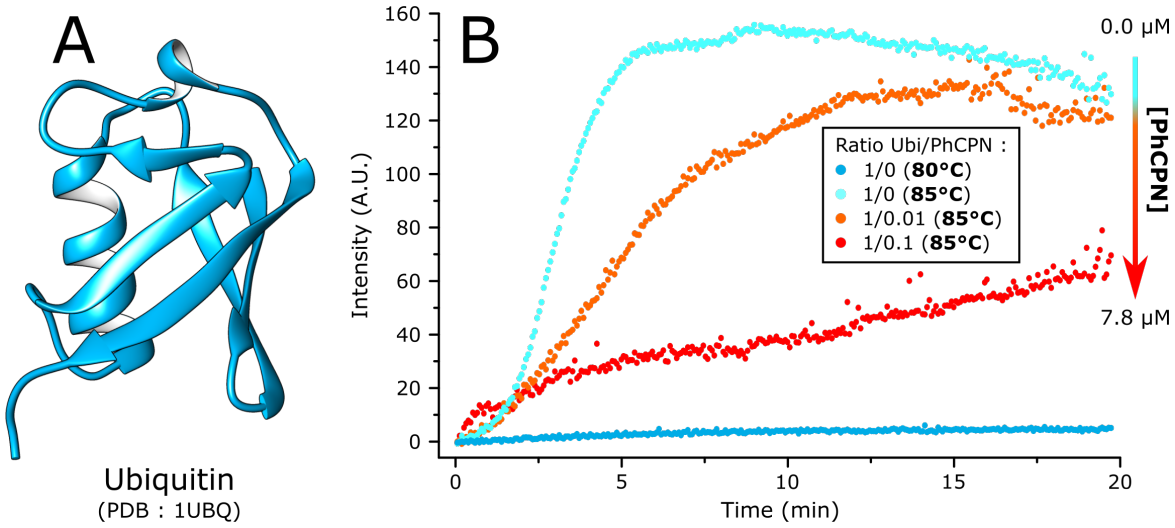


Fig. 4.1 : First interaction trials with Ubiquitin. (**A**) X-ray structure of *Homo sapiens* Ubiquitin (PDB : 1UBQ). (**B**) Light scattering measured at various temperatures and in the absence/presence of PhCPN. Residual background scattering from both buffer and PhCPN were subtracted before plotting.

The influence of PhCPN, even at low concentration (1 for 100 Ubiquitin units), on Ubiquitin aggregation is clearly visible at 85°C. Aggregation is not completely prevented but the amount of aggregated protein present in solution is dramatically decreased. Next step was to try to isolate a complex between PhCPN and Ubiquitin by gel filtration. Unfortunately, the complex could not be isolated at room temperature, likely because of the

fast exchange between PhCPN-bound and free unfolded Ubiquitin or the spontaneous refolding of interacting Ubiquitin at room temperature.

4.1.2 rsFolder

The second protein tested was rsFolder, a synthetic protein produced by another group of the IBS [El Khatib et al., 2016] (**figure 4.2, A**). rsFolder is a reversibly switchable fluorescent protein designed by hybridization of two GFP variants, Superfolder-GFP (a GFP variant engineered for fast folding kinetics) and rsEGFP2 (an EGFP variant switching from an on to an off state efficiently at low illumination power and able to sustain numerous on-off switching cycles). GFP interaction with GroEL/ES was already proven [Makino et al., 1997] but GFP has a tendency to aggregate at NMR concentration. It was therefore replaced by rsFolder, a particularly thermostable variant of GFP, which gives a great quality NMR spectrum considering its size (27 kDa), and can be methyl labelled (personal communication, Bernhard Brutscher and Isabel Ayala, NMR group). This protein would also have been particularly interesting to study because of its β-barrel tertiary fold, which differs greatly from the previously studied Lysozyme.

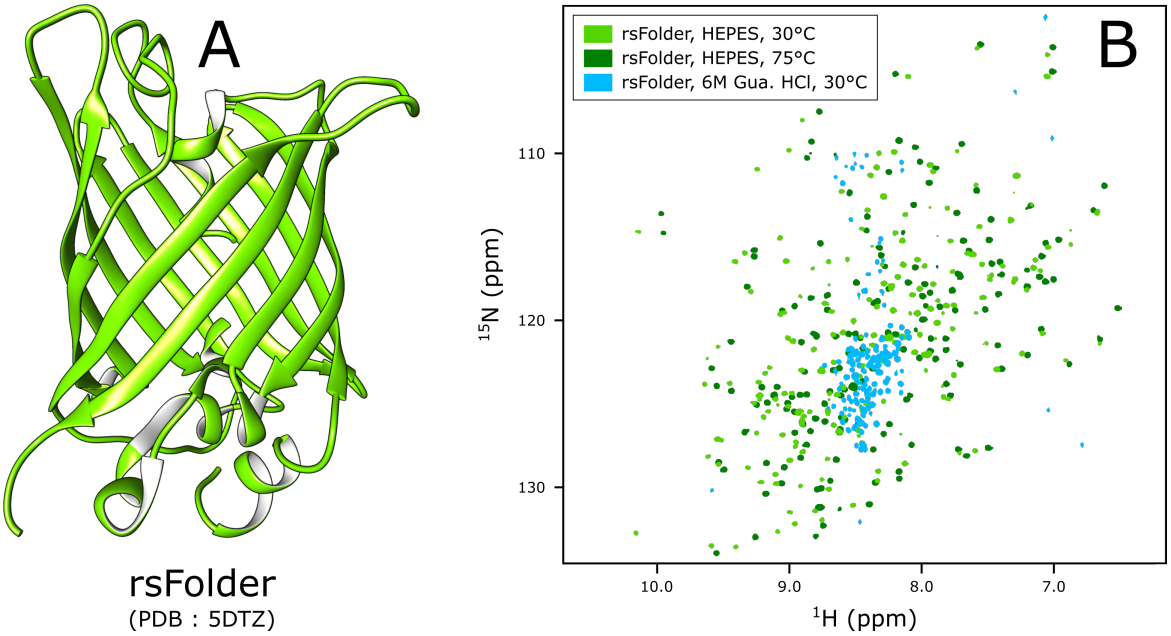


Fig. 4.2 : Unfolding tests on rsFolder. (**A**) Crystal structure of rsFolder in the fluorescent state (PDB : 5DTZ) from [El Khatib et al., 2016]. (**B**) BEST-TROSY $_1$H-$_{15}$N spectra of 100 µM (^{13}C,^{15}N)-rsFolder at different temperatures and in 50 mM HEPES pH 7.5 buffer and 6 M Guanidinium chloride buffer.

Remarkably, rsFolder is so stable that it was impossible to get the protein to unfold and aggregate within a reasonable temperature range, as checked by NMR (**figure 4.2, B**) and light scattering (**data not shown**). Nevertheless, another test was set to determine more drastic conditions under which the protein would unfold. Different amounts of urea, as well as guanidine-HCl, were added to ^{15}N-labelled rsFolder and its NMR spectrum was recorded to determine its folding state (**figure 4.2, B**). rsFolder could not be unfolded at an acceptable temperature range and needed to be put in presence of substantial amounts of denaturant (Guanidine 6 M) to unfold, conditions which are incompatible with PhCPN.

4.1.3 Malate Synthase G

The next protein model tested was *E. coli* MSG. MSG is the largest monomeric protein (82 kDa) which NMR backbone spectrum was nearly fully assigned to date (>95%) [Tugarinov et al., 2002] (**figure 4.3, A**). Its solution fold was also determined by NMR [Tugarinov et al., 2005] (**figure 4.3, B**) and by a combination of SAXS and NMR data by the same group a few years later [Grishaev et al., 2008]. This protein is extensively used in the NMR group to optimize methyl labelling conditions, as yields obtained from *E. coli* expression are above 100 mg per L of culture in M9-D_2O medium. MSG is 82 kDa and sensitive to temperature changes. The protein is not stable if frozen and is highly aggregation-prone when concentrated above 10 mg/mL or heated above 50°C. It is thus an adequate candidate to monitor the eventual protection effect of PhCPN on aggregation-prone proteins with fast and irreversible unfolding and aggregation.

A simple preliminary test was to heat MSG in the absence or presence of PhCPN at 60°C, centrifuge the sample and load its supernatant on an SDS-PAGE gel. Strikingly MSG was still soluble in the sample containing PhCPN, even after 30 minutes of heating, showing an active effect of PhCPN (**data not shown**). The next test consisted in monitoring the aggregation temperature of MSG in presence or absence of PhCPN by light scattering followed with a fluorometer. The aggregation temperature depends on the protein concentration, which was kept constant for all the different measurements. A concentration of 0.3 µM was used to avoid signal saturation of the fluorometer during the aggregation process. All experiments were recorded in the same conditions, with temperature monitored by steps of 2°C and an equilibration of 30 sec prior to light scattering measurements. Baseline was adjusted with a blank containing the equivalent PhCPN concentration, as PhCPN

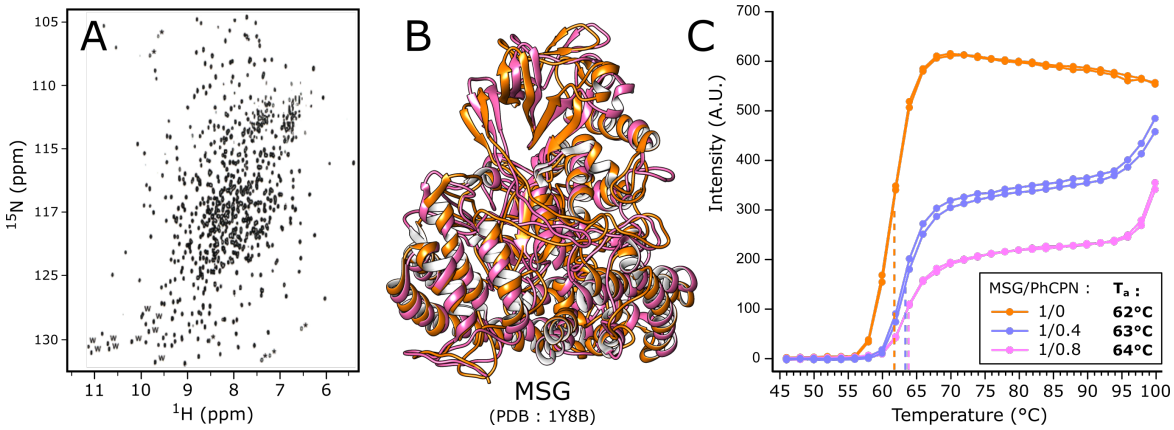


Fig. 4.3 : NMR Spectrum, structure and aggregation of the *E. coli* MSG protein. (**A**) ^{1}H-^{15}N TROSY-HSQC NMR spectrum of MSG at 37°C, adapted from [Tugarinov et al., 2002]. (**B**) Two examples of structures determined using angular restraints derived from chemical shift information [Shen et al., 2009, Tugarinov et al., 2002] and sparse ^{1}H-^{1}H nOe between NH and methyl probes [Tugarinov et al., 2005], (PDB : 1Y8B). (**C**) Monitoring of MSG aggregation temperature in presence or absence of different PhCPN amounts, with a constant MSG concentration of 0.3 µM. Residual background scattering from both buffer and PhCPN were subtracted before plotting. MSG and PhCPN concentrations were corrected with measured molar extinction coefficients (see **Suppl. A.1.2** and **Suppl. A.1.1**).

alone tends to slightly scatter due to its large size. The aggregation temperature is slightly increased by the addition of PhCPN, while the amount of aggregated MSG is decreased (**figure 4.3, C**). In fact, at 80°C all the MSG is aggregated when alone while half of it is still in solution in presence of 0.15 µM PhCPN. This behaviour seems at first in contradiction with the results obtained with HEWL, since PhCPN lowers the unfolding temperature of HEWL by 5°C [Mas et al., 2018]. The main difference is that this experiment does not report the unfolding but the aggregation of MSG. PhCPN effect starts to fade above 90°C, as PhCPN itself is destabilized by temperature and cannot ensure its anti-aggregation function anymore.

4.1.4 MSG, the selected model protein

These preliminary results, allowed to select the most interesting substrate for further investigations. Briefly :

- rsFolder is resistant to high temperatures. In addition, no PhCPN effect on rsFolder unfolding was observed in presence of Guanidinium chloride, possibly because of the denaturants' effect on PhCPN itself.

- PhCPN protects Ubiquitin from aggregation above 75°C, a temperature which cannot be reached with the available NMR probes. In addition, the complex Ubiquitin/PhCPN is in fast exchange with PhCPN, thus making its isolation problematic.
- MSG aggregation is, as in ubiquitin, diminished by PhCPN, despite its large size (82 kDa). Moreover, the interaction mechanism between MSG and PhCPN seems different from the one of HEWL.

From these initial results, MSG was logically chosen as a model substrate and its interaction with PhCPN was characterized by NMR, and by other complementary biophysical and biochemical techniques.

4.2 Characterization of the interaction

Once the model protein was selected, the goals of the study were the following :

- Evaluating the effect of PhCPN buffer composition on the aggregation kinetics
- Gaining information on the aggregation process by NMR
- Isolating an hypothetical MSG/PhCPN complex
- Determining the stoichiometry of such a complex
- Mapping the interaction site and precisely locate MSG inside or outside PhCPN

Characterizing such an interaction would allow us to gain knowledge on the behaviour of PhCPN and equivalent chaperonins when in interaction with large proteins, since the cavity size of PhCPN is very close to the size of MSG. This model is, in fact, relevant, as proteins of more than 82 kDa have already been reported to interact with chaperonins [Hirtreiter et al., 2009].

4.2.1 Chaperonin prevents aggregation and precipitation

Prior to any other experiments, the salt content of the working buffer was optimized to make sure that the conditions were optimal for MSG, since the standard MSG buffer (25 mM TRIS pH 7.0, 20 mM $MgCl_2$ and 5 mM DTT) was different from PhCPN buffer [Ayala et al., 2009].

Buffer salt content

An experimental design was used to test several buffer compositions, depending on the presence (+) or absence (-) of : 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$ respectively (**figure 4.4, A**). The buffering reagent consisting of 25 mM MES pH 6.5 was tested and replaced with 25 mM TRIS pH 6.5 but showed no influence on the aggregation kinetics (**data not shown**). For all experiments, the 25 mM MES pH 6.5 buffer reagent was preferred, as TRIS pH depends highly on temperature and can perturb experiments involving temperature changes.

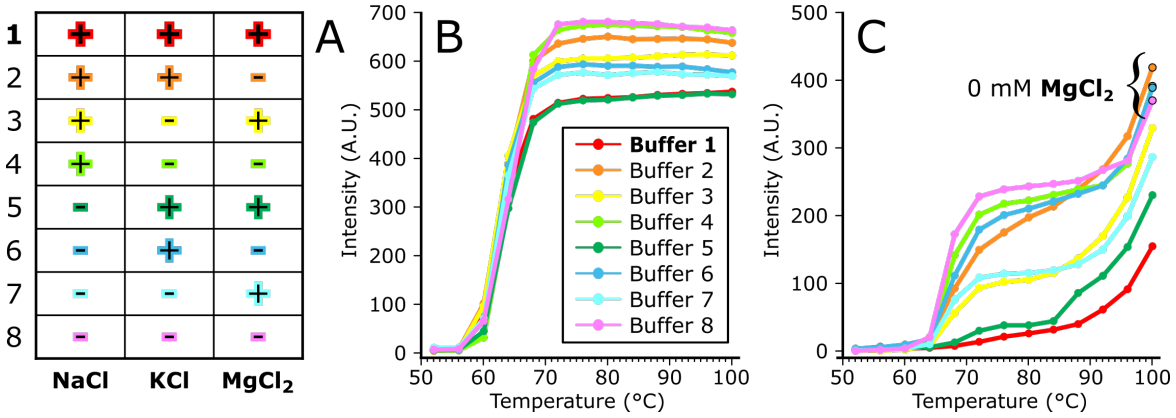


Fig. 4.4 : Design of experiments to evaluate the effect of the buffer salt content on MSG aggregation. (**A**) Experimental design. For NaCl, the + level corresponds to 50 mM NaCl, the - level to no NaCl. For KCl, the + level corresponds to 100 mM KCl, the - level to no KCl. For $MgCl_2$, the + level corresponds to 25 mM $MgCl_2$, the - level to no $MgCl_2$. (**B**) Aggregation of MSG monitored by light scattering at 500 nm and at different temperatures and in different buffers. Residual background scattering from both buffer and PhCPN were subtracted before plotting. (**C**) Aggregation of MSG in presence of PhCPN (1/1) monitored by light scattering at 500 nm and at different temperatures and in different buffers. Residual background scattering from both buffer and PhCPN were subtracted before plotting. Residual background scattering from both buffer and PhCPN were subtracted before plotting.

A first thing that can be noted is that no matter the composition of MSG buffer, the aggregation temperature is identical and around 65°C (**figure 4.4, B**). This temperature is different from the previous results presented in (**figure 4.3, C**) because the number of temperature steps is decreased here, thus changing the aggregation kinetics. In fact, the aggregation temperature measured by this technique is highly sensitive to the number of steps, as the fluorimeter chamber is kept at a fixed temperature for 30 sec prior to the scattering measurement. For this reason, the sample issued from the experiment described in (**Section 4.3**) was heated longer than these, resulting in different stability and

aggregation kinetics. The amount of aggregated MSG in the present experiment seems nevertheless independent on NaCl concentration (buffers 1 and 5) but varies greatly with the presence or absence of $MgCl_2$ (buffers 1 and 2 respectively) and KCl (buffers 1 and 3 respectively). The absence of these two salts altogether leads to a total aggregation of MSG (Buffer 4, (**figure 4.4, A & B**), corresponding to the maximal light scattering intensity. The addition of equimolar amounts of PhCPN in solution changes drastically the aggregation pattern (**figure 4.4, C**). Identically as before, one could notice that the aggregation kinetics are different from the ones presented in (**figure 4.3, C**), for the same reasons mentioned above. Aggregation is drastically lowered in PhCPN buffer (Buffer 1), as the buffer was optimized for maximal PhCPN activity. Aggregation is also highly dependant on the presence of $MgCl_2$, since an absence of $MgCl_2$ causes PhCPN to disassemble back into monomers (Buffers 2, 4, 6 and 8, see **Chap. 3**). The existence of PhCPN cavities play a large role in the anti-aggregation effect. A similar but attenuated effect is observed with the absence of KCl, as proved by the different aggregation patterns in Buffer 7 and 5, both containing $MgCl_2$ but either KCl or no KCl respectively. The presence of NaCl does not appear critical either, as proven by Buffer 5. Since the presence of these high amounts of salt did not influence the aggregation of MSG alone, it was decided to keep PhCPN buffer for the rest of the experiments involving MSG.

Inhibition of MSG aggregation

Prior to any further investigations, molar extinction coefficients (noted ϵ) were experimentally determined in order to estimate correctly the protein concentrations during the rest of the experiments. In fact, molar extinction coefficients are routinely computed using the ProtParam tool (https://web.expasy.org/protparam/), resulting in imprecision. Especially, PhCPN does not include Tryptophans in its sequence (**Suppl. A.1.1**), resulting in a low molar extinction coefficient and an high uncertainty on its predicted value by the software. The determination of the extinction coefficient consists in measuring the optical density of a given sample and analyzing its protein content by amino acid degradation and hydrolysis [Walsh and Brown, 1962]. In this way, the initial protein concentration of the sample, as well as the molar extinction coefficient, can be back calculated. A detailed protocol is developed in **Chap. 6.5.3**. Predicted values for ϵ are of 93195 $M^{-1}cm^{-1}$ for MSG and 10430 $M^{-1}cm^{-1}$ for PhCPN (**Suppl. A.1.1**).

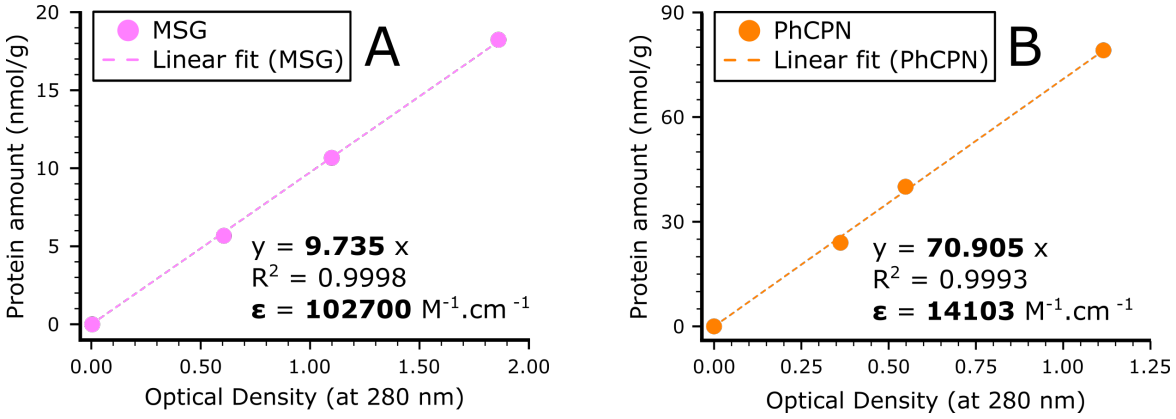


Fig. 4.5 : (**A**) & (**B**) Measurement of the extinction coefficient of MSG and PhCPN respectively, with help from amino acid analysis. The extinction coefficient is determined by linear regression.

Experimentally determined values of ϵ are of 102700 $M^{-1}cm^{-1}$ and 14103 $M^{-1}cm^{-1}$ for MSG and PhCPN respectively (**figure 4.5, A & B**). As expected, the measured value of ϵ for PhCPN is 26.0% higher than the predicted value, while the difference is just 9.3% for MSG. Such variations would cause an overestimation of PhCPN concentration compared to MSG in solution and result in wrong MSG/PhCPN ratios (ratios given in (**figure 4.3, C**) were corrected with the new ϵ values after the experiment).

Once the protein ratio could be determined precisely, the effect of the PhCPN concentration on MSG aggregation kinetics was monitored by light scattering. Both proteins were heated at a given temperature directly in the machine, separately or together at different molar ratios. MSG concentration was kept constant (0.3 µM) and the temperature was fixed at 60°C to observe the effect of PhCPN, in a similar fashion as described for Ubiquitin (**figure 4.1, B**). The error was calculated from a set of 3 measurements for each ratio.

A short lag phase can be observed prior to aggregation, due to the temperature equilibration time after the cuvette introduction inside the fluorometer (**figure 4.6, A**). Once the cuvette reaches the right temperature, MSG alone starts to aggregate and precipitates completely within minutes. Interestingly, this phenomenon is partially to completely prevented when PhCPN was introduced along with MSG (**figure 4.6, A**), similarly to what was observed with Ubiquitin before (**figure 4.1, C**). However, in this case, the kinetics of the interactions are unknown while Ubiquitin was in fast exchange with PhCPN. Increasing amounts of PhCPN decrease the apparent aggregation speed as well as the aggregated amount of MSG in solution. Aggregation is completely prevented for a ratio MSG/PhCPN of 1/3.2. The amount of PhCPN needed to observe the same anti-aggregation activity on

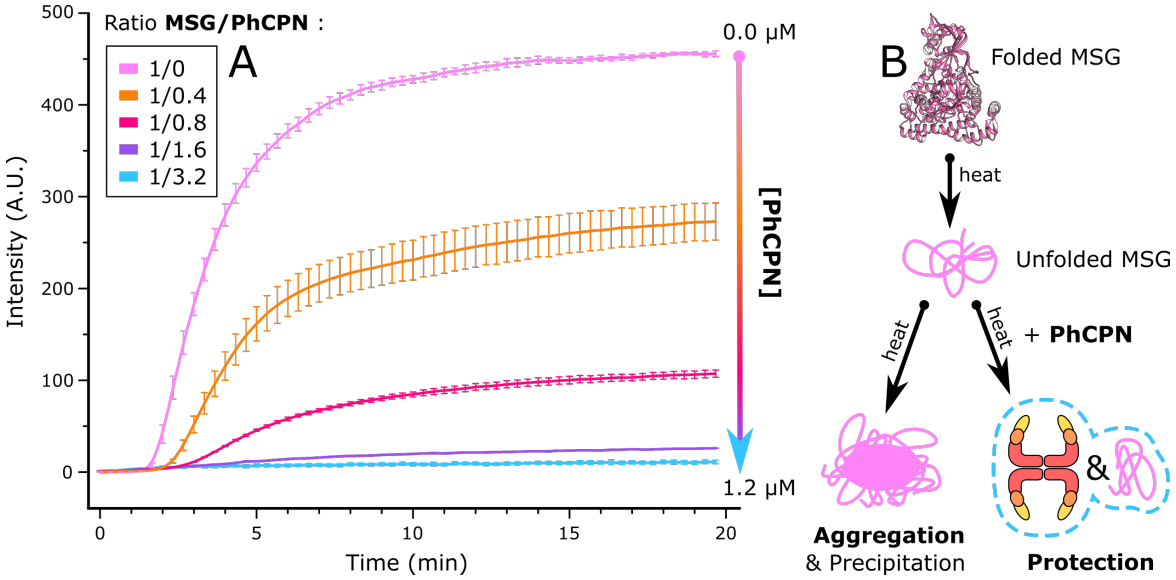


Fig. 4.6 : MSG aggregation in presence of PhCPN. (**A**) Aggregation of MSG monitored by light scattering at 60°C. 0.3 µM MSG in PhCPN buffer are incubated in the absence (pink) and presence of different molar ratios of PhCPN (orange MSG/PhCPN : 1/0.4 ; dark pink MSG/PhCPN : 1/0.8 ; violet MSG/PhCPN : 1/1.6 ; light blue MSG/PhCPN : 1/3.2). (**B**) Putative scheme of MSG unfolding in the absence or presence of PhCPN.

MSG is larger than with Ubiquitin. It can be conjectured that PhCPN is able to host at least one substrate protein per cavity at the time when the size of the substrate is not too large, as it is the case with HEWL [Mas et al., 2018], or that the kinetics of the interaction are different with Ubiquitin than with MSG. In the case of MSG, the number of protected protein per PhCPN assembly would be decreased, thus the need for more PhCPN per MSG unit to observe very similar effects as with Ubiquitin. The aggregation rates, when aggregation occurs, is independent of the presence of PhCPN. In fact, when the curves are artificially scaled to reach the same the maximal light scattering intensity, their rate of increase is almost constant.

The differences in anti-aggregation effect could be either explained by different interaction stoichiometry (increased amounts of protected Ubiquitin compared to MSG, possibly because of the large size of MSG), or different interaction dynamics (Faster exchange between the free and bound forms of Ubiquitin compared to MSG). When the curves are scaled as a function of the introduced amounts of MSG, the apparent reaction speeds are identical. This can be explained by an irreversible interaction between MSG and PhCPN (or very slow release of MSG & k_{off}), resulting in the sequestration of unfolded MSG

by PhCPN. The apparent concentration of MSG in solution is consequently decreased and so is the aggregation. The aggregation rate constant remains identical but since the apparent concentration of MSG is lower and the apparent aggregation rate depends on the concentration of MSG, the apparent aggregation rate is decreased.

However, from the experiment, a precise estimation of the exact number of MSG protected per PhCPN is not possible. This experiment also does not give details on the folding state of MSG while in interaction with PhCPN. In fact, MSG could still be folded or unfolded in solution but stabilised by the hydrophobic environment of PhCPN's apo cavity or directly interacting with the outer surface of PhCPN. Besides, the light scattering test only bears low protein concentrations far from native conditions. Interaction tests by NMR were carried out to address the folding state of MSG and investigate its aggregation behaviour at higher protein concentrations

MSG folding state in interaction with PhCPN

MSG was methyl-specifically labelled on Isoleucines-δi to study its folding at high temperature in the presence of PhCPN labelled on Methionines and Valines. The methyl labelling protocol can be found in **Chap. 6.2**. MSG Isoleucine resonance assignement can be found online on the BMRB under entry number 5471 (**figure 4.7, B**)[Tugarinov et al., 2002]. PhCPN Valines and Methionines were previously assigned in our group (see **Chap. 2.3.4**, [Mas, 2015]). Both proteins were observed simultaneously, as their methyl signals do not overlap (**figure 4.7, A**), in a MSG/PhCPN 2/1 ratio. Final concentrations of MSG and PhCPN were respectively of 50 µM and 25 µM in the different samples : MSG alone, PhCPN alone and MSG/PhCPN. The samples were heated in the NMR spectrometer at identical successive temperature steps and sets of 2D methyl HMQC spectra were recorded at each temperature to follow unfolding or peak disappearance.

At 40°C (**figure 4.7, A**), MSG methyl spectrum corresponds to the characteristic spectrum of a folded protein with dispersed chemical shifts, and does not exhibit any changes when PhCPN is added, as expected (**figure 4.7, D**). In fact, PhCPN was shown to interact with unfolded substrates only [Mas, 2015]. The differences start to appear when the temperature is raised to the vicinity of the aggregation temperature of MSG. The intensity of MSG spectrum start to decrease at 50°C, 52°C and 55°C (**figure 4.7, B**), the weak peaks shift and fully disappear at 60°C (**figure 4.7, C**), while a precipitate appears in the NMR

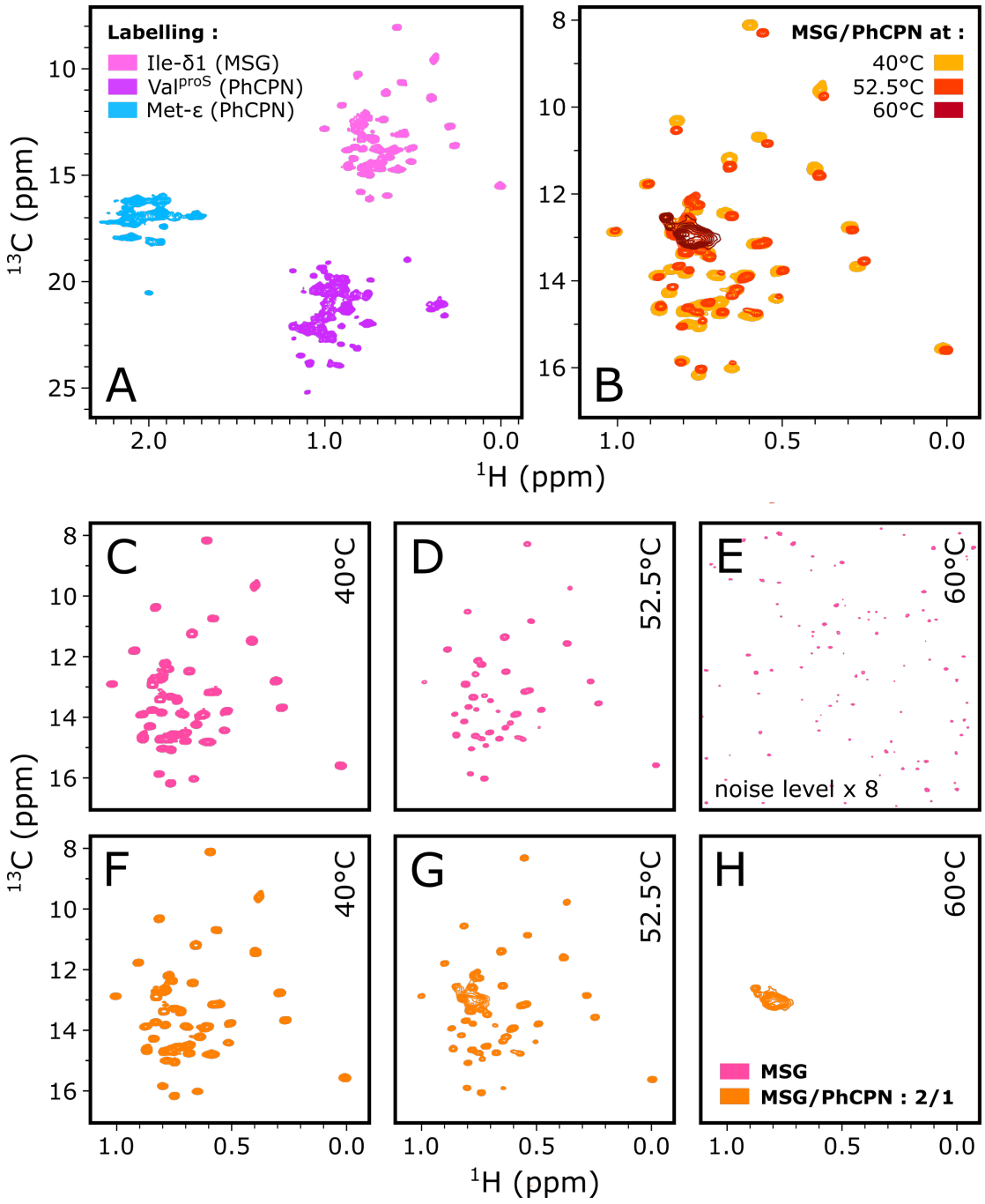


Fig. 4.7 : Prevention of thermal aggregation of MSG by PhCPN monitored by NMR. (**A**) Example of the ^{1}H-^{13}C methyl SOFAST HMQC spectrum acquired on the methyl labelled sample of mixed PhCPN and MSG, specifically labelled on Isoleucines (MSG), Methionines (PhCPN) and Valines (PhCPN). (**B**) Overlaid spectra of the MSG/PhCPN sample at different temperatures. Individual 2D HMQC spectra of 50 µM [U-D], (Ileδ1)-$^{13}CH_3$-labelled MSG at 40°C, 55°C and 60°C in the absence (**C**, **D** & **E**) and presence (**F**, **G** & **H**) of 25 µM unlabelled PhCPN. Spectra were recorded in D_2O PhCPN buffer, on an NMR spectrometer operating at a proton frequency of 850 MHz.

tube. The fact that these changes appear at a lower temperature compared to the previous light scattering experiments can be caused by both longer heating times in the case of NMR to acquire sequential spectra (**figure 4.7, A**) and increased proteins concentrations for NMR. Interestingly, in the presence of PhCPN and above the melting temperature of MSG, a new signal appears in the characteristic unfolded Isoleucine region of the spectrum, corresponding to unfolded MSG stabilized in solution by PhCPN (**figure 4.7, E**). Overlaid spectra can also be found in (**figure 4.7, B**). These peaks are still observable at 60°C whereas the MSG alone signal disappears completely (**figure 4.7, C & F**). It is thus clear that PhCPN enables MSG to remain unfolded in solution when exposed to high temperatures.

The series of spectra were acquired with 15 minutes steps and added together in order to get a high signal to noise spectra. However, the intensity of the signal of folded MSG decreases over time due to thermal aggregation, as previously shown. The unfolding rate is however independent on the presence of PhCPN as proven by (**figure 4.8, A**). In fact, at both 50 and 52°C the intensity disappearance rate with or without PhCPN is quite identical. The spectra at different temperatures were taken in the exact same conditions and the same overall heating time, using a temperature ramp. This finding was consistent with previously observed behaviour in (**figure 4.6, A**). Since the unfolding rate is independent of the presence of PhCPN, it can be deduced that the interaction of MSG and PhCPN is irreversible and MSG is sequestered by PhCPN (**figure 4.8, C**). In fact, if it was not the case, the quantities of unfolded MSG would be perturbed by the presence of PhCPN since the equilibrium between the folded and unfolded states would be modified. Oppositely, it was shown that the quantity of unfolded HEWL at high temperature is influenced by the presence of PhCPN, as the equilibrium between folded and unfolded HEWL is displaced towards unfolded HEWL [Mas et al., 2018].

This technique allows to gaining knowledge on the unfolded state of MSG interacting with PhCPN, since PhCPN was shown to interact with unfolded MSG only. It is however difficult to quantify the amount of MSG interacting with PhCPN from the intensity of the unfolded Isoleucines region of the spectrum. In addition, even if MSG is partially protected by PhCPN, the protein still aggregates slightly in the tube, resulting in protein loss. Tests were also carried out, with the aim of quantifying the aggregates of MSG to back-calculate the amount of MSG left in solution. Known amounts of MSG were voluntarily precipitated by heat and the resulting aggregated protein was resuspended in various amounts of 6

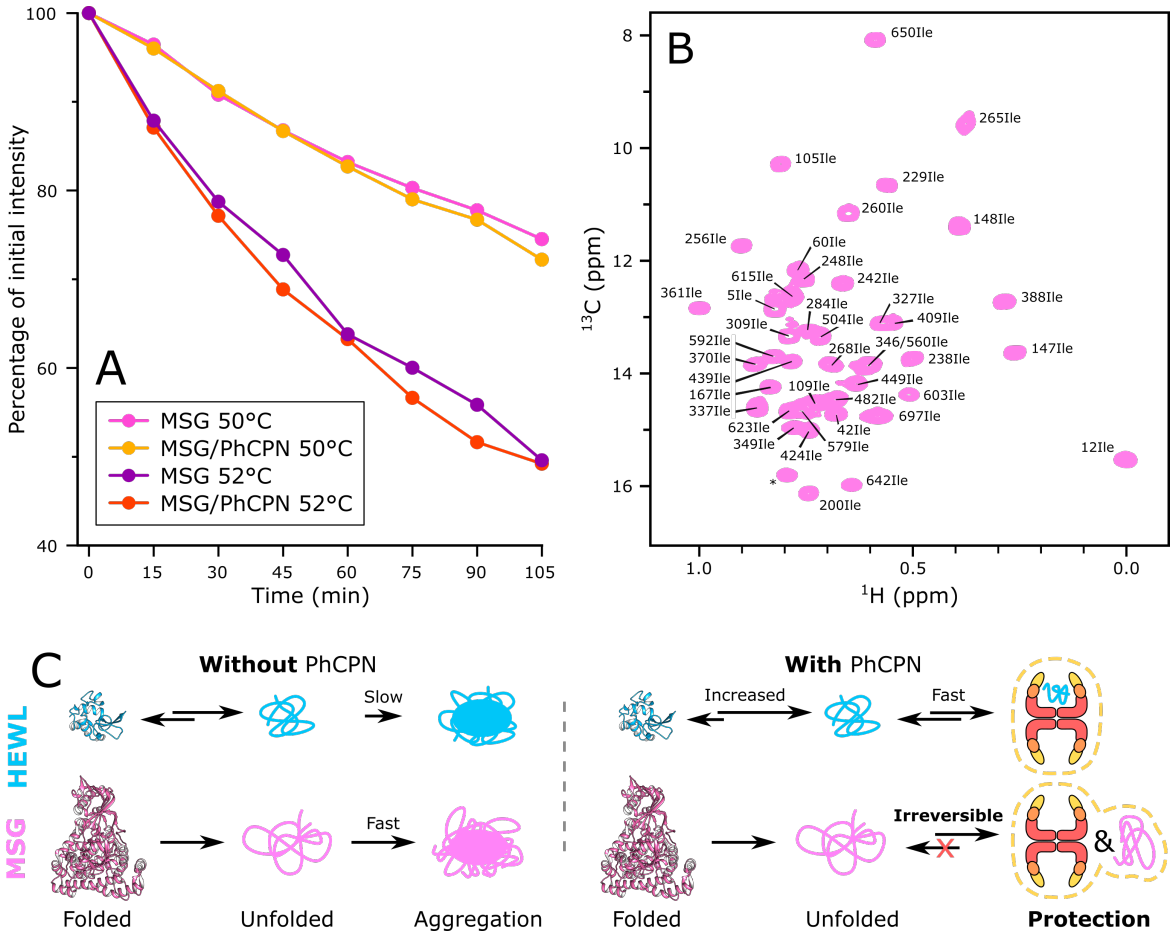


Fig. 4.8 : MSG aggregation kinetics in presence of PhCPN monitored by NMR. (**A**) Peaks intensity decrease over time due to unfolding. The peaks which were considered correspond to peaks which not completely fade over time, as well as peaks which are not included in the region of the spectra where the unfolded region of MSG appears in presence of PhCPN. (**B**) 2D methyl SOFAST HMQC spectra of 50 µM [U-D], (Ileδ1)-$^{13}CH_3$-labelled MSG at 40°C attributed from [Tugarinov et al., 2002]. Spectra were recorded in D_2O PhCPN buffer, on an NMR spectrometer operating at a proton frequency of 850 MHz. (**C**) Scheme of PhCPN protection mechanism in presence of two very different substrates, MSG and HEWL.

M urea (**figure 4.9, B & C**). The amount of resuspended protein was then estimated by monitoring the UV absorbance of the mixture (**figure 4.9, A**).

This quantification of the aggregates failed. In fact, the calculated amount of MSG was systematically overestimated and it depended on the urea volume (**figure 4.9, B**) as well as on the initial amount of MSG in the pellet (**figure 4.9, C**). For instance, 1.9 mg of aggregated MSG re-suspended in 200 µL urea had an absorption corresponding to 1,8 times its initial absorbance.

These results clearly show the protective role of PhCPN on MSG but, however, do not provide information on the location of MSG. In fact, it was still unclear whether it was hosted or bound by the PhCPN or not, or just unfolded in solution and in transient interaction with PhCPN. In addition, since the quantification by NMR or UV absorption was not reliable, there was still no indication on the stoichiometry of the complex, hence the need to isolate it.

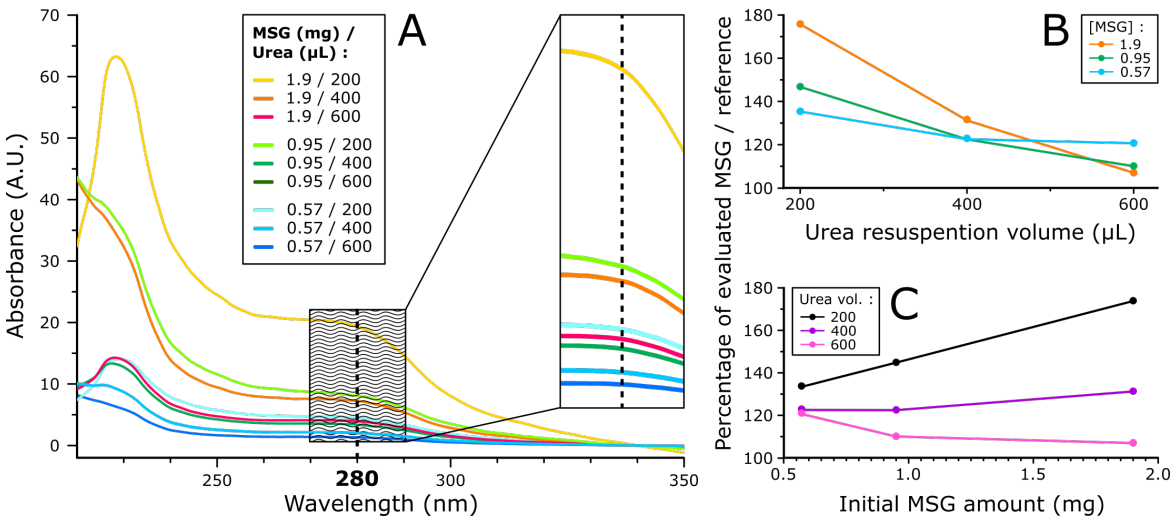


Fig. 4.9 : Quantification of the amount of MSG in the precipitated pellet by resuspension in urea. (**A**) UV-vis spectra of different amount of precipitated MSG resuspended in urea. The first number in the legend corresponds to the amount of MSG in the pellet, in mg. The second number corresponds to the volume of urea used for resuspension, in µL. A zoom of the different absorbances at 280 nm is provided on the right side of the figure. (**B**) Percentage of calculated resuspended MSG depending on the urea resuspension volume. (**C**) Percentage of calculated resuspended MSG as a function of the initial amount of MSG in the pellet.

4.2.2 Identification and isolation of a stable complex

As stated above, PhCPN has an anti-aggregation effect on MSG and allows to keep the protein unfolded and in solution at high temperatures. However, the existence of a MSG/PhCPN complex was still to prove. To this extent, DOSY NMR experiments (see **Chap. 6.3.1**) were first used to identify the complex, followed by analytical gel filtration combined with SDS-PAGE analysis to isolate it.

Diffusion properties of the complex

The DOSY experiment is a powerful way to determine whether or not MSG and PhCPN are diffusing together at high temperature, by comparing their individual translational diffusions. Like with the previous NMR experiments, methyl-specific labelling is needed to record signal from both MSG and PhCPN. The DOSY temperature-dependent diffusion coefficients of U-[D, ^{12}C], (Met-ϵ)-$^{13}CH_3$-stereospecifically labelled 25 µM PhCPN alone and U-[D, ^{12}C], (Ile-δ1, Val^{proS})-$^{13}CH_3$-stereospecifically labelled MSG mixed with non-isotope labelled PhCPN in 4/1 ratio were measured at different temperatures and compared. These coefficient are calculated from non-linear curve fit of the recovered signal intensity as a function of the gradient strength (see **Chap. 6.3.1**). Data prior fitting is displayed in (**figure 4.10, A, B & C**).

The reference diffusion is the one of PhCPN (**figure 4.10, D, orange curve**). As expected, its diffusion increases with temperature. In fact, the viscosity is a parameter of paramount importance since diffusion depends on it. In the case of MSG mixed with PhCPN, the situation is different. Below the melting temperature of MSG, MSG and PhCPN do not interact together and MSG diffusion increases with temperature. Its diffusion is compatible with a soluble monomeric 82 kDa protein (**figure 4.10, D, pink curve**) and is logically faster than PhCPN's (about 1 MDa). Interestingly, when MSG melting temperature is outreached (55°C measurement point), the diffusion of MSG slows down to match PhCPN's one. Diffusion keeps matching when the temperature is lowered (**figure 4.10, D, violet curve**) and increased again during a second heating cycle (**figure 4.10, D, blue curve**). The remaining diffusion difference between PhPCN and PhCPN with MSG after heating is due to a different viscosity between the two samples because residual MSG aggregates were removed from the NMR tube prior to the second set of measurements. These observations confirm that MSG is irreversibly confined by PhCPN and that this

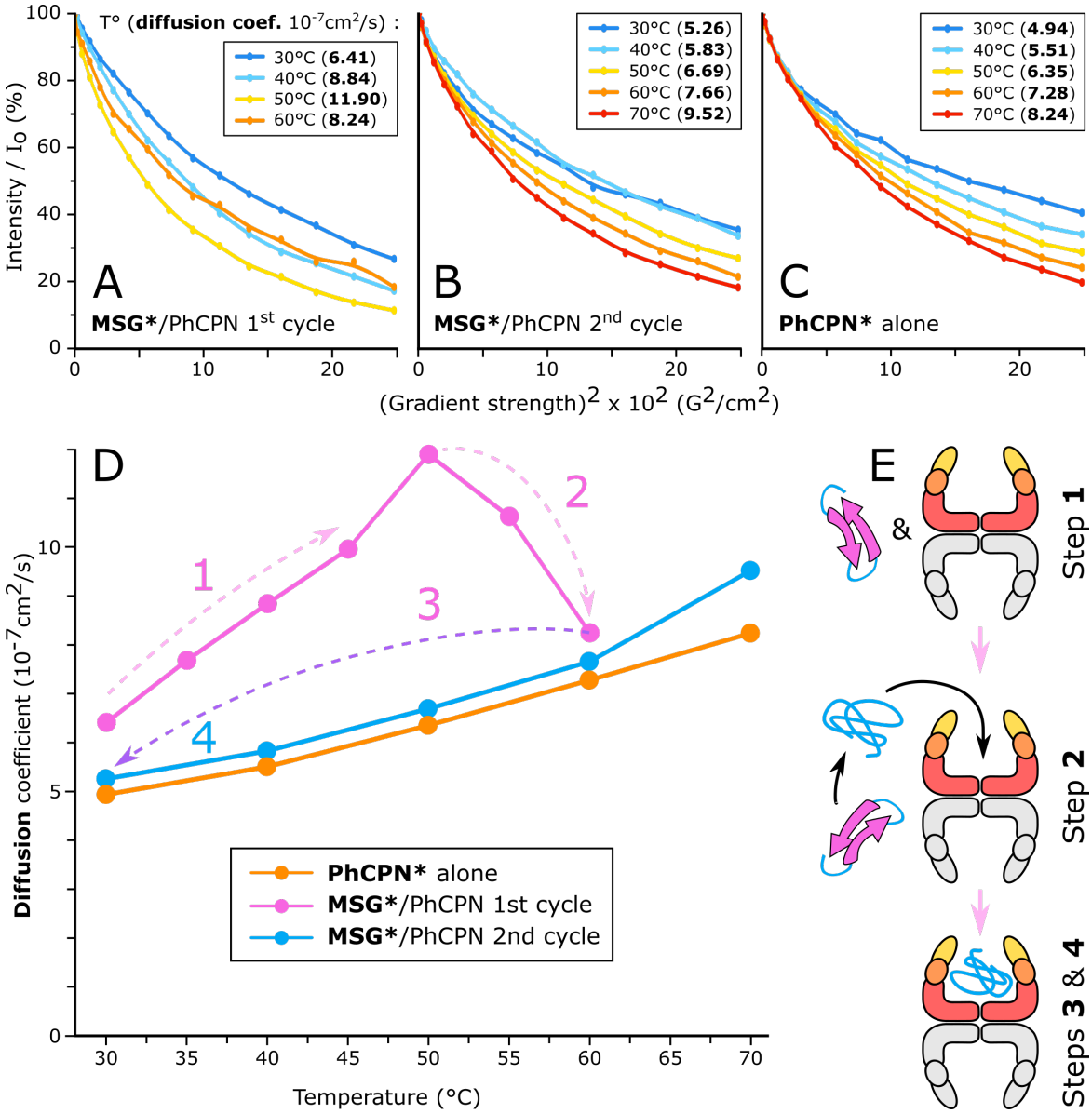


Fig. 4.10 : Identification of an MSG/PhCPN complex by DOSY NMR. (**A**), (**B**) & (**C**) Recovered signal intensity as a function of the gradient strength prior to fitting. Calculated diffusion coefficients are highlighted in bold. (**D**) Temperature-dependent diffusion coefficients of [U-D], (Met-ϵ-[$^{13}CH_3$]-labelled PhCPN (orange) and [U-D], (Ileδ1,Val^{proS})-[$^{13}CH_3$]-labelled MSG with non-isotope labelled PhCPN (pink and blue). The values were obtained from fitting of 1D DOSY signal decay profiles of the labelled methyl regions. (**E**) Schematic representation of MSG and PhCPN interacting together during the heating (**1**-**3**) and cooling process (**4**).

complex is stable with temperature. The next step was logically to purify and isolate this complex in order to gain knowledge on its precise stoichiometry

Gel filtration

Gel filtration can be used to isolate a complex formed by two supposedly strongly interacting proteins. MSG and PhCPN were incubated separately (**figure 4.11, A & B**) or together (**figure 4.11, C & D**) at different temperatures for 1 hour, in PhCPN buffer. The experimental conditions used here are similar to NMR ones, as a high protein concentration (20 mg/mL PhCPN before loading the sample on the column) is needed to get a measurable signal from the purifier detector. In fact, the samples were centrifuged to remove precipitated protein and only 15 µL of the supernatant were loaded on an HPLC analytical gel filtration column. Absorbances of the different samples were measured at 280 nm (**figure 4.11, A**) as a function of the elution volume. The region corresponding to PhCPN peak is enlarged for clarity. The region also corresponds to the fractions collected and loaded on SDS-PAGE gels for further analysis (**figure 4.11, E**).

Results show that, as expected, the soluble MSG elution peak disappears when the mixture is heated up because of aggregation and precipitation (**figure 4.11, A**). When MSG and PhCPN are mixed together, the intensity of PhCPN peak is similar between 20 and 50°C, and starts to increase above 50°C, suggesting that MSG interacts strongly with PhCPN and both proteins are eluted together. This is proven by the SDS-PAGE analysis of the PhCPN peak above 50°C, which contains a significant amount of MSG (**figure 4.11, C**). Doubling the concentration of MSG does not affect significantly the absorbance of PhCPN at elution volume 1.4 mL nor the intensity of the band on the SDS-PAGE gels (**figure 4.11, D**). It can be conjectured that the maximal PhCPN capacity of binding is already reached for an MSG/PhCPN ratio of 2/1. These results allowed a first coarse evaluation of the stoichiometry of the complex, which is described in the next section.

4.2.3 Stoichiometry of the complex

Evaluation of the complex stoichiometry is of prevalent importance to gain knowledge on the binding modes of PhCPN and particularly on its ability to hold substrates in its cavities. A rough evaluation of the complex composition was already possible after gel filtration and SDS-PAGE, by two different but complementary methods.

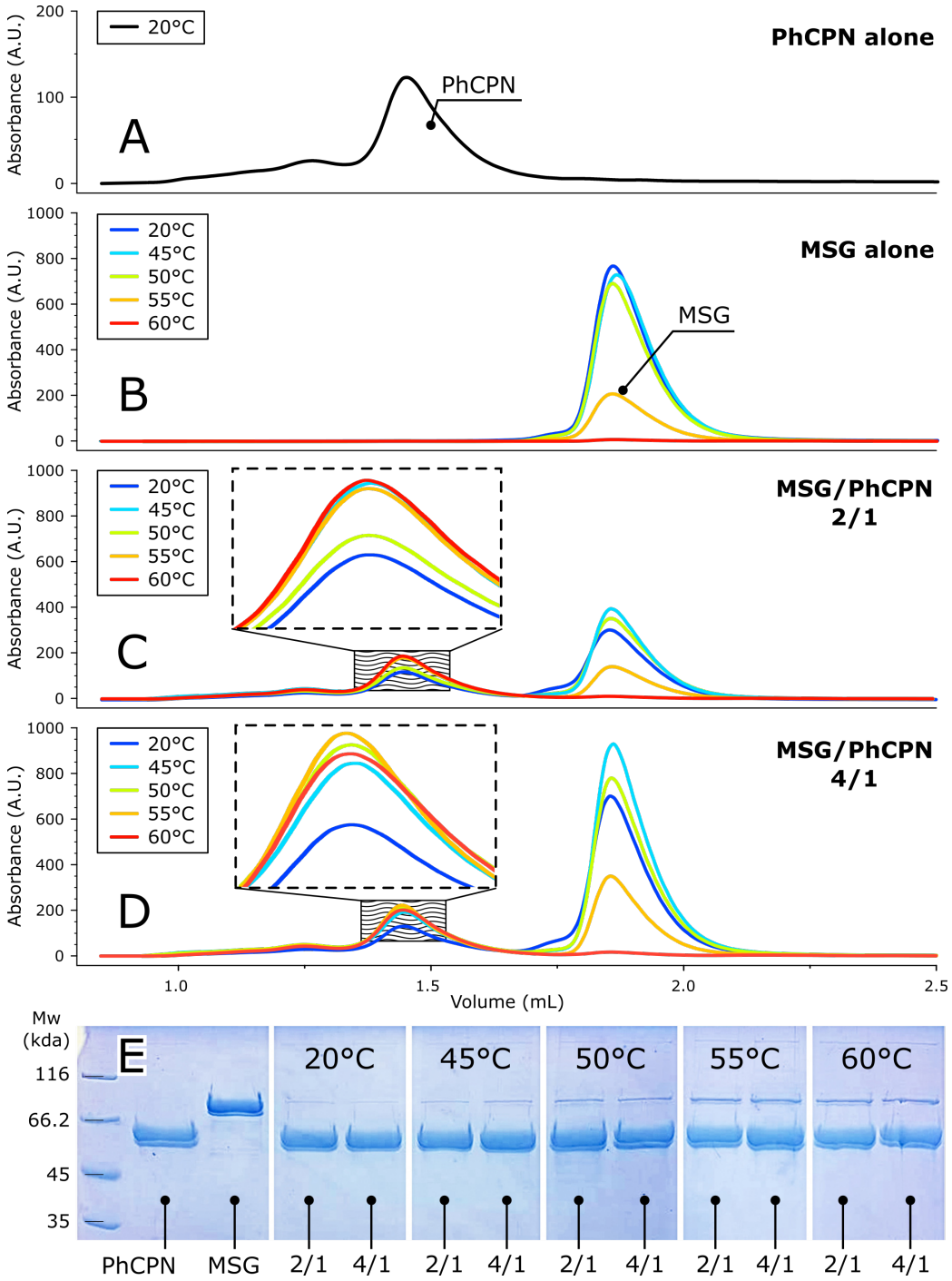


Fig. 4.11 : Isolation of a MSG/PhCPN complex by gel filtration. (**A** & **B**) Gel filtration chromatograms of MSG incubated at different temperatures and PhCPN incubated at 20°C. (**C** & **D**) Gel filtration chromatograms of a respectively 2/1 and a 4/1 mixture of MSG and PhCPN, incubated at different temperatures. The box filled with waves at elution volume 1.45 mL shows the fractions used for further SDS-PAGE. PhCPN migrates at its monomer size since the complex is destroyed by SDS. (**E**) SDS-PAGE analysis of PhCPN, MSG and the MSG/PhCPN complex at various temperatures, purified by the gel filtrations shown in (**A**), (**B**) and (**C**). The initial MSG/PhCPN molar ratios are indicated at the bottom of each lane.

Coarse grain evaluation

The first method consisted in measuring the intensity of PhCPN peak (1,3 mL elution volume on (**figure 4.11, A, B, C & D**) on the gel filtration profile and plotting it as a function of temperature (**figure 4.12, A**). By subtracting the reference intensity of the control PhCPN sample, it is possible to estimate the contribution of MSG to the total absorbance. The amount of MSG in the complex can be then calculated with the molar extinction coefficient of MSG (obtained in (**figure 4.5, A**) and finally the ratio MSG/PhCPN can be plotted as a function of temperature (**figure 4.12, C**). The second method consists in evaluating the relative intensity of the protein bands obtained on SDS-PAGE gels at different temperatures (**figure 4.11, E**). The ImageJ software was used for this purpose (**figure 4.12, B**) and allowed to quantify the bands of PhCPN (monomer) and MSG. The final calculated ratio was also plotted on (**figure 4.12, C**).

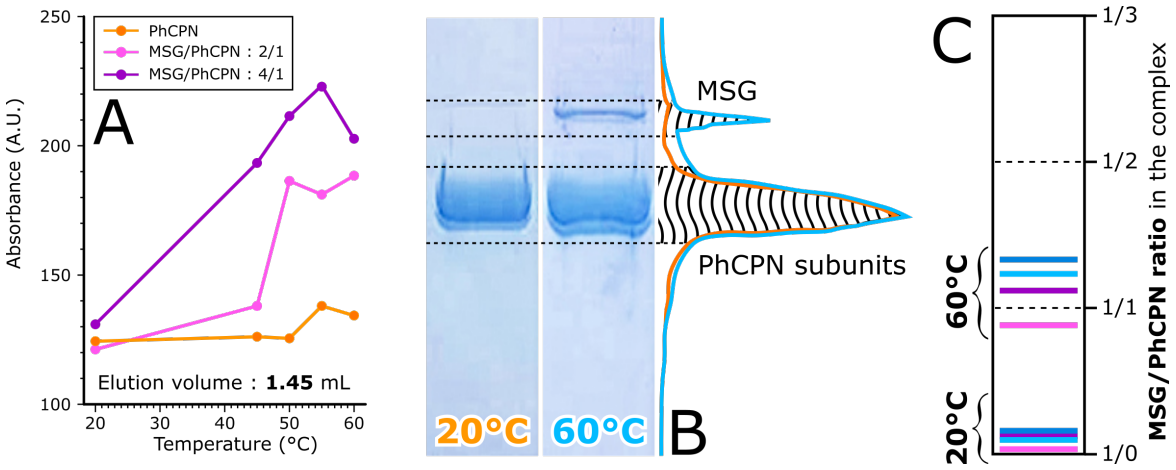


Fig. 4.12 : Quantification of the MSG/PhCPN ratio in the MSG/PhCPN complex. (**A**) Absorbance at 280 nm of the PhCPN peak in the gel filtration profiles provided in (**figure 4.11, E**), in the presence or absence of different molar ratios of MSG, with corresponding blanks substracted. (**B**) Schematic representation of the quantification on a SDS-PAGE gel, provided by UV absorbance at 280 nm of the different Gel filtration chromatograms of MSG incubated at different temperatures and PhCPN incubated at 20°C and 60°C. (**C**) Evaluation of the MSG/PhCPN ratio in the complex provided by different techniques. Light and dark pink lines were extracted from gel filtration chromatograms of a respectively 2/1 and a 4/1 mixture of MSG and PhCPN, incubated at 20°C and 60°C. Light and dark blue pink lines were extracted from PAGE gels of a respectively 2/1 and a 4/1 mixture of MSG and PhCPN, incubated at 20°C and 60°C.

Interestingly, these two approaches result in fairly similar estimations of the MSG/PhCPN ratio and indicate a stoichiometry close to 1/1. To confirm these preliminary observations, further investigations involving more rigorous techniques were carried out.

Evaluation by SEC-MALLS

A second technique used to precisely determine MSG/PhCPN stoichiometry was SEC-MALLS. Three samples containing MSG, PhCPN and the MSG/PhCPN complex respectively previously purified by gel filtration were run on the SEC column (**figure 4.13, A, dashed lines**). The MSG/PhCPN complex was prepared by mixing MSG and PhCPN in a 4/1 ratio and heating the sample at 60°C for 10 min. The refractive index was analyzed afterwards to monitor the molecular weight of the proteins in solution and get a picture of the molecular weight of each species (**figure 4.13, A & B**). The detailed protocol of this experiment can be found in **Chap. 6.5.3**.

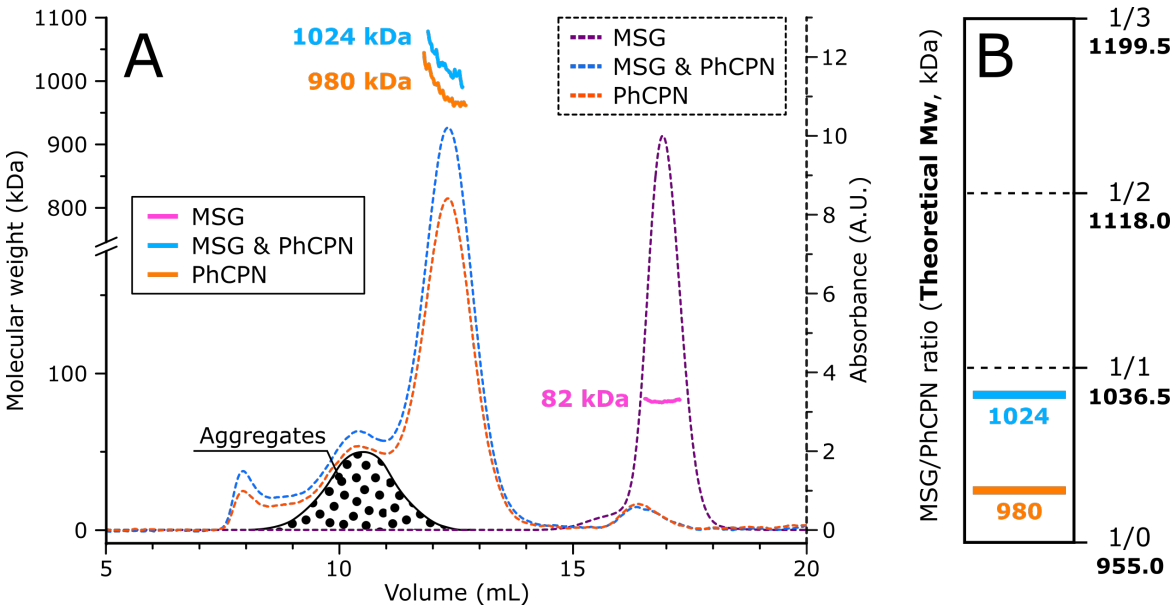


Fig. 4.13 : SEC-MALLS analysis of PhCPN, MSG and MSG/PhCPN complex. (**A**) SEC-MALLS profiles of three samples : MSG, PhCPN, and the MSG/PhCPN complex previously purified by gel filtration. (**B**) Experimental values of the molecular weights of the protein in solution in the different samples. (**C**) Theoretical values of the molecular weights of the protein in solution in the different samples.

Validating previous results, the SEC-MALLS calculated values are also in agreement with a 1/1 complex. Although the molecular weight obtained with SEC-MALLS for MSG and PhCPN is slightly different from the theoretical values (1036 kDa), the error is still very

acceptable (1.2%). It fits better with the theoretical value of the molecular weight of a 1/1 complex (**figure 4.13, B**).

Amino-acid analysis evaluation

Finally, the complex obtained after mixing MSG and PhCPN in a 4/1 ratio, heating at 60°C and gel filtration was submitted to amino acid hydrolysis to get a clear composition in amino acids of the sample. An internal standard reference composed of Norleucine, an artificial amino acid, is added to the samples prior to hydrolysis (see **Chap. 6.5.3**). The obtained values, expressed in picomoles of amino acids, are shown in the table (**figure 4.14, A**). Some amino acids are omitted because they are destroyed during the hydrolysis reaction.

A	MSG		PhCPN		MSG / PhCPN 1 / 1		2 / 1	3 / 1
Residue	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.	Calc.	Calc.
D/N	85	86	789	784	**875**	**870**	956	1042
T	32	33	365	400	**418**	**433**	466	499
G	53	51	708	672	**759**	**723**	774	825
A	76	73	1044	992	**1120**	**1065**	1138	1211
V	44	46	719	736	**751**	**782**	828	874
I	43	44	809	832	**855**	**876**	920	964
L	72	70	759	752	**842**	**822**	892	962
Y	18	18	109	112	**133**	**130**	148	166
F	20	19	80	80	**99**	**99**	118	137
H	21	23	91	96	**108**	**119**	142	165
K	32	32	907	896	**927**	**928**	960	992
R	44	44	308	336	**373**	**380**	424	468
Sum	540	539	6688	6688	**7252**	**7227**	7766	8305

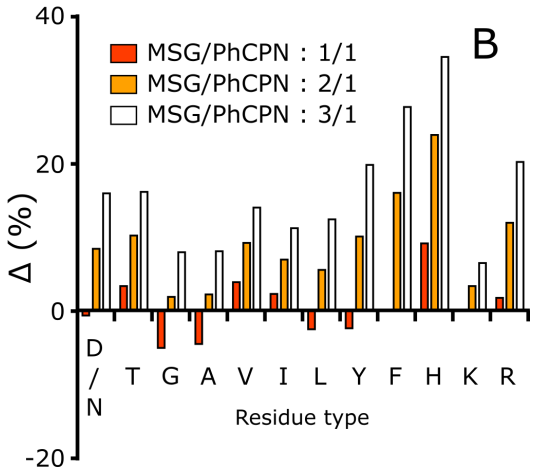


Fig. 4.14 : Amino-acid analysis of the MSG/PhCPN complex composition. Some amino acids are excluded for the calculation because they are destroyed during acid hydrolysis. (**A**) Table detailing the experimental and theoretical amino acid compositions of the different samples : MSG, PhCPN, and the MSG/PhCPN complex. The ratio best fitting the experimental composition is colored in red. All the values are expressed in picomoles. (**B**) Differences of the amino acid compositions between the experimental values of the MSG/PhCPN complex from acid hydrolysis and the calculated values for ratios 1/1 , 2/1 and 3/1.

When comparing the experimental value with the theoretical ones, it is clear that the protein ratio in the complex is 1/1. In fact, the difference between the expected quantity (in bold red) and the experimental quantity (in bold black) for the ratio 1/1 is really low while it increases dramatically for the other possible ratios (2/1 and 3/1). This difference is expressed in percent and summarized in (**figure 4.14, B**). This last experiment confirms previous observations and the stoichiometry of the MSG/PhCPN complex. Interestingly,

PhCPN is not capable of hosting two substrate proteins in both its cavities at the same time if the substrate protein is too large. The only information yet to determine was the localization of MSG inside or outside PhCPN, which is elaborated in the next section.

4.2.4 Substrate location

To complete this study on the interaction between MSG and PhCPN, the final goal was to determine the binding site of MSG on PhCPN. This is particularly interesting as several theories exist on how the substrates bind to Type II Chaperonins (see **Chap. 2.3.3**). In addition, MSG volume (140x10^3 Å^3, measured with Chimera, [Pettersen et al., 2004] should theoretically fit in PhCPN cavity (approximately 500x10^3 Å^3, measured with Chimera, (**figure 4.15, A**)), but unfolded MSG could protrude from the chamber and interact with residues outside PhCPN cavities. Several experimental techniques were employed to address this question.

Electron microscopy

The first approach for determining MSG binding site involved electron microscopy. PhCPN pictures were taken by EM long before this study and gave satisfactory and exploitable images. The drawback of this technique, however, is that PhCPN orients perpendicularly to the carbon layer and it is infrequent to observe side views of PhCPN on EM images [Mas, 2015]. Samples were taken before and after heating, and then negatively stained prior to EM analysis. The detailed protocol can be found in **Chap. 6.5.2**.

These pictures first show that PhCPN is well formed in its characteristic two rings shape (**figure 4.15, B**). All the subunits are well distinguishable. MSG is however too small to be clearly recognized (**figure 4.15, C**). Nevertheless, the sample is monodisperse and does not contain aggregates (samples were not filtered prior to deposition and imaging). The mixture of PhCPN and MSG before heating is really similar to PhCPN alone (**figure 4.15, D**). The width of PhCPN is of 17.2 nm with a variance of 1.0, taken on the double measurement of 10 PhCPN particles. Some of the pictures of the mixture of MSG and PhCPN prior to heating show extra densities inside and outside the cavity (**figure 4.15, D**) but it is impossible to conclude anything from them. These densities seem to have disappeared in the pictures of MSG and PhCPN after heating. This time, the width of PhCPN is of 16.9 nm with a variance of 1.5, taken on the double measurement of 10 PhCPN particles.

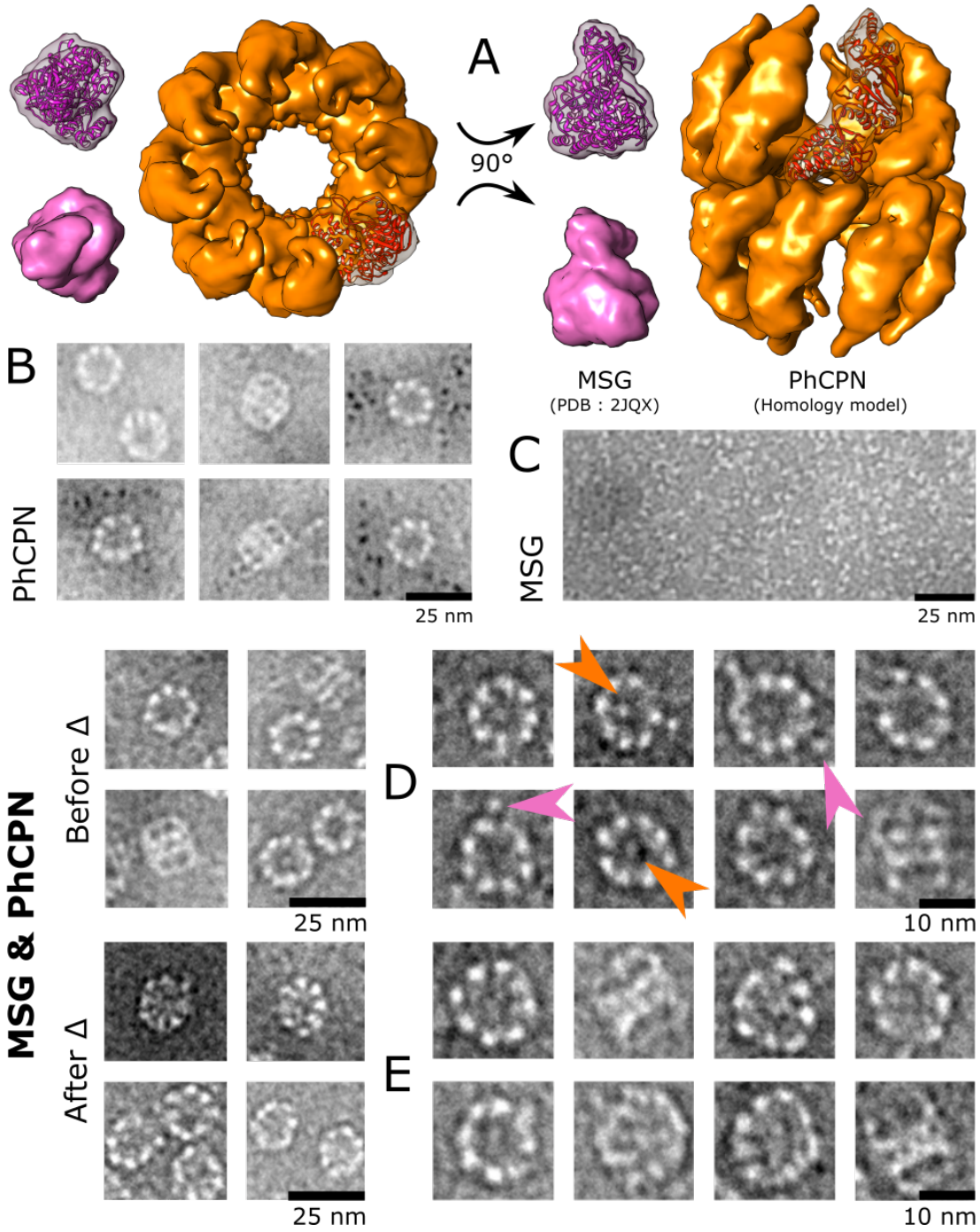


Fig. 4.15 : EM pictures of the MSG/PhCPN complex. (**A**) Models of PhCPN and MSG with relative sizing. MSG and one of PhCPN subunits are depicted with the ribbon-tracing mode. (**B**) Pure PhCPN EM pictures. (**C**) Pure MSG EM picture. (**D**) Pictures of a mixture of PhCPN and MSG prior to heating. Orange and pink arrows enlight additional densities respectively inside and outside the cavity. (**E**) Pictures of PhCPN and MSG after heating.

The variance is increased of 50% and the width is slightly different. The shape of PhCPN even seems altered and irregular, even though the monomers are still visible (**figure 4.15, E**). The control sample of PhCPN alone after heating did not show such modifications (**data not shown**). This could explain why PhCPN can only accommodate one MSG per PhCPN unit. If the protein, once inserted in the cavity, perturbs the shape of PhCPN, it could as well alter its ability to bind additional substrate proteins in the opposite cavity. This experiment was repeated several times and the results depicted on the pictures were reproducible. Samples were also tested by cryo-EM to eventually build a 3D-model of PhCPN to map the structures differences before and after heating. Unfortunately, the preferential orientations of PhCPN make the 3D reconstruction impossible. Finally, gold labelling was attempted on MSG, as it possesses a non-cleavable His-tag at its Nter. Regrettably, the gold particle did not attach on the tag. Nothing could be concluded on the position of MSG from these EM experiments, thus the necessity to use NMR has risen.

CSP mapping of the interaction interface

The goal of this experiment was to monitor the chemical shift perturbations (CSP, see **Chap. 6.3.1**) changes on the PhCPN spectrum caused by the interaction with MSG at high temperature after unfolding of MSG. These experiments were, in fact, the same ones as in **Section 4.2.1**, but Methionine and Valine regions of the spectra (corresponding to PhCPN) were monitored instead of Ile (corresponding to MSG).

At low temperature, the PhCPN spectra were identical in the presence or absence of MSG, as expected since PhCPN is was previously shown to interact exclusively with the unfolded form of MSG. However, at 60°C, small peak shifts can be observed when comparing the samples with or without MSG (**figure 4.16**). These shifts are observed in the Methionine region of the spectra while the Valine region remains mostly identical.

Such a result can be expected as the observed signal is the average of the signals emerging from 16 identical monomers. Even if MSG interacts with one or several of them, the signal difference will be averaged and the peak shifts will remain small or unnoticeable. These peak shifts correspond to Met27 and Met51, situated in the region of PhCPN that was shown to interact with unfolded HEWL [Mas et al., 2018]. Since the peak shifts were small and only visible for two residues, PRE-NMR was used.

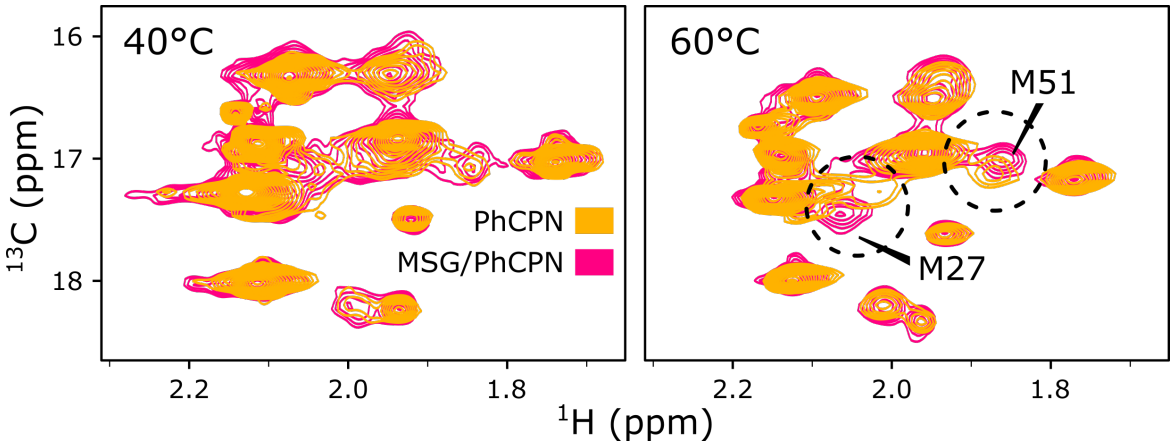


Fig. 4.16 : PhCPN 2D methyl SOFAST HMQC spectra in presence or absence of MSG. 2D HMQC spectra of 25 µM [U-D], (Met-ϵ,Val^{proS})-$^{13}CH_3$-labelled PhCPN at 40°C and 60°C in the absence (pink) and presence (orange) of 50 µM MSG. Spectra were recorded in D_2O PhCPN buffer, on an NMR spectrometer operating at a proton frequency of 850 MHz.

Determination of the interaction site by PRE

PRE NMR is a technique of choice to gain knowledge on the substrate location. Indeed, attaching paramagnetic tags on MSG and introducing it into methyl-labelled PhCPN would allow one to picture the PhCPN residues affected by the presence of a closeby paramagnetic center and yield information of MSG's position. As stated previously, the assignment for the Methionines (100%) and the Valines (70%) of PhCPN were available and provided us with accurate and localised probes to picture PhCPN globally. First, MSG was randomly spin labelled on its lysine residues with OXYL-1-NHS (**figure 4.17, A**). The goal was to determine experimental labelling conditions that would leave no label-free MSG after spin label reaction. The free spin label was 281 Da and would cause a mass increase of 166 Da per spin label attached (MSG possesses 32 lysine residues). Several labelling conditions were tested, different initial amounts of spin labels per lysine residues, different reaction times, and the presence or absence of DTT. Theoretical and experimental details can be found in **Chap. 6.3.1** and **Chap. 6.4** respectively.

One can observe that there is some MSG amino acids oxidation already present in the reference MSG sample (without spin labels, **figure 4.17, B**). The peak shape is different from the one in (**figure 4.17, C**) but the buffer conditions are different as well. In the presence of DTT in the reaction buffer, no spin label is attached (**figure 4.17, C**), probably because the nucleophile thiol function of DTT reacted with the esther group of OXYL-

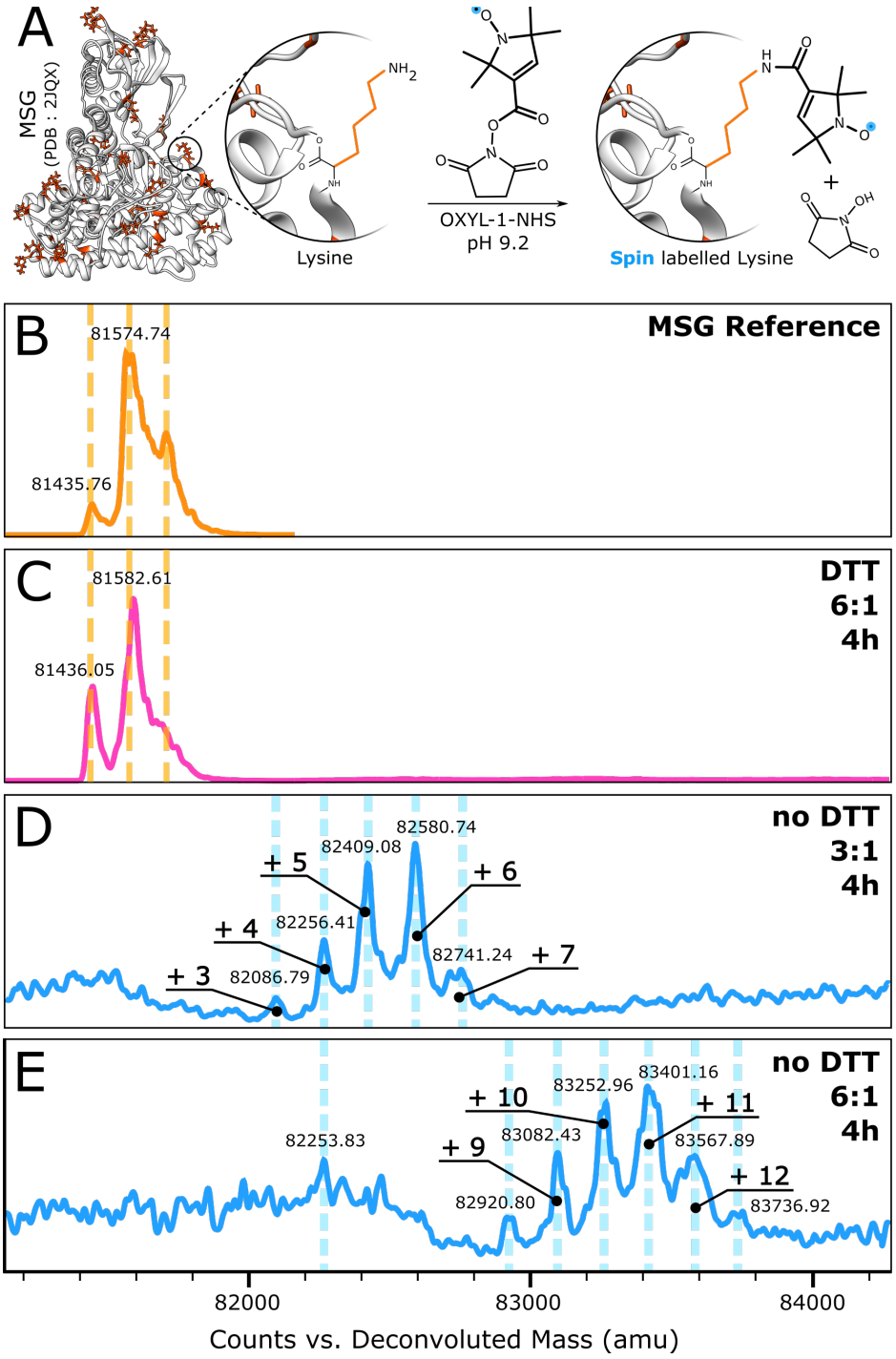


Fig. 4.17 : Spin labelling of MSG lysines by OXYL-1-NHS. (**A**) Scheme of the spin-labelling reaction. (**B**) Reference MS spectrum of MSG in labelling buffer without spin label and DTT. (**C**) Spectrum of MSG after spin label reaction, in presence of DTT. (**D**) Spectrum of MSG after spin label reaction, in absence of DTT and with a 3-fold molar excess of OXYL-1-NHS. (**E**) Spectrum of MSG after spin label reaction, in absence of DTT and with a 6-fold molar excess of OXYL-1-NHS.

1-NHS and made the labelling reaction impossible. In (**figure 4.17, D & E**) however, DTT was removed from the labelling buffer and spin labels are successively attached to MSG. In presence of 6-fold excess of OXYL-1-NHS, an average of 11 spin labels are successively attached to MSG. These conditions were kept for the final experiment with PhCPN. In the second part of the experiment, MSG was spin labelled and heated in presence of [U-D], (Met-ϵ, Val^{proS})-[$^{13}CH_3$]-labelled PhCPN. Precipitated protein was removed by centrifugation and ^{1}H-^{13}C HMQC spectrum was recorded (**figure 4.18, A & B, para spectra**). After the first HMQC experiment, the spin labels were reduced by using Na-ascorbate and a second ^{1}H-^{13}3C HMQC spectrum was recorded (**figure 4.18, A & B, dia spectra**). The spectra were assigned and intensities ratios plotted (**figure 4.18, E**), with overlapped residues indicated by asterisks. Some parts of the structure of PhCPN are clearly affected by the presence of the spin labels. These regions are reported on the monomer (**figure 4.18, D & E**) of PhCPN homology model (**figure 4.18, C**). Peaks situated in the so-called "sensor loop" of PhCPN are the most attenuated [Mas, 2015]. Two of these peaks, Met27 et Met51 were already perturbed by MSG in the first NMR experiments (**figure 4.18, C**). These results prove that MSG is directly bound to the bottom of the cavity. These PRE-induced CSP were used to locate MSG in PhCPN cavity and the results were included in the article published in 2018 in Science Advances [Mas et al., 2018] and copied in **Suppl. A.2**.

4.2.5 Refolding

The final step of this study was to investigate whether or not PhCPN was able to refold unfolded MSG hosted in its cavity. Mild and intense heating conditions were tested. One sample was heated progressively to reach 60°C after 10 minutes (**figure 4.19, orange curve**) while the other was directly heated at 60°C for the same amount of time (**figure 4.19, green curve**). ATP was added to the two types of samples after heating and its presence is characterized by its large absorption peak at 2.0 mL on the gel filtration profile.

Small traces of MSG can be still detected in solution after being heated mildly while it disappeared completely after the heat shock. MSG was inserted in PhCPN in the two samples, as shown by the increase of PhCPN elution peak intensity. However, no increase of intensity of the MSG peak was noticed after the addition of 1 mM ATP. Refolding of MSG in presence of ATP was also attempted by NMR. MSG was inserted into PhCPN cavity

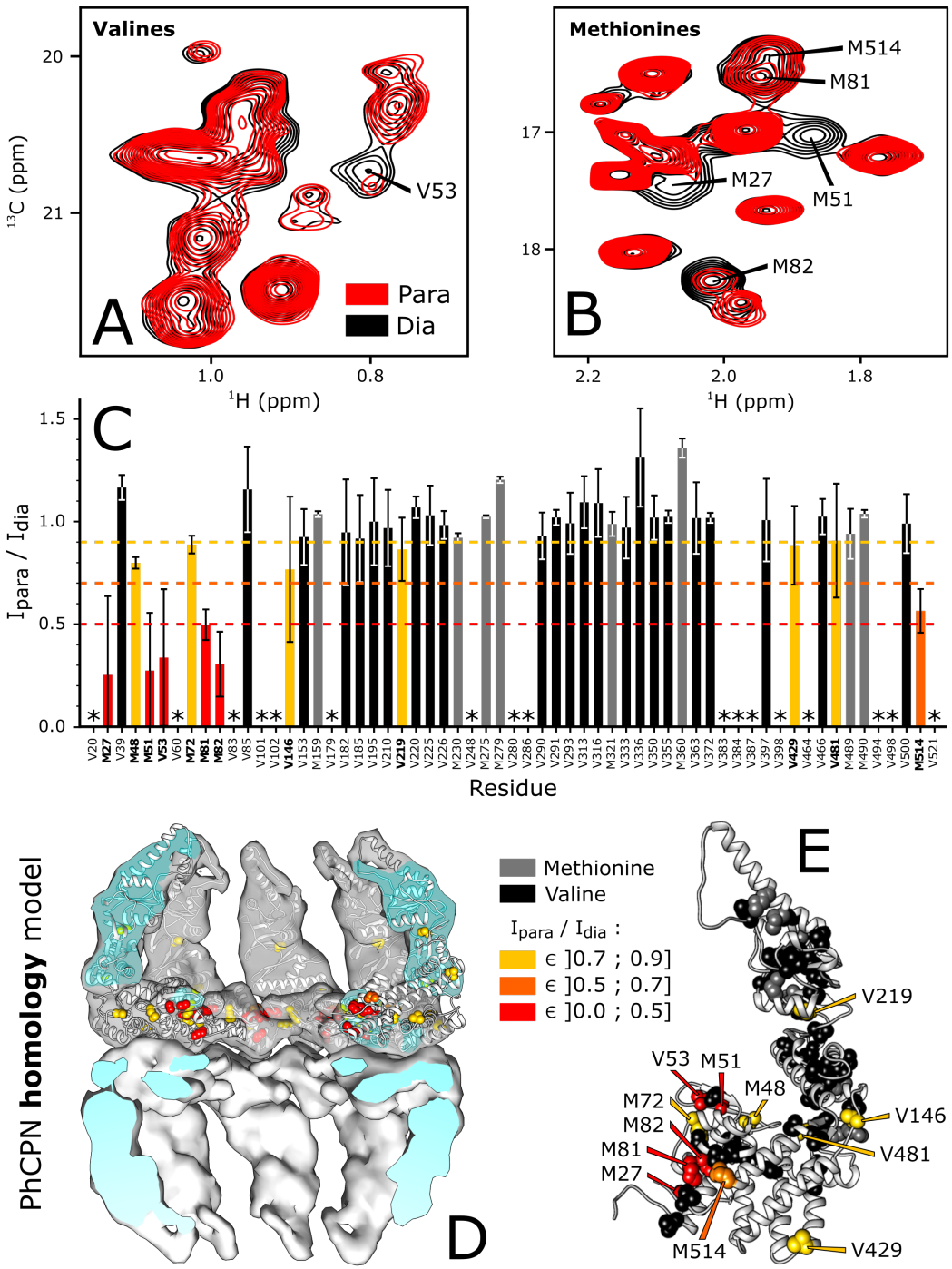


Fig. 4.18 : PRE-NMR to map PhCPN binding site. (**A**) 2D HMQC, Valine region and Methionine region (**B**) of the [U-D], (Met-ϵ, Val^{proS})-[$^{13}CH_3$]-labelled PhCPN in the presence of spin-labelled MSG (red) and reduced spin-labelled MSG (black). The residues most affected by PRE are highlighted. (**C**) PRE histogram of the [U-D], (Met-ϵ, Val^{proS})-[$^{13}CH_3$]-labelled PhCPN in the presence of the randomly spin-labelled MSG on the lysine residues. Unassigned and overlapped residues are indicated by asterisks (**D**) Homology model of PhCPN, see **Chap. 6.7**. (**E**) Mapping of the PRE effects on the homology model of PhCPN monomer.

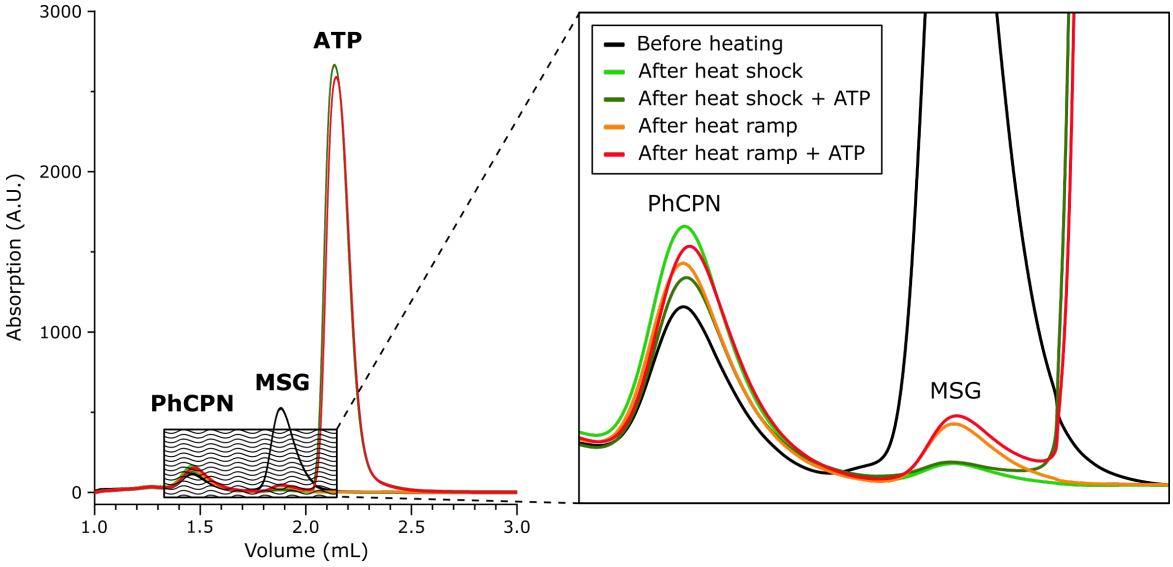


Fig. 4.19 : Influence of the presence of 1 mM ATP on MSG after heating, in presence of PhCPN. Different techniques were used to heat the samples prior to the experiment. Samples were analyzed after injection on a SEC-5 analytical gel filtration column.

similarly as in **Section 4.2.1**, and 10 mM ATP were inserted in the NMR tube just before spectra recording. The ATP hydrolysis capacity of PhCPN was monitored at the same time as the intensity of the MSG methyl spectrum, to eventually spot refolding or signal intensity loss, which would have corresponded to aggregation. Interleaved methyl SOFAST HSQC and ^{31}P 1D NMR spectra were recorded at 72°C to follow the open and closed states of PhCPN at the same time as ATP hydrolyses into ADP. The amount of PhCPN in the open or closed conformations was monitored by extracting the intensities of the Methionine residues M275 and M279 from the methyl SOFAST NMR spectra, reporting for the different structural states of PhCPN, following protocols described in [Mas et al., 2018] (**figure 4.20, D**). These data were then compared to those obtained in the same article (**figure 4.20, B, C, D & F**).

No signal of folded MSG could be monitored throughout the complete hydrolysis of the 10 mM ATP. At the same time, the signal corresponding to unfolded MSG trapped in PhCPN cage remained constant, showing that the energy provided by ATP hydrolysis was not enough to release MSG from the hydrophobic core of PhCPN cavities (data not shown). Interestingly, the amount of open/closed conformation at the ratio ATP/ADP : 58%/42% (**figure 4.20, A**) was measured and compared to what was obtained in [Mas et al., 2018]. In absence of substrate, PhCPN is able to close 57% of its monomers in

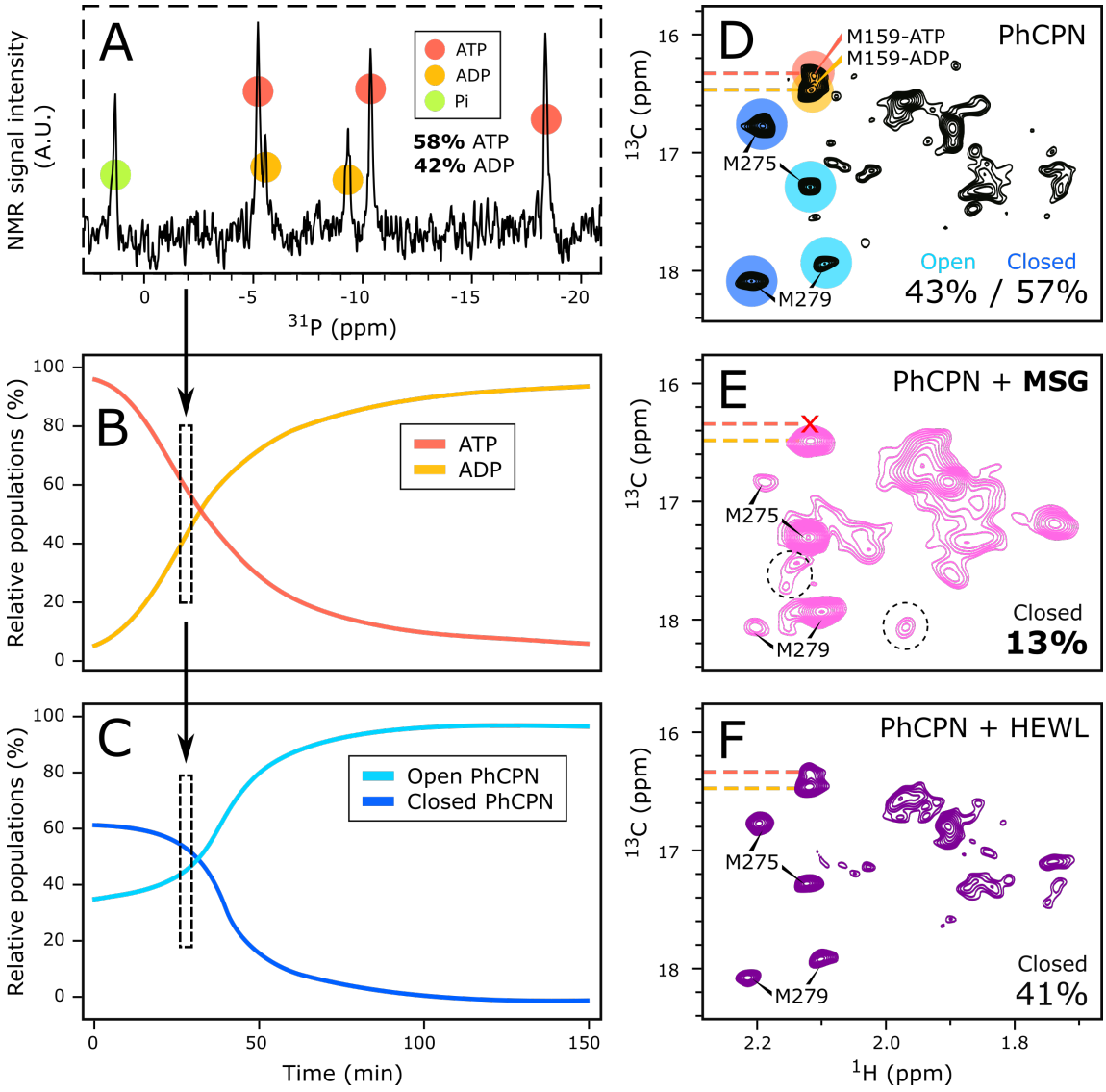


Fig. 4.20 : Open and closed conformations of PhCPN during ATP hydrolysis, in the absence or presence of substrate proteins inserted within its cavities. (**A**) $_{31}$P NMR spectrum of a mixture of PhCPN and ATP during ATP hydrolysis, corresponding to the ATP/ADP ratio at which spectra (**D**), (**E**) and (**F**) were recorded. (**B**) Percentage of ATP and ADP as a function of time during ATP hydrolysis by PhCPN at 75°C, in absence of substrate protein. (**C**) Percentages of open and closed forms of PhCPN, reported by NMR, during ATP hydrolysis at 75°C, in absence of substrate. (**B**) & (**C**) are adapted from [Mas et al., 2018]. (**D**) ^{1}H-^{13}C methyl SOFAST HMQC spectrum of PhCPN Methionines during ATP hydrolysis (approximately 60% ATP left , 75°C), in absence of substrates. 57% of PhCPN monomers are in a closed conformation. The light blue peaks correspond to the open conformation of PhCPN, the dark blue peaks to the closed conformation of PhCPN. (**E**) ^{1}H-^{13}C methyl SOFAST HMQC spectrum of PhCPN Methionines during ATP hydrolysis (58% ATP left, 72°C), in presence of MSG within the cavity. 13% of PhCPN monomers are in a closed conformation. (**F**) ^{1}H-^{13}C methyl SOFAST HMQC spectrum of PhCPN Methionines during ATP hydrolysis (approximately 60% ATP left, 75°C), in presence of HEWL within the cavity. 41% of PhCPN monomers are in a closed conformation. Spectra in figure (**D**) & (**F**) were recorded and kindly provided by Dr. Guillaume Mas.

presence of 58% ATP (**figure 4.20, B, C & D**), while the amount of closed PhCPN decreases in presence of HEWL within its cavity for a similar ATP/ADP ratio (**figure 4.20, F**). However, in presence of MSG inserted inside the cavity, the amount of closed monomers drops to 13% only (**figure 4.20, E**), and additional peaks can be observed in the Methionine spectrum (**figure 4.20, E, black circles**). Moreover, the signal of ATP-bound M159 is not present, which could indicate ATP binding impairment. This result shows that the ATP induced closure mechanism of PhCPN was very likely impaired by the presence of MSG in the cavity, possibly due to its large size, causing structural changes.

4.3 Chapter summary and conclusion

In this chapter, the interaction between MSG and PhCPN was characterized. It was successively shown that :

- PhCPN does not interact with folded MSG (**figure 4.21, A**) but when MSG unfolds with heat (**figure 4.21, B**), PhCPN prevents the aggregation of MSG by hosting it into its cavity and forming a stable complex (**figure 4.21, C**). The aggregation kinetics of MSG are only dependent on the free MSG concentration, proving that MSG does not exchange from a bound to a free form when unfolded with PhCPN. The complex is stable even at low temperatures
- The stable MSG/PhCPN complex is composed of one MSG per PhCPN particle and MSG is hosted in the core of PhCPN cavity. The structure of PhCPN is affected by the presence of MSG in one of its cavities, as proven by irregular shapes observed on EM images.
- PhCPN is unable to refold MSG at temperatures from 20 to 75°C, even in presence of ATP (**figure 4.21, D**)
- The structure of PhCPN and its ATP-fueled structural changes from an open to a closed state are impaired by the presence of MSG in the cavity, possibly due to steric perturbations

These results are partially in agreement with previously published data on the interaction between HEWL and PhCPN [Mas et al., 2018]. First, PhCPN only interacts with

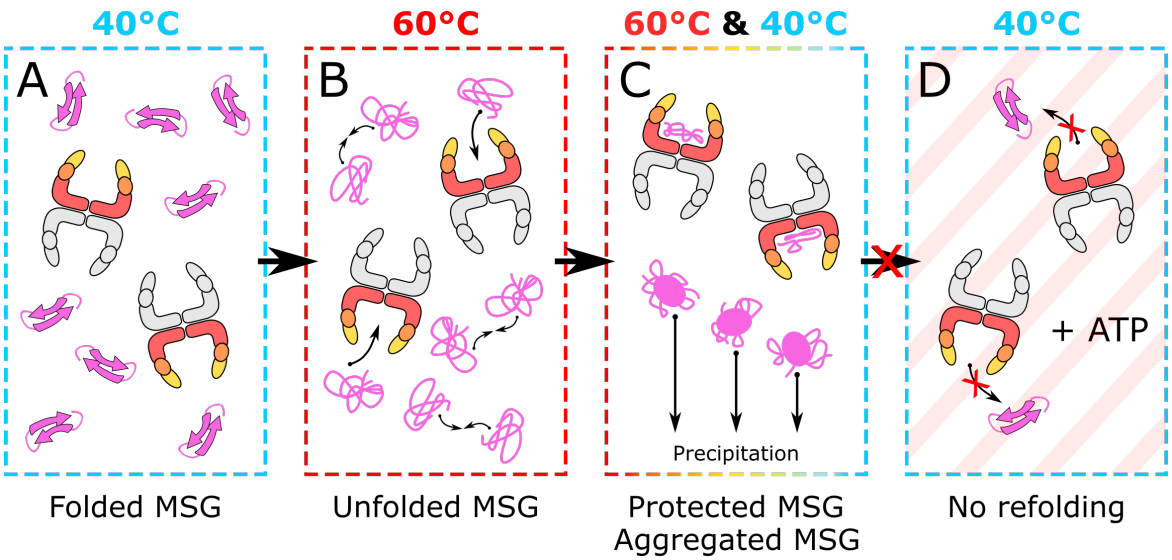


Fig. 4.21 : Scheme describing the interaction between PhCPN and MSG at different temperatures. (**A**) MSG is well folded at low temperature and does not interact with PhCPN. (**B**) At 60°C, MSG is immediately unfolded by temperature. (**C**) If MSG is in excess compared to PhCPN, two different things happen. A fraction of unfolded MSG is hosted by PhCPN while the rest aggregates and precipitates. (**D**) The fraction of MSG protected by PhCPN cannot be refolded, even in presence of ATP.

unfolded MSG, as reported for HEWL. MSG and HEWL were also shown to interact in the same region of PhCPN. Even though other studies mentioned interaction with the apical domains of PhCPN monomers, the substrate is located next to the equatorial domain. The similarities end here, as the rest of the results are quite different from previously observed ones with HEWL. The ratio Substrates/PhCPN is different in both cases, with one substrate per cavity with HEWL, thus two HEWL per one PhCPN, and only one substrate per PhCPN with MSG. This can be explained by the large size of MSG damaging the cavity, or simply the cavity adapting to MSG size and changing the allostery of the second cavity. The melting temperature of HEWL is also lowered in presence of PhCPN, as it stabilizes its unfolded state and HEWL is in a permanent exchange between its folded and unfolded state at high temperature. It is not the case for MSG, while it was previously reported that GroEL could help the refolding of MSG denatured by Guanidine [Dahiya and Chaudhuri, 2014]. This is possibly because the thermal unfolding of MSG is different from chemical unfolding and is not a reversible process, even in presence of ATP (**figure 4.21, D**). Several factors might explain this incapacity to refold MSG. First, the ATP-dependant closure mechanism of PhCPN cavity is induced by heat, optimally around 70°C [Mas, 2015]. This temperature is too elevated for MSG, thus the need to stay below 40°C for refolding. This temperature

could be too low for PhCPN to refold such a large substrate. Another justification would be that MSG is too large for proper refolding by a machinery like PhCPN and might need other additional chaperones to properly fold again. A last and more likely explanation would be that PhCPN is not functional anymore because of the presence of MSG in one of its cavities, and the lack of space to refold MSG within the cavity. The presence of the substrate in the cavity can also stabilize the open conformation of PhCPN and the energy provided by ATP hydrolysis is not sufficient to close the cavity anymore.

These results shed light on a new behaviour of PhCPN in presence of large aggregating substrates. PhCPN is able to host substrates of the same order of size compared to its cavities when folded, as it is the case with MSG, see **Section 4.2.4**. In this fashion, PhCPN reduces the amount of unfolded protein in solution by protecting the host protein in its cavity while losing its functionality. At the same time, PhCPN loses its ability to refold proteins since its closing mechanism is impaired by the presence of the large substrate in its cavity. The situation is probably different *in vivo*, as other chaperones are present to help folding and refolding large physiologic substrates with a size comparable to MSG. Moreover, large proteins and therefore potential substrates represent only a small portion of the native Proteasome and such interaction is unlikely. For instance, MSG is 731 AA long while the median protein size in archaea is only of 247 AA [Brocchieri and Karlin, 2005].

Chapter 5

Amyloidogenic protein model

Ce chapitre résume cette fois l'étude du Thermosome de l'archée Pyrococcus horikoshii, PhCPN, en interaction avec une protéine amyloïde, l'Amyline, impliquée dans le diabète de type II. Ce substrat est très différent de MSG car il est de très petite taille et possède la caractéristique de pouvoir s'auto-associer et former des fibres amyloïdes, d'où son instabilité. La RMN permet de montrer que PhCPN interagit préférentiellement avec la région N-terminale de l'Amyline et que l'Amyline est en échange rapide entre sa forme liée et sa forme libre. De plus, il est conjecturé que PhCPN est capable de piéger des oligomères d'Amyline, réduisant considérablement sa capacité à former des fibres. Ce chapitre montre également que l'état oligomérique de PhCPN influe grandement sur sa capacité à interagir avec ses substrats et montre que la cavité de PhCPN est nécessaire pour inhiber la formation de fibres amyloïdes.

Another type of proteins studied during this thesis was amyloidogenic proteins. As stated in **Chap. 1.3.2**, these proteins are of particular interest because they are involved in numerous human diseases. As most of them are intrinsically disordered in solution, hence exposing their most hydrophobic patches, they are potential candidates for PhCPN substrates. This chapter summarizes the stepwise study of PhCPN interaction with an amyloidogenic protein, including :

- The amyloidogenic protein substrate selection
- The assignment of the amyloid protein in the working buffers
- The characterization of the interaction between the amyloid protein and PhCPN by NMR, in different buffers
- The differences in aggregation kinetics of the amyloid protein in the absence or presence of PhCPN
- The observation and characterization of the fibrils by imaging techniques

As PhCPN is a very large assembly (1 MDa) and amyloidogenic proteins are usually very sensitive to experimental conditions, the study methods differ greatly from those of the previous chapter and are precisely detailed in **Chap. 6**.

5.1 Suitable protein model

As previously mentioned, PhCPN is a ubiquitous protein and is able to interact with a various set of substrates, such as unfolded proteins or intrinsically disordered proteins (or **IDP**, see **Chap. 4** and **Chap. 2.3.3**), provided that they possess or expose hydrophobic residues. Especially, the PhCPN prokaryotic homologous GroEL was already showed to interact an amyloidogenic peptide by NMR [Libich et al., 2013].

The goal of the second part of this project was to investigate the interaction between PhCPN and an amyloidogenic protein by NMR using a set of various biophysical techniques. As for aggregating proteins, the substrate protein should meet certain requirements prior to the beginning of the study :

- Fibrillating kinetics which can be followed by NMR at high concentrations (between 25 to 200 µM)

- Compatible with PhCPN buffer
- No high sensitivity to heat. The fibrillating protein should stay soluble for at least a few hours at temperatures above 30°C
- Handleable in an L2 level laboratory
- Available in significant quantities for NMR

It was also recommendable but not necessary that the protein would exhibit a small size this time, to complement the work and observations obtained prior to this thesis with Lysozyme (14.3 kDa ; [Mas et al., 2018] and the data obtained with MSG (82 kDa ; **Chap. 4**). Several proteins fulfilled the above requirements. In fact, as fibrillating proteins are involved in numerous aging diseases, literature relevant for this study is proficient. Particularly, prokaryotic GroEL/ES was already shown to interact with both α-Synuclein and Amyloid-β 1-40 by the group of Pr. Koichi Kato [Nishida et al., 2013, Yagi-Utsumi et al., 2013] and with Amyloid-β 1-40 by the group of Dr. Marius Clore [Libich et al., 2013]. Collaborators of the Institute for Physical Biology at the Heinrich-Heine-University in Düsseldorf posses state of the art knowledge in amyloidogenic protein biochemical and structural studies, as well as an adapted facility for their production. They kindly provided three different amyloids to be tested for this study, α-Synuclein, Amyloid-β 1-42, and another fibrillating protein, the Amylin (**or IAPP**).

5.1.1 α-Synuclein

The human α-Synuclein protein (α**Syn**) is composed of 140 amino acids, for a molecular weight of 14.5 kDa. Even though this protein is intrinsically disordered, its residues from 1 to 60 show α-helix propensity when interacting with lipidic surfaces [Rao et al., 2010]. αSyn is abundant at the tips of neurons, in presynaptic terminals ensuring the transmission of the nervous signal [Jakes et al., 1994]. This protein is also found in other tissues in lower amounts, such as in the heart or muscles, as well as red blood cells. The function of this protein is still not yet completely understood [Ghiglieri et al., 2018] but some mutations of the gene coding for αSyn increase the risk to develop Parkinson's disease. Fibrillated and aggregated αSyn proteins are found in disorders known as synucleinopathies. They are present in Lewy bodies, located in the brains of patients with Parkinson's disease

[Spillantini et al., 1997]. This protein is particularly fit for NMR studies as it is stable in various sets of buffers and its structure was already determined by NMR, when interaction with lipid micelles [Rao et al., 2010] (**figure 5.1, A**). Since this protein was previously shown to interact with GroEL/ES by NMR, the first experiment consisted in reproducing that data using PhCPN [Nishida et al., 2013],(**figure 5.1, B, red bars**)). Assignment of 76% of the NH frequencies was transposed directly using previously reported assignment from the BMRB database (BMRB entry n°18857). Even though the assignment was not complete, it was sufficient for first interaction trials as it extended throughout the whole structure. 20 µM ^{15}N isotopically labelled αSyn and 40 µM unlabelled PhCPN in PhCPN buffer (25 mM MES pH 6.5, 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$) were mixed at 4°C and (^{15}N)-BEST-TROSY NMR experiments were recorded [Solyom et al., 2013].

The residues numbers were plotted against the ratio of the intensity of the αSyn peaks with and without PhCPN (**figure 5.1, B**), in PhCPN buffer. A significant intensity loss was reported in the literature for the Nter region of αSyn as well as around the 40th residue, in pH 7.0, 50 µM phosphate buffer containing 100 mM NaCl (**figure 5.1, B, red**). The

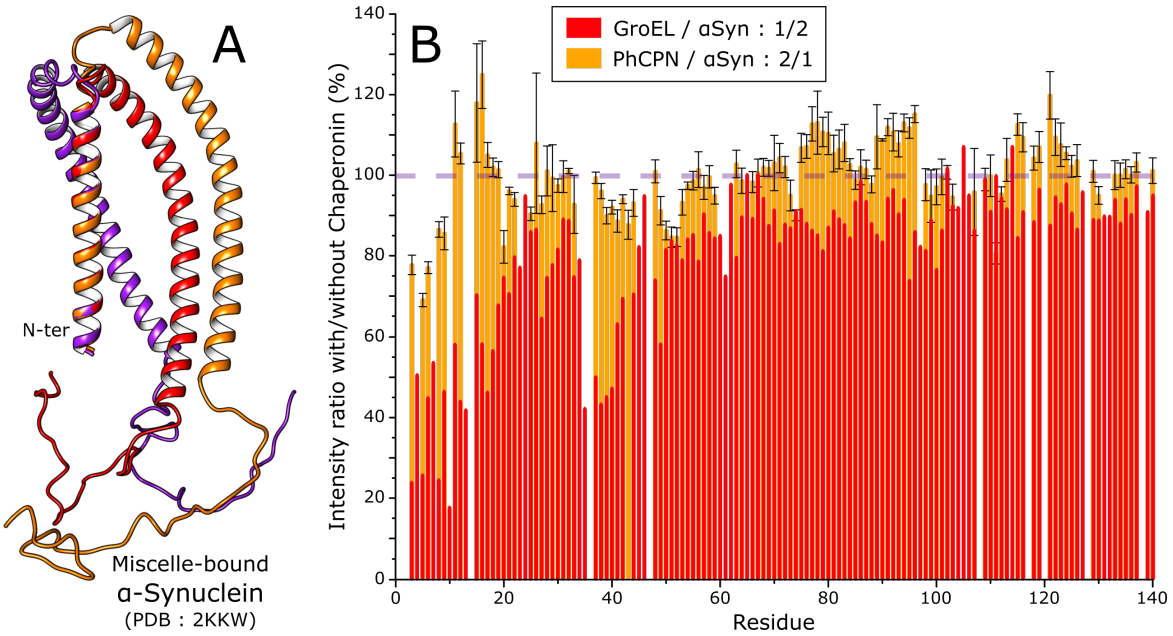


Fig. 5.1 : Structure and interaction of αSyn with GroEL and PhCPN Chaperonins. (**A**) Three NMR structures of α-Synuclein inserted in model membranes (PDB : 2KKW [Rao et al., 2010]. (**B**) Intensity ratios of the NMR signal of ^{15}N-labelled αSyn in presence or absence of PhCPN or GroEL. Data provided for GroEL were adapted from [Nishida et al., 2013]. BEST-TROSY data for PhCPN were recorded on a 700 MHz Bruker spectrometer, equipped with a 5 mm Cryo-Probe.

interaction with PhCPN was weaker in the Nter region and absent in the other region (**figure 5.1, B, orange**). In addition, the ratio PhCPN/αSyn was 4 times higher than the sample with GroEL and no peak shifts were observed, reinforcing the idea that both proteins were not tightly interacting.

5.1.2 Amyloid-β 1-42

Amyloid-β (**A**β) is an intrinsically disordered peptide cleaved from the amyloid precursor protein. Aβ is mainly produced in neurons but is also found in skin and skeletal muscles [Puig and Combs, 2013]. Aβ can aggregate to form seeds, or auto assemble into oligomers and amyloid fibrils, which structure varies greatly depending on the experimental conditions [Petkova et al., 2005]. A recent high-resolution CryoEM study of Aβ42 revealed a screw-like organisation of two intertwined protofilaments, forming the fibrils [Gremer et al., 2017], (**figure 5.2, B**). Nevertheless, other valid and different structures are available as well. Oligomers are largely admitted to being toxic [Cleary et al., 2005] but seeds were also shown to be responsible for fibrillation of Aβ and other proteins such as Tau, in a prion-like mechanism [Nussbaum et al., 2013].

The length of the Aβ peptide is variable, from 36 to 43 AA, the most prevalent species being Aβ40 and Aβ42. The Aβ40 peptide is more common but Aβ42 shows a high propensity to form aggregates or amyloid fibrils and is thus associated with diseases [Jarrett et al., 1993]. Aβ is involved in cognitive impairment diseases, the two main ones being Alzheimer's and Down syndrome. In fact, Aβ is found in amyloid plaques in the brain of Alzheimer's patients, similar to Lewy bodies in the case of αSyn and Parkinson's disease. As for αSyn, its function is not completely understood to date and there is still a debate over the causal link between Aβ and Alzheimer's. The initial belief, also called the "amyloid hypothesis" by Georges Glenner in 1984, was that since misfolded Aβ is present in abnormal quantities in the brain of Alzheimer's patients, Aβ was the central cause of the disease and needed to be targeted or depleted. This theory is now largely questioned since Aβ was found to be highly conserved [Tharp and Sarkar, 2013] and to play various essential protective roles in the human body, reviewed in [Brothers et al., 2018]. In fact, Aβ was proven to be an antimicrobial peptide, to promote brain recovery in cases of brain injuries, and to act as a tumor suppressor, since patients with Alzheimer's have 50% fewer chances to develop some forms of cancer [Aiello Bowles et al., 2017]. Most recent

studies seem to satisfy both theories, as the formation of oligomers could be triggered as a response to infection by the Herpes simplex Virus 1 [Eimer et al., 2018]. The oligomers and fibrils would form a physical barrier against the virus but in the meantime accelerate the development of Alzheimer's disease.

Aβ42 was slightly more challenging to study than αSyn due to its higher instability and fibrillation propensity. Nevertheless, it fit for NMR studies as it is a small peptide, allowing the acquisition of high-quality ^{1}H-^{15}N 2D HMQC NMR spectrum in short experimental time. Since Aβ40 was also proven to interact with GroEL/ES by NMR, the first experiment consisted in reproducing the data with PhCPN [Yagi-Utsumi et al., 2013] (**figure 5.2, A, dark blue bars**). 50 µM ^{15}N isotopically labelled Aβ42 and 25 µM unlabelled PhCPN in PhCPN buffer (25 mM MES pH 6.5, 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$) were mixed at 4°C and sequential (^{15}N)-BEST-TROSY NMR experiments were recorded. Aβ42 strongly reacted to PhCPN buffer and fibrillated within a few minutes, making the recording of a proper NMR spectrum impossible. Buffer conditions were exchanged for a "salt-free" buffer composed of 50 mM Tris pH 7.5, lacking Magnesium required to maintain the hexadecameric structure of PhCPN. Sequential NMR spectra were recorded again and this time, 93% of the resonances could be assigned by transferring the existing assignment table provided by collaborators of the Heinrich-Heine-University. The 7% percent left corresponded to peaks which were not detected. The residue number was then plotted against the ratio of the intensity of the Aβ42 peaks with or without PhCPN (**figure 5.2, A, light blue bars**).

On the first recorded spectrum, the "interaction pattern" was highly similar to the one observed with GroEL. Residues corresponding to the most hydrophobic regions of Aβ42 were strongly impacted by the presence of PhCPN and their resonances intensity reduced (**figure 5.2, A, light blue bars**). In addition, no peak shifts were observed. The rest of the peaks are less or not affected since the non-interacting parts of the protein can tumble more freely as it would in solution without monomeric PhCPN. The results could however not be reproduced due to the apparition of new sets of peaks, which could be characteristic of peptide degradation or indicate the presence of small Aβ oligomers (**figure 5.2, C, violet spectrum**). To identify the underlying cause of these observations, the sample was submitted to electrospray-mass analysis (**figure 5.2, D**). Two sets of fragments were detected : fragments at m/z = 1366.99 (3+) & m/z = 1025.49 (4+) corresponding to a loss of 466 Da (approximately 4 AA) and fragment at m/z = 1432.25 (2+) corresponding

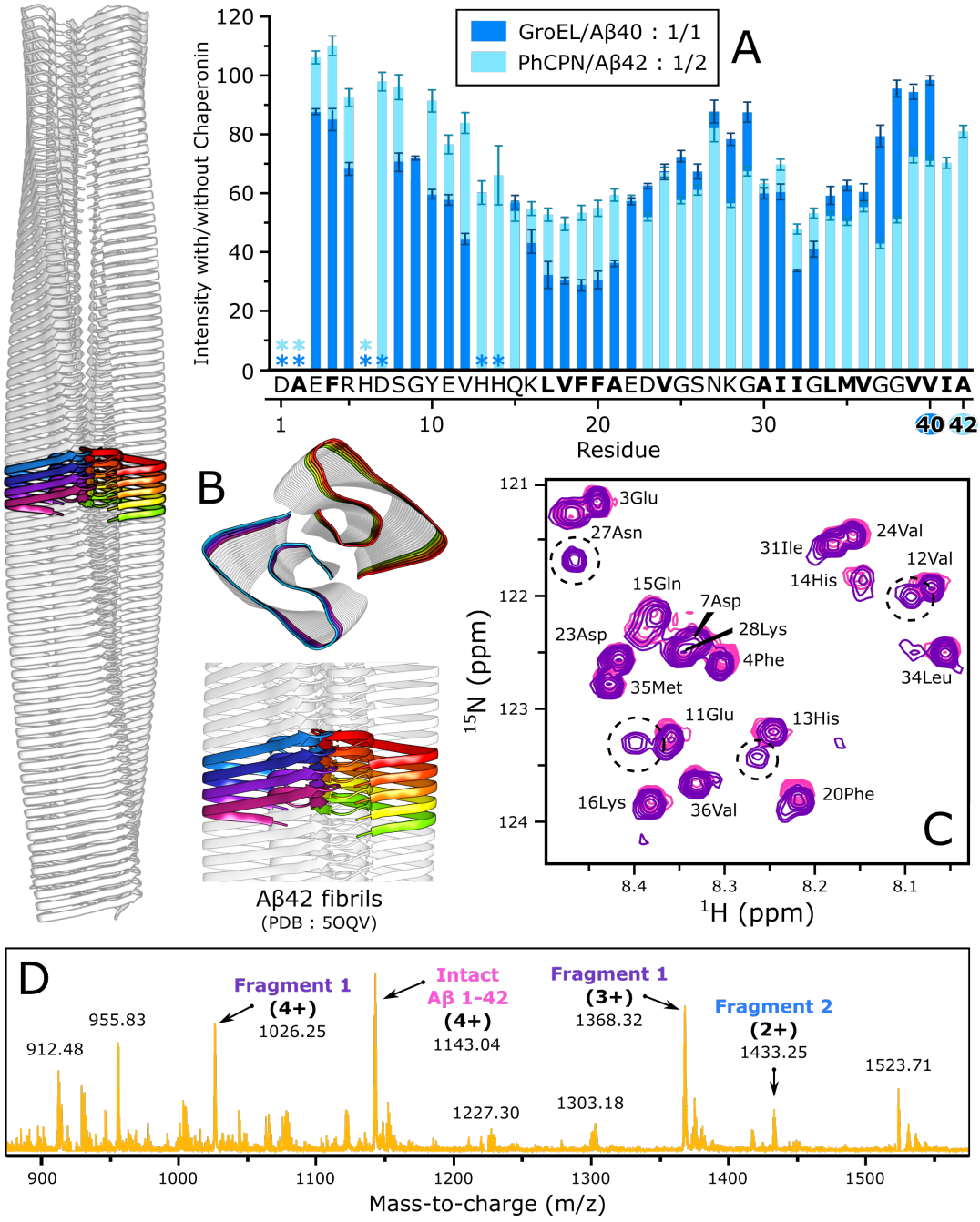


Fig. 5.2 : Study of Aβ42. (**A**) Intensity ratios of the NMR signal of (^{15}N)-Aβ40 or Aβ42 in presence or absence of 20 µM GroEL or 25 µM PhCPN respectively. The sequence of Aβ42 is reported under the graph, hydrophobic residues are in bold. Data provided for GroEL were extracted from [Yagi-Utsumi et al., 2013]. BEST-TROSY data for PhCPN were recorded on a 700 MHz Bruker spectrometer, equipped with a 5mm Cryo-Probe. (**B**) Different views of the EM structures of Aβ42 fibrils, (PDB : 5OQV ; [Gremer et al., 2017]. (**C**) Degradation of Aβ42 over the time observed by BEST-TROSY NMR. The violet spectrum showing degradation was taken 30 minutes after the pink one, without degradation. (**D**) MS spectrum Aβ42 in the violet sample in (**C**).

to a loss of 1700Da (approximately 15 AA). From the mass spectrum, it was clear that the sample was subjected to degradation.

5.1.3 Amylin

Amylin, or Poly Islet Amyloid Peptide (**IAPP**), is a pancreatic hormone (**figure 5.3, A**) first isolated and purified from amyloid plaques found in a patient's pancreatic tumor in 1986 [Westermark et al., 1986]. Amylin is derived from an 89 amino acids long precursor known as preProIAPP, which is cleaved and post-translationally converted into Amylin [Krampert et al., 2000]. This 4 kDa and 37 amino acid long amyloidogenic peptide is amylated at its Cter and exhibits a disulfide bridge between its residues 2 and 7, forming a Nter loop [Rink et al., 1993]. The residues from position 20 to 29 were found to be responsible for the amyloidogenic properties of Amylin [Westermark et al., 1990]. Amylin is co-stored and co-expressed with insulin by the pancreas, and specifically by pancreatic β-cells found in islets of Langerhans (reviewed in [Hay et al., 2015] (**figure 5.3, B & C**)).

Nevertheless, Amylin is also secreted by the central nervous system and specific regions of the brain involved in metabolic control [Li et al., 2015]. The peptide circulates in the bloodstream to activate receptors in the brain, trigger glucagon release from the pancreas and produce a satiation signal. In addition, Amylin controls the production of β-cells by the pancreas, depending on the blood glucose content. In presence of high amounts of glucose, the β-cells proliferation is reduced while it is inversely increased by the lack of glucose [Visa et al., 2015]. Amylin is also linked to diabetes, as Amylin amyloid fibrils are involved in disruption of pancreatic β-cells and the development of type II diabetes *mellitus*. In fact, the Amylin oligomers are lethal to pancreatic β-cells, leading to a decrease in insulin production and thus causing type II diabetes [Zraika et al., 2010] (**figure 5.3, D, E & F**)). The toxicity mechanism seems to be caused by different and complementary molecular mechanisms. Amylin oligomers are able to disrupt cell membranes by forming pores and disrupting ionic homeostasis, leading to cell death [Mirzabekov et al., 1996]. In parallel, different studies point out the fact that Amylin oligomers can generate reactive oxygen species toxic for the organism and also activate apoptotic pathways of the cells. Interestingly, oligomerization of Amylin can be induced in the healthy pancreas by injecting Amylin seeds, triggering type II diabetes, implying that Amylin can act as a prion, in the exact same fashion as Aβ [Mukherjee et al., 2017].

The functional study of Amylin is indeed challenging since the protein is both a cure, used to treat obesity (rev in [Boyle et al., 2018], and a problem, involved in type II diabetes *mellitus*. On a technical point of view, Amylin is similar in size to Aβ42 and fibrillates easily. In addition, collaborators have previously set up an efficient strategy for the recombinant expression and purification of this protein [Mirecka et al., 2014b].

At the beginning of this thesis, there were no studies published on the interaction of Amylin with a Chaperonin. The experimental setup used with the two other amyloidogenic proteins was reproduced and the concentration of Amylin was similar to the one used for Aβ42. 50 µM ^{15}N-labelled Amylin and 3.3 µM & 6.6 µM unlabelled PhCPN were mixed together in PhCPN buffer (25 mM MES pH 6.5, 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$) at 10°C and BEST-TROSY NMR experiments were recorded. A clear interaction between PhCPN and Amylin was observed. In fact, the intensity of the Amylin peaks decreases dramatically when the PhCPN is added, so that 21 out of 33 visible peaks disappear at the plotted spectrum contour level (**figure 5.3, G & H**). Preliminary assignment of the protein spectrum was transferred from the BMRB (BMRB entry n°18795) the chemical shifts significantly differed as the experiments were carried out in another buffer. For proper study of Amylin in PhCPN buffer, the protein had to be fully assigned again. In fact, the assignment by successive titrations from one buffer to another would have been too much protein consuming and could have caused protein fibrillation due to sample perturbation. The stability of Amylin alone was checked by leaving the sample in the spectrometer at 30°C for several hours. Degradation similar to the one observed with Aβ42 was observed (**data not shown**) and therefore experimental conditions needed to be optimized to work in proper and reproducible conditions.

5.1.4 Amylin, the selected model substrate

From these preliminary experiments, the Aβ42 and Amylin amyloidogenic substrates seemed fit for further investigations, while αSyn gave mitigated results. The results could be summarized as follows :

- αSyn was stable in the experimental buffer. However, the interaction between αSyn and PhCPN was weaker than expected, even in presence of higher amounts of PhCPN relative to prior studies. In addition, the interaction between αSyn and a

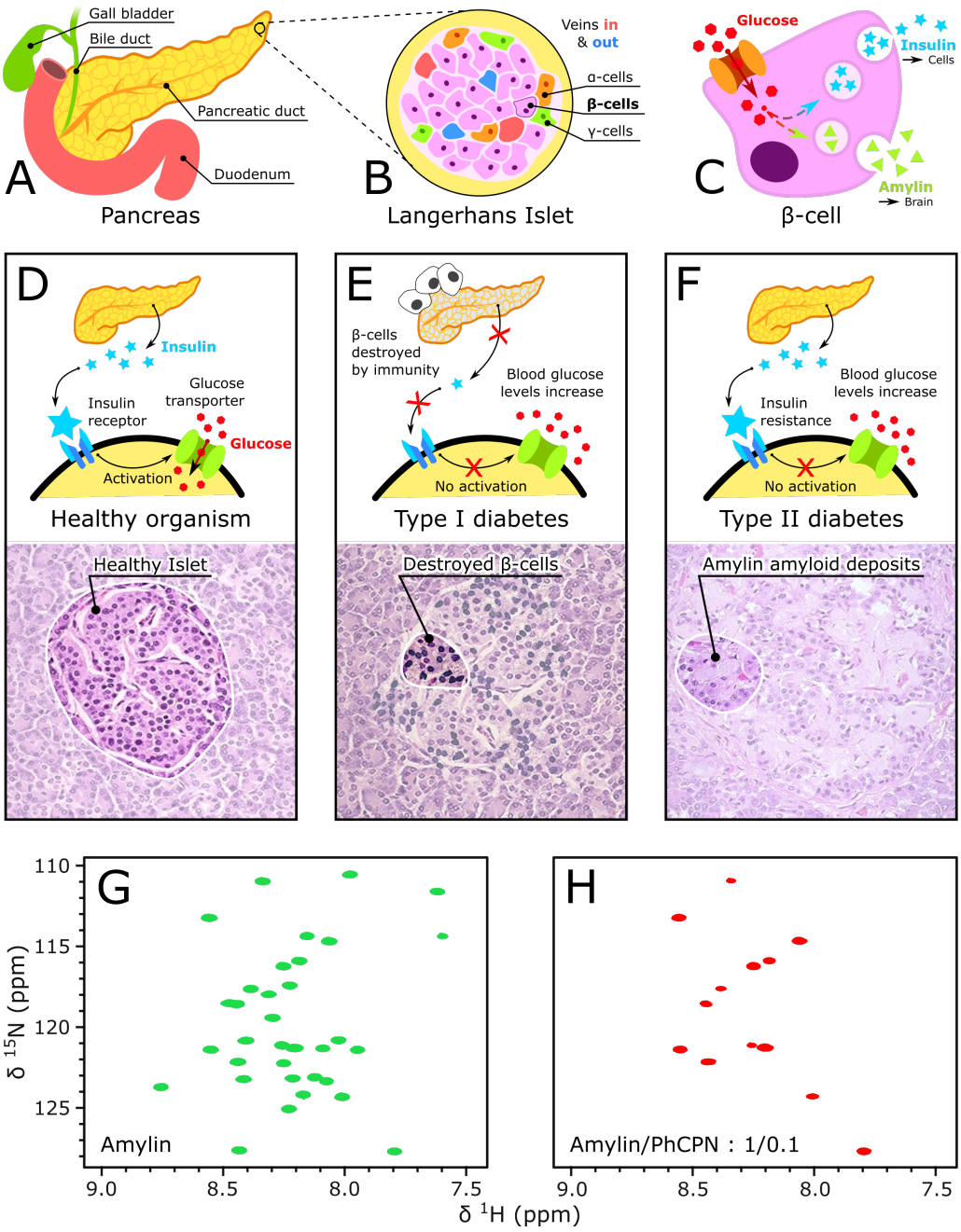


Fig. 5.3 : Diabetes mechanism and first interaction test. (**A**) Scheme of the human pancreas. (**B**) Scheme of the Langerhans islets, found in the pancreas. (**C**) The principle of the β-cell function. (**D**) Regulation of glucose levels by a healthy organism and a section from a normal pancreatic islet of Langerhans. (**E**) Glucose regulation malfunction and a section from an islet being destroyed by lymphocytes in an autoimmune mechanism, characteristic of type I diabetes. (**F**) Glucose regulation malfunction and a section from an islet containing amyloid deposits constituted of derivatives of Amylin, characteristic of advanced type II diabetes. (**G**) and (**H**) NMR ^{1}H-^{15}N BEST-TROSY signal of 50 µM (^{15}N)-Amylin in the absence (**G**) or presence of PhCPN (**H**), recorded on a 600 MHz Bruker spectrometer, equipped with a 5mm Cryo-Probe. Histological slides in (**D**), (**E**) and (**F**) were adapted from the medical library of Utah website, https://library.med.utah.edu/.

chaperonin was already investigated before [Nishida et al., 2013], and αSyn size was really similar to the one of HEWL (14 kDa and 14.5 kDa respectively).

- Aβ42 was extremely sensitive to the buffer salt content and pH, as shown by its very fast fibrillation propensity. However, the peptide had reasonable stability in a buffer lacking $MgCl_2$ in which PhCPN is mainly monomeric. Aβ42 was also sensitive to degradation but its interaction with PhCPN was comparable to previously observed results with Aβ40 and GroEL.

- Amylin is really stable in different buffers. The protein is also sensitive to degradation but it can be prevented using proper experimental conditions. The protein was never studied and its purification and isolation were recently optimized when the project was started, making it a challenging substrate protein. The interaction pattern is slightly different from the one of Aβ42.

To conclude, the results with Amylin were more encouraging than the ones with Aβ42 and αSyn but some problems needed to be addressed in order to carry on studying this protein. Especially, a way to prevent degradation needed to be found and the assignment of NMR spectrum of Amylin had to be performed again. This model was then selected for further studies, aiming at characterizing the interaction between chaperonin and amyloidogenic proteins.

5.2 Characterization of the interaction

Once the model protein was selected, the stewise goals of the study were the following :

- Prevent Amylin degradation
- Assign the NMR spectrum of Amylin in PhCPN and MES buffer
- Characterize the interaction between Amylin and PhCPN by NMR, in different buffers
- Monitor the aggregation kinetics of Amylin in the absence or presence of PhCPN
- Observe and characterize the fibrils by imaging techniques, in different buffers and in the presence or absence of PhCPN.

Characterizing such an interaction would allow us to gain knowledge on the behaviour of PhCPN and equivalent chaperonins when in interaction with small substrates (4 kDa), to complement prior studies with HEWL (14.5 kDa ; [Mas, 2015] and MSG (82 kDa ; **Chap. 4**). In fact, little is known about the behaviour of a Chaperonin in presence of an amyloidogenic substrate. The chaperonins could for instance simply lower the apparent concentration of amyloidogenic proteins in solution like with MSG (see **Chap. 4**. They also could be directly involved in the fibrillation process by interacting specifically with a folding intermediate or the fibrils. In this chapter, NMR and complementary biophysical and biochemical tools were used to address these points.

5.2.1 Optimization of the experimental conditions

Several problems emerged during the preliminary experiments and needed to be addressed before further studies. The first and main concern was regarding peptide degradation, which drastically reduces the intact protein content and yields highly inaccurate results. Especially, degradation is favored when working with short and intrinsically disordered proteins or peptides lacking tertiary/quaternary structure, since degradation is enhanced by the accessibility of the peptidic chains [Parsell and Sauer, 1989].

Prevention of the degradation

In order to prevent the degradation of the peptide, the source of the contamination needed to be determined. It was found that both PhCPN sample was contaminated by proteases. In fact, proteases are studied by other scientists in the NMR group, especially TET (a P*yrococcus horikoshii* protease) and ClpP (an *E. coli* protease), and are purified on the same columns as PhCPN, since columns are shared within projects. The purification process of PhCPN was modified by adding a thorough column cleaning before each purification. Filtration steps were also added after gel filtration in order to prevent eventual bacterial contamination prior to -20°C storage. Even with the exceptional stability of PhCPN, samples were kept frozen and aliquoted. PhCPN samples were unfrozen individually prior to every experiment and the remaining protein was not re-used. The rest of the sample preparation protocol was also modified, encouraging the use of cones equipped with filters to pipette the proteins and buffers. The buffer content was slightly modified by adding an extra 0.02% w/w of sodium azide as a bactericide. As a supplementary control,

all the samples, buffers and materials (syringes and pipettes) used for the preliminary tests were also deposited on antibiotic-free LB agar Petri dishes to eventually eliminate materials in case of potential bacterial contamination.

All these additional precautions allowed to prevent completely the degradation of Amylin, as assessed by Mass Spectrometry (**MS**) (**figure 5.4**). In fact, before the sample optimization, the peptide was so degraded that the peak corresponding to Amylin was absent on the spectrum. After optimization, the sample was pure and only composed of intact 15N Amylin (theoretical mass of 3956 Da, see **Suppl. A.1.3**). Nevertheless, in the case of NMR experiments, NMR spectra were recorded sequentially to follow the evolution of the samples and confirm the absence peptide degradation in the individual spectra, then summed up to obtain spectra with higher signal to noise ratio.

Assignment of the NMR spectrum

The resonances transfer of Amylin from the BMRB was uncertain due to the differences in buffer composition and working pH. A new assignment was achieved in **PhCPN buffer** (25 mM MES pH 6.5, 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$) and 25 mM MES pH 6.5 buffer (abbreviated **MES Buffer**), both from 10°C to 30°C by temperature progressive

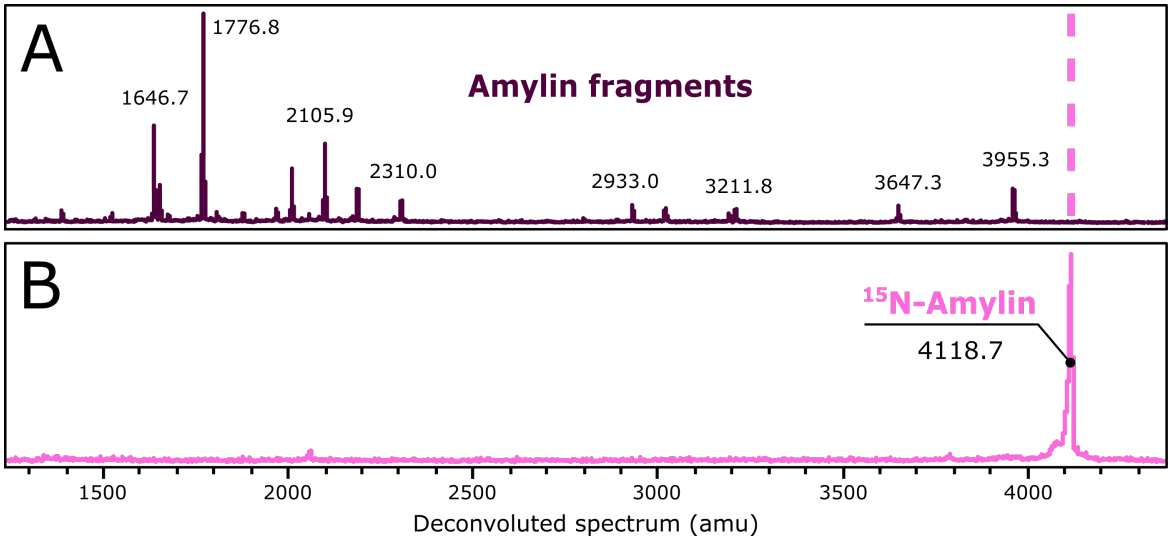


Fig. 5.4 : Elimination of Amylin degradation in presence of PhCPN. (**A**) MALDI-TOF MS spectrum of degraded 50 µM ^{15}N-Amylin, kept at 4°C overnight after mixing with 2.5 µM PhCPN. The original peptide was so degraded that it was not even present in the sample anymore. (**B**) MALDI TOF MS spectrum of ^{15}N-Amylin after modification of the experimental setup. The degradation was completely prevented and only the monoisotopic peak of intact ^{15}N-Amylin was present.

increase. Different 2D and 3D NMR experiments were recorded at a total (^{13}C,^{15}N)-Amylin concentration of 200 µM for assignment : 2D HMQC SOFAST, 2D BEST-TROSY, 3D BEST-HNCA, 3D BEST-HNCO, 3D BEST-HNCACAB, 3D BEST-HNCOCA and 3D BEST-HNCACO [Lescop et al., 2007, Schanda et al., 2005, 2006].

The backbone resonances were assigned without ambiguities at up to 94.6% at 10°C PhCPN with the CcpNmr Analysis software [Vranken et al., 2005], the AA n°1 and 2 being invisible on the spectra (**figure 5.5**). At 30°C, the peaks tend to disappear, especially in MES Buffer, due to increased amide hydrogen exchange with the protons from the solvent, especially since Amylin is an IDP and most of its residues are highly solvent-accessible.

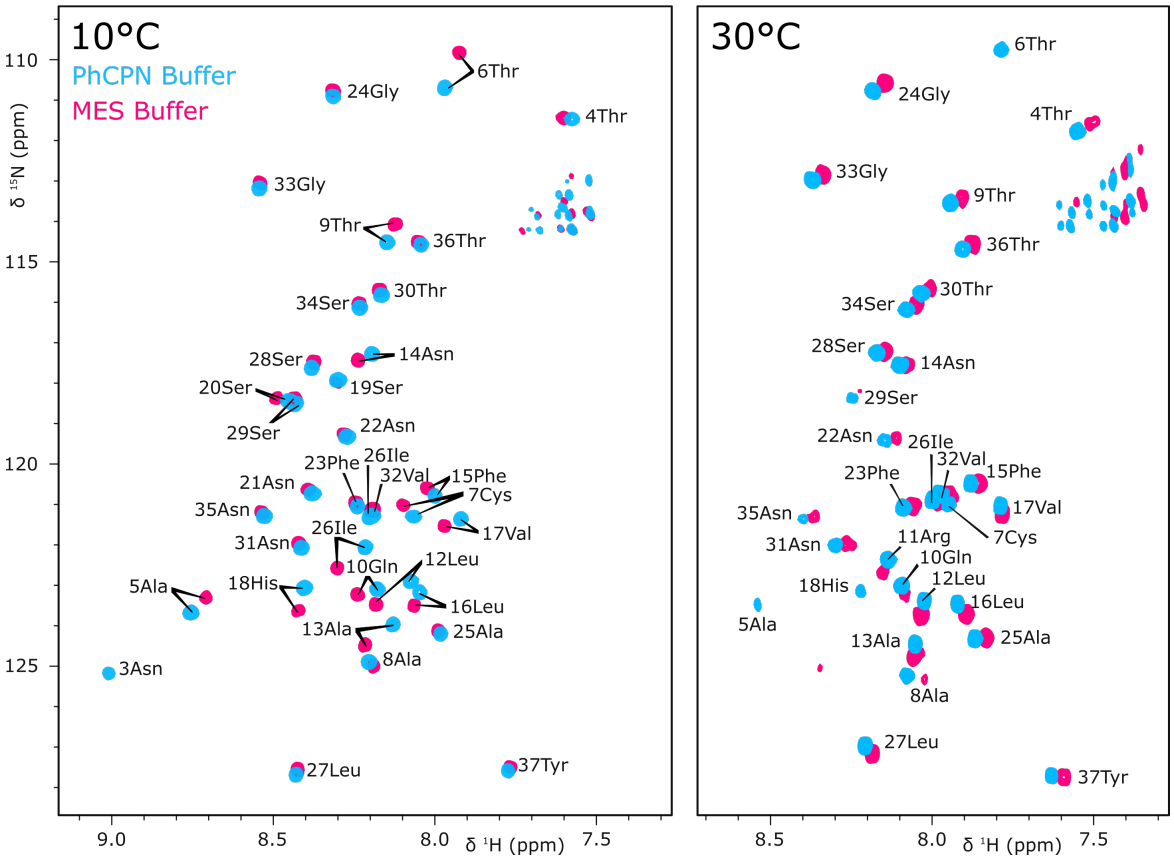


Fig. 5.5 : BEST-TROSY ^{1}H-^{15}N NMR spectra of 50 µM Amylin in both PhCPN and MES Buffer, at 10 and 30°C.

5.2.2 Interaction of Amylin with PhCPN

Once the experimental conditions were optimized to get the best quality data possible, the interaction between Amylin and PhCPN was investigated in detail by NMR. The interaction was monitored in the same two buffers, PhCPN buffer and MES buffer, at 10°C, in presence of different amounts of PhCPN.

Hexadecameric PhCPN

NMR BEST TROSY ^{1}H-^{15}N spectra of 50 µM ^{15}N-Amylin were recorded at 10°C (**figure 5.6, A**). The intensity ratio of the Amylin signal with or without PhCPN was plotted for each amino acid residue (**figure 5.6, B**). As presented in the assignment section, the first two residues are not visible on the spectra. The titration spectra were also recorded at 30°C but most of the peaks corresponding to the middle of the protein sequence disappeared, due to the fast exchange of the amide proton with solvent, giving incomplete interaction data.

Most of the Amylin signal decreases drastically upon PhCPN addition, especially in the Nter region and at the region that contains most of the hydrophobic residues of the protein, while the Cter remains almost unaffected (**figure 5.6, B**). The most hydrophobic regions of Amylin are also affected by the presence of PhCPN, but the intensity differences are smaller (**figure 5.6, B & C, residues 23 and 27-28**). Interestingly, all the positively charged residues of Amylin are also located in the interacting region, namely K1 and R11 (H18 is not charged at pH 6.5 considering its pKa of 6.0). No chemical shift perturbations were observed on the spectra, which can be explained by the large size of PhCPN. In fact, PhCPN causes the disappearance of the signal corresponding to the interacting region, as it tumbles slower (approximately 500 times slower than Amylin, with expected τ_C near 1 µs for PhCPN compared to approximately 2 ns for free Amylin, see **Section 5.2.2**), and thus signal relaxes faster (**figure 5.6, D**). The Cter region of PhCPN-bound Amylin behaves similarly to free Amylin, thus explaining its high signal intensity. By comparing the exchange dynamics between Aβ40 and GroEL (fast exchange, [Libich et al., 2013] and the overall relative hydrophobicity of both Aβ40 and Amylin, the free and bound form of the Amylin are very likely in fast exchange.

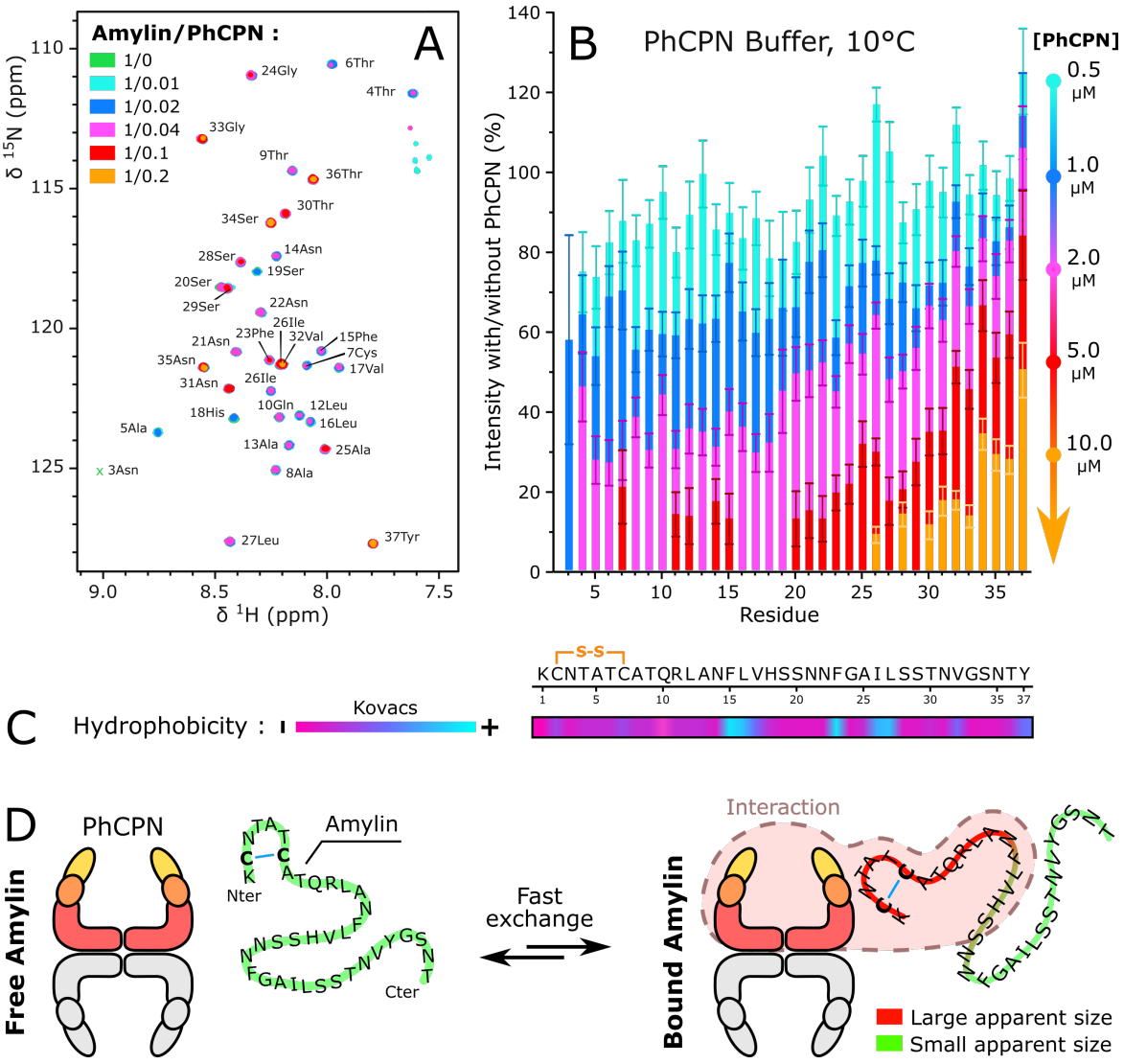


Fig. 5.6 : NMR titration of Amylin in presence of different amounts of hexadecameric PhCPN at 10°C, in PhCPN buffer pH 6.5. (**A**) Overlayed BEST-TROSY ^{1}H-^{15}N NMR spectra of 50 µM Amylin in PhCPN buffer, in the absence or presence of different amounts of PhCPN. (**B**) Plot of the intensity of Amylin signal with or without PhCPN, for different PhCPN concentrations. (**C**) Hydrophobicity plot of the Amylin (see **Chap. 6.7**). (**D**) Scheme of the Amylin/PhCPN interaction.

Monomeric PhCPN

In the same way as before, NMR BEST TROSY ^{1}H-^{15}N spectra of 50 µM ^{15}N-labelled Amylin were recorded at 10°C (**figure 5.7, A**) and the intensity ratios were plotted for each amino acid residue (**figure 5.7, B**), this time in MES Buffer. The number of points in the indirect dimension was decreased for all the spectra, including the reference, in order to add scans as the signal was relaxing faster compared to previous experiments and to keep an overall experimental time of the same length as previously. The first three residues were invisible on the recorded spectra.

The interaction was stronger as proven by the large decrease of peak intensity in presence of 10 times smaller amounts of PhCPN (all the PhCPN concentrations are expressed as 16-mer concentrations for clarity). This large interaction difference could be caused by the lack of salt and increased electrostatic interactions between Amylin and PhCPN. Since PhCPN is monomeric in MES buffer, the increased interaction could also be explained by additional interaction surfaces usually shielded by the presence of the other neighbouring monomers in the intact PhCPN 16-mer. The interaction pattern is almost identical as in

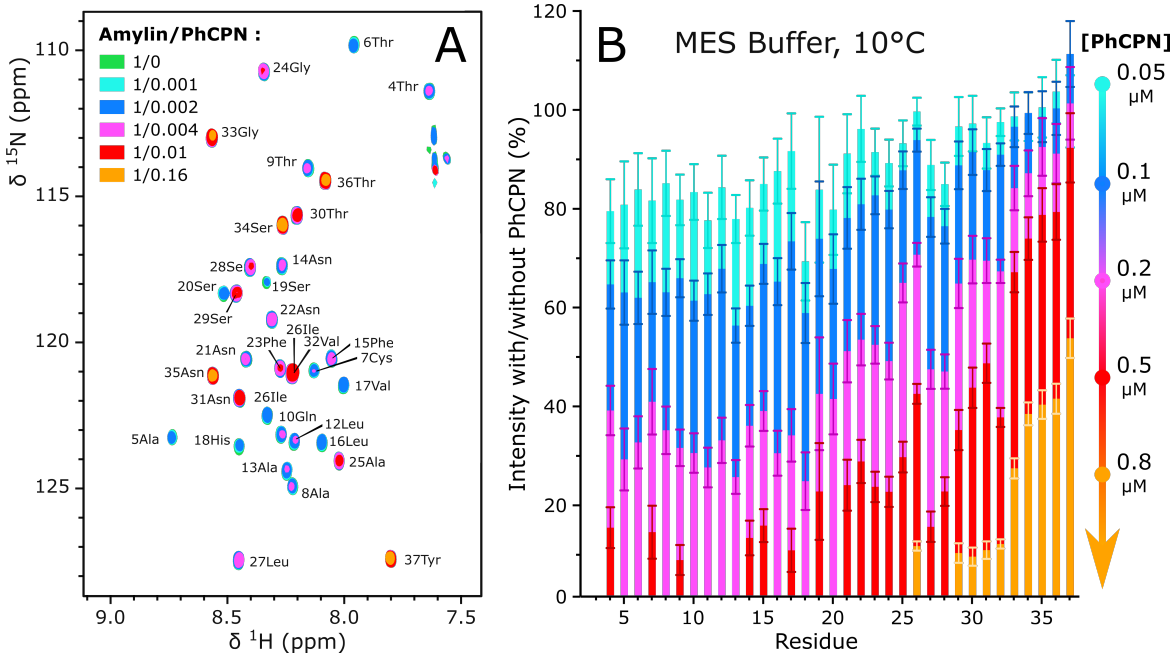


Fig. 5.7 : NMR titration of Amylin in presence of different amounts of monomeric PhCPN at 10°C, in MES buffer. (**A**) Overlayed BEST-TROSY ^{1}1H-^{15}N NMR spectra of 50 µM Amylin in MES buffer, in the absence or in presence of different amounts of PhCPN. (B) Plot of the intensity of Amylin signal with or without PhCPN, for different PhCPN concentrations.

the PhCPN buffer, with a clearly extinct Nter region while the Cter remains almost unaffected. Interestingly, peaks shifts are still not discernable while the size of the monomers is this time compatible with the observation of chemical shifts if present, which could indicate different exchange mechanism (see **Chap. 6.3.1**).

Amylin-CC

A synthetic Amylin was also produced by collaborators and its interaction with PhCPN was monitored. This new construct, Amylin-CC, is an engineered Amylin that is supposed to be unable to fibrillate due to additional disulfide bridge. This bridge between residues C13 and C25 is introduced into Amylin-CC through point mutations A13C and A25C, thus located in the region responsible for fibrillation (residues 20 to 29). Since Ala and Cys have a similar hydrophobicity, these mutations do not impact the hydrophobicity pattern of the whole peptide. The main interest of this experiment is to investigate whether or not the compactness of the peptide affects its interaction pattern with PhCPN and the K_D of the interaction. In fact, the structures compactnesses of Amylin and Amylin-CC are drastically different (**figure 5.8, B**).

The ^{1}H-^{15}N BEST-TROSY NMR spectrum of Amylin-CC in PhCPN buffer is very different from the one of Amylin, as most of the residues shift due to conformational changes induced by the presence of the new disulfide bridge (**figure 5.8, A**). The spectrum assignment was performed by collaborators, who also confirmed by HPLC and NMR that only one disulfide topology is present, namely the one with the natural C2-C7 bridge and the additional C13-C25 disulfide bridge. The interaction patterns of Amylin and Amylin-CC when both proteins are mixed separately with PhCPN are very similar. The slight intensity differences come from a difference of the PhCPN quantity introduced in the samples. The loss of signal in the Amylin-CC sample is of 67.7% while it is of 76.6% in the Amylin one, a difference of 8.9%. When both proteins are mixed together without the addition of PhCPN, the spectrum of the mix corresponds to the sum of the spectra of each isolated protein and overlap perfectly (**figure 5.8, A, "invisible" blue spectrum in the image background**). Especially, the peaks corresponding to the Cter of Amylin and Amylin-CC cannot be differentiated. No additional peaks are visible, thus showing that there is no interaction between Amylin and Amylin-CC. When PhCPN is introduced in the sample, the interaction pattern is once again very similar, with strong Nter interaction

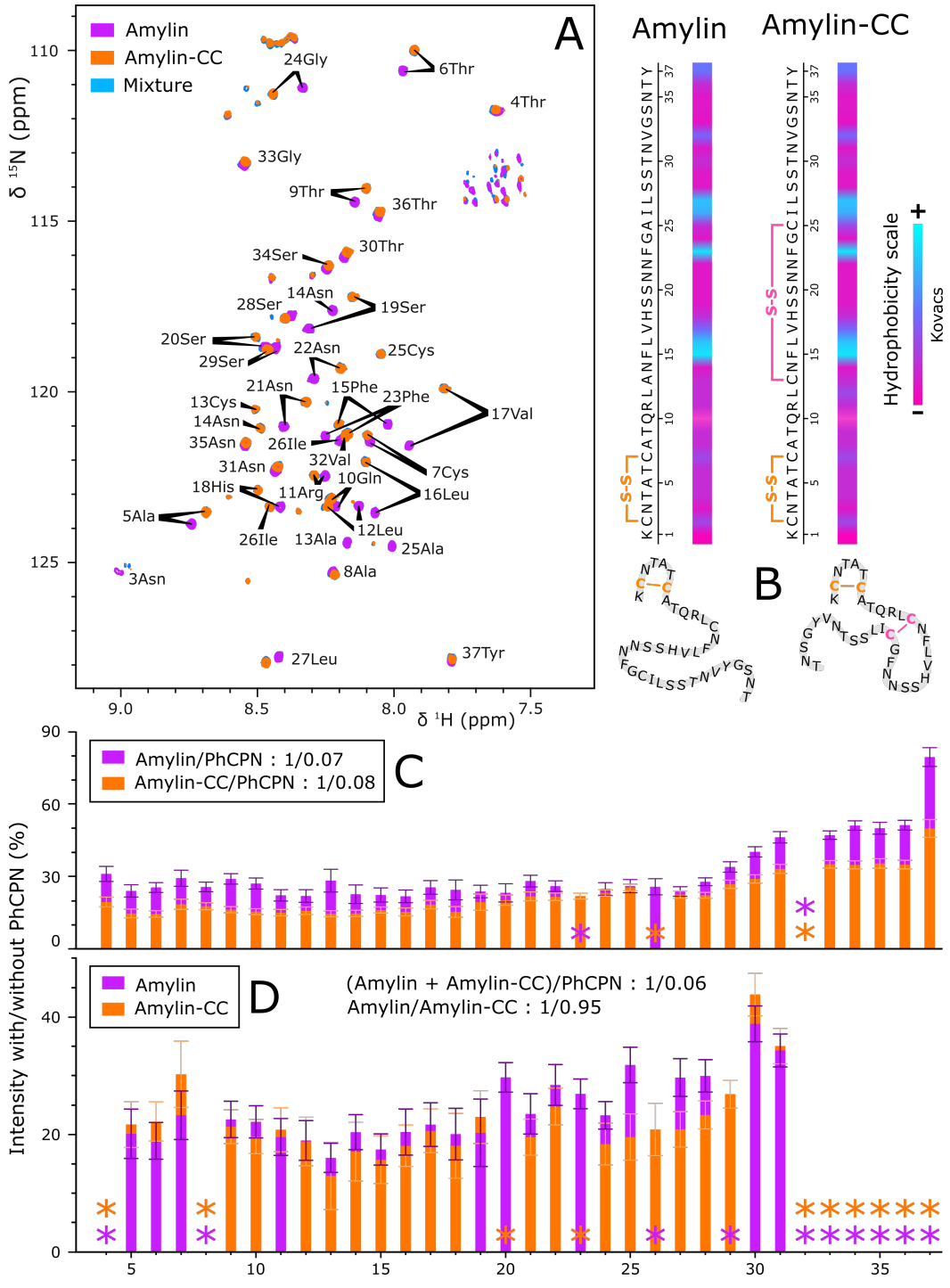


Fig. 5.8 : Interaction tests with Amylin-CC. (**A**) Comparison of 15 µM Amylin and Amylin-CC NMR ^{1}H-^{15}N BEST-TROSY spectra. (**B**) Sequence and hydrophobicity plot of both Amylin and Amylin-CC. The detailed hydrophobicity scale is presented in **Suppl. 6.7**. (**C**) Comparison of the signal loss induced by addition of PhCPN in separated Amylin and Amylin-CC samples. (**D**) Comparison of the signal loss induced by addition of PhCPN in a mixed Amylin and Amylin-CC sample. Overlapping peaks are indicated with asterisks.

and an approximately identical K_D (**figure 5.8, C & D**). The intensity loss is of 78.0% and 80.4% for Amylin and Amylin-CC respectively. It is possible to conclude that both peptides interact in a similar way with PhCPN, no matter their conformation. There is also no "preference" for one peptide compared to the other when they are mixed together in presence of PhCPN, even if their overall compactness and structural features are different.

K_D estimation

Determination of K_D is very challenging due to the size of PhCPN and the supposedly weak interaction between Amylin and PhCPN. A first technique was used to study the binding affinity of both proteins was isothermal titration calorimetry (**ITC**). Reference ITC cell was filled with 400 µL PhCPN buffer at 10°C. A titration of 400 µL of 10 µM Amylin in the ITC chamber, by 12 µM PhCPN in a 80 µL chamber at 10°C was performed but no significant thermodynamic changes were observed. The opposite titration was also tested but the result was similar, which prevented the measure of the affinity by ITC.

In the case of liquid-state NMR, when Amylin and PhCPN interact, the signal of the region of Amylin in interaction is completely extinct because of PhCPN's slow tumbling and fast relaxation (see **Chap. 6.3.2**). A possibility to determine a K_D in these conditions would be to perdeuterate and methyl label Amylin in a similar way as in the case of HEWL [Mas et al., 2018] and MSG (see **Chap. 4.2.1** and **Chap. 6.2**). Unfortunately, the fibrillating nature of Amylin makes perdeuteration difficult as it perturbs and modifies the stability of the Amylin construct. In addition, the yield of Amylin production is already very low in protonated conditions and would dramatically decrease upon perdeuteration. Nevertheless, a coarse grain estimation of the K_D from the titration data was carried out, assuming fast exchange between free and PhCPN-bound Amylin (see **Chap. 6.3.1**). In fact, the intensity loss of the Amylin signal (I/I_o) is proportional to the quantity of Amylin interacting with PhCPN, the stoichiometry of the complex and the rotational correlation times (τ_C) of both PhCPN and Amylin (**figure 5.9, formulae (b) & (c)**).

The rotational correlation time of Amylin was estimated from [Danielsson et al., 2006], which measured the rotational correlation time of Aβ40 at different temperatures. Since Aβ40 is similar in size with Amylin and equally intrinsically disordered, τ_A was estimated by extrapolating the rotational correlation time at 10°C and remove the volume corresponding to the 3 extra amino acids in Aβ40 (Amylin is 37-AA long). The estimated

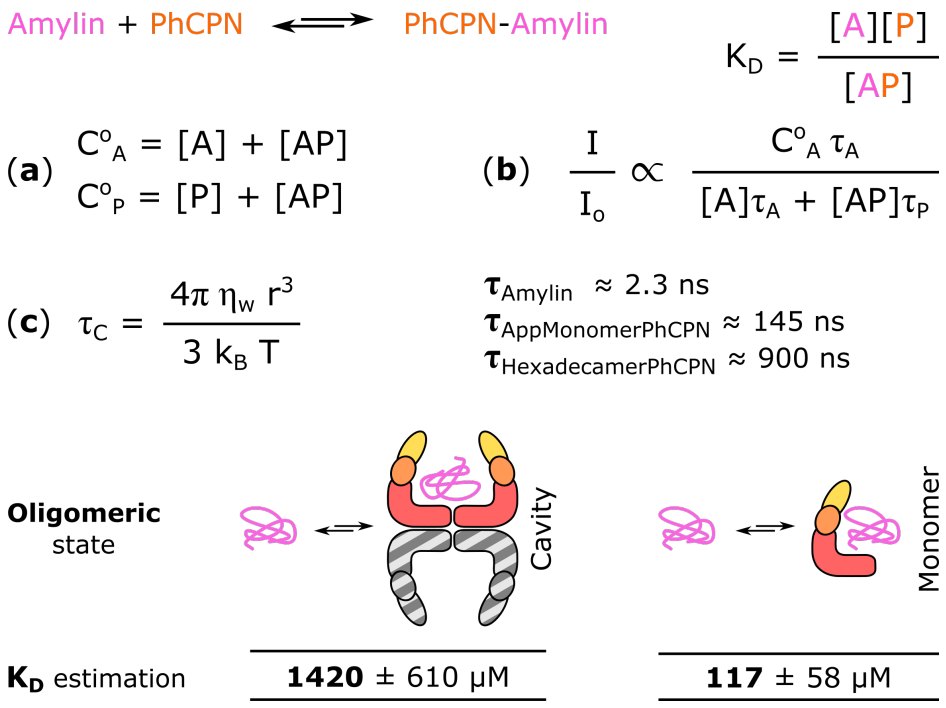


Fig. 5.9 : Formulas to calculate the K_D of the Amylin/PhCPN interaction. [A] is the free Amylin concentration ; [P] is the free PhCPN concentration ; [AP] is the complex concentration ; C_oA is the initial Amylin concentration ; CoP is the initial PhCPN concentration ; η_W is the solvent viscosity ; r the hydrodynamic radius of the particle and k_B the Boltzmann constant.

rotational correlation time of Amylin is 2.3 ns. The rotational correlation time of GroEL is 310 ns at 35°C [Riek et al., 2002]. Extrapolation with the volume of hexadecameric PhCPN compared to the volume of GroEL as well as a temperature and viscosity correction using the formula from (**figure 5.9, C**) [Cavanagh et al., 2007] gave access to τ_P, approximately equal to 900 ns for hexadecameric PhCPN and 35 ns for monomeric PhCPN at 10°C. In MES buffer, most of PhCPN is monomeric but residual hexadecamers are still present, representing approximately 10% of the total PhCPN. τ_C thus needs to be corrected to an apparent τ_C corresponding of 90% monomeric PhCPN and 10% hexadecameric PhCPN, of 145 ns, assuming fast exchange between the monomer and the hexadecameric forms of PhCPN.

To calculate K_D, several approximations were used. First, the τ_C change of PhCPN upon binding of Amylin was neglected, since Amylin size is negligible compared to the one of hexadecameric and monomeric PhCPN. Secondly, hexadecameric PhCPN concentration was expressed in half particles (or cavities) to simplify the equations, since it is very likely that there is no cooperative effect between both open apo cavities when binding a peptide, in absence of ATP/ADP. The K_D are evaluated with both monomeric or hexadecameric PhCPN for different I/I_O ratios from **Section 5.2.2** and **Section 5.2.2**.

The calculated K_D values are reported in (**figure 5.9, A**) and are of 1.42 mM and 117 µM, for one PhCPN cavity and monomeric PhCPN respectively. This result is consistent with the intensity changes observed upon addition of PhCPN, since the intensity loss is 10 times higher in presence of monomeric PhCPN compared to hexadecameric PhCPN. The stronger interaction in presence of monomeric PhCPN in MES buffer can be explained by the additional presence of electrostatic interactions, shielded by the salts in the case of hexadecameric PhCPN, in PhCPN buffer. In addition, the interaction surface is different when PhCPN is monomeric, possibly exposing additional interaction sites with increased affinity for Amylin.

These values can be compared to the K_D of HEWL and PhCPN, of 1.6 µM for one cavity [Mas et al., 2018]. These data are coherent since HEWL is larger than Amylin and exposes more hydrophobic patches when unfolded, thus increasing its interaction strength when binding PhCPN. Despite the size of PhCPN cavity, it is very unlikely that two Amylin are hosted simultaneously because of the weak affinity between the two species and the concentration used for the tests (Amylin is free in solution 99.7% of the time in presence of PhCPN, according to estimations). Nevertheless, these values remain very hypothetical

and would need to be measured more precisely with the previously cited techniques such as perdeuteration and methyl labelling.

5.2.3 Aggregation kinetics

Several techniques can be used to investigate the aggregation kinetics of amyloid proteins. Aggregation kinetics can be monitored by liquid state NMR, by recording a series of successive NMR spectra and observing the signal intensity loss as the aggregates form, since their corresponding liquid-state NMR signal is broadened, due to slower tumbling of the oligomers or fibrils. In this case, a succession of BEST-TROSY ^{1}H-^{15}N spectra was recorded and the average intensity of 11 fixed peaks was plotted as a function of time (**figure 5.10**). The peaks disappeared as the free Amylin monomers were aggregating, as expected.

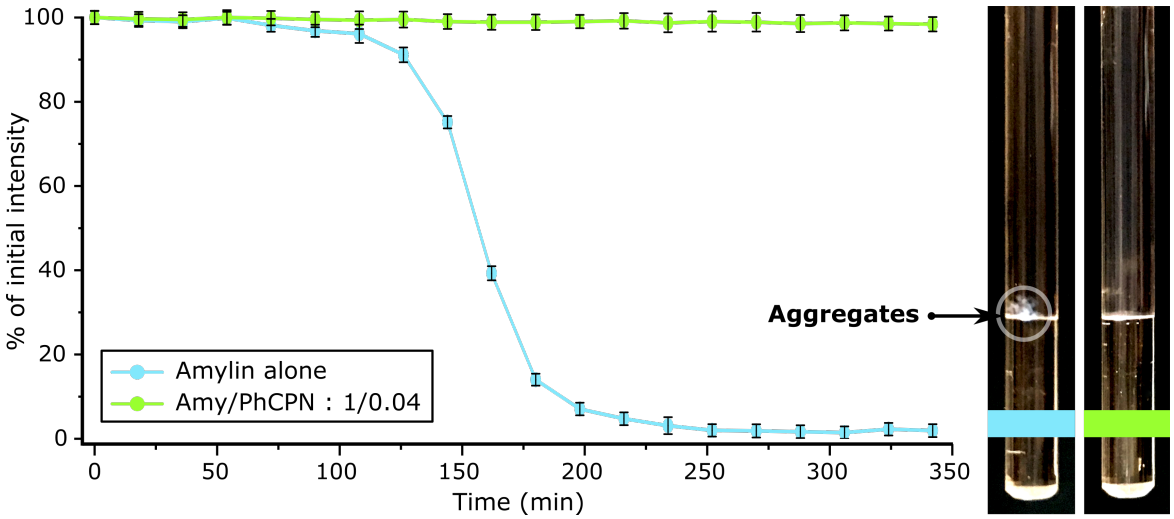


Fig. 5.10 : Amylin aggregation monitored by NMR. Average intensity of 50 µM ^{15}N-Amylin assigned peaks overtime, in presence or absence of 2 µM PhCPN, in PhCPN Buffer, and the corresponding pictures of the 3 mm Shigemi NMR tubes after the series of experiments.

In the pure Amylin sample, it was clearly possible to observe the disappearance of the signal of monomeric Amylin, presumably due to the formation of large oligomers or aggregates since no degradation could be spotted on the individual NMR spectra. The Amylin monomers disappearance rate was approximately 0.85 µmol/min and the lag phase of 125 min. In the presence of PhCPN however, the signal remained approximately constant

during the time of the experiment, showing an important aggregation/fibrillation inhibition. Nevertheless, this experiment was not perfectly reproducible. In fact, aggregation occurred randomly, probably because of the lack of shaking of the sample in the spectrometer, preventing the reproducible formation of seeds. The experiment was repeated several times and fibrillation was not always observed, even after 12 hours of recording. This technique lacking reproducibility and consuming large amounts of the protein sample was abandoned and replaced by a widespread technique to observe amyloid aggregation, monitoring of the Thioflavin T (**ThT**) dye fluorescence. In fact, the ThT dye possesses a high affinity for β-sheets, found in amyloid fibrillar structures [Biancalana and Koide, 2010]. In the following section, the variation of ThT fluorescence intensity in presence of Amylin was studied as a function of time. The ThT fluorescence experiment and its principle are precisely described in **Chap. 6.5.1**.

Optimization of the experimental conditions

Many factors can influence the aggregation assays, especially in the case of amyloidogenic proteins as the kinetics needs to start by the formation of an aggregation nucleus, a rather stochastic event (reviewed in [Giehm and Otzen, 2010]). The impact of several factors on aggregation needed to be investigated to find the best possible conditions to obtain reproducible aggregation data. These important parameters include :

- Temperature (aggregation rate generally increases with temperature)
- Protein stock (different batches contain different initial amounts of residual seeds)
- Concentration (aggregation increases with amylogenic monomer concentration)
- Plate shaking (the 96-well plate needs to be shaken prior to fluorescence measurement, in order to homogenize the samples)
- Plate type (different types of plates are available for these types of experiments, with special a coating of the wells to prevent binding of the aggregates)
- Sample preparation conditions (samples are kept freeze-dried in aliquots and need to be resuspended prior to aggregation assays), presence of glass beads to accelerate fibrillation, preparation of the Amylin prior to introduction on the 96-well plate.

In the following experiments, the fluorescence intensity of the aggregates was measured with two gain values in case of unexpected signal intensity increase. All experiments were recorded in identical experimental conditions except the variating parameter. In the case of different shaking duration and plate type, experiments were recorded separately.

The first tests were carried out in simple conditions : continuous shaking, 17 µM Amylin, 30°C, medium binding 96-well plate and scans every 5 minutes. Four different Amylin aliquots were used and dissolved separately, and their fibrillation was followed by ThT fluorescence (**figure 5.11, A**). The reproducibility in these conditions was low, probably due to sample heterogeneity caused the use by different Amylin stocks. Especially, lag phases were very heterogeneous, possibly because of different initial amounts of residual amyloid seeds in the different aliquots. For the following experiments, the aliquots were all mixed together prior to the experiments to homogenize the samples and successively avoid such bias (**figure 5.11, B, C & D**). The aggregation increases rapidly and progressively decreases again for three samples out of four, which could be due to a first fast aggregation followed by a rearrangement of the monomers into a more organized structure. This particularity was observed in every other aggregation experiments with Amylin alone and was reported in the literature [Colby et al., 2007]. The orbital shaking duration was also proven to be very important (**figure 5.11, B**). The conditions were the following : various shaking periods between fluorescence measurements, 8.5 µM Amylin, 30°C, medium binding 96-well plate and scans every 5 minutes. The shaking increased the aggregation rate, which can be explained by the formation of additional fibrils ends through fibrils breakage [DePace et al., 1998]. It was also conjectured that the rotation can dissociate large complexes, enabling fibrils ends to elongate again [Serio et al., 2000] ; and that shaking increases the air-water interface, stabilizing partially folded amyloidogenic oligomers [Sluzky et al., 1991]. The shaking, however, needed to be kept at a minimum to avoid breakage of the aggregates or the fibrils, but still homogenize the samples before fluorescence measurements. Some groups even advise not to shake at all, because it could change the aggregation reaction mechanisms [Meisl et al., 2016]. In this case, even if the aggregation speed and lag phase change, the overall shape of the curve was similar, thus indicating a similar aggregation mechanism. Minimum shaking of 3 seconds was therefore conserved for all the following experiments.

The salt content of the buffer is also of paramount importance and affects greatly the fibrillation (**figure 5.11, C**). Salt effect reported in the literature for Aβ42 [Meisl et al., 2017]

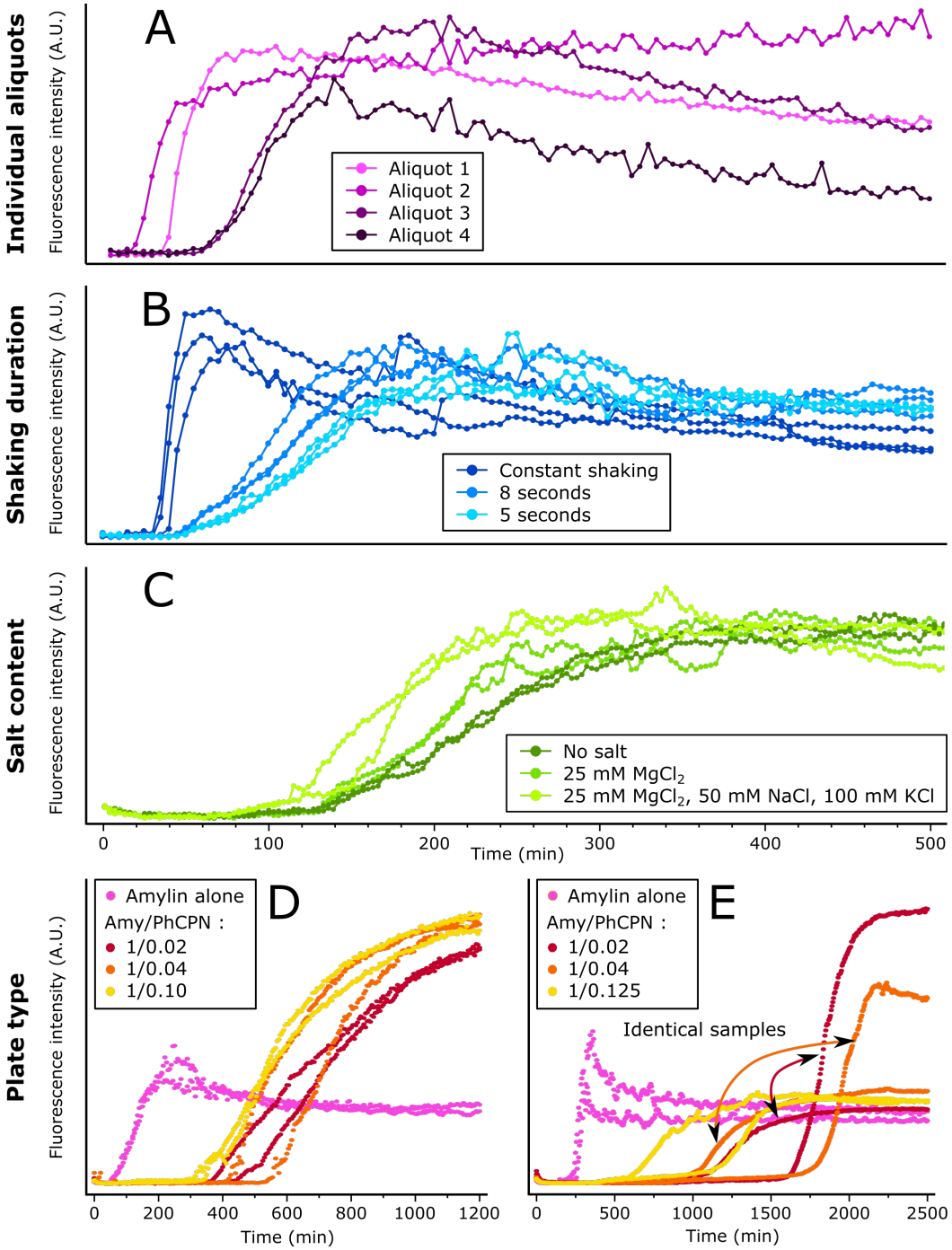


Fig. 5.11 : Optimization of the fibrillation experiments. (**A**) Amylin samples from the same purification batch but kept in separate aliquots ; 17 µM Amylin, 30°C, medium binding plate, constant shaking. (**B**) Influence of the plate shaking duration on the fibrillation speed. 8.5 µM Amylin, 30°C, medium binding plate. (**C**) Influence of the total amount of salts in the buffer. 6.5 µM Amylin, 30°C, 3 seconds shaking, medium binding plate. (**D**) Fibrillation on a regular medium binding 96-well plate. 8.5 µM Amylin, 30°C, 5 seconds shaking. (**E**) Fibrillation on a low binding 96-well plate. 8 µM Amylin, 30°C, 5 seconds shaking.

could be reproduced in the following conditions : 3 seconds shaking before fluorescence measurements, 6.5 µM amylin, 30°C, medium binding 96-well plate and scans every 8 minutes. The 0 mM salt condition corresponds to MES buffer pH 6.5 ; the 25 mM salt condition to MES buffer pH 6.5 with 25 mM $MgCl_2$; and finally the 175 mM salt condition to PhCPN buffer. The presence of salts increased aggregation speed, as previously reported [Meisl et al., 2017]. This result is particularly important to keep in mind for the analysis of the data acquired in different buffers in the next paragraph.

Different types of 96-well plates are available to carry out aggregation assays. The surface coating of the wells can differ, from low binding, medium binding to high binding surface properties, depending on needs. In order to get more reproducible data, a different well coating was tested, in the following conditions : 5 seconds shaking before fluorescence measurements, 8.5 µM Amylin, 30°C, medium and low binding 96-well plate and scans every 5 minutes. Interestingly, low binding plates (**figure 5.11, E**) yield less reproducible data than medium binding plates (**figure 5.11, D**). In fact, the lag phase and the fluorescence are dramatically heterogeneous for the two last PhCPN concentrations (**figure 5.11, E, orange and red curves**) compared to fluorescence obtained from medium binding plates (**figure 5.11, D**). One could notice a slight variation of the signal at the beginning of the experiments (**figure 5.11, C, D & E**). In fact, the plates were prepared on ice and their temperatures needed to equilibrate after introduction in the fluorometer, and the fluorescence efficiency of ThT varies with temperature. This variation is however not present in (**figure 5.11, A & B**) because of a different and less efficient cooling method used during plate preparation. The addition of glass beads to increase the reproducibility of the fibrillation experiments was also tested [Giehm and Otzen, 2010]. An increase in aggregation rate was observed but came with a great addition of noise caused by the presence of the bead during fluorescence measurements (**data not shown**). The glass bead solution was logically abandoned.

To summarize, plate shaking largely influences the fibrillation speed and needed to be kept at a minimum to homogenize the samples without breaking the fibrils. Aggregation/fibrillation speed also increases with the salt content of the buffer and is a parameter to consider, even though its influence is less important than the one of shaking. The different conditions tested also showed that reproducibility could be increased by working with amyloidogenic protein samples homogenized prior to the aggregation/fibrillation experiments instead of working with single vials. Oppositely, glass beads and low-binding

96-well plates do not increase reproducibility and their use should be avoided. The adapted and corrected protocol is described in detail in **Chap. 6.5.1**.

Curves processing

Once reproducible aggregation curves could be obtained in acceptable conditions, the data needed to be fitted and processed to extract various information, such as fibrillation rate and lag time prior to fibrillation (**figure 5.12, A**). In literature, several fitting methods can be found to obtain these information and are subject to controversy [Gade Malmos et al., 2017, Meisl et al., 2016]. The main debate is centered around whether the fibrillation data should be normalized prior to analysis or not, especially in the case of a dosage by an interacting partner. In fact, the modification of the experimental conditions (reviewed in [Fändrich et al., 2018, Ow and Dunstan, 2014]), including the addition of an interacting partner, can affect the morphology of the fibrils and modify the affinity of ThT for the fibrils, thus changing the fluorescence intensity. In the case of Amylin and PhCPN for instance, the fluorescence intensities are highly heterogeneous with a dramatic increase of the final fluorescence intensity in presence of small amounts of PhCPN (**figure 5.12, B ; pink and dark blue curves**), thus the need for additional processing and normalization. This difference can be explained by a different binding mode of ThT to the fibrils, possibly caused by the presence of different fibrils morphology [Lindberg et al., 2015]. On the other hand, when the fluorescence is normalized, the data corresponding to samples with a low Amylin/PhCPN ratio are very noisy and aberrant when compared to the other samples, and do not plateau within the experimental time (**figure 5.12, D**). Another correction is also possible, consisting in scaling the Amylin intensity to the maximum fluorescence intensity of Amylin in presence of PhCPN (**figure 5.12, C**). This solution allows to still monitor the differences of intensities in presence of PhCPN which could be due to incomplete fibrillation while removing the bias caused by the large intensity differences in presence or absence of PhCPN. This correction was chosen to process all the following fibrillation data.

Once the curves were corrected according to the model used in (**figure 5.12, C**), the fibrillation rates and lag phase of Amylin in presence of PhCPN were extracted from the fluorescence curves. The scaling factor of the pure Amylin sample is indicated in the caption of each figure.

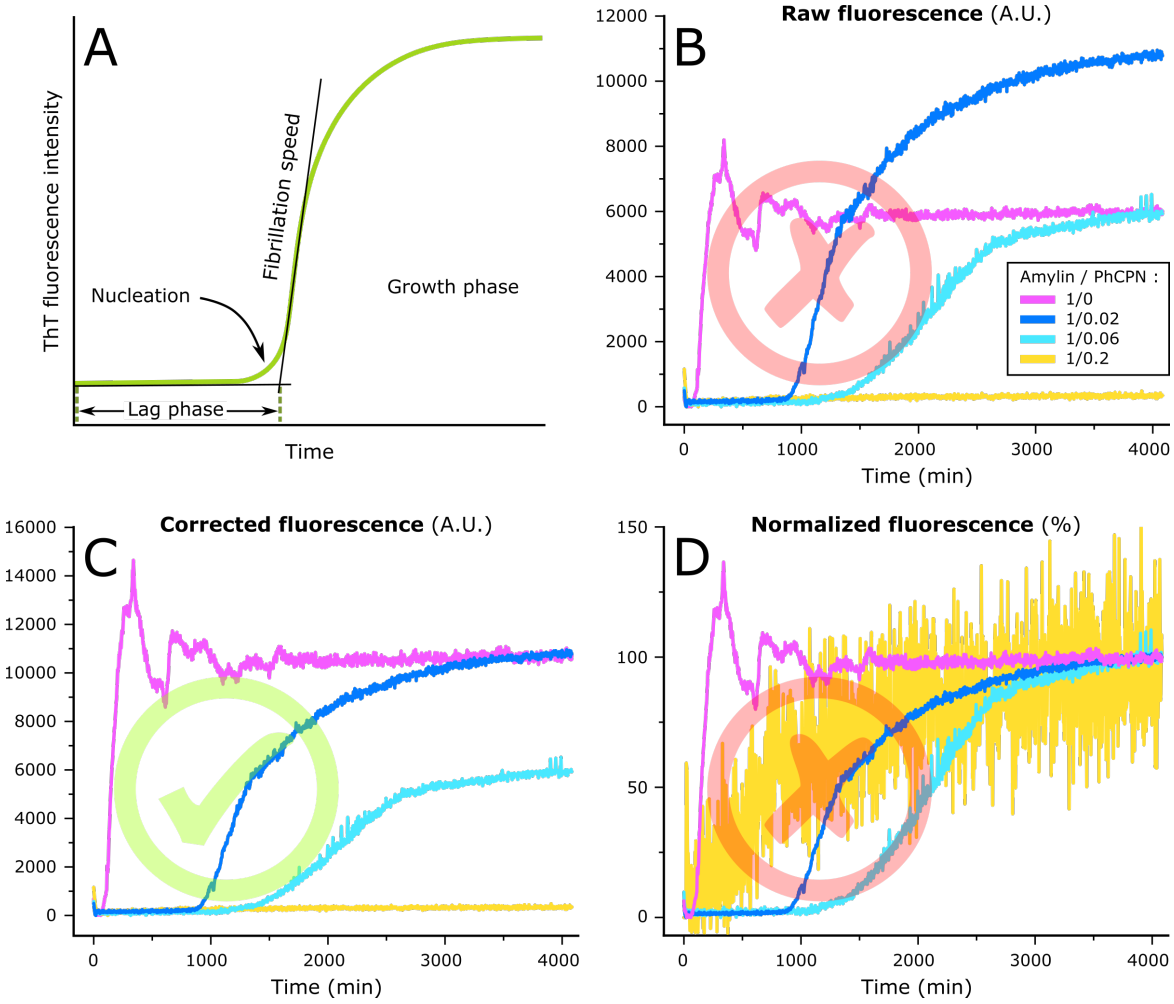


Fig. 5.12 : Fibrillation curves processing. (**A**) Scheme of a fibrillation curve and the different information which can be extracted from it. (**B**) Raw fluorescence data of Amylin in the absence or presence of different amounts of PhCPN. (**C**) Partially corrected fluorescence data, with Amylin equilibrium intensity scaled to 100% of the final maximum intensity in presence of PhCPN. (**D**) Corrected fluorescence with all curves normalized. Fibrillation data in (**B**), (**C**) & (**D**) were obtained with samples of 6.5 µM Amylin, heated at 30°C and agitated for 5 seconds prior to fluorescence measurement. The background fluorescence of PhCPN was subtracted from the curves.

PhCPN effect on the fibrillation

Once the experimental conditions were optimized, the influence of PhCPN on Amylin fibrillation could be monitored. Two buffers were used to observe the influence of the oligomeric state of PHCPN as well. 6.5 µM Amylin were introduced on a 96-well plate with 3 seconds shaking to limit fibrils breakage, in presence of diverse amounts of PhCPN in MES (monomeric PhCPN) and PhCPN buffer (hexadecameric PhCPN). Fibrillation speeds are expressed in A.U./min since the quantity of final fibrils cannot be thoroughly quantified by this technique.

The aggregation speed of Amylin alone in PhCPN buffer is faster (87.4 A.U./min) compared to MES buffer (74.2 A.U./min) as previously observed, since PhCPN contains more salts (175 mM total) compared to MES buffer (0 mM total), as aggregation increases with salt concentration (**figure 5.13, A, B, E & F**) [Meisl et al., 2017]. In the case of hexadecameric PhCPN in PhCPN buffer (**figure 5.13, A**), the effect of PhCPN on fibrillation is clear. The final fluorescence intensity decreases with increasing concentrations of PhCPN, which indicates that a fraction of the total Amylin does not aggregate or fibrillate. The shape of the aggregation or fibrillation curves is also dependant on the concentration of PhCPN. In fact, it seems that the reaction mechanism is different for lower PhCPN concentrations compared to high PhCPN concentration. This can be observed at the initial aggregation points. In the case of low PhCPN concentrations, the fibrillation starts spontaneously and increases very quickly while in presence of higher amounts of PhCPN, the fibrillation increases more progressively (**figure 5.13, A, violet curve versus light blue curve**). Such differences can be explained by different nucleation processes [Meisl et al., 2016]. Nevertheless, the fibrillation speed is much slower in presence of PhCPN and decreases dramatically as the concentration of PhCPN is increased (**figure 5.13, E**). The lag phase increases dramatically with PhCPN concentration (**figure 5.13, G**). Even low amounts of PhCPN are enough to obtain a large delay of the aggregation process (870 minutes delay at 0.065 µM PhCPN). However, this effect is less pronounced in presence of monomeric PhCPN in MES buffer (207 minutes delay at 0.065 µM PhCPN ; (**figure 5.13, H**)). The final fluorescence intensity is not much affected by the presence of PhCPN and it seems that all the Amylin aggregates in both cases (**figure 5.13, B**). The fibrillation rate is nevertheless decreased by the presence of monomeric PhCPN, in a smaller proportion than with hexadecameric PhCPN (**figure 5.13, F**). The effect of monomeric PhCPN on

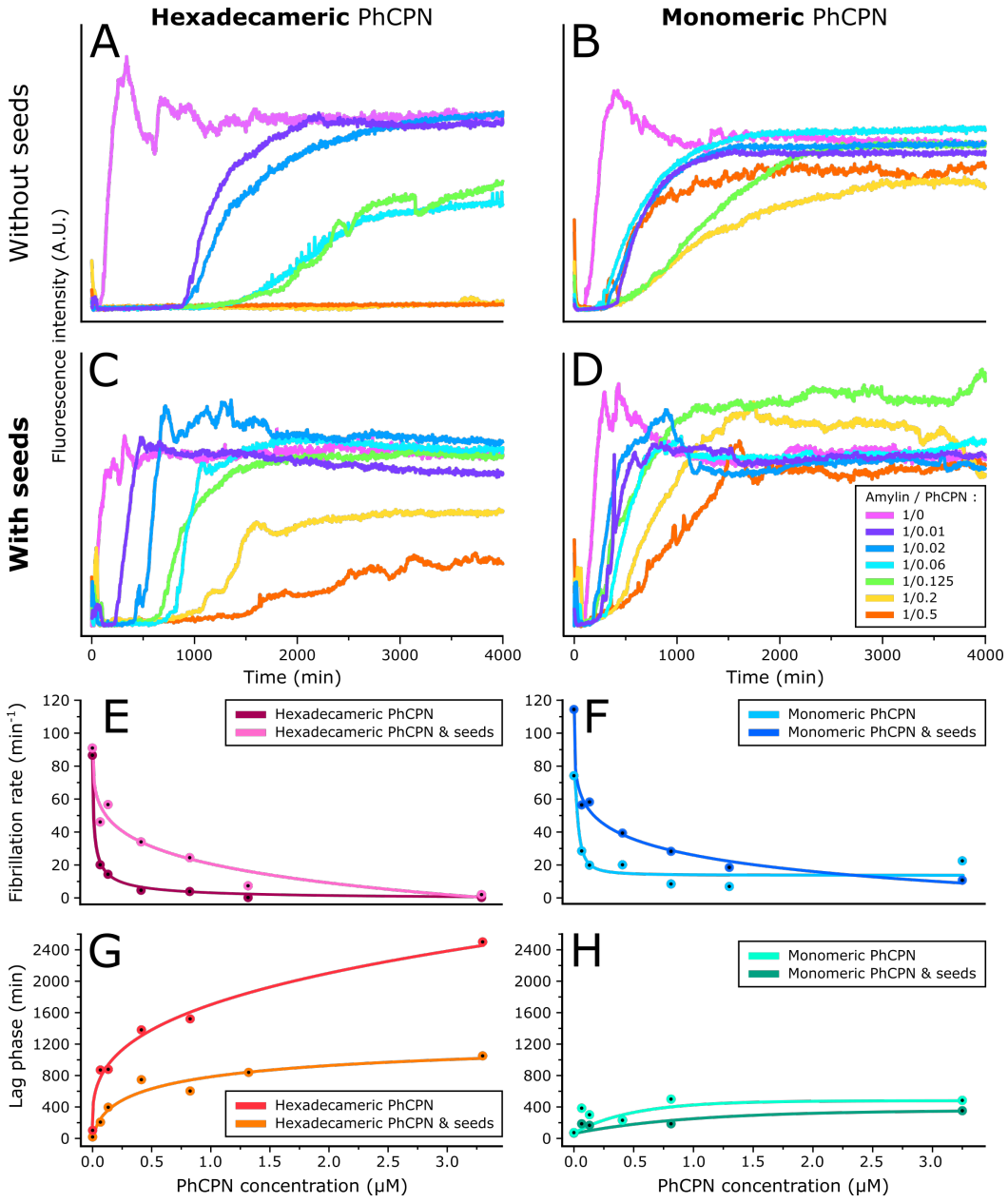


Fig. 5.13 : Influence of different PhCPN ratios on Amylin fibrillation followed by ThT fluorescence. (**A**) Influence of hexadecameric PhCPN on Amylin, in PhCPN buffer. Amylin alone was scaled by a factor x1,4. (**B**) Influence of monomeric PhCPN on Amylin, in MES buffer. Amylin alone was scaled by a factor of x1,2. (**C**) & (**D**) Influence of the introduction of 10% seeds on the fibrillation of Amylin in PhCPN and MES buffer, respectively. Fluorescence intensity scales were removed for clarity since the curves intensities were artificially scaled. Experiments were carried out on a medium binding 96-well plate, at 30°C, with measurements every 8 minutes and 3 seconds shaking prior to each measurement, as described in **Chap. 6.5.1**. Amylin alone was scaled by a factor x1,3 in PhCPN buffer and x1,9 in MES buffer. (**E**) & (**F**) Fibrillation rates extracted from the fibrillation curves in PhCPN buffer and MES buffer, respectively. (**G**) & (**H**) Lag phases extracted from the fibrillation curves in PhCPN buffer and MES buffer, respectively.

the lag phase is also way less pronounced but still follows an increasing tendency with increasing amounts of PhCPN (**figure 5.13, H**). Note that in the case of the highest PhCPN concentration in MES buffer, however, aggregation is increased again (**figure 5.13, B, orange curve**). This observation was reproducible but could not be properly explained. The PhCPN monomers have a clearly lower impact on aggregation compared to the hexadecamers at similar concentrations. Monomers were proven to interact with Amylin by NMR. Interestingly, the interaction with free Amylin was stronger in MES buffer, but the aggregation is dramatically reduced in PhCPN buffer.

Seeding in presence of PhCPN

The same experiments were reproduced after the addition of seeded Amylin to initiate the aggregation reaction. In fact, in presence of high amounts of seeds, the lag phase is supposed to be suppressed and aggregation is only driven by elongation mechanisms [Cukalevski et al., 2015]. In presence of low amounts of seeds, however, the aggregation curves report the eventual presence of secondary nucleation processes. Seeds were prepared by dissolving lyophilized Amylin to a concentration of 100 µM in PhCPN or MES buffer depending on the experiment, complemented with a sterile glass bead and agitated for 48 h at 37°C. 2 µL of this mixture was added to each well to reach a fibril amount of 10% (w/w) and aggregation was followed by fluorescence as before. In presence of hexadecameric PhCPN in PhCPN buffer and the seeds, fibrillation delay can still be observed (**figure 5.13, C**). The lag phase is globally diminished approximately by 50% compared to the case when the seeds are absent, but nevertheless still remain detectable (**figure 5.13, G**). The fibrillation rate is also less decreased compared to PhCPN in PhCPN buffer and absence of seeds (**figure 5.13, E**). The final fluorescence intensities are also decreased by the addition of increasing amounts of PhCPN. In absence of seeding, aggregates were not detected within experimental time for the smallest Amylin/PhCPN ratios, while after addition of the seeds the lag phase was globally reduced, allowing to observe aggregation, confirming the reduced effect of PhCPN on the fibrillation rate and the lag phase. Aggregation of Amylin with monomeric PhCPN in MES buffer in presence of seeds is also different (**figure 5.13, D**), especially because the lag phases are almost completely canceled (**figure 5.13, H**). When comparing the behaviour of the monomeric and the hexadecameric PhCPN in presence of seeds and Amylin, it seems that for each case inter-

action mechanism is different. While the concentration of seeds remained the same in all the experiments, the shape and behaviour of the curves were dramatically distinctive. One could conjecture that in presence of intact PhCPN, the main aggregation mechanism is secondary aggregation, possible because PhCPN might sequestrate the seeds, thus the need for a secondary aggregation to start the overall aggregation process and explaining the presence of the lag phase even in presence of amyloid seeds. On the contrary, the aggregation starts right at the beginning of the experiments in presence of monomeric PhCPN, showing that the main aggregation process is elongation. Since monomers are unable to sequester the seeds, due to their lack of cavity, the fibrillation starts and its speed is decreased by the presence of the monomers of PhCPN, that transiently interact with the Amylin monomers and thus slowing down the fibrillation process.

5.2.4 Characterization of the fibrils

AFM and EM techniques are widely used to characterize the shape and dimensions of the fibrils. Both approaches were utilized in PhCPN and MES Buffer, to observe the effect of PhCPN on the fibril formation.

Atomic Force Microscopy

The fibrils were first imaged using atomic-force microscopy (**AFM**). In the case of AFM, the dimension which is measured precisely is the height of the objects and not the width. Height measurements were carried out on different samples in MES buffer. The same conditions could not be scanned in PhCPN buffer since the buffer was containing too much salts and perturbed the AFM tip and cantilever. Imaging in PhCPN buffer was latterly carried out by EM (see **Section 5.2.4**).

It was possible to observe the presence of a large number of fibrils in MES buffer (**figure 5.14, A**). However, it was impossible to observe the twists on the fibrils and differentiate the different strands. The diameter of the fibrils is of approximately 2 nm, error included. The length of the fibrils was heterogeneous, ranging from 1 to several µM. Similar dimensions were reported in the literature, with fibrils lengths from 100 nm to several µM [Bhattacharya et al., 2007, Fukuma et al., 2008]. The fibrils do not seem to be branched and also form seed-like structures. No fibrils could be observed in presence of PhCPN (**figure 5.14, B**). Additional dots are visible on the pictures, very likely correspond-

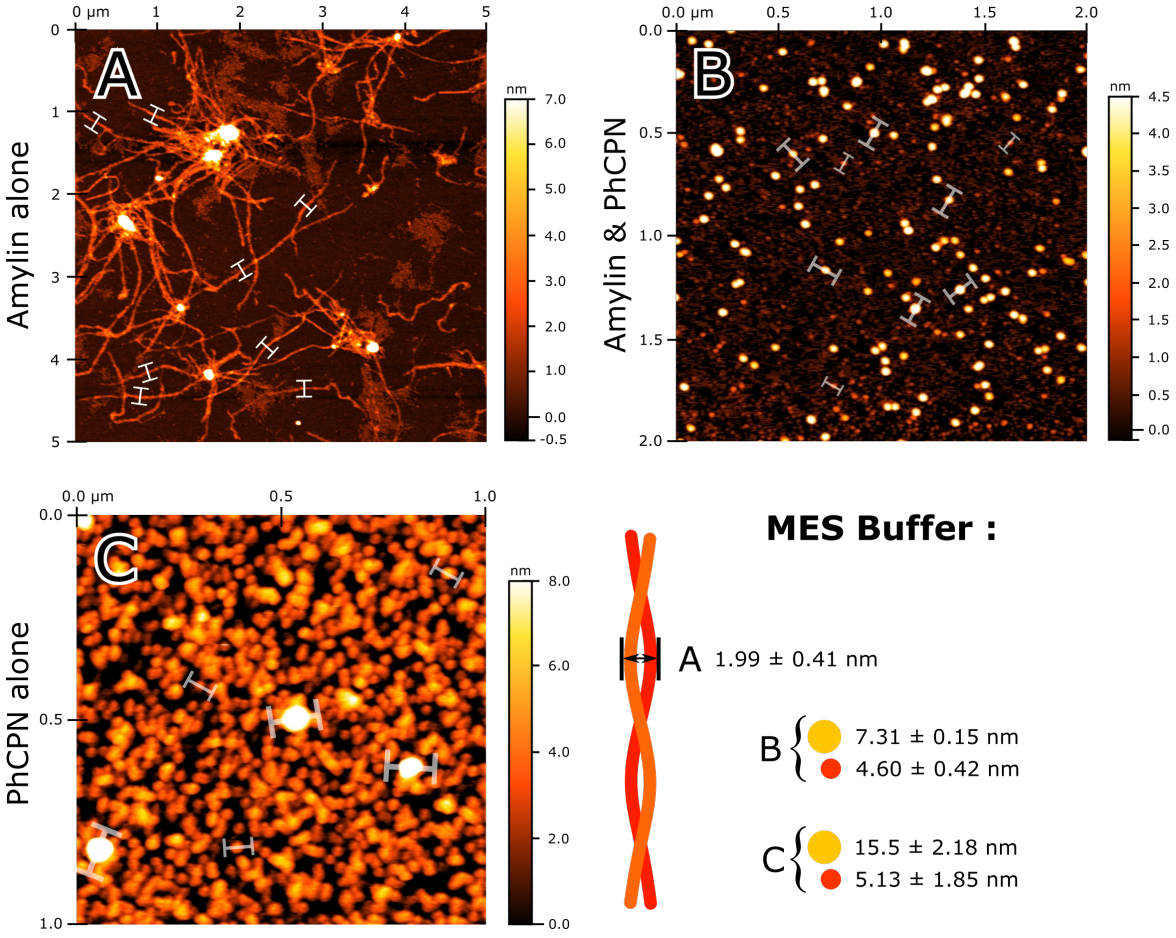


Fig. 5.14 : AFM pictures of Amylin in MES buffer only, in absence or presence of PhCPN, and their corresponding measurements. (**A**) 100 µM Amylin in MES buffer. (**B**) 100 µM Amylin in presence of 5 µM PhCPN in MES buffer. (**C**) 5 µM PhCPN in MES buffer.

ing to PhCPN monomers (7.31 nm ; (**figure 5.14, B**)), and possibly few hexadecamers (15.5 nm ; (**figure 5.14, C**)). These sizes probably correspond to PhCPN monomers and hexadecamers seen from the top. Small particles were also noticeable in the background of the images (**figure 5.14, B & C, red dots**). They could correspond either to small Amylin oligomers or parts of PhCPN stacked at the background of the images but there was no proper way to confirm this assumption.

Electron Microscopy

After the first evaluation by AFM, the fibrils were also observed and measured by negative staining Electron Microscopy (**EM**). The different samples were prepared in low binding safe Eppendorfs and composed of 100 µM Amylin in the presence or absence of 5 µM PhCPN in different buffers and left steady to fibrillate for one week at room temperature. Samples were taken from the tubes after gentle mixing and either diluted or directly deposited for negative staining EM analysis. EM images were also used to measure the width of the fibrils and the twist periodicity, the resolution being high enough to observe the twists on the fibrils. Measurements were performed with the Gwyddion software (**figure 5.15, C**). The plot intensity was arbitrarly fixed while the width was precise and defined as a function of the image original scale. The width was measured either at the base of the peaks or with the full width at half maximum (**FWHM**). By comparison with the theoretical height of the simulated PhCPN open structure (**figure 2.13, A & B** and **Chap. 6.7**), the most accurate measurement was taken at the base of the peaks. The different dimensions already measured for various amyloidogenic protein fibrils are reviewed in [Li et al., 2016].

The pictures were more resolved than with AFM and the different twist periods could be observed in MES buffer, as shown on (**figure 5.15, A, n°5**). The heterogeneity of the fibrils explains the high uncertainty on the intertwist value of 125.2 ± 34.3 nm (**figure 5.15, A, C & D**). Literature values are quite different, of 4-40 nm and were measured by AFM, in PBS buffer [Fukuma et al., 2008]. The fibrils are interleaved and form a compact network, making a precise measurement of their average length particularly challenging (**figure 5.15, A & B**). Fibrils can extend from a few hundred nm to several µm, as previously observed by AFM and reported in the literature [Bhattacharya et al., 2007, Fukuma et al., 2008]. The diameter of the fibrils strands was of 7.2 nm. This measurement seemed

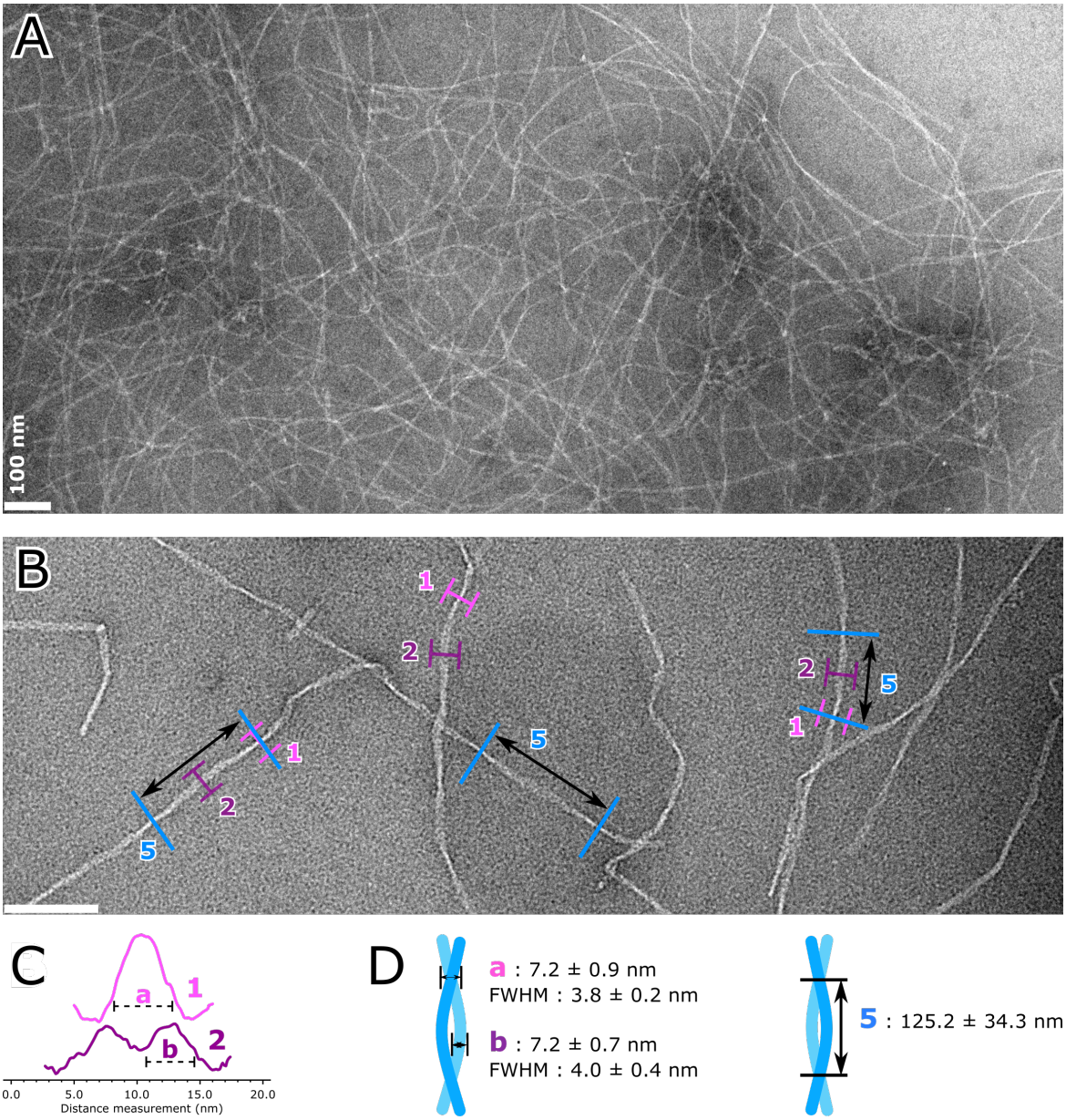


Fig. 5.15 : Negative staining EM picture of Amylin fibrils in absence of PhCPN, in MES Buffer. 100 µM were left to fibrillate at room temperature for 7 days in a low binding Eppendorf tube and deposed on the copper grid. (**A**) Ensemble view of Amylin amyloid fibrils in MES buffer. (**B**) Close up of the Amylin fibrils in MES buffer. (**C**) Intensity plots of the shades of white on the pictures to precisely determine the width of the different fibrils on the EM pictures. (**D**) Dimensions of Amylin fibrils in MES buffer. Measurements of the fibrils were carried out with the Gwyddeon software.

reliable, as the measurements come from the side view (**figure 5.15, A, n°1**) as well as the front view of the fibrils (**figure 5.15, A, n°2**). However, these dimensions are very different from the ones observed by AFM, with a diameter measuring approximately 2 nm.

In MES buffer (**figure 5.16, B**), fibrillation is favored and long and homogenous fibrils can be observed. Some fibrils seem to start from a common origin point, which could be Amylin seeds, as previously observed by AFM (**figure 5.14, A**). The fibrils are made of

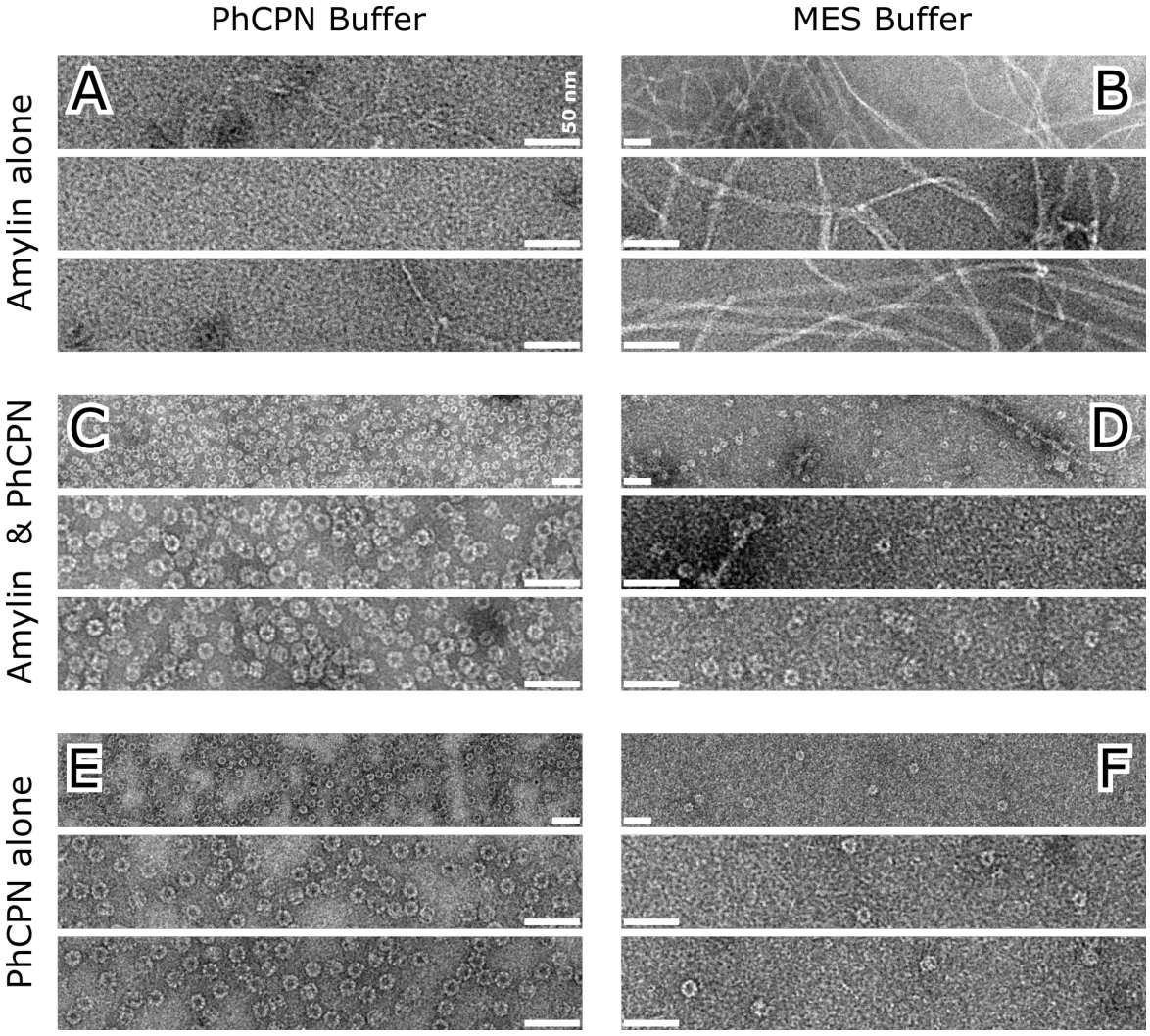


Fig. 5.16 : Negative staining EM pictures of Amylin in presence or absence of PhCPN, in different buffers. (**A**) & (**B**), 100 µM Amylin in PhCPN and MES buffer respectively. (**C**) & (**D**) 100 µM Amylin in presence of 5 µM PhCPN in PhCPN and MES buffer respectively, diluted 10 times prior to deposition on the copper grid. (**E**) & (**F**) 5 µ M PhCPN in PhCPN and MES Buffer respectively, diluted 10 times prior to deposition on the copper grid. Samples are diluted 10 times because of the large size of PhCPN. Maximum PhCPN concentration for acceptable quality negative staining EM images is about 0.5 µM.

two intertwined protofibrils. Observable fibrillation is not favored by the PhCPN buffer (**figure 5.16, A**) probably because of high salt content and lack of interactions between the monomers to form the fibrils. Nevertheless, a few fibrils are still present in the sample but exhibit slightly smaller diameter, and could not be precisely measured because of the lack of contrast and the presence of background noise in the images (**figure 5.16, D**). Their approximate diameter is 7 nm and is homogeneous throughout the fibrils, unlike what is observed in absence of PhCPN in the same buffer. This could show that the fibrils are constituted of a unique Amylin protofilament, contrary to the two intertwined protofilaments observed in absence of PhCPN (**figure 5.16, A**), thus proving an effect of PhCPN on the fibrils structure. In PhCPN buffer and in presence of PhCPN (**figure 5.16, C**), no fibrils could be observed over the whole grid. In MES buffer, PhCPN is mainly monomeric, which explains the very few visible PhCPN particles (**figure 5.16, D & F**). The monomers of PhCPN can be seen in the background of the images, as shown on the PhCPN reference sample. Very few fibrils can be observed in MES buffer in presence of PhCPN, which confirms the anti-fibrillation activity of PhCPN, even when monomeric. These fibrils seem to be made of only one protofilament.

5.3 Chapter summary and discussion

In this chapter, the interaction between Amylin, PhCPN and PhCPN monomers was characterized. It was successively shown that :

- PhCPN, no matter if monomeric or hexadecameric, interacts preferentially with the Nter region of Amylin.
- Different conformation and compaction of the peptide caused by additional disulfide bridge do not alter the interaction of Amylin with PhCPN.
- Both proteins are very likely in fast exchange but the precise K_D could not be determined due to technical difficulties to methyl-label Amylin.
- PhCPN monomer has more affinity for Amylin than hexadecameric PhCPN, approximately 100 µM compared to 1 mM respectively, as estimated by NMR.

- The aggregation of Amylin is dramatically influenced by the presence of PhCPN, especially when hexadecameric, possible due to cooperativity between the monomers forming the rings.
- The fibrils could be pictured and characterized in MES buffer by EM and AFM, but not in PhCPN buffer.

These results are in agreement with previously published data on the interaction between HEWL and PhCPN [Mas et al., 2018] and the previous observations with MSG. PhCPN was proved to interact with the two selected disordered peptides, Aβ42 and Amylin.

Peptide structure - In the case of Amylin, the structure of the peptide does not seem to matter, since Amylin and Amylin with an extra disulfide bridge interact in an almost identical way with PhCPN. Nevertheless, the binding mode of PhCPN for Amylin and Aβ42 is different since PhCPN interacts mainly with Nter region of Amylin, whereas in the case of Aβ42, it interacts with the hydrophobic regions only. This could be due to the presence of the induced loop at the Nter of Amylin caused by the presence of the disulfide bridge, forming a large hydrophobic center. Interestingly, this region is not believed to be involved in the fibrillation of Amylin [Westermark et al., 1990]. Also, PhCPN does not interact with αSyn, which could indicate a form of substrate selection from PhCPN as well.

Dynamics of the complex - As just stated, PhCPN, whether monomeric or hexadecameric, interacts preferentially with the Nter of Amylin. The complex and the free proteins are very likely in fast exchange, as previously reported in the literature [Libich et al., 2013], with a weak interaction, as proven by its K_D in the mM range estimated from NMR spectra. The interaction between monomeric PhCPN is stronger compared to the interaction with complete PhCPN as demonstrated by NMR. This can be explained by different interacting modes and especially interactions surfaces since the substrate-accessible surface of monomeric PhCPN is sensitively different than in the case of hexadecameric PhCPN. As stated previously, this large difference could also be explained by the presence of additional salts to maintain PhCPN hexadecameric, masking electrostatic interactions since the Nter of Amylin is also its most charged region.

Possible interaction with seeds - At the same time, the aggregation rate of Amylin is slower and less important in presence of hexadecameric PhCPN than with monomeric PhCPN. Full PhCPN particles artificially reduce the concentration of free Amylin in solution by hosting them in this cavity, reducing aggregation. It would also seem that integral PhCPN interacts with oligomers or seeds and delays aggregation. In fact, as proven by ThT assays, monomeric PhCPN only influences the aggregation speed in presence of seeds and does not delay aggregation, possibly because monomeric PhCPN does not interact with seeds. If monomeric Amylin is in competition with Amylin seeds in integral PhCPN, that could explain why the interaction observed is weaker than with monomeric PhCPN. The cavity would be already occupied by the seeds and the monomeric Amylin interaction surface would be reduced, thus increasing the apparent K_D.

Picturing the fibrils - Another interesting fact is the difference of aggregation and fibrillation of Amylin in the two different experimental buffers used in this study, PhCPN buffer (high salt content) and MES buffer (low salt content). The aggregation, monitored by ThT fluorescence is increased in presence of salts, as reported in the literature [Meisl et al., 2017]. Unfortunately, fibrils were extremely difficult to find in the negative staining EM experiments in PhCPN buffer. These two observations seem contradictory. A first explanation would be that the electrostatic interactions between the copper grid (coated with carbon or not) for EM and the mica for AFM were masked by the presence of salts and that the fibrils were washed away during the preparation of the samples prior to imaging, even after numerous attempts. A second and more likely hypothesis would be that long fibrils are not formed during the ThT assays, due to shaking. In fact, samples from the 96-well plates were taken at different times before, during and after aggregation and no fibrils could be pictured by EM either. These oligomers would be impossible to picture because of their small size. AFM supports this idea since the dots visible at the very background of the images taken in presence of PhCPN buffer are presumably not PhCPN monomers. The GroEL chaperonin was recently reported to decorate amyloid fibrils in a regular manner [Wälti et al., 2017]. This behaviour could not be pictured in the two working buffers since no fibrils were formed in PhCPN buffer or at least could not be pictured. It is also difficult to conclude anything in MES buffer since PhCPN is mainly monomeric but PhCPN does not seem to decorate the fibrils.

Conclusion - These results demonstrate the anti-aggregation activity of both monomeric and hexadecameric PhCPN in presence of fibrillating amyloidogenic proteins. In addition, the behaviour of the monomer alone or the hexadecameric particle could be treated separately, emphasizing the importance and effect of PhCPN cavity. PhCPN reduces the amount of free Amylin in solution by hosting the amyloid monomers, in fast exchange. These Amylin monomers are not free to interact with other free monomers in solution, thus delaying the fibrillation.

This situation is more comparable to the study with HEWL, since the substrate is in a permanent exchange between its PhCPN-bound and free state, unlike with MSG. However, the location of Amylin within PhCPN cavity could not be determined to date and remains to be investigated. Despite the large size of PhCPN cavity, PhCPN seems to host one Amylin per cavity at the time, similarly like with HEWL, even though the precise stoichiometry could not be determined. In fact, the weak affinity between PhCPN and Amylin makes it very unlikely that two Amylin interact with the cavity at the same time, since Amylin is more than 99% of the time free in solution.

5.4 Perspectives

The study of the interaction between PhCPN and an amyloidogenic substrate protein described in this thesis has raised additional questions which need to be further investigated.

5.4.1 Remaining questions

One of the main interrogations at the end of this project is the location of Amylin within PhCPN. In fact, no shift upon liaison with Amylin could be observed by liquid-state NMR without methyl labelling, due to the small size of Amylin and the dynamics of the interaction, similarly with HEWL [Mas, 2015]. In addition, the amyloidogenic nature of Amylin limited the reachable temperature range for the study by NMR, resulting in low signal to noise for PhCPN spectra. Unfortunately, methyl labelling is really challenging and not adapted to Amylin because of its purification procedure and especially deuteration. To compensate, PRE experiments were considered to yield similar data like with MSG in the previous chapter. Using PRE could also prove or not that the interaction site on the

monomers or hexadecamers of PhCPN is different if the assignment of the monomers of PhCPN is carried out first. An identical interaction site could reinforce the theory that oligomers interact with the cavity, and would explain the differences in observed K_D.

PRE is also challenging, as Amylin needs to be paramagnetically tagged. First trials with TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl) to label Amylin were carried out by collaborators but the nitroxide spin probe was not paramagnetic when bound to Amylin, possibly due to reduction by the Cysteines. Another solution was to create a synthetic Amylin bearing a covalently linked paramagnetic tag, the DOTA (1,4,7,10-Tetraazacyclododecane-1,4,7,10-tetraacetic acid). The DOTA group is attached to Amylin by a two β-Alanine-long linker (**figure 5.17, A**). The DOTA can be loaded with Gd (III), chelated in the center of the aza-crown motif and become paramagnetic.

At the time of redaction, the synthetic peptide was ordered from a company and needs to be tested in presence of PhCPN. The assignment of methyl groups of PhCPN Methionines and Valines has been transferred at low temperature and the conditions optimized to get the best compromise between experimental time and resolution, to picture intensity changes in presence of the paramagnetically tagged Amylin (**figure 5.17, B & C**).

Another promising aspect of this work is the possible interaction between PhCPN and amyloid seeds/oligomers. Proving more thoroughly this interaction by completing the work with seeds from different amyloidogenic proteins could pave the way for further study of the interaction in the context of disease treatment since seeds are believed to be

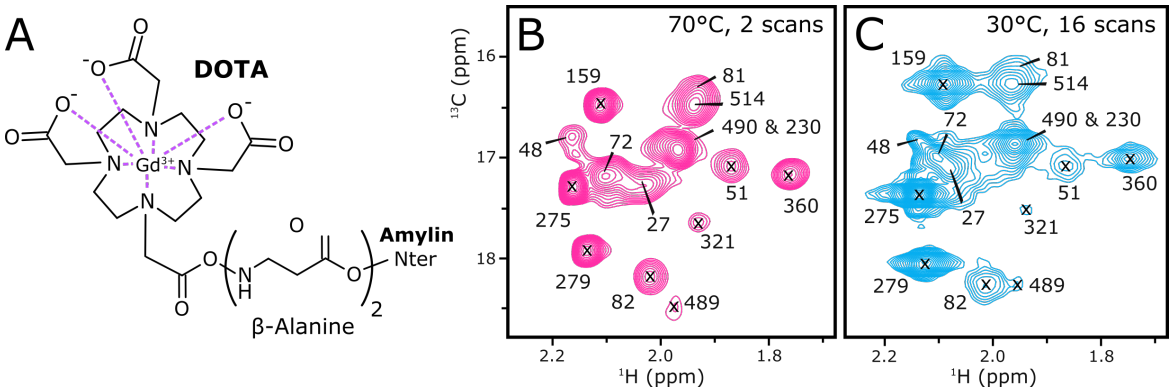


Fig. 5.17 : Preparation of the PRE experiments. (**A**) Scheme of the synthetic DOTA-Amylin paramagnetic compound, loaded with Gd(III). (**B**) (^{1}H-^{13}C) methyl HSQC spectrum of 25 µM PhCPN at 70°C, 2 scans. (**C**) (^{1}H-^{13}C) methyl HSQC spectrum of 25 µM PhCPN at 30°C, 16 scans.

the toxic species in most amyloidosis. PhCPN could also be used as a tool to encapsulate and study individual seeds/oligomers. In fact, isolation of the seeds is particularly challenging since their lifetime is very short. Protein engineering to trap the seeds inside the cavity with the cavity closing with ATP could theoretically be possible. One could argue that these seeds/oligomers were not observed by NMR in the presented work but that could logically be explained by the short lifetime of the oligomers, since recording one spectrum takes approximately 15 minutes. Last but not least, most experiments were carried out in both PhCPN and MES buffer. Even though the obtained results are completely valid, it would still be interesting to evaluate the effect of the difference of salt concentration on the interaction with hexadecameric PhCPN. In fact, in **Chap. 3**, it was determined that PhCPN is hexadecameric in MES buffer complemented with 25 mM $MgCl_2$ only. Repeating some of the experiments in such a buffer could enlighten the effects of salts on the fibrillation and thus define more precisely the effect of PhCPN.

5.4.2 Future of the project

As side tests during this thesis, Amylin was also proven to interact with PhCPN co-chaperonin, the Prefoldin from *Pyrococcus horikoshii* (**PhPFD**). Prefoldin and Amylin were mixed together in PhCPN buffer, at different Amylin/Prefoldin ratios, in conditions very similar to those reported in **Section 5.2.2** and **Section 5.2.2**.

A clear interaction between the two proteins could be observed by titration. Peaks intensity decrease because of the increasing apparent size of Amylin when associated with PhPFD (**figure 5.18, A & B**). The signal of bound Amylin was still observable, therefore the chemical shifts too (plotted as a function of the AA residue (**figure 5.18, C**). Such a result opened an avenue on the more complete study of the behaviour of PhCPN in presence of its co-chaperonin and the substrate transfer between both chaperones. MSG could not be used for these type of experiments but Amylin is the perfect candidate protein and is also of therapeutic interest. This aspect of the project is further investigated by a new PhD student of the NMR group, Ricarda Toerner. The aim is to be able, to term, to study the whole chaperonin machinery in interaction with a fibrillating substrate, like the Amylin.

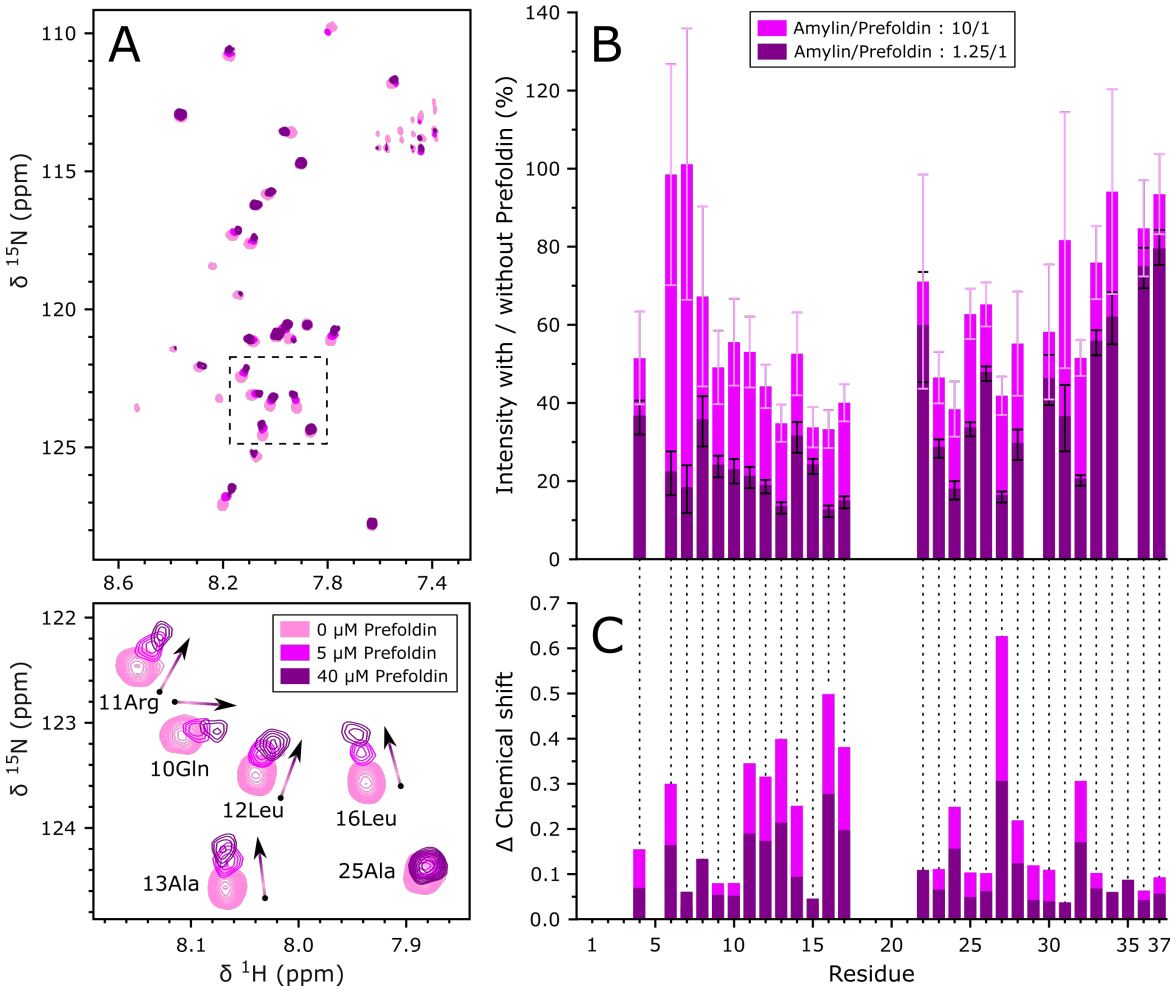


Fig. 5.18 : Monitoring the interaction of Amylin with PhPFD by NMR, at 30°C. (**A**) Overlayed BEST-TROSY ^{1}H-^{15}N NMR spectra of 50 µM Amylin in PhCPN buffer, in the absence or presence of different amounts of PhPFD. (**B**) Plot of the intensity of Amylin signal with or without PhPFD, for different PhPFD concentrations. (**C**) Plot of the chemical shift of Amylin peaks as a function of PhPFD concentration.

Chapter 6

Materials and Methods

Ce chapitre décrit précisément le matériel et les protocoles opératoires employés pour produire les protéines nécessaires à ce travail de thèse. Il regroupe notamment des éléments de théorie rattachés aux expériences biochimiques et biophysiques clefs utilisées pour caractériser l'interaction de PhCPN avec ses protéines substrats. Il développe plus particulièrement les différentes techniques de RMN associées à la biochimie structurale qui sont évoquées dans les différents chapitres introductifs et qui ont été employées durant ce doctorat.

6.1 Culture media and buffers

All standard chemicals were purchased from Merck/Sigma-Aldrich and Euromedex. Isotopes and D_2O were purchased from Merck/Sigma-Aldrich, Euriso-top and Cambridge Isotope Laboratories. Methyl specific labelling precursors were purchased from NMRBio.

LB Medium - Lysogeny Broth medium (LB) was directly purchased from Sigma-Aldrich and complemented with appropriate antibiotics.

M9 H_2O Medium - The base M9 composition was of 40 mM Na_2HPO_4, 20 mM KH_2PO_4, 8,5 mM NaCl and 20 mM NH_4Cl. After autoclave, Glucose was added to a concentration of 10 mM and the following concentrations in complements were added to the medium : 1 mM $MgSO_4$, 0,1 mM $CaCl_2$, 0,1 mM $MnCl_2$, 50 µM $ZnS0_4$, 0,1 mM $FeCl_3$ and a cocktail of Pyridoxine, Biotin, D-Pantothenic acid hemicalcium salt, Folic acid, Choline chloride, Niacinamide, Riboflavin, Thiamine vitamins. NH_4Cl and glucose 13C, 15N and 2H isotopes were adapted depending on the desired labelling for NMR.

M9 D_2O Medium - The composition of the deuterated M9 was identical as the previously described one except that solvent was substituted by D_2O (99.85%). The vitamins were resuspended and lyophilized in D_2O two times prior addition to the vitamin cocktail. Amounts of NH_4Cl and glucose ^{13}C, ^{15}N and ^{2}H isotopes were adapted depending on the desired labelling for NMR (for methyl labelling, see below section expression and purification section).

PhCPN Buffer - PhCPN buffer was composed of 25 mM MES pH 6.5, 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$ and was optimized for the ATPase activity of PhCPN by Dr. Guillaume Mas **(Mas, 2015)**.

Deuterated PhCPN Buffer - Deuterated PhCPN buffer was composed of 25 mM MES-d13 pH 6.5 (read 6.9 on the pH-meter), 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$

MES Buffer - MES Buffer was composed of 25 mM MES pH 6.5.

Deuterated MES Buffer - Deuterated MES buffer was composed of 25 mM MES-d13 pH 6.5 (read 6.9 on the pH-meter).

6.2 Protein expression and purification

Pyrococcus horikoshii PhCPN - For expression of the *Pyrococcus horikoshii* Thermosome (**PhCPN**), the recombinant plasmid pET21b-PhCPN was transformed into the *E. coli* strain BL21(DE3)-RIL. Pre-cultures were grown in LB medium complemented with 1 mM Kanamycin and 0.5 mM Chloramphenicol antibiotics overnight at 37°C, harvested and transferred in M9 medium for the final culture. In fact, PhCPN yield is increased in M9 compared to LB, even for cultures grew in H_2O. The temperature was set up to 37°C and culture was started at the optical density 0.8 (O.D.) with rotary shaking. When O.D. reached 0.8, expression was induced by addition of IPTG to a final concentration of 1 mM. The culture was incubated for 4 h at 37°C and cells were collected by centrifugation, cells pellets were stored at -20°C until use. A cell pellet corresponding to 1 L of culture was thawed and mixed on a mixing wheel in 50 mL of a first lysis buffer containing 20 mM NaCl, 5 mM $MgCl_2$, 20 mM TRIS pH 8, Chicken Lysozyme (0,3 mg/mL, Euromedex), EDTA-free protease inhibitors (Roche Applied Sciences), DNase (0,05 mg/mL, Euromedex) and RNase (0,02 mg/mL, Euromedex). Once disrupted by sonication (5 min effective lysis, rounds of 5 sec on, 5 sec off (**figure 6.1, A**), the resulting mixture was heated at 80°C for 30 minutes in a water bath and centrifuged at 16,000 x g for 30 min to pellet cell debris and precipitated proteins. The sample was filtered and loaded onto an anion exchange Resource Q column (GE Healthcare), previously equilibrated with Buffer A (20 mM TRIS pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$). After injection of all the sample, the column was washed with 20 column volumes (CV) of Buffer A at 3 ml/min to remove unbound or weakly bound PhCPN and residual unwanted proteins. PhCPN was then eluted with a 2 CV linear gradient of 0% to 100% elution Buffer B (20 mM TRIS pH 8.0, 1 M NaCl, 5 mM $MgCl_2$) and collected (**figure 6.1, B**). The fractions containing PhCPN were then pooled and applied to a 120 mL custom SuperOse6 size-exclusion chromatography column (GE Healthcare, homemade packing by Isabel Ayala), equilibrated, and eluted with PhCPN Buffer, containing 25 mM MES pH 6.5, 50 mM NaCl, 100 mM KCl and 25 mM $MgCl_2$ (**figure 6.1, C**). After purification, PhCPN was either kept at 4°C for rapid use or stored at -20°C (**figure 6.1, D**). Yield fluctuates around 20 mg to 30 mg PhCPN per litre of culture in H_2O. All purification steps were controlled by an Äkta Purifier system (GE Healthcare) and carried out at room temperature. The protein elution was monitored by measuring the absorbance at 280 and 260 nm.

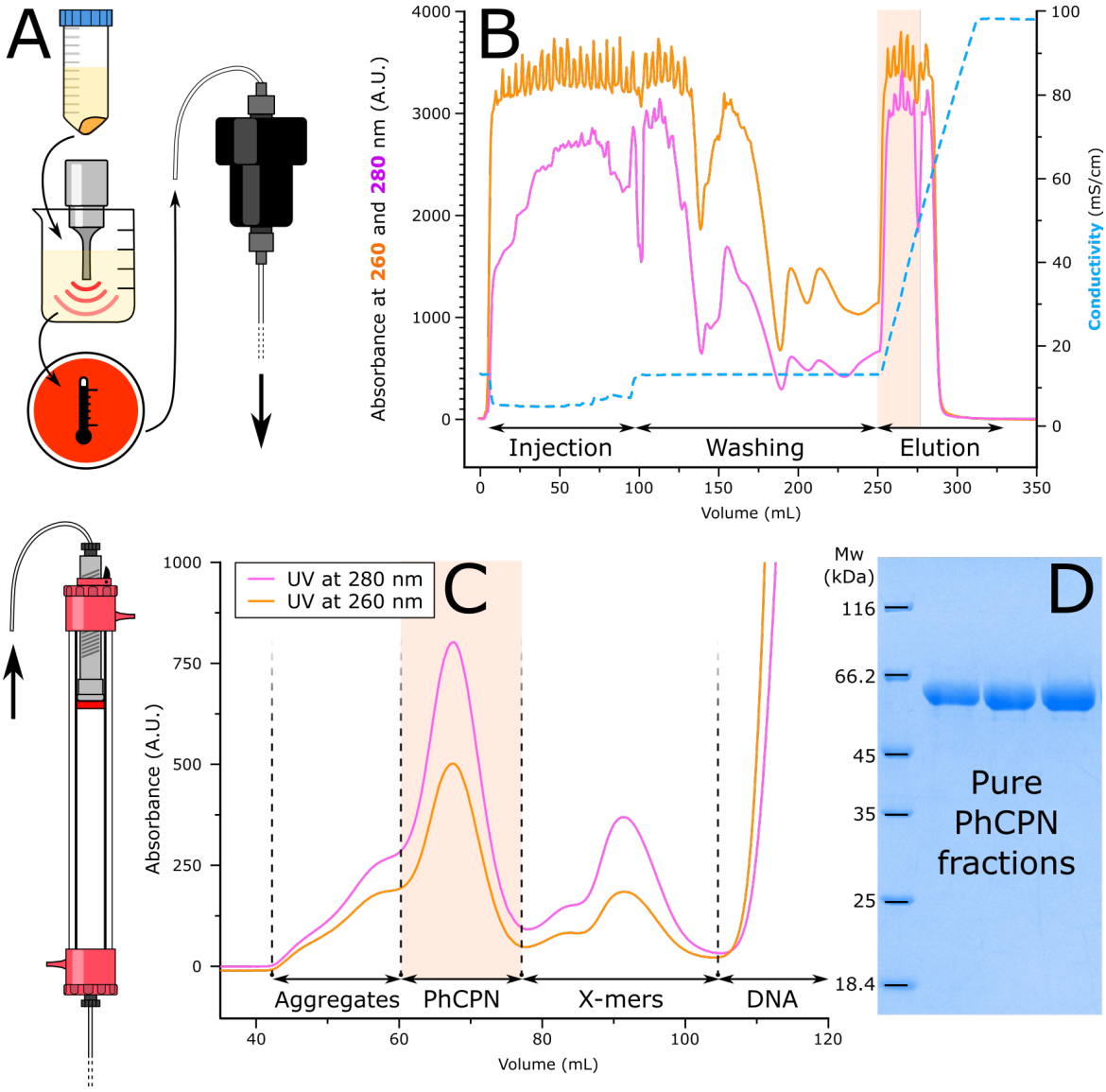


Fig. 6.1 : Purification of PhCPN. (**A**) Initial purification steps of PhCPN. The cells are lysed and heated at 80°C. The resulting mixture is injected into an RQ column. (**B**) RQ profile of PhCPN purification. The fractions kept for further purification are enclosed in salmon color. (**C**) The fractions from RQ are concentrated and injected on the SuperOse preparative column. The aggregates, PhCPN and X-mers are separated. The final pure fractions used for the rest of the experiments are enclosed in salmon color. (**D**) Purity check of the salmon-colored fractions.

Escherichia coli MSG - For expression of the *Escherichia coli* Malate Synthase G (**MSG**), the recombinant plasmid pET21b-MSG was transformed into the *E. coli* strain BL21(DE3). Precultures and the final culture were grown in LB medium complemented with Kanamycin at 1 mM. The culture was started at O.D. 0.3 and 37°C with rotary shaking. When O.D. reached 0.8, the temperature was lowered to 20°C and expression was induced by addition of IPTG to a final concentration of 1 mM. The culture was incubated overnight at 20°C and collected by centrifugation, cells pellets were stored at -20°C until use. A cell pellet corresponding to 500 mL of culture was thawed and mixed on a mixing wheel in 20 mL of a lysis buffer containing 100 mM NaCl, 10 mM $MgCl_2$, 50 mM TRIS pH 8.0 and 5 mM β-mercaptoethanol. Once disrupted by sonication (2 min effective lysis, rounds of 2 sec on, 8 sec off), the resulting mixture was filtered and loaded onto a chelating Sepharose column (40 mL of resin for 1 L of culture), previously equilibrated with lysis buffer. The column and bound protein were rinsed with 5 CV of lysis buffer, then with 4 CV of lysis buffer complemented with 20 mM imidazole. 4 CV of lysis buffer with 50 mM imidazole were used to unbind and collect MSG and, finally, 5 CV of lysis buffer with 500 mM imidazole were employed to remove all the remaining proteins from the column. Collected fractions were loaded on SDS-PAGE gels to spot the ones containing MSG and pool them together. The sample was concentrated with 30 kDa cutoff Amicon (Merck) and loaded on an HiLoad 16/60 Superdex 200 size-exclusion chromatography column (GE Healthcare) previously equilibrated with 25 mM Tris pH 7.0, 20 mM $MgCl_2$ and 5 mM DTT buffer. After purification, the protein was kept at 4°C to prevent precipitation (folded MSG cannot be frozen for storage). Yield is about 200 mg MSG per litre of culture in H_2O. Before mixing with PhCPN, the buffer was exchanged for PhCPN buffer. All purification steps were controlled by an Äkta Purifier system (GE Healthcare) and carried out at room temperature. The protein elution was monitored by measuring the absorbance at 280 and 215 nm.

Amylin (IAPP) - Expression and purification of the Amylin protein were carried out by Dr. Lothar Gremer. The protocol is originally described in [Mirecka et al., 2014a] and copied here for clarity. For expression of the HI18-IAPP fusion protein, the recombinant plasmid pET302-HI18-IAPP was transformed into the *E. coli* strain BL21(DE3). Pre-cultures were grown in LB medium complemented with Ampicilin overnight at 37°C, harvested and transferred in M9 medium for the final culture. The culture was started

at O.D. 0.3 and 37°C with rotary shaking. When O.D. reached 0.6, the temperature was lowered to 27°C and expression was induced by addition of IPTG to a final concentration of 1 mM. The culture was incubated overnight at 27°C and cells were collected by centrifugation. A cell pellet corresponding to 1 L of culture was resuspended in 20 mL of a lysis buffer containing 20 mM Imidazole, 500 mM NaCl, 20 mM sodium phosphate pH 8.0 and containing EDTA-free protease inhibitors (Roche Applied Sciences). Cells were lysed using a cell disruptor (Constant Systems). Cells debris and precipitated proteins were pelleted by centrifugation at 28,000 x g. The sample was filtered and loaded onto an HisTrap FF 5 ml column at a flow rate of 5 ml/min. The column was washed with 5 CV of 20 mM imidazole, 500 mM NaCl and 20 mM sodium phosphate at pH 8.0. The HI18-IAPP fusion protein was eluted with 5 CV of 250 mM imidazole in 500 mM NaCl, 20 mM sodium phosphate, pH 8.0. Pooled elution fractions were loaded onto a HiLoad 16/60 Superdex 75 size-exclusion chromatography column (GE Healthcare) equilibrated with 20 mM Tris–HCl pH 8.0, 50 mM NaCl, and the protein was eluted at a flow rate of 1 ml/min. Monomeric fractions were pooled and stored at -80°C. All purification steps were controlled by an Äkta Purifier system (GE Healthcare) and carried out at room temperature. The protein elution was monitored by measuring the absorbance at 280 and 215 nm. HI18-IAPP fusion protein was cleaved with Factor Xa (New England Biolabs) added at 25 U/mg of fusion protein in 20 mM Tris-HCl pH 8.0 containing 10 mM $CaCl_2$, for 16 h at 16°C. The cleavage reaction products were subjected to reverse phase high-pressure liquid chromatography (RP-HPLC) on a semi-preparative Zorbax 300SB-C8 RP-HPLC column (Agilent) connected to an Agilent 1260 Infinity system. IAPP was eluted isocratically with 27% (vol/vol) acetonitrile in water, 0.1% (vol/vol) trifluoroacetic acid at 80°C for 15 min and a flow rate of 4 ml/min. Peptide elution was detected by UV absorption at 214 nm. After each run, the column was washed with 80% (vol/vol) acetonitrile in water, 0.1% (vol/vol) trifluoroacetic acid, at 80°C for 5 min. Fractions containing IAPP were lyophilized and stored at -20°C. To assure a uniform starting material for each experiment, 2 mg of the lyophilized peptide was dissolved in 6 M GdnHCl, 50 mM NaCl, 20 mM sodium phosphate, pH 6.0. The peptide was applied on a Superdex 75 10/300 GL column equilibrated in 20 mM sodium phosphate, pH 6.0. The monomeric fraction was collected and stored at -80°C. Protein fractions from each purification step were analyzed by Tricine-SDS-PAGE using a 16.5% Tris-Tricine peptide gel on a Criterion gel system (Bio-Rad). Coomassie Brilliant Blue R-250 was used to visualize the protein bands. The concentration of IAPP

was determined from UV absorbance at 280 nm, considering a molar extinction coefficient of 1400 M^{-1} cm^{-1}. The purity of recombinant IAPP was verified by reversed phase HPLC.

Amyloid-β 1-42 & α-Synuclein - Expression and purification of Aβ42 & αSyn protein were carried out by Dr. Lothar Gremer. The protocols are highly similar to the one of Amylin and will not be further described. They can be found in [Hoyer et al., 2008] and [Mirecka et al., 2014b] respectively.

Methyl-specific labelling - Cells transformed with the plasmid of choice were progressively adapted to M9/D_2O medium in three stages over 48 h. The medium contained 1 g/L NH_4Cl and 2 g/L d-glucose-d7 (Sigma-Aldrich). Finally, bacteria were grown at 37 °C in M9 medium prepared with 99.85% D_2O (Eurisotop). The following NMR-Bio kits were added one hour prior to induction : DLAM-Me/V_{proS} for Methionine and Valine labelled PhCPN and SLAM-Iδ1 for Isoleucine MSG [Kerfah et al., 2015]. When the O.D. at 600 nm reached 0.8, the protein production was induced by the addition of 1 mM IPTG. Purification protocols were identical to the ones previously described. For NMR, proteins buffers were exchanged for PhCPN

6.3 NMR methods and application to chaperones and their substrates

The interactions between proteins are extensively studied by Nuclear Magnetic Resonance (**NMR**). NMR is, in fact, an *ad-hoc* technique to characterize transient interactions and is logically one of the core disciplines of structural biology. This part of the manuscript will present the principle of the different NMR techniques used to study protein-protein interactions and especially chaperone-substrate interactions. Proteins are highly dynamic entities, constantly going through changes such as structural modifications or interaction with other proteins. Nuclear Magnetic Resonance spectroscopy, or NMR, based on the quantum mechanical properties of atomic nuclear spins, is a tool of choice to study such transformations. In fact, resonance frequencies, also expressed as chemical shifts, are highly sensitive to the changes of the nuclei environment, as long as they are detectable (spin unequal to zero, such as ^{1}H, ^{13}C or ^{15}N). The basic physics lying under the concept of NMR will not be explained further and can be found in reference books, such as [Keeler,

2010, Levitt, 2008]. In the next section, various NMR experiments are introduced, along with their application to the study of protein interactions. Specific labelling techniques are also presented, used to overcome the size limitation of the NMR technique. The presented techniques were either exploited during the thesis or used in the papers presented in **Chap. 2**.

6.3.1 NMR to study proteins and their interactions

Chemical Shift Perturbation & Line shape analysis

The goal of 2D heteronuclear NMR spectroscopy is to correlate the chemical shift of protons to their attached heteronucleus (such as ^{13}C or ^{15}N). The widely used heteronuclear single and multiple quantum coherence experiments, HSQC [Bodenhausen and Ruben, 1980], HMQC [Mueller, 1979] and ^{15}N SOFAST-HMQC [Schanda and Brutscher, 2005] are based on magnetization transfer from protons to their attached heteronuclei, through $^{1}JH_{C/N}$ scalar couplings. These experiments are particularly fit for small to medium-sized proteins (up to 30 kDa). In the case of larger proteins, however, specific protonation of HN or methyl labelled groups in otherwise perdeuterated protein background are needed (see **Section 6.3.2**) and adapted heteronuclear NMR experiments are used, namely methyl TROSY [Ollerenshaw et al., 2003] and ^{15}N TROSY [Pervushin et al., 1997].

The study of chemical shift perturbations (**CSP**) consists in mapping the differences of chemical shifts of a protein in two different states. These chemical shift changes are caused by modifications of the local environment around a given nucleus (^{1}H, ^{13}C or ^{15}N). These structure modifications can arise from protein-protein or protein-ligand interactions as well as conformational changes. Quantifying these CSP in terms of chemical shifts can give access to thermodynamic measurements upon titrations with proteins and ligands, including K_D determination [Waudby et al., 2016, Williamson, 2013] (**figure 6.2, A & B**). Mapping the CSP on the structure of the protein can also help to identify binding interfaces or structural rearrangements caused by binding of a partner protein or a substrate. One can also study the modifications in lineshape of a protein's resonances by measuring the width at middle height of its characteristic peaks, when observable. They are usually caused by a variation of the protein relaxation due to changes in its apparent size and its tumbling rate, upon binding to another protein. Lineshape modifications can also be caused by an exchange between different conformations, occurring in the 10 µs to 100 ms

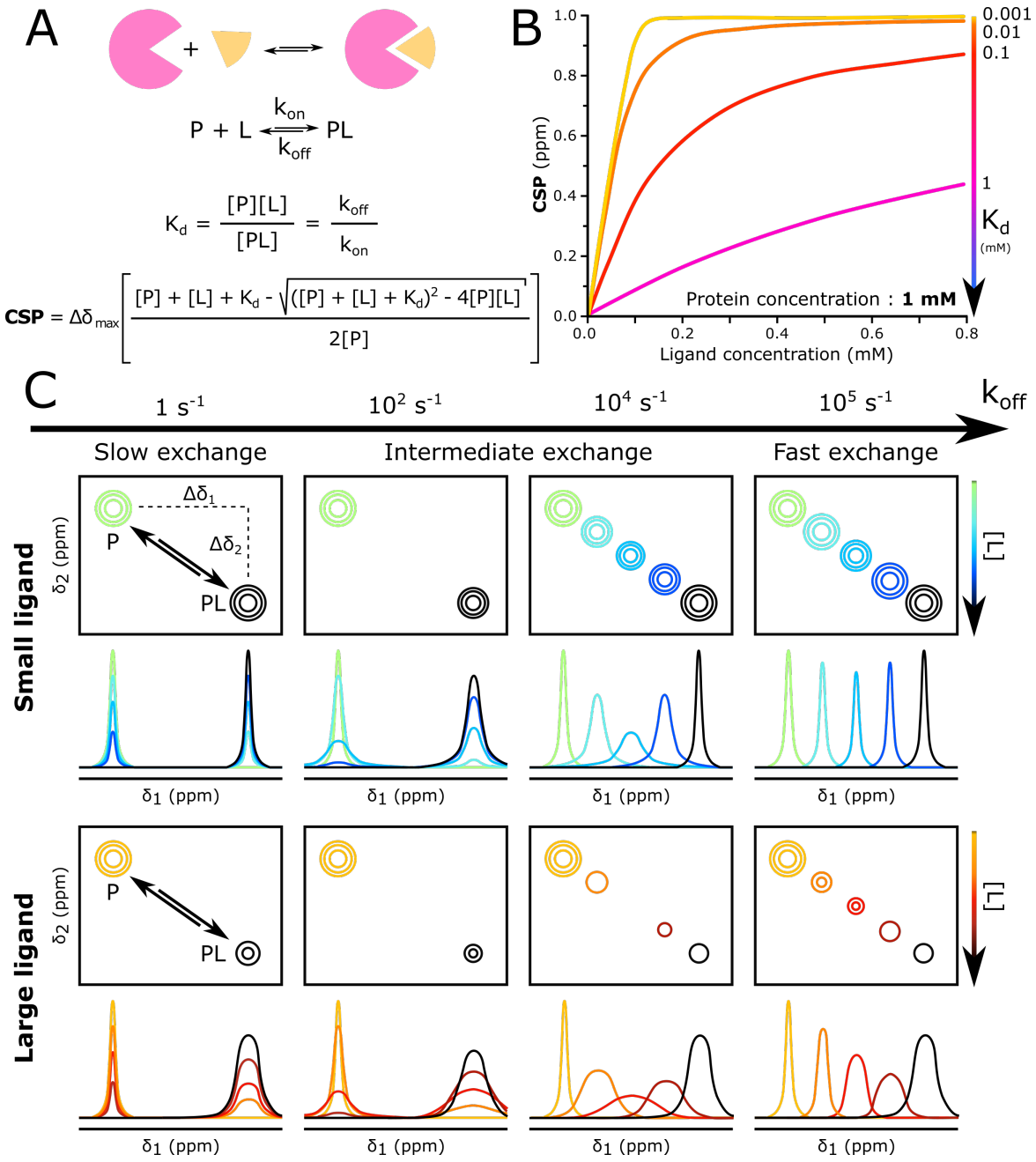


Fig. 6.2 : Chemical shift perturbations. (**A**) Schematic of protein in interaction with a ligand and mathematical relations between protein concentration, CSP, K_D and k_{off}. (**B**) Schematic of K_D values as a function of CSP, modified from [Williamson, 2013]. (**C**) Schematic of 2D NMR spectra as a function of the exchange rate between bound and free protein, k_{off}, in presence of small or large ligand, modified from [Waudby et al., 2016].

timescale. Fits of the peaks lineshape can allow obtaining qualitative information on the kinetic parameters of binding events or conformational changes. For more quantitative observation, $T_{1\rho}$ or CPMG (**next paragraph**) NMR experiments need to be carried out. In the case of chaperones, both CSP and lineshape analysis can be used to map interaction sites on the chaperone or directly on its substrate. It also gives information on the different chaperones states, as most of them are fueled by ATP hydrolysis and exhibit various conformations. Unfortunately, chaperones are often large (more than 30 kDa) and their binding to substrates cause increases in lineshape, coupled with a decrease in intensity (**figure 6.2, C & D**).

Hydrogen-Deuterium exchange

Amino-acids amide protons inter-exchange at a wide range of rates, from ms to days, depending on various parameters, including the solvent accessibility of the residues [Zhang et al., 1995]. In fact, the amide proton of a residue buried in the core of a protein exchanges dramatically slower than the one of an amino acid exposed to the solvent on the surface of the protein. This exchange also depends on the hydrogen-bonding network as well as the secondary structure of the protein, the temperature, the pH [Baxter and Williamson, 1997] and the salt content [Tadeo et al., 2007] of the buffer. These amide exchange rates yield precise information on the solvent-exposed surfaces of a protein and indirectly on its folding state ; as well as protein dynamics (**figure 6.3, A & B**). They also allow to map contact sites in protein-protein complexes, since they are protected from the solvent and their amide exchange rate is decreased [Dyson et al., 2008] (**figure 6.3, C**). Amide exchange rates can be monitored by ^{1}H-^{15}N 2D NMR using hydrogen/deuterium (**H/D**) exchange, by observing the decrease or increase of amide signals consecutive to rapid solvent changes, from H_2O to D_2O or D_2O to H_2O respectively. H/D exchange can be simply monitored by addition of D_2O or H_2O in protonated or deuterated protein samples respectively. For slow exchanging protons, H/D exchange can also be checked by rapid dissolution of a lyophilized deuterated or protonated protein in H_2O or D_2O respectively. These changes can be recorded with fast NMR experiments such as ^{1}H-^{15}N SOFAST HMQC [Schanda et al., 2007, 2005].

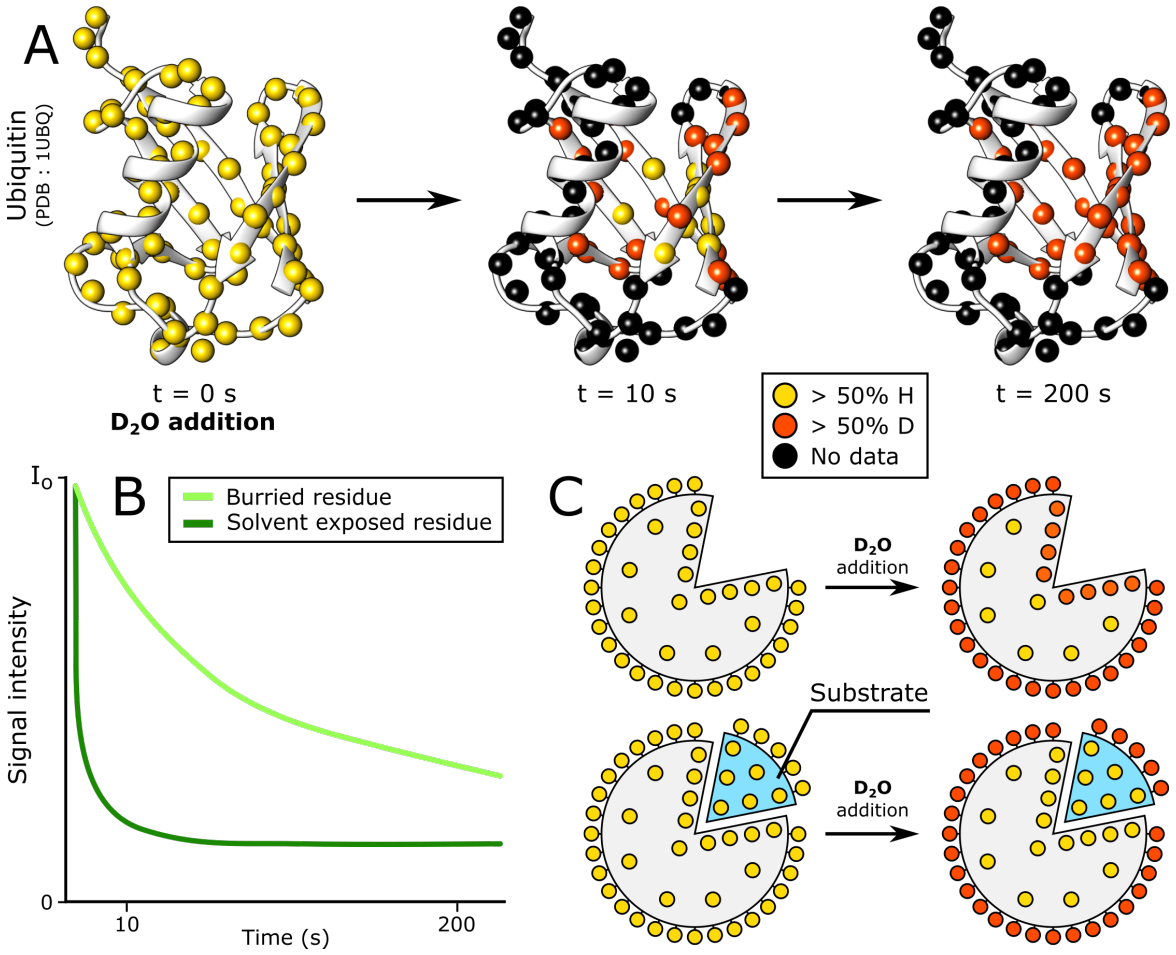


Fig. 6.3 : Hydrogen-deuterium exchange. (**A**) Hydrogen-deuterium is exchange monitored on the Ubiquitin protein, data extracted and adapted from [Schanda et al., 2007]. D_2O is added at T_0 and ^1H-^{15}N SOFAST 2D spectra were sequentially recorded, allowing the monitoring of the signal intensity decrease corresponding to hydrogen-deuterium exchange. No data was recorded for the residues depicted in black, due to fast H/D exchange. (**B**) Schematic of the signal intensity decrease of an amino acid as a function of its solvent accessibility, in the case of hydrogen/deuterium exchange ^1H-^{15}N 2D NMR experiment after addition of D_2O. (**C**) Schematic of the use of hydrogen/deuterium exchange to map interaction sites on a protein complex.

CPMG

The Carr-Purcell-Meiboom-Gill (**CPMG**) experiment allows to measure spin-spin or transverse relaxation dispersion of nuclei [Carr and Purcell, 1954, Meiboom and Gill, 1958]. This NMR experiment allows probing the exchange of a nucleus between two states, with an exchange rate in a µs to ms timescale [Anthis and Clore, 2015]. Particularly, the chemical shift difference between the two states can be measured, the exchange rate between the two states estimated and quantification of the population of the species in exchange performed. The CPMG experiment gives access to the values of R_{2eff} (**figure 6.4, B**), the sum of R_2 (the transverse relaxation rate) and R_{ex} (the contribution due to the exchange process). The CPMG NMR experiment consists in an initial 90° pulse to transfer the magnetisation in the xy transverse plane. A series of a variable number of periodic 180° pulses to refocus the magnetisation is then applied (**figure 6.4, C**). The frequency of these refocusing pulses is termed the CPMG frequency. If the considered nuclei are in an exchange between two different states, their magnetisation will precess around the z axis with different frequencies (**figure 6.4, A**). The magnetization will not be properly refocused by the 180° pulses, thus resulting in an exchange enhancement of the transverse relaxation of the nuclei and line broadening of the recorded signal (**figure 6.4, C, orange**). Still in the case of a nucleus exhibiting two states, the value of R_{2eff} decreases with increasing frequency of the 180° refocusing pulses, to finally be equal to R_2 for frequency values above the exchange speed between the two states. However, if the nucleus exhibits only one population, the magnetization is completely refocused after each 180° pulse, resulting in a final flat R_{2eff} profile, independently of the number of 180° pulses applied and their frequency (**figure 6.4, C, green**).

STD

Saturation transfer difference (**STD**) NMR is a technique used to detect protein-protein or protein-ligand interactions, with a K_D range of 10^{-3} to 10^{-8} M [Mayer and Meyer, 1999]. It also allows to directly map the interaction sites. The initial step of STD consists in recording a first spectrum of the putatively interacting proteins or protein and ligand, with an "off" irradiation frequency, used as a reference (**figure 6.5, A**). Then, a second spectrum is recorded, this time with selective saturation of a spectral region which contains the signals of only one of the proteins (**figure 6.5, B**). If this protein is in interaction with another

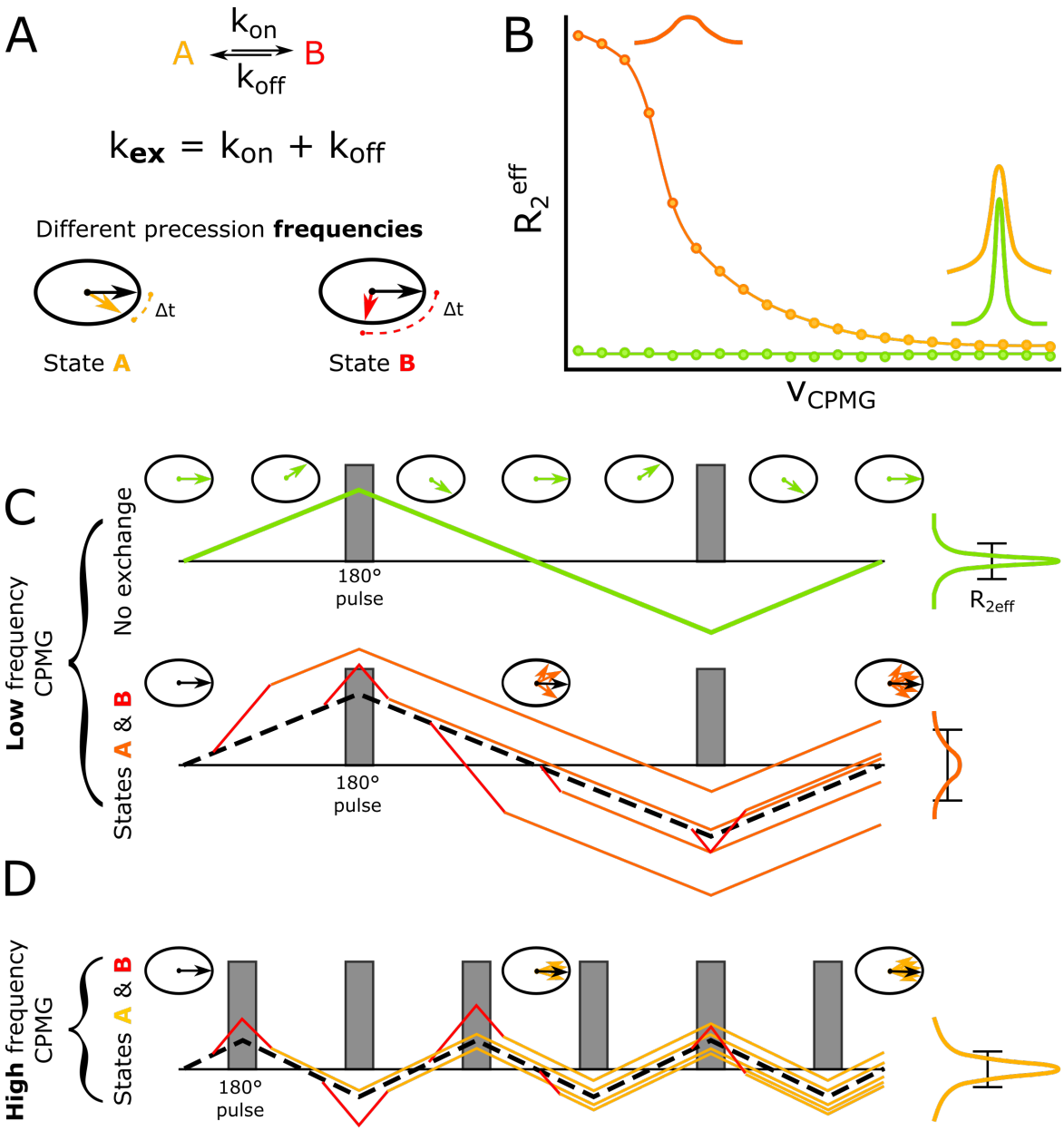


Fig. 6.4 : The CMPG NMR experiment. (**A**) Scheme of a nucleus in an exchange between two states. (**B**) Profiles of a CPMG curve in the cases of large exchange (**orange**) or no exchange (**green**) between two states. (**C**) Scheme of the CPMG refocusing pulses, adapted from [Mittermaier and Kay, 2006]. The first case (**green**) represents the magnetization of a nucleus exhibiting only one state, in the case of low-frequency CPMG. The second case (**orange**) represents the magnetization of a nucleus in exchange between two states of different relaxation properties, in the case of low-frequency CPMG. (**D**) The scheme of the last case represents the magnetization of a nucleus in an exchange between two states of different relaxation properties, in the case of high-frequency CPMG.

protein or a ligand, the saturation will spread to its protons via spin diffusion through intermolecular nOes and its signal will decrease. Subtracting the saturated signal to the reference signal gives the STD spectrum, exhibiting signals of the residues interacting with the saturated protein only (**figure 6.5, C**).

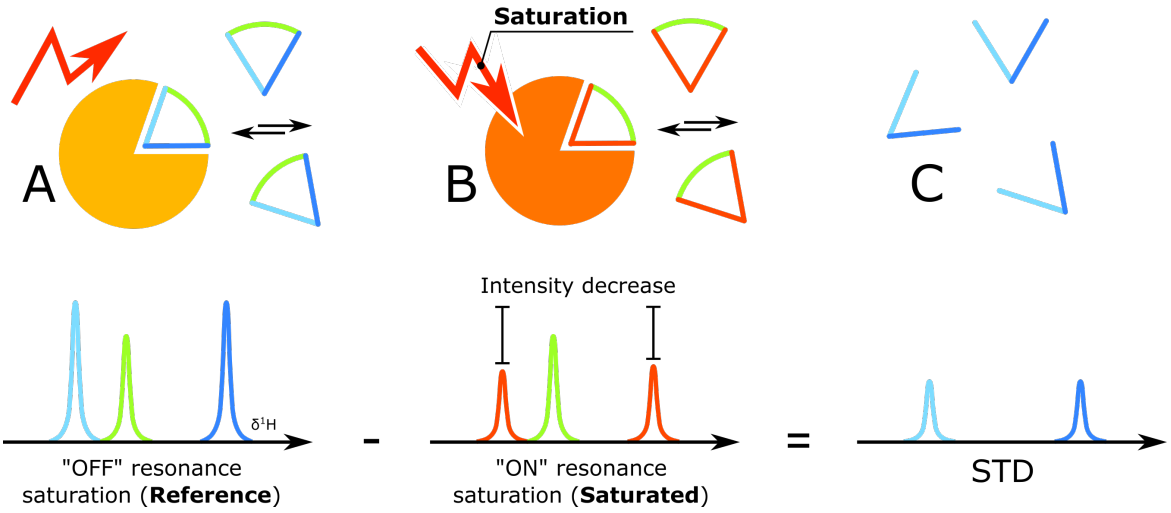


Fig. 6.5 : Saturation transfer difference principle. (**A**) Off-resonance saturation pulse. The signals of the substrates or the interacting protein are not affected. (**B**) On-resonance saturation pulse. The residues close to the interaction site are affected and their intensity decreased. (**C**) Difference of the off and on spectra, the STD experiment.

Transient dark states

As explained in the first section, proteins signal can be broadened beyond detection when they start to interact with a large protein due to slow tumbling or conformational exchange and increased transversal relaxation. In the same manner, a uniformly 15N 13C labelled large protein will be invisible for classical NMR, in a so-called “dark state”. Two specific NMR experiments are useful to investigate these low populated and "invisible" states. Dark state exchange saturation transfer (**DEST**) and chemical exchange saturation transfer (**CEST**) profiles allow computing the exchange rates between each state, as well as transverse relaxation rates (reviewed in [Anthis and Clore, 2015]). DEST is used to gain knowledge on slow tumbling large assemblies or aggregates in an exchange with a smaller and visible state of a lifetime of 500 µs to 1s. This technique is of particular use in the case of two interacting protein of dramatically different sizes in fast exchange. On a

similar fashion, CEST is useful to characterize exchanges with existence of the visible state from 2 to 100 ms, provided that the rotational time of both dark and visible species are comparable and they exhibit different chemical shifts.

PRE

Paramagnetic relaxation enhancement or **PRE** is a technique based on the measurement of NMR signal relaxation differences of nuclear spins in the close vicinity of paramagnetic centers [Battiste and Wagner, 2000]. These centers are either natural or artificially fixed on proteins and usually constituted of chelated lanthanide ions, metal ions or organic radicals. The signal relaxation of atoms in the vicinity of these centers is enhanced, and the corresponding peaks are broadened or completely bleached on the recorded NMR spectrum (**figure 6.6, B**). The measurements of the signal intensity loss or R_2 linewidth changes provides distance constraints in the 10 to 30 Å range, useful for structure calculations as well as interaction sites mapping [Anthis and Clore, 2015, Battiste and Wagner, 2000] (**figure 6.6, A**). PRE mainly allows to map interaction interfaces in protein-protein or protein-ligand complexes, by decreasing the signal of the amino acids in the vicinity of a paramagnetic center, usually covalently linked on the interaction partner or the ligand. The paramagnetic centers can also be attached on a protein to provide distance constraints for structure calculation or map intramolecular structural changes. PRE has been proven particularly useful in the study of large ensembles, such as chaperones, interacting with substrates of a medium to small size. In fact, chemical shifts on the side of the chaperone are usually averaged out due to protein dynamics, while PRE dramatically influences the relaxation of the residues interacting or closeby to the paramagnetic center.

Homonuclear nOe

Homonuclear Nuclear Overhauser effect (**nOe**) measurements allow to obtain information on small $_1$H-$_1$H intermolecular distances (up to 10 Å in case of methyl specifically labelled proteins) from dipolar interactions between adjacent nuclei [Sounier et al., 2007]. They are extensively used in structure calculations and also to determine regions in close proximity due to protein-protein interactions.

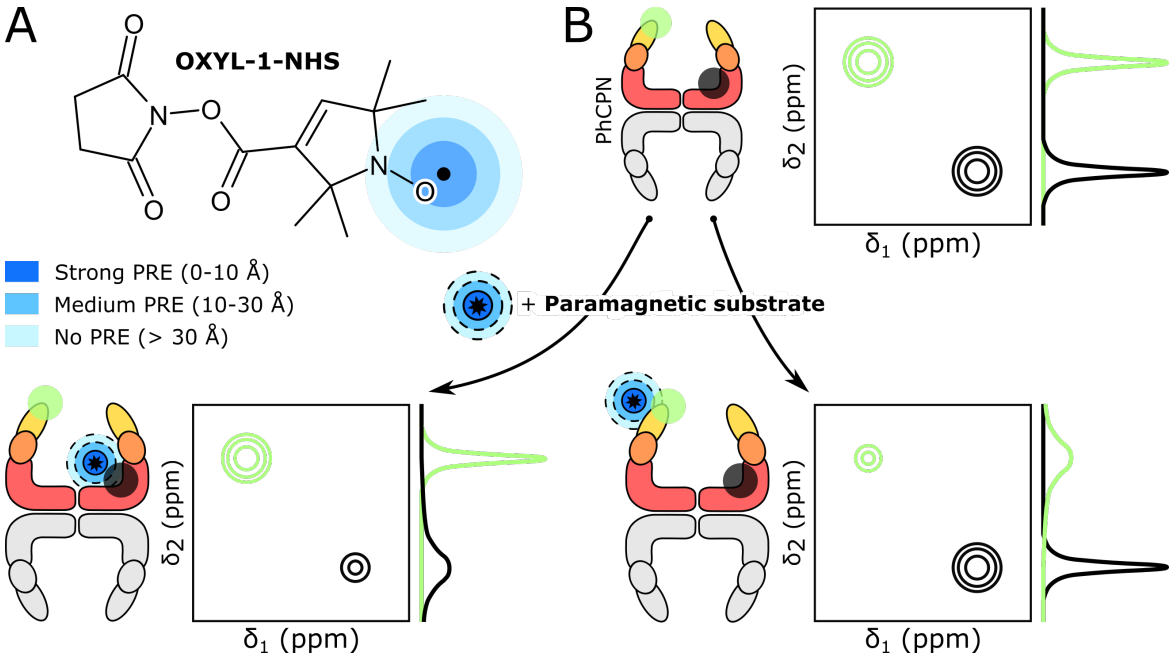


Fig. 6.6 : PRE NMR. (**A**) Molecular sketch of OXYL-1-NHS, a paramagnetic agent which can be used to tag lysine residues. (**B**) Schematic of the addition of a paramagnetically tagged substrate to PhCPN, and its effect on a 2D NMR spectrum as a function of the substrate binding site.

Diffusion DOSY

Diffusion NMR techniques enable to compare the size and shape of proteins in a given media. The translational diffusion of a molecule in solution is directly linked to its hydrodynamic radius and can be measured by applying two consecutive pulsed gradients to the proteins. In details, the Diffusion-ordered spectroscopy (**DOSY**) NMR experiment consists in an initial 90° pulse to transfer the magnetisation in the xy transverse plane [Morris and Johnson, 1992, Stejskal and Tanner, 1965]. The magnetisation is then dephased by using a pulsed gradient according to the z-direction, to "label" proteins spins depending on their position in the sample according to the z-axis (**figure 6.7, A**). After a delay of 50 to 200 ms during which the protein undergoes translational diffusion (**figure 6.7, B**), a second pulsed gradient is applied to rephase the magnetization and recover the spins position indirectly (**figure 6.7, C**). If the proteins are steady in solution during the diffusion delay, the dephasing and rephasing gradients compensate each other and the signal is not impacted. However, if the proteins diffuse translationally during Δt, the magnetisation is not completely rephased after the second gradient and the recorded signal is weakened (**figure 6.7, D**). The same experiment is recorded at different gradient strengths and the

obtained resulting signals attenuation allows to back-calculate the coefficient of diffusion of the proteins by the relation given in (**figure 6.7, D**), proportionally to their size(s).

In the study of protein complexes, such an experiment is particularly useful to determine if two species are interacting strongly together, exhibiting the same diffusion. This techniques also allows to monitor the oligomeric state of a protein, as well as the sample's integrity, to detect degradation, resulting in modifications of apparent protein size.

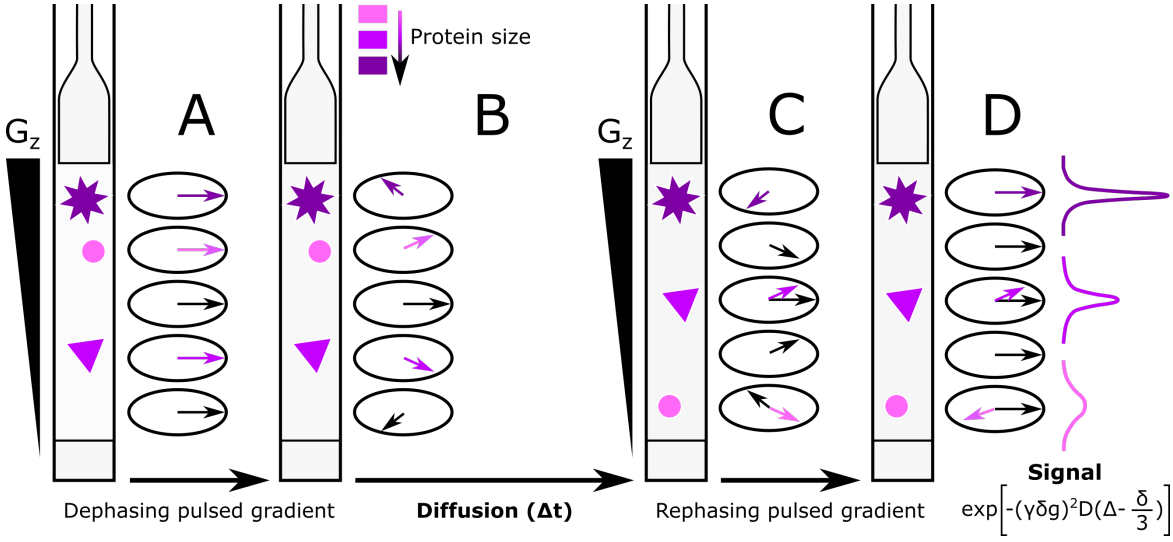


Fig. 6.7 : The DOSY NMR experiment. (**A**) The magnetization is dephased by a pulsed gradient according to z. (**B**) Delay during which the proteins diffuse in solution. (**C**) The magnetization is refocused by a second pulsed gradient. (**D**) The signal is recorded and depends on the proteins diffusion. Its intensity is proportional to the diffusion coefficient D, the diffusion delay Δt, the gradient length δ and strength g, and γ the gyromagnetic ratio of the considered spins.

MAS-NMR & DNP MAS

Solid-state NMR spectroscopy (**SS-NMR**) under magic angle spinning (**MAS**) is a technique allowing the averaging of anisotropic interactions due to low or absent molecular motions. In solution, rapid isotropic tumbling averages the anisotropic interactions spatial component to zero while this component remains in solids, broadening the recorded signal. To remediate this problem, the sample can be either artificially oriented by addition of an alignment media (static solid state) or spun in a rotor inclined at an angle of 54.74° relative to the external magnetic field (the magic angle). This rotation allows to remove the orientation dependence of the molecules and only keep the isotropic contribution, narrowing the peaks and allowing the recording of well-resolved multi-dimensional spectra. The

sensitivity and resolution of the recorded spectra are enhanced with the rotation speed (up to 100 kHz) but also requires special equipment. This technique is particularly useful in the study of large protein systems with low molecular tumbling, such as fibrillar or amorphous aggregates, membrane-associated proteins or simple large protein complexes. It provides information on the molecular level, of structural and dynamic information. In fact, even if the samples are sedimented in rotors in a pseudo-solid state by ultracentrifugation, they remain hydrated, which allows the measurement of molecular motions. MAS NMR requires isotopic labelling and is compatible with methyl-specific labelling.

MAS NMR can be supplemented with dynamic nuclear polarization (**DNP**), providing increased sensitivity and allowing to work at natural isotopic abundance. Such a technique is based on the transfer of the inherently much higher polarization of electron spins to nuclear spins by microwave irradiation. By doing so, DNP increases the sensitivity of MAS NMR by several orders of magnitude and allows to reduce drastically the experimental time compared to classical MAS NMR. The bases of the technique, as well as its whole range of application, are reviewed in [Lilly Thankamony et al., 2017]. Nevertheless, working with DNP requires some special experimental conditions as well as adapted equipment, notably a microwave source. The samples need to be complemented with paramagnetic centers, used as polarization sources. Moreover, DNP currently requires to work at really low temperatures (around 100 K), thus restraining the native dynamics of the proteins.

6.3.2 Specific labelling to study large assemblies

The previously depicted NMR experiments are well fitted for small or medium size proteins, up to 30 kDa. Nevertheless, in the case of larger proteins, several problems appear and need to be overcome. A first problem is the crowding of the NMR spectrum. For instance, a 2D heteronuclear spectrum of a protein of more than 100 AA is already quite crowded and the complexity increases with the length of the protein. In the case of globular proteins which can reach 80 kDa, it becomes completely inextricable, with lots of spectral overlaps. A second issue which needs to be addressed is the decrease in molecular tumbling caused by the protein size. Tumbling being directly connected to transversal relaxation, the relaxation speed will increase and the peaks will be broadened, beyond detection in the case of large macromolecular assemblies such as the PhCPN. An elegant answer to these problems is the use of isotope labelling strategies.

Deuteration

Spectral linewidth broadening is mainly due to a large amount of proton dipolar interaction in proteins. Protein labelling using perdeuteration and selective reintroduction of the labile protons on NH after expression is a way to get rid of most of these interactions. In fact, the protons quantity in the proteins are reduced by 80% and proton-proton interactions are mostly replaced by lower magnitude proton-deuterium interactions resulting in slower transversal relaxation rates (narrower linewidth). The spectra obtained with NMR TROSY experiments are less crowded and allow assignment of proteins up to 100 kDa.

Selective methyl labelling

Methyl group bearing amino acids, Ala, Ile, Met, Val, Leu and Thr, represent on average approximately 30% of the total AA content of a protein and particularly present in hydrophobic regions (**figure 6.8, A**). These methyl group located at the end of the side of the amino acids chains show increased flexibility, except for Ala, physically close to the backbone. In addition, the methyl groups have very fast rotation around the C-C axis freely, which had a favorable effect on transversal relaxation, even when part of a large protein. Their relaxation rates decrease, which makes them observable by NMR, even in the conditions exposed previously (**figure 6.8, B & D**). A selective protonation of methyl groups combined with perdeuteration to remove dipolar interactions and TROSY NMR to preserve favorable relaxation of methyl coherence due to dipole-dipole interference involving 4 spins in isolated methyl groups (**figure 6.8, C**) yield NMR spectra with good resolution and intensity (**figure 6.8, D**). The selective methyl group protonation of perdeuterated proteins was first started on Val, Leu and Ile by the group of Pr. Lewis Kay [Gardner and Kay, 1997] and is now extended to all types of methyl groups, with possibilities of individual or combined labelling to reduce signal overlap (reviewed in [Kerfah et al., 2015]).

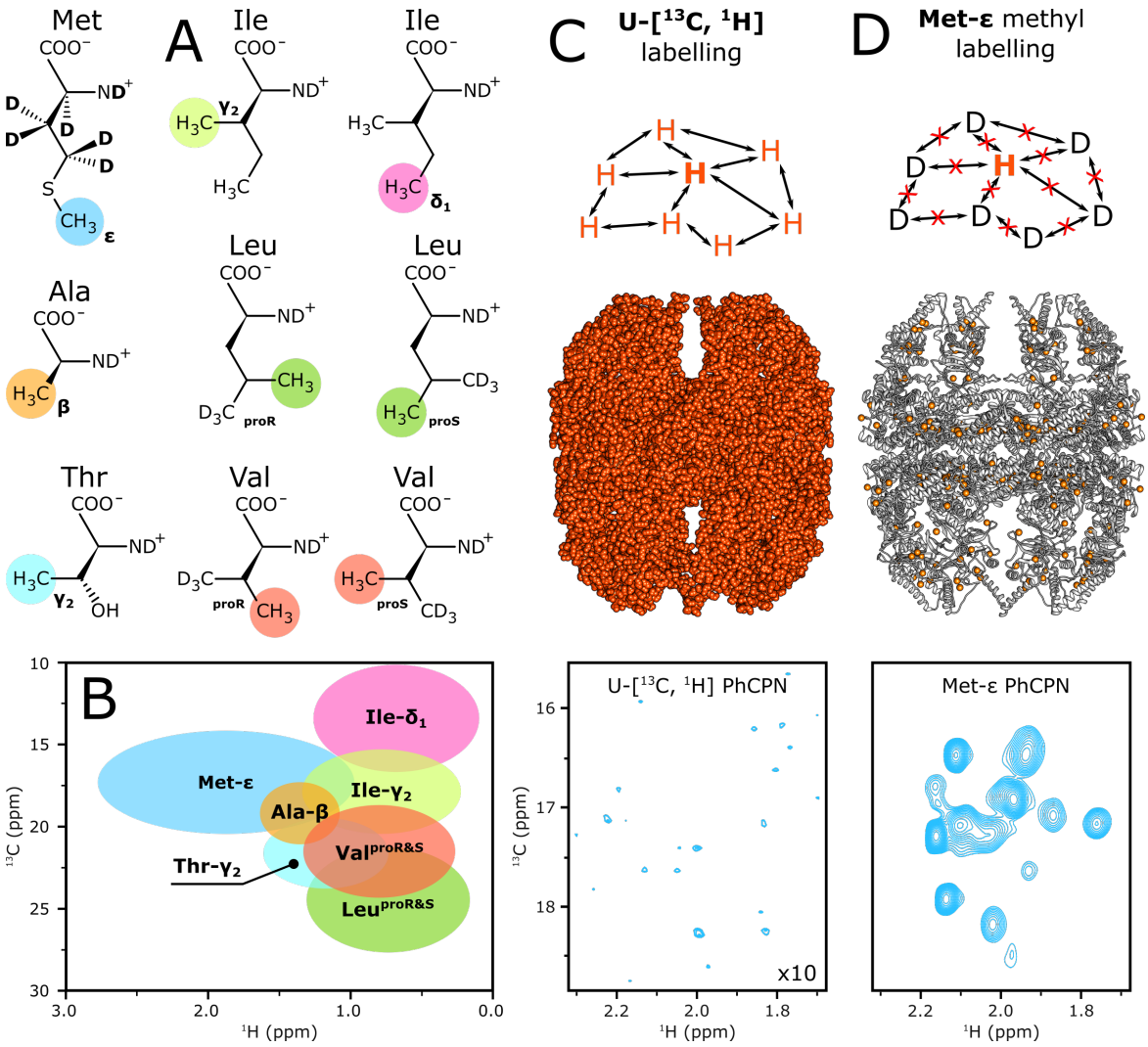


Fig. 6.8 : Specific methyl labelling for NMR. (**A**) Amino-acids which can labelled and their methyl groups. (**B**) The colored circles correspond to the expected location of methyl resonances for the considered methyl groups. (**C**) Uniform labelling of 15 µM PhCPN and the resulting 2D SOFAST methyl HMQC spectrum, lacking signals. (**D**) Met-ϵ labelling of 15 µM PhCPN (approximately 1 MDa) and the resulting 2D SOFAST methyl HMQC spectrum with well-resolved peaks.

Methyl-labelling combined with adequate NMR pulse sequences provides information on the global fold of large proteins up to more than 1 MDa (**figure 6.8, D**), map substrates binding and determine interaction kinetics. It is therefore particularly suitable for the study of chaperones.

6.4 NMR experimental parameters

NMR Samples

Samples were transferred in 3 mm Shigemi tubes (ShigemiNMR) made of glass matching the magnetic susceptibility of D_2O using an eVol syringe (SGE Analytical Science), closed with appropriate plungers and sealed with parafilm.

DOSY

Temperature-dependent diffusion coefficients of [U-D], (Met-ϵ)-$^{13}CH_3$-labelled PhPCN (black open circles) and [U-D], (Ileδ1,Val^{proS})-$^{13}CH_3$-labelled MSG with non-isotope labelled PhPCN were measured by DOSY (Diffusion Ordered Spectroscopy, see **Section 6.3.1**). The diffusion values were obtained from fitting of 1D DOSY signal decay profiles of the labelled methyl regions. All measurements were performed on a Bruker Advance 850 MHz spectrometer, equipped with a 5 mm cryoprobe. The standard Bruker pulse sequence dosy_cdec was applied to record 1D ^1H diffusion experiments with ^{13}C filter and decoupling. The gradient magnitude was ranging from 0.963 to 47.187 G/cm (16 points) and the diffusion delay set to 100 ms. For each of 16 gradient magnitudes, 256 transients of 1438 complex data points were acquired. The diffusion coefficients were obtained by fitting the experimental curves using non-linear least square method and the equation from **figure 6.7, D**. To check the integrity of the samples at each temperature, 2D methyl SOFAST HMQC experiments were interlaced with the 1D DOSY measurements.

PRE Samples

Random spin labelling of the ϵ-amino groups of the MSG lysine residues with OXYL-1-NHS (1-oxyl-2,2,5,5-tetramethylpyrroline-3-carboxylate N-hydroxysuccinimide ester ; Toronto Research Chemicals), was carried out following a published protocol [Hartl et al., 2012]. MSG was solubilized in 10 mM Na_2CO_3 pH 9.2 buffer at a concentration of 40 µM. A

stock solution was prepared by dissolving 10 mg of OXYL-1-NHS in 100 µL of DMSO. After addition of 6-fold molar excess of OXYL-1-NHS over lysine residues to the protein solution, the reaction mixture was incubated for 1h at room temperature followed by 4 h at 4°C. The excess of spin label was removed by washing the sample with approximately 20 volumes of NMR-buffer using centrifugal filter devices (Vivaspin 15, 5000 MWCO, Vivascience). The average number of spin-labelled lysines was determined by mass spectrometry.

PRE NMR

Paramagnetic relaxation enhancement values were measured, providing qualitative information on the interactions between the MSG and PhCPN by comparing peak intensities (theory described in **Chap. 6.3.1**). 2D SOFAST-methyl-TROSY NMR experiments were recorded at 75°C with 16 µM PhCPN (256 µM monomer) U-[^{2}H] Met-[$^{13}CH_3$], Val-[13CH3]ProS encapsulating either 32 µM (OXYL-1-NHS)-labelled $_{1}$H MSG (I_{ox}, oxidized) or $_{1}$H MSG (I_{red}, reduced). The I_{red} sample spectrum was recorded after reducing the spin label of the encapsulation sample by adding Na-ascorbate to a final concentration of 2.2 mM and incubate for 24 h at 4°C.

Titration by PhCPN

Unlabelled PhCPN was prepared according to the previous protocol. PhCPN was concentrated to a final concentration of 20 µM and diluted depending on the final ratio Amylin/PhCPN expected. Once prepared, samples were kept on ice before measurements to avoid fibrillation. For each NMR measurement, samples were prepared with a constant concentration of 50 µM $_{15}$N-labelled Amylin and one sample was prepared for each Amylin/PhCPN ratio. BEST-TROSY spectra were recorded at 10°C on an 850 MHz Bruker spectrometer equipped with a 5mm CryoProbe. NMR spectra were processed using NMRPipe and analyzed with CcpNMR.

6.5 Additional biophysical and biochemical techniques

6.5.1 Monitoring amorphous and amyloid aggregation

Light scattering (amorphous aggregates)

The light scattering technique consists in measuring the amount of aggregates present in solution by Rayleigh scattering (**figure 6.9, A & B**). It depends on the degree of aggregation, the protein or protein assembly molecular weight, temperature and protein concentration. At a fixed temperature, aggregation tends to increase with protein concentration. The scattering angle is kept constant and for this reason, static light scattering does not give information on the size distribution of the particles in solution, unlike dynamic light scattering. In this case, the scattering is measured with a fluorometer at a 90° angle from the incident laser beamline. Several modes of measurements are routinely used with such a machine including determination of the aggregation temperature (dependant on concentration, (**figure 6.9, C**)) and evaluation of the aggregation kinetics at a given temperature (also dependant on protein concentration, (**figure 6.9, D**)).

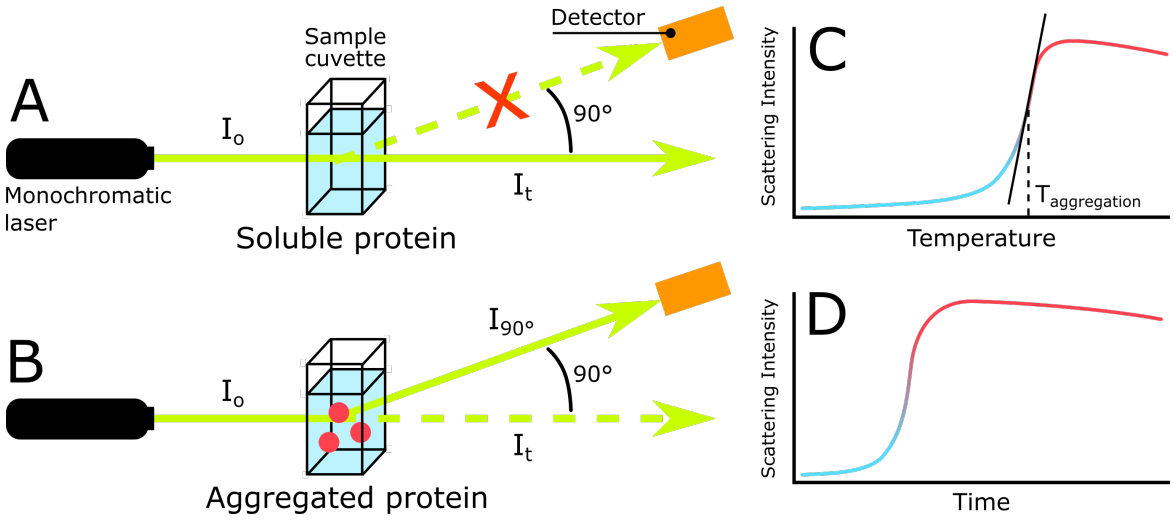


Fig. 6.9 : The principle of light scattering. (**A**) A monochromatic laser goes through the sample cuvette and when proteins remain in monomeric state, light scattering is negligible and no signal is detected at 90° from the incident direction of the light. (**B**) In presence of aggregates in the cuvette, the incident laser light is scattered by the floating particles and signal can be detected at 90°. (**C**) Schematic scattering as a function of temperature. (**D**) Schematic scattering as a function of time, at a fixed temperature.

In this study, aggregation kinetics were monitored by light scattering at 500 nm by a Cary Varian Eclipse fluorometer using a 400 µL quartz cuvette with a path length of 1 cm. Excitation and emission wavelengths were set at 500 nm with 5 and 2.5 slit widths each [Dahiya and Chaudhuri, 2014]. Measurements were done in triplicate, the same day and in the same cuvette to avoid bias and lamp fluctuations.

ThT assay (amyloid aggregates)

The presence of amyloid fibrils can be highlighted by different techniques, including the Thioflavin T (**ThT**) fluorescence assay. ThT was first used by Vassar and Culling in 1959 [Vassar and Culling, 1959] to stain amyloid fibrils in histological samples and showed that ThT fluorescence was colocalized with amyloid deposits. It was then tested to stain fibrils in vitro in 1989 [Naiki et al., 1989] and nowadays ThT fluorescence assays are used by many to monitor real-time fibrillation.

Thioflavin T is an organic dye, composed of both a polar benzothiazole head and an hydrophobic dimethylamino group, able to bind amyloid fibrils [LeVine, 1993, Pillay and Govender, 2013] (**figure 6.10**). ThT is believed to bind preferentially to the amyloid fibrils while being oriented parallel to the fibril axis, inserted between the protofilaments constituting the fibrils [Groenning et al., 2007, Krebs et al., 2005, Lindberg et al., 2017]. The ThT has a maximal excitation and emission wavelengths of 385 nm and 445 nm respectively. When bound to fibrils, its fluorescence is enhanced and these maxima change to respectively 450 and 485 nm, enabling to follow the progression of fibrillation. ThT fluorescence can be explained by the "molecular rotor" behaviour of the ThT molecules [Krebs et al., 2005, Stsiapura et al., 2010, Sulatskaya et al., 2017]. In fact, in water and in absence of β-sheet rich fibrils, the excitation energy provided to free ThT is believed to be dissipated through rotational motion of the benzothiazole relative to the aminobenzene rings, therefore the absence of fluorescence. However, when ThT is bound to fibrils, this rotational motion is restrained and the energy cannot be dissipated by rotational motion anymore, resulting in a preserved excited state and an enhanced fluorescence.

Some of the interesting advantages of the ThT dye is that it does not alter the structure of the forming fibrils [Lindberg et al., 2017], nor it does significantly affect the fibrillation kinetics at working concentrations (around 30 µM, [Groenning et al., 2007, Xue et al., 2017]. Additionally, ThT fluorescence intensity is dependant on the fibrils conformations and

can thus report structural changes of the fibrils at the atomistic level [Lindberg et al., 2015]. However, ThT can also bind non-fibrillar structures and especially hydrophobic pockets [Biancalana and Koide, 2010], hence the need for proper blanks measurements prior to fibrillation assays. Several protocols are available in literature and describe precisely how to prepare, record and process ThT assays data in 96-well plates [Gade Malmos et al., 2017, Meisl et al., 2016, Wördehoff and Hoyer, 2018]. ThT fibrillation experiments are challenging because of the poor reproducibility of the technique, partly because of the stochastic effect of nucleation (see **Chap. 5.2.3**). To prevent it, glass beads can be added to increase reproducibility and the plates can also be orbitally shaken directly in the plate reader prior to fluorescence measurements ([Wördehoff and Hoyer, 2018] ; see **Chap. 5.2.3**).

The 96-well plate was prepared on a plastic freezer block to avoid fibrillation and evaporation prior to the beginning of the experiment. ThT was used to monitor the

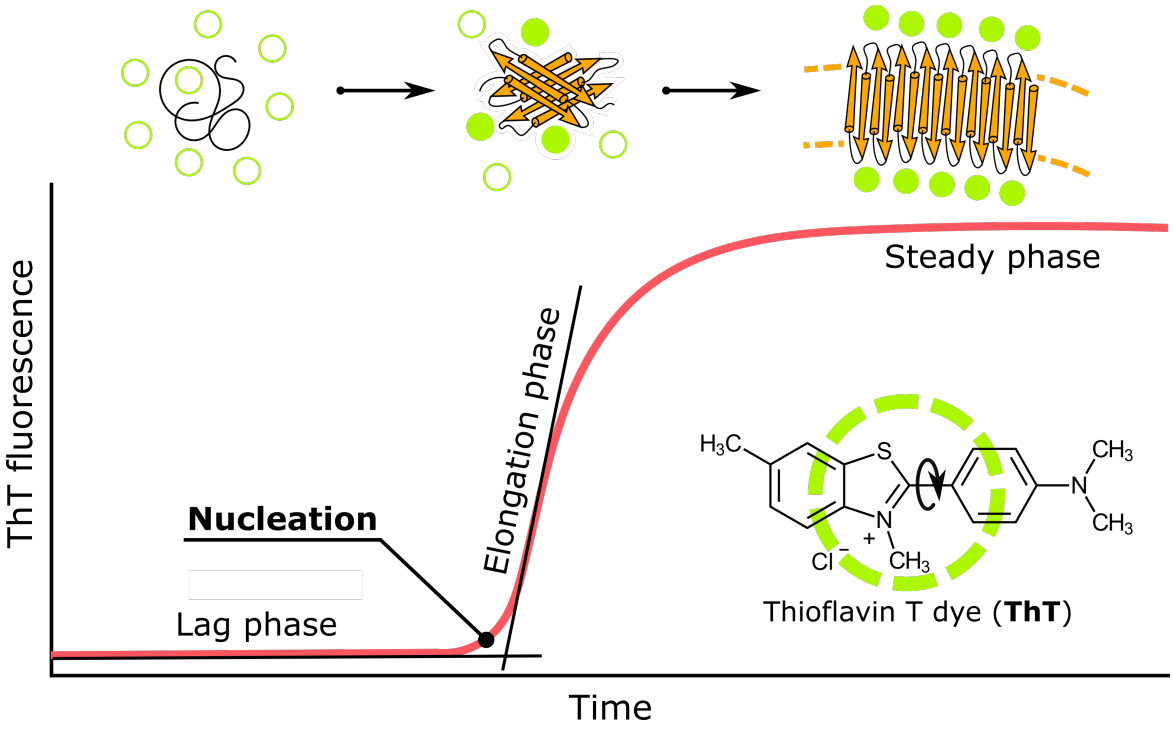


Fig. 6.10 : Scheme of the ThT molecule and the ThT fluorescence assay. During the lag phase, no fibrils nor β-sheets are present, thus the absence of fluorescence. Fluorescence starts to increase after the fibrillation initiation and reaches a maximum when all the dye is bound to the fibrils or the fibrils stop growing due to lack of available monomer to increase the length of the fibrils and therefore the overall amount of β-sheets in solution.

formation of the fibrils in each well, at a fixed concentration of 30 μM. Measurements were replicated and corrected with corresponding blanks. Fluorescence was followed with a Synergy H4 (BioTek) plate reader, using a black 96-wells plate with μ-clear bottom (Greiner Bio-One), sealed with non-porous adhesive lid to avoid evaporation. The temperature was fixed at 30°C and excitation and emission were respectively set at 440 nm and 487 nm, with monochromator slit widths of 9 nm. The plate was shaken using low orbital shaking for a 3 seconds before each measurement to homogenize the samples. The Amylin concentration was fixed to 6.5 μM for each well and measurements were taken every 8 minutes.

6.5.2 Imaging techniques

Negative staining electron microscopy

Samples were absorbed to the clean side of a carbon film on mica, stained with a solution of 2% ammonium molybdate and transferred to a 400-mesh copper grid. The images were taken under low dose conditions (<10 e-/Å2) at a magnification of x23000 and x49000 times with defocus values between 1.2 and 2.5 μm on a Tecnai 12 LaB_6 electron microscope at 120 kV accelerating voltage using CCD Camera Gatan Orius 1000. Measurements on EM pictures were taken with the Gwyddion software.

Atomic force microscopy

5 μl of the sample were adsorbed for 10 minutes to a freshly cleaved muscovite mica surface in presence of humid atmosphere to avoid drying, followed by washing with milliQ water (100 μL, 3 times) and drying with N_2 gas. Imaging was performed in intermittent contact mode (AC mode) in a JPK Nano Wizard 3 atomic force microscope using a silicon cantilever with silicon tip (OMCL-AC160TS-R3, Olympus) with a typical tip radius of 9 ± 2 nm, a force constant of 26 N/m and resonance frequency around 300 kHz. The images were processed using JPK DP Data Processing Software and Gwyddion.

6.5.3 Other techniques

Analytical gel filtration

Prior to analysis, samples were spun down at 16000g to remove eventual aggregated proteins. 15 µL of each sample were injected on the 10 µL loop of a Bio SEC-5 HPLC size exclusion column (500 Å pores, Agilent) installed on a ÄKTApurifier purification system (GE healthcare). Samples were eluted at a flow rate of 0.2 mL/min and absorbance was measured at 280 nm.

SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gels (**SDS-PAGE**) were manually cast. The stacking region of the gels was approximately 1 cm high and its composition is reported figure X. The running region was containing 15% Bis-Tris and its composition is reported in (**figure 6.11**). 2 µL of Unstained PageRuler protein ladder (Ficher Scientific) was used as a migration standard on each gel. SDS-PAGE gels were run at 180 V for approximately 45 min in a buffer composed of 25 mM Tris pH 8.2, 0.1% SDS and 192 mM Glycine. The gels were stained with Coomassie blue stain and destained with a solution containing acetic acid and ethanol.

Composition	**Stacking (5%)**	**Running (15%)**
Water	4.4 mL	4.125 mL
2M Tris pH 8.8	-	2.05 mL
2M Tris pH 8.8	0.73 mL	-
10% SDS	55 µL	110 µL
50% Glycerol	-	143 µL
40% Acrylamide (29:1)	826 µL	4.125 mL
TEMED	10.3 µL	11 µL
10% APS	30 µL	80 µL
Total volume	6.05 mL	11 mL

Fig. 6.11 : Composition of the stacking and running portions of homemade SDS-PAGE gels. Percentages are expressed in (w/v).

Amino-acid analysis

First, L-Norleucine (Sigma) control samples were prepared in order to ensure no sample loss. Protein sample of known O.D. was weighed to obtain a molar quantity of between 500 and 6000 pmoles of each amino acid. It was then inserted into an hydrolysis tube previously pyrolysed and complemented with 5 to 6 nmoles L-Norleucine (internal reference, (**figure 6.12, A**)). The sample was vacuum dried, and 200 µL of a mixture of 1 mL HCl 6 M (Pierce) and 10 µL liquefied Phenol (Aldrich) were added to the tube. The tube was closed, submitted to vacuum and purged with Argon to remove traces of oxygen. The tube was then heated for 24 hours at 110°C in a PicoTag (Waters) hydrolysis oven, cooled down, opened and vacuum dried. The hydrolysed sample (**figure 6.12, A**) was dissolved in Sodium Citrate Loading buffer (Biochrom) and centrifuged for 30 min to remove any particles in suspension. 145 µL of the samples were introduced into a glass tube (Chromacol 021-NV), itself inserted in another glass flask closed with a screwed rubber lid (Chromacol 2SV P829). The sample was lastly inserted in a Midas sample passer, separated and analysed with a Biochrom 30 amino acid analyzer (**figure 6.12, B**) and data were treated with the EZChrom Elite v.3.1.4 software (**figure 6.12, C**).

Sec-MALLS

Size Exclusion Chromatography (**SEC**) was performed with a Superdex 200 column (GE Healthcare) equilibrated with PhCPN buffer prior to the experiment. For accurate size determination, the column was calibrated with standard proteins of known Stokes radius. Runs were carried out at 20°C with a flow rate of 0.5 mL/min. An online DAWN-HELEOS II (Wyatt Technology Corp.) detector equipped with a laser emitting at 690 nm was used for Multiple Angle Laser Light Scattering (**MALLS**) detection. Protein concentration was measured by differential refractive index measurements using an Optilab T-rEX detector (Wyatt Technology Corp.) settled with a refractive index increment of 0,185 mL/g (dn/dc). Molecular weights were then calculated by the ASTRA software (Wyatt Technology Corp.).

Mass spectrometry (Electrospray)

4 µL of 100 µM Aβ42 in 50 mM TRIS-HCl pH 7.5 buffer were needed for analysis. For spin labelling analysis, 4 µl of each sample (5 µM dilution) were needed. All solvents used were

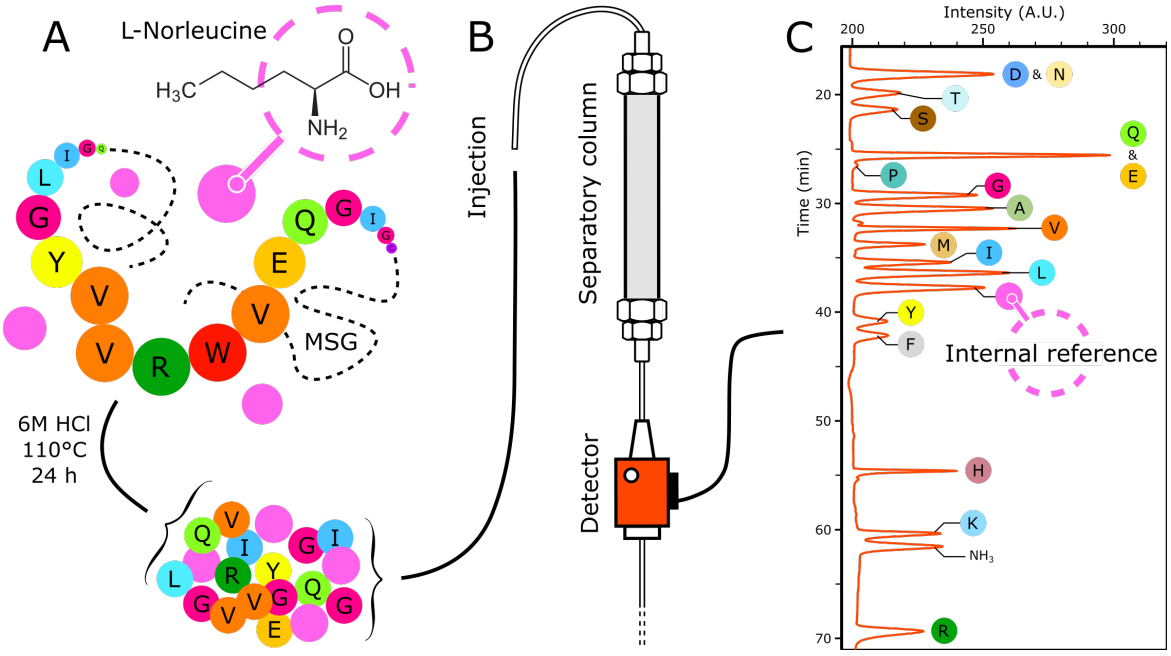


Fig. 6.12 : The principle of amino acid analysis. (**A**) MSG is hydrolysed back into single amino acids. A given amount of L-Norleucine, a non-natural synthetic amino acid, is added to MSG prior to hydrolysis as an internal standard to quantify the amino acids. (**B**) The amino acid mixture is injected on a separative gel filtration column mounted on an HPLC purification system. (**C**) Spectrogram obtained after analysis of an MSG sample, containing an L-Norleucine peak.

HPLC grade. Solvent A was composed of 0.03% TFA, 5% ACN ; and solvent B of 95% ACN , 5% H_2O and 0.03% TFA. The samples were injected on an on-line desalting column and protein trap (Zorbax 300SB-C8, 5 µm, 5x0.3 mm, Agilent Technologies), running 3 min at 100 µl/min with 100% of solvent A, then eluted at 50 µµl/min with 70% solvent B. The samples were analyzed on a 6210 LC-TOF mass spectrometer interfaced with LC pump system (Agilent Technologies).

Mass spectrometry (Matrix Assisted Laser Desorption Ionisation)

100 µM ^{15}N-Amylin in PhCPN buffer were diluted 1:10 in α-Cyano-4-hydroxycinnamic acid matrix and 2 µL were deposited directly on the target. The sample was analyzed on a MALDI TOF MS (Autoflex, Bruker Daltonics), operated in reflectron or linear positive mode.

ITC

The isothemal calorimetry (**ITC**) binding assays were performed with a MicroCal iTC200 (GE healthcare). A solution of 10 µM Amylin in PhCPN buffer was prepared and 400 µL were placed in the sample cell, the reference cell was filled with PhCPN buffer and the syringe with 12 µM PhCPN in PhCPN buffer. Titrations were carried out at 10 and 30°C by progressive additions of 5 µL PhCPN in the reaction cell and variable waiting time between each addition. The inverse titration was also carried out, with 100 µM Amylin in the syringe and 5 µM PhCPN in the sample cell.

6.6 Images

Protein graphics

Molecular graphics and analyses were performed with the UCSF Chimera package [Pettersen et al., 2004]. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

Other illustrations

Images and schematics were made using Inkscape and Adobe Photoshop software.

6.7 Physical and chemical theoretical parameters

General parameters

Extinction coefficient was computed with the ProtParam tool (https://web.expasy.org/protparam/). The other computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition and grand average of hydropathicity. Extinction coefficients are expressed in $M^{-1}cm^{-1}$, at 280 nm measured in water.

Hydrophobicity plots

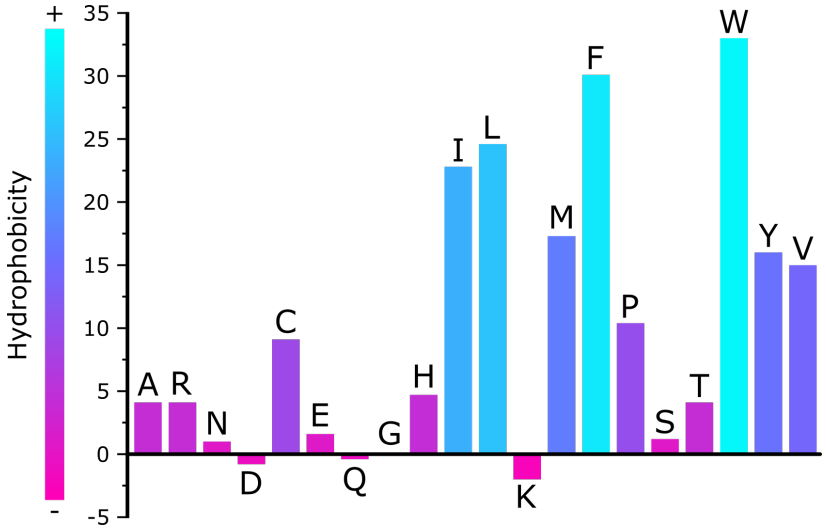


Fig. 6.13 : Hydrophobicity scale of Kovacs [Kovacs et al., 2006].

Many hydrophobicity scales are available, taking into account different biophysical parameters of the amino acids. In this manuscript, hydrophobicity was plotted according to the hydrophobicity scale of Kovacs (**figure 6.13**) [Kovacs et al., 2006]. This scale is particularly relevant to characterize the hydrophobicity of amino acids included relatively short to medium size peptides chains as Amylin and Aβ42.

Swiss Homology model of PhCPN

Target template alignment using the structure of the Thermosome from *Methanosarcina mazei* (MmCPN, PDB : 3IZH) as a template for the open-conformation of PhCPN (see **figure 6.14, A**). Predictions were computed by the Swiss website (https://swissmodel.expasy.org/) [Benkert et al., 2011, Bertoni et al., 2017, Guex et al., 2009]. More detailed calculation methods are available on the cited website. Sequence identity is of 67.1%, providing a good model with high nativeness fidelity scores (**figure 6.14, B**), with a GMQE of 0.81 : "GMQE (Global Model Quality Estimation) is a quality estimation which combines properties from the target–template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target. Higher numbers indicate higher reliability (**figure 6.14, C**)". The global and per-residue model quality has been calculated using the QMEAN scoring function [Benkert et al., 2011].

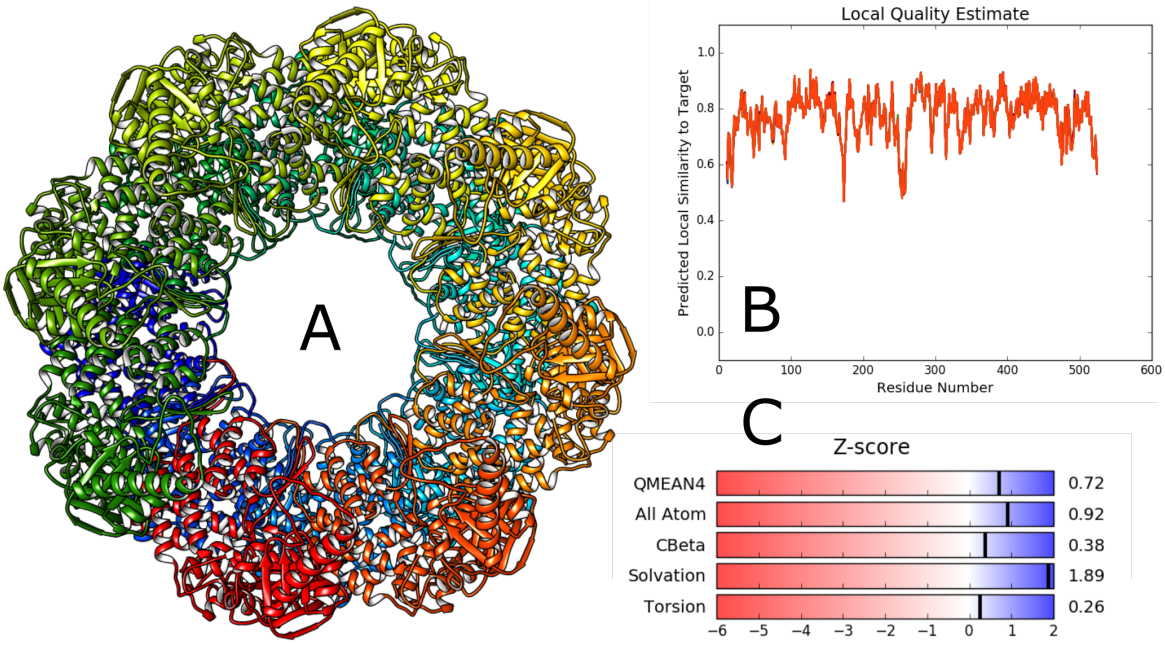


Fig. 6.14 : Homology model of PhCPN. (**A**) Homology structure of PhCPN in the open conformation, obtained from the template structure of MmCPN, registered under PDB number 3IZH. (**B**) Local estimation of the quality of the provided structure, provided by the swissProt website. (**C**) Z-score of the calculated structure. For detailed information about structure calculations, see https://swissmodel.expasy.org/.

References

Ahmad, A., Bhattacharya, A., McDonald, R. A., Cordes, M., Ellington, B., Bertelsen, E. B., and Zuiderweg, E. R. P. (2011). Heat shock protein 70 kDa chaperone/DnaJ cochaperone complex employs an unusual dynamic interface. *Proceedings of the National Academy of Sciences*, 108(47):18966–18971.

Aiello Bowles, E. J., Walker, R. L., Anderson, M. L., Dublin, S., Crane, P. K., and Larson, E. B. (2017). Risk of Alzheimer's disease or dementia following a cancer diagnosis. *PLoS ONE*, 12(6):1–11.

Ananthan, J., Goldberg, A. L., and Voellmy, R. (1986). Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes. *Science (New York, N.Y.)*, 232(4749):522–4.

Ando, S., Ishida, H., Kosugi, Y., and Ishikawa, K. (2002). Hyperthermostable endoglucanase from Pyrococcus horikoshii. *Applied and Environmental Microbiology*, 68(1):430–433.

Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 47(9):1309–1314.

Anthis, N. J. and Clore, G. M. (2015). Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics*, 48(1):35–116.

Ayala, I., Sounier, R., Usé, N., Gans, P., and Boisbouvier, J. (2009). An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *Journal of Biomolecular NMR*, 43(2):111–119.

Back, R., Dominguez, C., Rothé, B., Bobo, C., Beaufils, C., Moréra, S., Meyer, P., Charpentier, B., Branlant, C., Allain, F. H. T., and Manival, X. (2013). High-resolution structural analysis shows how Tah1 tethers Hsp90 to the R2TP complex. *Structure*, 21(10):1834–1847.

Baker, L. A., Sinnige, T., Schellenberger, P., de Keyzer, J., Siebert, C. A., Driessen, A. J., Baldus, M., and Grünewald, K. (2018). Combined1H-Detected Solid-State NMR Spectroscopy and Electron Cryotomography to Study Membrane Proteins across Resolutions in Native Environments. *Structure*, 26(1):161–170.e3.

Baldwin, A. J., Hilton, G. R., Lioe, H., Bagnéris, C., Benesch, J. L. P., and Kay, L. E. (2011). Quaternary dynamics of αb-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *Journal of Molecular Biology*, 413(2):310–320.

Baldwin, A. J. and Kay, L. E. (2012). Measurement of the signs of methyl 13C chemical shift differences between interconverting ground and excited protein states by R 1ρ: An application to αB-crystallin. *Journal of Biomolecular NMR*, 53(1):1–12.

Baldwin, A. J., Walsh, P., Hansen, D. F., Hilton, G. R., Benesch, J. L., Sharpe, S., and Kay, L. E. (2012). Probing dynamic conformations of the high-molecular-weight αb-crystallin heat shock protein ensemble by NMR spectroscopy. *Journal of the American Chemical Society*, 134(37):15343–15350.

Barbet-Massin, E., Huang, C. T., Daebel, V., Hsu, S. T. D., and Reif, B. (2015). Site-specific solid-state NMR studies of "trigger factor" in complex with the large ribosomal subunit 50S. *Angewandte Chemie - International Edition*, 54(14):4367–4369.

Barnett, M. E., Nagy, M., Kedzierska, S., and Zolkiewski, M. (2005). The amino-terminal domain of ClpB supports binding to strongly aggregated proteins. *Journal of Biological Chemistry*, 280(41):34940–34945.

Bascos, N. A. D., Mayer, M. P., Bukau, B., and Landry, S. J. (2017). The Hsp40 J-domain modulates Hsp70 conformation and ATPase activity with a semi-elliptical spring. *Protein Science*, 26(9):1838–1851.

Battiste, J. L. and Wagner, G. (2000). Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry*, 39(18):5355–5365.

Baumketner, A., Jewett, A., and Shea, J. E. (2003). Effects of confinement in chaperonin assisted protein folding: Rate enhancement by decreasing the roughness of the folding energy landscape. *Journal of Molecular Biology*, 332(3):701–713.

Baxter, N. J. and Williamson, M. P. (1997). Temperature dependence of 1H chemical shifts in proteins. *Journal of Biomolecular NMR*, 9(4):359–369.

Beckmann, R. P., Mizzen, L. E., and Welch, W. J. (1990). Interaction of Hsp 70 with newly synthesized proteins: implications for protein folding and assembly. *Science*, 248(4957):850–854.

Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3):343–350.

Bertelsen, E. B., Chang, L., Gestwicki, J. E., and Zuiderweg, E. R. P. (2009). Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proceedings of the National Academy of Sciences*, 106(21):8471–8476.

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., and Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*, 7(1):1–15.

Bhattacharya, S., Naveena Lavanya Latha, J., Kumresan, R., and Singh, S. (2007). Cloning and expression of human islet amyloid polypeptide in cultured cells. *Biochemical and Biophysical Research Communications*, 356(3):622–628.

Biancalana, M. and Koide, S. (2010). Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804(7):1405–1412.

Bodenhausen, G. and Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1):185–189.

Boyle, C. N., Lutz, T. A., and Le Foll, C. (2018). Amylin – Its role in the homeostatic and hedonic control of eating and recent developments of amylin analogs to treat obesity. *Molecular Metabolism*, 8(November 2017):203–210.

Bracher, A. and Verghese, J. (2015). GrpE, Hsp110/Grp170, HspBP1/Sil1 and BAG Domain Proteins: Nucleotide Exchange Factors for Hsp70 Molecular Chaperones. In *Subcellular Biochemistry*, volume 78, pages 1–34. Springer.

Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U., and Hayer-Hartl, M. (2001). Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*, 107(2):223–233.

Brocchieri, L. and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Research*, 33(10):3390–3400.

Bross, P. (2015). *The Hsp60 Chaperonin*. Springer.

Brothers, H. M., Gosztyla, M. L., and Robinson, S. R. (2018). The physiological roles of amyloid-β peptide hint at new ways to treat Alzheimer's disease.

Brutscher, B., Brüschweiler, R., and Ernst, R. R. (1997). Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by 1H, 13C, and 15N nuclear magnetic resonance spectroscopy. *Biochemistry*, 36(42):13043–13053.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21(3):167–195.

Buchner, J. and Li, J. (2013). Structure, Function and Regulation of the Hsp90 Machinery. *Biomedical Journal*, 36(3):106–117.

Burmann, B. M. and Hiller, S. (2012). Solution NMR Studies of Membrane-Protein–Chaperone Complexes. *CHIMIA International Journal for Chemistry*, 66(10):759–763.

Burmann, B. M., Wang, C., and Hiller, S. (2013). Conformation and dynamics of the periplasmic membrane-protein-chaperone complexes OmpX-Skp and tOmpA-Skp. *Nature Structural and Molecular Biology*, 20(11):1265–1272.

Cabrita, L. D., Cassaignau, A. M., Launay, H. M., Waudby, C. A., Wlodarski, T., Camilloni, C., Karyadi, M. E., Robertson, A. L., Wang, X., Wentink, A. S., Goodsell, L. S., Woolhead, C. A., Vendruscolo, M., Dobson, C. M., and Christodoulou, J. (2016). A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nature Structural and Molecular Biology*, 23(4):278–285.

Cabrita, L. D., Hsu, S.-T. D., Launay, H., Dobson, C. M., and Christodoulou, J. (2009). Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proceedings of the National Academy of Sciences*, 106(52):22239–22244.

Cai, K., Frederick, R. O., Kim, J. H., Reinen, N. M., Tonelli, M., and Markley, J. L. (2013). Human mitochondrial chaperone (mtHSP70) and cysteine desulfurase (NFS1) bind preferentially to the disordered conformation, whereas co-chaperone (HSC20) binds to the structured conformation of the iron-sulfur cluster scaffold protein (ISCU). *Journal of Biological Chemistry*, 288(40):28755–28770.

Callon, M., Burmann, B. M., and Hiller, S. (2014). Structural mapping of a chaperone-substrate interaction surface. *Angewandte Chemie - International Edition*, 53(20):5069–5072.

Cameron, R. T., Quinn, S. D., Cairns, L. S., MacLeod, R., Samuel, I. D., Smith, B. O., Carlos Penedo, J., and Baillie, G. S. (2014). The phosphorylation of Hsp20 enhances its association with amyloid-β to increase protection against neuronal cell death. *Molecular and Cellular Neuroscience*, 61:46–55.

Carr, H. Y. and Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94(3):630–638.

Carver, J. A., Aquilina, J. A., and Truscott, R. J. W. (1993). An investigation into the stability of α-crystallin by NMR spectroscopy; evidence for a two-domain structure. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 1164(1):22–28.

Carver, J. A., Guerreiro, N., Nicholls, K. A., and Truscott, R. J. W. (1995). On the interaction of α-crystallin with unfolded proteins. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 1252(2):251–260.

Carver, J. A., Lindner, R. A., Lyon, C., Canet, D., Hernandez, H., Dobson, C. M., and Redfield, C. (2002). The interaction of the molecular chaperone α-crystallin with unfolding α-lactalbumin: A structural and kinetic spectroscopic study. *Journal of Molecular Biology*, 318(3):815–827.

Cassaignau, A. M., Launay, H. M., Karyadi, M. E., Wang, X., Waudby, C. A., Deckert, A., Robertson, A. L., Christodoulou, J., and Cabrita, L. D. (2016). A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nature Protocols*, 11(8):1492–1507.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Skelton, N. J., and Rance, M. (2007). *Protein NMR Spectroscopy*. Elsevier.

Chandel, T. I., Zaman, M., Khan, M. V., Ali, M., Rabbani, G., Ishtikhar, M., and Khan, R. H. (2018). A mechanistic insight into protein-ligand interaction, folding, misfolding, aggregation and inhibition of protein aggregates: An overview. *International Journal of Biological Macromolecules*, 106:1115–1129.

Chou, Y. T. and Gierasch, L. M. (2005). The conformation of a signal peptide bound by Escherichia coli preprotein translocase SecA. *Journal of Biological Chemistry*, 280(38):32753–32760.

Chou, Y. T., Swain, J. F., and Gierasch, L. M. (2002). Functionally significant mobile regions of Escherichia coli SecA ATPase identified by NMR. *Journal of Biological Chemistry*, 277(52):50985–50990.

Cleary, J. P., Walsh, D. M., Hofmeister, J. J., Shankar, G. M., Kuskowski, M. A., Selkoe, D. J., and Ashe, K. H. (2005). Natural oligomers of the amyloid-β protein specifically disrupt cognitive function. *Nature Neuroscience*, 8(1):79–84.

Colby, D. W., Zhang, Q., Wang, S., Groth, D., Legname, G., Riesner, D., and Prusiner, S. B. (2007). Prion detection by an amyloid seeding assay. *Proceedings of the National Academy of Sciences*, 104(52):20914–20919.

Cowan, N. J. and Lewis, S. A. (2001). Type II chaperonins, prefoldin, and the tubulin-specific chaperones.

Csermely, P. (1999). Chaperone-percolator model: A possible molecular mechanism of anfinsen-cage-type chaperones. *BioEssays*, 21(11):959–965.

Cukalevski, R., Yang, X., Meisl, G., Weininger, U., Bernfur, K., Frohm, B., Knowles, T. P., and Linse, S. (2015). The Aβ40 and Aβ42 peptides self-assemble into separate homomolecular fibrils in binary mixtures but cross-react during primary nucleation. *Chemical Science*, 6(7):4215–4233.

Dahiya, V. and Chaudhuri, T. K. (2014). Chaperones GroEL/GroES accelerate the refolding of a multidomain protein through modulating on-pathway intermediates. *Journal of Biological Chemistry*, 289(1):286–298.

Danielsson, J., Andersson, A., Jarvet, J., and Gräslund, A. (2006). 15N relaxation study of the amyloid β-peptide: Structural propensities and persistence length. *Magnetic Resonance in Chemistry*, 44(7 SPEC. ISS.):114–121.

Deckert, A., Waudby, C. A., Wlodarski, T., Wentink, A. S., Wang, X., Kirkpatrick, J. P., Paton, J. F. S., Camilloni, C., Kukic, P., Dobson, C. M., Vendruscolo, M., Cabrita, L. D., and Christodoulou, J. (2016). Structural characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor. *Proceedings of the National Academy of Sciences*, 113(18):5012–5017.

del Río, J. A., Ferrer, I., and Gavín, R. (2018). Role of cellular prion protein in interneuronal amyloid transmission. *Progress in Neurobiology*, 3.

Delaye, M. and Tardieu, A. (1983). Short-range order of crystallin proteins accounts for eye lens transparency. *Nature*, 302(5907):415–417.

Dempsey, B. R., Wrona, M., Moulin, J. M., Gloor, G. B., Jalilehvand, F., Lajoie, G., Shaw, G. S., and Shilton, B. H. (2004). Solution NMR structure and X-ray absorption analysis of the C-terminal zinc-binding domain of the SecA ATPase. *Biochemistry*, 43(29):9361–9371.

DePace, A. H., Santoso, A., Hillner, P., and Weissman, J. S. (1998). A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell*, 93(7):1241–1252.

Ditzel, L., Löwe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R., and Steinbacher, S. (1998). Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT.

Douglas, N. R., Reissmann, S., Zhang, J., Chen, B., Jakana, J., Kumar, R., Chiu, W., and Frydman, J. (2011). Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. *Cell*, 144(2):240–252.

Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, 18(6):756–764.

Dyson, H. J., Kostic, M., Liu, J., and Martinez-Yamout, M. A. (2008). Hydrogen-deuterium exchange strategy for delineation of contact sites in protein complexes. *FEBS Letters*, 582(10):1495–1500.

Eichmann, C., Preissler, S., Riek, R., and Deuerling, E. (2010). Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proceedings of the National Academy of Sciences*, 107(20):9111–9116.

Eimer, W. A., Vijaya Kumar, D. K., Navalpur Shanmugam, N. K., Rodriguez, A. S., Mitchell, T., Washicosky, K. J., György, B., Breakefield, X. O., Tanzi, R. E., and Moir, R. D. (2018). Alzheimer's Disease-Associated β-Amyloid Is Rapidly Seeded by Herpesviridae to Protect against Brain Infection. *Neuron*, 99(1):56–63.e3.

Eisenberg, D. and Jucker, M. (2012). The amyloid state of proteins in human diseases. *Cell*, 148(6):1188–1203.

El Khatib, M., Martins, A., Bourgeois, D., Colletier, J.-P., Adam, V., and Read, R. J. (2016). Rational design of ultrastable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. *Scientific Reports*, 6(1):18459.

Ellgaard, L., McCaul, N., Chatsisvili, A., and Braakman, I. (2016). Co- and Post-Translational Protein Folding in the ER. *Traffic*, 17(6):615–638.

Ellis, R. J. (1977). Protein synthesis by isolated chloroplasts. *Biochimica et Biophysica Acta*, 463:185–215.

Ellis, R. J. (1996). Revisiting the Anfinsen cage. *Folding and Design*, 1(1):R9–R15.

Ellis, R. J. (2001). Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Current Opinion in Structural Biology*, 11(1):114–119.

Esposito, G., Garvey, M., Alverdi, V., Pettirossi, F., Corazza, A., Fogolari, F., Polano, M., Mangione, P. P., Giorgetti, S., Stoppini, M., Rekas, A., Bellotti, V., Heck, A. J., and Carver, J. A. (2013). Monitoring the interaction between β2-microglobulin and the molecular chaperone αB-crystallin by NMR and mass spectrometry: αB-crystallin dissociates β2-microglobulin oligomers. *Journal of Biological Chemistry*, 288(24):17844–17858.

Ewalt, K. L., Hendrick, J. P., Houry, W. A., and Ulrich Hartl, F. (1997). In Vivo Observation of Polypeptide Flux through the Bacterial Chaperonin System. *Cell*, 90:491–500.

Fändrich, M., Nyström, S., Nilsson, K. P., Böckmann, A., LeVine, H., and Hammarström, P. (2018). Amyloid fibril polymorphism: a challenge for molecular imaging and therapy. *Journal of Internal Medicine*, 283(3):218–237.

Ferreira, S. T., Vieira, M. N. N., and De Felice, F. G. (2007). Soluble protein oligomers as emerging toxins in Alzheimer's and other amyloid diseases. *IUBMB Life*, 59(4-5):332–345.

Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Current opinion in structural biology*, 7:3–9.

Finkelstein, A. V., Bogatyreva, N. S., and Garbuzynskiy, S. O. (2013). Restrictions to protein folding determined by the protein size. *FEBS Letters*, 587(13):1884–1890.

Fohlman, J., Eaker, D., Karlsson, E., and Thesleff, S. (1976). Taipoxin, an Extremely Potent Presynaptic Neurotoxin from the Venom of the Australian Snake Taipan (Oxyuranus s. scutellatus): Isolation, Characterization, Quaternary Structure and Pharmacological Properties. *European Journal of Biochemistry*, 68(2):457–469.

Freiburg, A., Trombitas, K., Hell, W., Cazorla, O., Fougerousse, F., Centner, T., Kolmerer, B., Witt, C., Beckmann, J. S., Gregorio, C. C., Granzier, H., and Labeit, S. (2000). Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circulation Research*, 86(11):1114–1121.

Fu, X. (2014). Chaperone function and mechanism of small heat-shock proteins. *Acta Biochimica et Biophysica Sinica*, 46(5):347–356.

Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T., and Taguchi, H. (2010). A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO Journal*, 29(9):1552–1564.

Fukuma, T., Mostaert, A. S., Serpell, L. C., and Jarvis, S. P. (2008). Revealing molecular-level surface structure of amyloid fibrils in liquid by means of frequency modulation atomic force microscopy. *Nanotechnology*, 19(38):384010.

Gade Malmos, K., Blancas-Mejia, L. M., Weber, B., Buchner, J., Ramirez-Alvarado, M., Naiki, H., and Otzen, D. (2017). ThT 101: a primer on the use of thioflavin T to investigate amyloid formation. *Amyloid*, 24(1):1–16.

Gardner, K. H. and Kay, L. E. (1997). Production and incorporation of 15N, 13C, 2H (1H-d1 methyl) isoleucine into proteins for multidimensional NMR studies. *Journal of the American Chemical Society*, 119(32):7599–7600.

Geissler, S., Siegers, K., and Schiebel, E. (1998). A novel protein complex promoting formation of functional alpha- and gamma-tubulin. *The EMBO journal*, 17(4):952–66.

Gelis, I., Bonvin, A. M. J. J., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., and Kalodimos, C. G. (2007). Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell*, 131(4):756–769.

Georgopoulos, C., Tilly, K., Drahos, D., and Hendrix, R. (1982). The B66.0 protein of Escherichia coli is the product of the dnaK+ gene. *Journal of Bacteriology*, 149(3):1175–1177.

Ghiglieri, V., Calabrese, V., and Calabresi, P. (2018). Alpha-Synuclein: From Early Synaptic Dysfunction to Neurodegeneration. *Frontiers in neurology*, 9(MAY):295.

Gianni, S. and Jemth, P. (2016). Protein folding: Vexing debates on a fundamental problem. *Biophysical chemistry*, 212:17–21.

Giehm, L. and Otzen, D. E. (2010). Strategies to increase the reproducibility of protein fibrillization in plate reader assays. *Analytical Biochemistry*, 400(2):270–281.

Gilbert, R. J. C., Fucini, P., Connell, S., Fuller, S. D., Nierhaus, K. H., Robinson, C. V., and Dobson, C. M. (2004). Three-Dimensional Structures of Translating Ribosomes by Cryo-EM of protein synthesis. *Molecular Cell*, 14:57–66.

Goldberg, M. S., Zhang, J., Sondek, S., Matthews, C. R., Fox, R. O., and Horwich, A. L. (1997). Native-like structure of a protein-folding intermediate bound to the chaperonin GroEL. *Proceedings of the National Academy of Sciences*, 94(4):1080–1085.

González, J. M., Masuchi, Y., Robb, F. T., Ammerman, J. W., Maeder, D. L., Yanagibayashi, M., Tamaoka, J., Kato, C., González, J. M., Masuchi, Y., Robb, F. T., Maeder, D. L., Ammerman, J. W., Yanagibayashi, M., Tamaoka, J., and Kato, ·. C. (1998). Pyrococcus horikoshii sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles*, 2:123–130.

Gosztyla, M. L., Brothers, H. M., and Robinson, S. R. (2018). Alzheimer's Amyloid-β is an Antimicrobial Peptide: A Review of the Evidence. *Journal of Alzheimer's Disease*, pages 1–12.

Gozu, M., Hoshino, M., Higurashi, T., Kato, H., and Goto, Y. (2002). The interaction of beta(2)-glycoprotein I domain V with chaperonin GroEL: The similarity with the domain V and membrane interaction. *Protein Science*, 11(12):2792–2803.

Grangeasse, C., Stülke, J., and Mijakovic, I. (2015). Regulatory potential of post-translational modifications in bacteria. *Frontiers in Microbiology*, 6(MAY):2014–2015.

Greene, M. K., Maskos, K., and Landry, S. J. (1998). Role of the J-domain in the cooperation of Hsp40 with Hsp70. *Proceedings of the National Academy of Sciences*, 95(11):6108–6113.

Gregersen, N., Bross, P., Vang, S., and Christensen, J. H. (2006). Protein Misfolding and Human Disease. *Annual Review of Genomics and Human Genetics*, 7(1):103–124.

Gremer, L., Schölzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R. B., Tusche, M., Lopez-Iglesias, C., Hoyer, W., Heise, H., Willbold, D., and Schröder, G. F. (2017). Fibril structure of amyloid-β(1–42) by cryo–electron microscopy. *Science*, 358(6359):116–119.

Grishaev, A., Tugarinov, V., Kay, L. E., Trewhella, J., and Bax, A. (2008). Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *Journal of Biomolecular NMR*, 40(2):95–106.

Groenning, M., Norrman, M., Flink, J. M., van de Weert, M., Bukrinsky, J. T., Schluckebier, G., and Frokjaer, S. (2007). Binding mode of Thioflavin T in insulin amyloid fibrils. *Journal of Structural Biology*, 159(3):483–497.

Guex, N., Peitsch, M. C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 30(SUPPL. 1):162–173.

Haber, E. and Anfinsen, C. B. (1962). Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *The Journal of biological chemistry*, 237(6):1839–1844.

Hagn, F., Lagleder, S., Retzlaff, M., Rohrberg, J., Demmer, O., Richter, K., Buchner, J., and Kessler, H. (2011). Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53. *Nature Structural & Molecular Biology*, 18(10):1086–1093.

Hall, D., Kardos, J., Edskes, H., Carver, J. A., and Goto, Y. (2015). A multi-pathway perspective on protein aggregation: Implications for control of the rate and extent of amyloid formation. *FEBS Letters*, 589(6):672–679.

Hanson, P. I. and Whiteheart, S. W. (2005). AAA+ proteins: Have engine, will work.

Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, 475(7356):324–332.

Hartl, M. J., Burmann, B. M., Prasch, S. J., Schwarzinger, C., Schweimer, K., Wöhrl, B. M., Rösch, P., and Schwarzinger, S. (2012). Fast mapping of biomolecular interfaces by random spin labeling (rsl). *Journal of Biomolecular Structure and Dynamics*, 29(4):793–798.

Haslbeck, M. and Vierling, E. (2015). A first line of stress defense: Small heat shock proteins and their function in protein homeostasis. *Journal of Molecular Biology*, 427(7):1537–1548.

Hay, D. L., Chen, S., Lutz, T. A., Parkes, D. G., Roth, J. D., Diego, S., Physiology, V., Pharmaceuticals, I., and Diego, S. (2015). Amylin: Pharmacology, Physiology, and Clinical Potential. *Pharmacol Rev*, 67(July):564–600.

He, L. and Hiller, S. (2018). Common Patterns in Chaperone Interactions with a Native Client Protein. *Angewandte Chemie - International Edition*, 57(20):5921–5924.

He, L., Sharpe, T., Mazur, A., and Hiller, S. (2016). A molecular mechanism of chaperone-client recognition. *Science Advances*, 2(11).

He, W. T., Xue, W., Gao, Y. G., Hong, J. Y., Yue, H. W., Jiang, L. L., and Hu, H. Y. (2017). HSP90 recognizes the N-terminus of huntingtin involved in regulation of huntingtin aggregation by USP19. *Scientific Reports*, 7(1).

Hemmingsen, S. M. (1992). What is a chaperonin ? *Nature*, 357:650.

Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature*, 333(6171):330–334.

Hirtreiter, A. M., Calloni, G., Forner, F., Scheibe, B., Puype, M., Vandekerckhove, J., Mann, M., Hartl, F. U., and Hayer-Hartl, M. (2009). Differential substrate specificity of group i and group II chaperonins in the archaeon Methanosarcina mazei. *Molecular Microbiology*, 74(5):1152–1168.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. *Biochimica et biophysica acta*, 1803(6):650–61.

Horne, B. E., Li, T., Genevaux, P., Georgopoulos, C., and Landry, S. J. (2010). The Hsp40 J-domain stimulates Hsp70 when tethered by the client to the ATPase domain. *Journal of Biological Chemistry*, 285(28):21679–21688.

Horst, R., Bertelsen, E. B., Fiaux, J., Wider, G., Horwich, A. L., and Wuthrich, K. (2005). Direct NMR observation of a substrate protein bound to the chaperonin GroEL. *Proceedings of the National Academy of Sciences*, 102(36):12748–12753.

Horst, R., Fenton, W. A., Englander, S. W., Wüthrich, K., and Horwich, A. L. (2007). Folding Trajectories of Human Dihydrofolate Reductase Inside the GroEL–GroES Chaperonin Cavity and Free in Solution. *Proceedings of the National Academy of Sciences*, 104(52):20788–20792.

Hoyer, W., Gronwall, C., Jonsson, A., Stahl, S., and Hard, T. (2008). Stabilization of a-hairpin in monomeric Alzheimer's amyloid-peptide inhibits amyloid formation. *Proceedings of the National Academy of Sciences*, 105(13):5099–5104.

Hsu, S.-T. D., Fucini, P., Cabrita, L. D., Launay, H., Dobson, C. M., and Christodoulou, J. (2007). Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proceedings of the National Academy of Sciences*, 104(42):16516–16521.

Hua, Q. X., Dementieva, I. S., Walsh, M. A., Hallenga, K., Weiss, M. A., and Joachimiak, A. (2001). A thermophilic mini-chaperonin contains a conserved polypeptide-binding surface: Combined crystallographic and NMR studies of the GroEL apical domain with implications for substrate interactions. *Journal of Molecular Biology*, 306(3):513–525.

Huang, C., Rossi, P., Saio, T., and Kalodimos, C. G. (2016). Structural basis for the antifolding activity of a molecular chaperone. *Nature*, 537:202–206.

Iizuka, R. and Funatsu, T. (2016). Chaperonin GroEL uses asymmetric and symmetric reaction cycles in response to the concentration of non-native substrate proteins. *Biophysics and physicobiology*, 13:63–69.

Jacobs, D. M., Langer, T., Elshorst, B., Saxena, K., Fiebig, K. M., Vogtherr, M., and Schwalbe, H. (2006). NMR Backbone Assignment of the N-terminal Domain of Human HSP90. *Journal of Biomolecular NMR*, 36(S1):52–52.

Jahn, T. R. and Radford, S. E. (2005). The Yin and Yang of protein folding. *FEBS Journal*, 272(23):5962–5970.

Jakes, R., Spillantini, M. G., and Goedert, M. (1994). Identification of two distinct synucleins from human brain. *FEBS Letters*, 345(1):27–32.

Jarrett, J. T., Berger, E. P., and Lansbury, P. T. (1993). The Carboxy Terminus of the β Amyloid Protein Is Critical for the Seeding of Amyloid Formation: Implications for the Pathogenesis of Alzheimer's Disease. *Biochemistry*, 32(18):4693–4697.

Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kühne, R., Stout, J. R., Higman, V. A., Klevit, R. E., Van Rossum, B. J., and Oschkinat, H. (2010). Solid-state NMR and SAXS studies provide a structural basis for the activation of αb-crystallin oligomers. *Nature Structural and Molecular Biology*, 17(9):1037–1042.

Jehle, S., van Rossum, B., Stout, J. R., Noguchi, S. M., Falber, K., Rehbein, K., Oschkinat, H., Klevit, R. E., and Rajagopal, P. (2009). αB-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer. *Journal of Molecular Biology*, 385(5):1481–1497.

Jiménez, B., Ugwu, F., Zhao, R., Ortí, L., Makhnevych, T., Pineda-Lucena, A., and Houry, W. A. (2012). Structure of minimal tetratricopeptide repeat domain protein Tah1 reveals mechanism of its interaction with Pih1 and Hsp90. *Journal of Biological Chemistry*, 287(8):5698–5709.

Jou, W. M., Haegeman, G., Ysebaert, M., and Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, 237(5350):82–88.

Kampinga, H. H. and Craig, E. A. (2010). The Hsp70 chaperone machinery: J-proteins as drivers of functional specificity NIH Public Access. *Nat Rev Mol Cell Biol*, 11(8):579–592.

Karagöz, G. E., Duarte, A. M. S., Akoury, E., Ippel, H., Biernat, J., Morán Luengo, T., Radli, M., Didenko, T., Nordhues, B. A., Veprintsev, D. B., Dickey, C. A., Mandelkow, E., Zweckstetter, M., Boelens, R., Madl, T., and Rüdiger, S. G. D. (2014). Hsp90-tau complex reveals molecular basis for specificity in chaperone action. *Cell*, 156(5):963–974.

Karagoz, G. E., Duarte, A. M. S., Ippel, H., Uetrecht, C., Sinnige, T., van Rosmalen, M., Hausmann, J., Heck, A. J. R., Boelens, R., and Rudiger, S. G. D. (2011). N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. *Proceedings of the National Academy of Sciences*, 108(2):580–585.

Karagöz, G. E. and Rüdiger, S. G. D. (2015). Hsp90 interaction with clients. *Trends in Biochemical Sciences*, 40(2):117–125.

Karplus, M. and Weaver, D. L. (1976). Protein-folding dynamics. *Nature*, 260(5550):404–406.

Kauzmann, W. (1959). Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry*, 14(C):1–63.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S.-I., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K.-I., Yoshizawa, T., Nakamura, Y., Robb, F. T., Horikoshi, K., Masuchi, Y., Shizuya, H., and Kikuchi, H. (1998). Complete Sequence and Gene Organization of the Genome of a Hyper-thermophilic Archaebacterium, Pyrococcus horikoshii OT3. *DNA RESEARCH*, 5:55–76.

Kayed, R. and Lasagna-Reeves, C. A. (2012). Molecular Mechanisms of Amyloid Oligomers Toxicity. *Journal of Alzheimer's Disease*, 33(s1):S67–S78.

Keeler, J. (2010). *Understanding NMR spectroscopy*. John Wiley and Sons.

Keramisanou, D., Biris, N., Gelis, I., Sianidis, G., Karamanou, S., Economou, A., and Kalodimos, C. G. (2006). Disorder-order folding transitions underlie catalysis in the helicase motor of SecA. *Nature Structural and Molecular Biology*, 13(7):594–602.

Kerfah, R., Plevin, M. J., Sounier, R., Gans, P., and Boisbouvier, J. (2015). Methyl-specific isotopic labeling: A molecular tool box for solution NMR studies of large proteins.

Khabirova, E., Moloney, A., Marciniak, S. J., Williams, J., Lomas, D. A., Oliver, S. G., Favrin, G., Sattelle, D. B., and Crowther, D. C. (2014). The TRiC/CCT chaperone is implicated in Alzheimer's disease based on patient GWAS and an RNAi screen in Aβ-expressing Caenorhabditis elegans. *PLoS ONE*, 9(7).

Khan, M. V., Zakariya, S. M., and Khan, R. H. (2018). Protein folding, misfolding and aggregation: A tale of constructive to destructive assembly. *International Journal of Biological Macromolecules*, 112:217–229.

Kim, J. H., Tonelli, M., Frederick, R. O., Chow, D. C. F., and Markley, J. L. (2012). Specialized Hsp70 chaperone (HscA) binds preferentially to the disordered form, whereas J-protein (HscB) binds preferentially to the structured form of the iron-sulfur cluster scaffold protein (IscU). *Journal of Biological Chemistry*, 287(37):31406–31413.

Klumpp, M., Baumeister, W., and Essen, L. O. (1997). Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell*, 91:263–270.

Kobayashi, N., Freund, S. M., Chatellier, J., Zahn, R., and Fersht, a. R. (1999). NMR analysis of the binding of a rhodanese peptide to a minichaperone in solution. *Journal of molecular biology*, 292(1):181–90.

Koculi, E., Horst, R., Horwich, A. L., and Wüthrich, K. (2011). Nuclear magnetic resonance spectroscopy with the stringent substrate rhodanese bound to the single-ring variant SR1 of the E. coli chaperonin GroEL. *Protein Science*, 20(8):1380–1386.

Kovacs, J. M., Mant, C. T., and Hodges, R. S. (2006). Determination of intrinsic hydrophilicity/hydrophobicity of amino acid side chains in peptides in the absence of nearest-neighbor or conformational effects. *Biopolymers*, 84(3):283–97.

Krampert, M., Bernhagen, J., Schmucker, J., Horn, A., Schmauder, A., Brunner, H., Voelter, W., and Kapurniotu, A. (2000). Amyloidogenicity of recombinant human pro-islet amyloid polypeptide (ProIAPP). *Chemistry and Biology*, 7(11):855–871.

Krebs, M. R. H., Bromley, E. H. C., and Donald, A. M. (2005). The binding of thioflavin-T to amyloid fibrils: Localisation and implications. *Journal of Structural Biology*, 149(1):30–37.

Kriehuber, T., Rattei, T., Weinmaier, T., Bepperling, A., Haslbeck, M., and Buchner, J. (2010). Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *The FASEB Journal*, 24(10):3633–3642.

Kurimoto, E., Nishi, Y., Yamaguchi, Y., Zako, T., Iizuka, R., Ide, N., Yohda, M., and Kato, K. (2007). Dynamics of group II chaperonin and prefoldin probed by 13C NMR spectroscopy. *Proteins: Structure, Function, and Bioinformatics*, 70(4):1257–1263.

Kurt, N., Rajagopalan, S., and Cavagnero, S. (2006). Effect of Hsp70 chaperone on the folding and misfolding of polypeptides modeling an elongating protein chain. *Journal of Molecular Biology*, 355(4):809–820.

Kyle, R. A. (2001). Amyloidosis: A convoluted story. *British Journal of Haematology*, 114(3):529–538.

Landry, S. J., Jordan, R., McMacken, R., and Gierasch, L. M. (1992). Different conformations for the same polypeptide bound to chaperones DnaK and GroEL. *Nature*, 355(6359):455–457.

Lange, S., Franks, W. T., Rajagopalan, N., Döring, K., Geiger, M. A., Linden, A., van Rossum, B. J., Kramer, G., Bukau, B., and Oschkinat, H. (2016). Structural analysis of a signal peptide inside the ribosome tunnel by DNP MAS NMR. *Science advances*, 2(8):e1600379.

Laskey, R. A., Honda, B. M., Mills, A. D., and Finch, J. T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature*, 275(275):415–420.

Le Breton, E. R. and Carver, J. A. (1996). Solution conformation of bovine lens alpha- and betaB2-crystallin terminal extensions. *International journal of peptide and protein research*, 47(1-2):9–19.

Lee, G. J. and Vierling, E. (2000). A Small Heat Shock Protein Cooperates with Heat Shock Protein 70 Systems to Reactivate a Heat-Denatured Protein. *Plant Physiology*, 122(1):189–198.

Lee, J. H., Zhang, D., Hughes, C., Okuno, Y., Sekhar, A., and Cavagnero, S. (2015). Heterogeneous binding of the SH3 client protein to the DnaK molecular chaperone. *Proceedings of the National Academy of Sciences*, 112(31):E4206–E4215.

Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M., and Tsai, F. T. (2003). The structure of ClpB: A molecular chaperone that rescues proteins from an aggregated state. *Cell*, 115(2):229–240.

Lee, Y. T., Jacob, J., Michowski, W., Nowotny, M., Kuznicki, J., and Chazin, W. J. (2004). Human Sgt1 Binds HSP90 through the CHORD-Sgt1 Domain and Not the Tetratricopeptide Repeat Domain. *Journal of Biological Chemistry*, 279(16):16511–16517.

Lescop, E., Schanda, P., and Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance*, 187(1):163–169.

LeVine, H. (1993). Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein science*, 2(3):404–10.

Levitt, M. and Warshel, A. (1975). Computer simulation of protein folding. *Nature*, 253:694–698.

Levitt, M. H. (2008). *Spin dynamics : basics of nuclear magnetic resonance*.

Lewis, V. A., Hynes, G. M., Zheng, D., Saibil, H., and Willison, K. (1992). T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. *Nature*, 358(6383):249–252.

Li, T., Lin, J., and Lucius, A. L. (2015). Examination of polypeptide substrate specificity for Escherichia coli ClpB. *Proteins: Structure, Function and Bioinformatics*, 83(1):117–134.

Li, Y., Liang, H., Zhao, H., Chen, D., Liu, B., Fuhs, T., and Dong, M. (2016). Characterization of Inter- and Intramolecular Interactions of Amyloid Fibrils by AFM-Based Single-Molecule Force Spectroscopy. *Journal of Nanomaterials*, 2016.

Libich, D. S., Fawzi, N. L., Ying, J., and Clore, G. M. (2013). Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. *Proceedings of the National Academy of Sciences*, 110(28):11361–11366.

Libich, D. S., Tugarinov, V., and Clore, G. M. (2015). Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. *Proceedings of the National Academy of Sciences*, 112(29):8817–8823.

Lilly Thankamony, A. S., Wittmann, J. J., Kaushik, M., and Corzilius, B. (2017). Dynamic nuclear polarization for sensitivity enhancement in modern solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 102-103:120–195.

Lindberg, D. J., Wenger, A., Sundin, E., Wesén, E., Westerlund, F., and Esbjörner, E. K. (2017). Binding of Thioflavin-T to Amyloid Fibrils Leads to Fluorescence Self-Quenching and Fibril Compaction. *Biochemistry*, 56(16):2170–2174.

Lindberg, D. J., Wranne, M. S., Gilbert Gatty, M., Westerlund, F., and Esbjörner, E. K. (2015). Steady-state and time-resolved Thioflavin-T fluorescence can report on morphological differences in amyloid fibrils formed by Aβ(1-40) and Aβ(1-42). *Biochemical and Biophysical Research Communications*, 458(2):418–423.

Lindner, R. A., Kapur, A., Mariani, M., Titmuss, S. J., and Carver, J. A. (1998). Structural alterations of α-crystallin during its chaperone action. *European Journal of Biochemistry*, 258(1):170–183.

Linnert, M., Lin, Y. J., Manns, A., Haupt, K., Paschke, A. K., Fischer, G., Weiwad, M., and Lücke, C. (2013). The FKBP-type domain of the human aryl hydrocarbon receptor-interacting protein reveals an unusual Hsp90 interaction. *Biochemistry*, 52(12):2097–2107.

Liu, Z., Wang, C., Li, Y., Zhao, C., Li, T., Li, D., Zhang, S., and Liu, C. (2018). Mechanistic insights into the switch of αB-crystallin chaperone activity and self-multimerization. *Journal of Biological Chemistry*, 293(38):14880–14890.

Lopez, T., Dalton, K., and Frydman, J. (2015). The Mechanism and Function of Group II Chaperonins. *Journal of Molecular Biology*, 427(18):2919–2930.

Lorenz, O. R., Freiburger, L., Rutz, D. A., Krause, M., Zierer, B. K., Alvira, S., Cuéllar, J., Valpuesta, J., Madl, T., Sattler, M., and Buchner, J. (2014). Modulation of the Hsp90 chaperone cycle by a stringent client protein. *Molecular Cell*, 53(6):941–953.

Mainz, A., Bardiaux, B., Kuppler, F., Multhaup, G., Felli, I. C., Pierattelli, R., and Reif, B. (2012). Structural and mechanistic implications of metal binding in the small heat-shock protein αB-crystallin. *Journal of Biological Chemistry*, 287(2):1128–1138.

Mainz, A., Jehle, S., Van Rossum, B. J., Oschkinat, H., and Reif, B. (2009). Large protein complexes with extreme rotational correlation times investigated in solution by magic-angle-spinning NMR spectroscopy. *Journal of the American Chemical Society*, 131(44):15968–15969.

Mainz, A., Peschek, J., Stavropoulou, M., Back, K. C., Bardiaux, B., Asami, S., Prade, E., Peters, C., Weinkauf, S., Buchner, J., and Reif, B. (2015). The chaperone αb-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nature Structural and Molecular Biology*, 22(11):898–905.

Makino, Y., Amada, K., Taguchi, H., and Yoshida, M. (1997). Chaperonin-mediated Folding of Green Fluorescent Protein. *Journal of Biological Chemistry*, 272(19):12468–12474.

Martin, J., Mayhew, M., Langer, T., and Hartl, U. (1993). The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature*, 366(6452):228–233.

Martín-Benito, J., Gómez-Reino, J., Stirling, P. C., Lundin, V. F., Gómez-Puertas, P., Boskovic, J., Chacón, P., Fernández, J. J., Berenguer, J., Leroux, M. R., and Valpuesta, J. M. (2007). Divergent Substrate-Binding Mechanisms Reveal an Evolutionary Specialization of Eukaryotic Prefoldin Compared to Its Archaeal Counterpart. *Structure*, 15(1):101–110.

Martinez-Yamout, M. A., Venkitakrishnan, R. P., Preece, N. E., Kroon, G., Wright, P. E., and Dyson, H. J. (2006). Localization of sites of interaction between p23 and Hsp90 in solution. *Journal of Biological Chemistry*, 281(20):14457–14464.

Mas, G. (2015). *Étude structurale et fonctionnelle par RMN d'une chaperonine de 1 MDa en action*. PhD thesis, Université Grenoble Alpes.

Mas, G., Guan, J. Y., Crublet, E., Debled, E. C., Moriscot, C., Gans, P., Schoehn, G., Macek, P., Schanda, P., and Boisbouvier, J. (2018). Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Science Advances*, 4(9):eaau4196.

Matousek, W. M. and Alexandrescu, A. T. (2004). NMR structure of the C-terminal domain of SecA in the free state. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1702(2):163–171.

Mayer, M. and Meyer, B. (1999). Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie - International Edition*, 38(12):1784–1788.

Mayer, M. P. (2013). Hsp70 chaperone dynamics and molecular mechanism. *Trends in Biochemical Sciences*, 38(10):507–514.

Mayer, M. P. and Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, 62(6):670–684.

Meehan, S., Knowles, T. P., Baldwin, A. J., Smith, J. F., Squires, A. M., Clements, P., Treweek, T. M., Ecroyd, H., Tartaglia, G. G., Vendruscolo, M., MacPhee, C. E., Dobson, C. M., and Carver, J. A. (2007). Characterisation of Amyloid Fibril Formation by Small Heat-shock Chaperone Proteins Human αA-, αB- and R120G αB-Crystallins. *Journal of Molecular Biology*, 372(2):470–484.

Meiboom, S. and Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29(8):688–691.

Meisl, G., Kirkegaard, J. B., Arosio, P., Michaels, T. C., Vendruscolo, M., Dobson, C. M., Linse, S., and Knowles, T. P. (2016). Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nature Protocols*, 11(2):252–272.

Meisl, G., Yang, X., Dobson, C. M., Linse, S., and Knowles, T. P. (2017). Modulation of electrostatic interactions to reveal a reaction network unifying the aggregation behaviour of the Aβ42 peptide and its variants. *Chemical Science*, 8(6):4352–4362.

Meng, W., Clerico, E. M., McArthur, N., and Gierasch, L. M. (2018). Allosteric landscapes of eukaryotic cytoplasmic Hsp70s are shaped by evolutionary tuning of key interfaces. *Proceedings of the National Academy of Sciences*, page 201811105.

Mirecka, E. A., Gremer, L., Schiefer, S., Oesterhelt, F., Stoldt, M., Willbold, D., and Hoyer, W. (2014a). Engineered aggregation inhibitor fusion for production of highly amyloidogenic human islet amyloid polypeptide. *Journal of Biotechnology*, 191:221–227.

Mirecka, E. A., Shaykhalishahi, H., Gauhar, A., Akgül, S., Lecher, J., Willbold, D., Stoldt, M., and Hoyer, W. (2014b). Sequestration of a β-hairpin for control of α-synuclein aggregation. *Angewandte Chemie - International Edition*, 53(16):4227–4230.

Mirzabekov, T. A., Lin, M. C., and Kagan, B. L. (1996). Pore formation by the cytotoxic islet amyloid peptide amylin. *Journal of Biological Chemistry*, 271(4):1988–1992.

Mittermaier, A. and Kay, L. E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science*, 312(5771):224–228.

Mogk, A., Bukau, B., and Kampinga, H. H. (2018). Cellular Handling of Protein Aggregates by Disaggregation Machines. *Molecular Cell*, 69(2):214–226.

Mogk, A., Deuerling, E., Vorderwülbecke, S., Vierling, E., and Bukau, B. (2003). Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Molecular Microbiology*, 50(2):585–595.

Morgado, L., Burmann, B. M., Sharpe, T., Mazur, A., and Hiller, S. (2017). The dynamic dimer structure of the chaperone Trigger Factor. *Nature Communications*, 8(1).

Morris, K. F. and Johnson, C. S. (1992). Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society*, 114(8):3139–3141.

Morshauser, R. C., Hu, W., Wang, H., Pang, Y., Flynn, G. C., and Zuiderweg, E. R. (1999). High-resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. *Journal of Molecular Biology*, 289(5):1387–1403.

Motojima, F. (2015). How do chaperonins fold protein? *Biophysics*, 11(0):93–102.

Mueller, L. (1979). Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence. *Journal of the American Chemical Society*, 101(16):4481–4484.

Mukherjee, A., Morales-Scheihing, D., Salvadores, N., Moreno-Gonzalez, I., Gonzalez, C., Taylor-Presse, K., Mendez, N., Shahnawaz, M., Gaber, A. O., Sabek, O. M., Fraga, D. W., and Soto, C. (2017). Induction of IAPP amyloid deposition and associated diabetic abnormalities by a prion-like mechanism. *The Journal of Experimental Medicine*, page jem.20161134.

Mulder, G. J. (1839). Ueber die Zusammensetzung einiger thierischen Substanzen. *Journal fur Praktische Chemie*, 16(1):129–152.

Murphy, M. E. (2013). The HSP70 family and cancer. *Carcinogenesis*, 34(6):1181–1188.

Naiki, H., Higuchi, K., Hosokawa, M., and Takeda, T. (1989). Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavine T. *Analytical Biochemistry*, 177(2):244–249.

Narayanan, S., Bosl, B., Walter, S., and Reif, B. (2003). Importance of low-oligomeric-weight species for prion propagation in the yeast prion system Sup35/Hsp104. *Proceedings of the National Academy of Sciences*, 100(16):9286–9291.

Narayanan, S., Walter, S., and Reif, B. (2006). Yeast prion-protein, Sup35, fibril formation proceeds by addition and substraction of oligomers. *ChemBioChem*, 7(5):757–765.

Neckers, L., Zuehlke, A. D., and Moses, M. A. (2017). Heat shock protein 90: its inhibition and function. *Philosophical Transactions B*.

Nieba-Axmann, S. E., Ottiger, M., Wüthrich, K., and Plückthun, A. (1997). Multiple cycles of global unfolding of GroEL-bound cyclophilin A evidenced by NMR. *Journal of Molecular Biology*, 271(5):803–818.

Nishida, N., Yagi-Utsumi, M., Motojima, F., Yoshida, M., Shimada, I., and Kato, K. (2013). Nuclear magnetic resonance approaches for characterizing interactions between the bacterial chaperonin GroEL and unstructured proteins. *Journal of Bioscience and Bioengineering*, 116(2):160–164.

Nussbaum, J. M., Seward, M. E., and Bloom, G. S. (2013). Alzheimer disease. *Prion*, 7(1):14–19.

Okochi, M., Yoshida, T., Maruyama, T., Kawarabayasi, Y., Kikuchi, H., and Yohda, M. (2002). Pyrococcus Prefoldin Stabilizes Protein-Folding Intermediates and Transfers Them to Chaperonins for Correct Folding. *Biochemical and Biophysical Research Communications*, 291:769–774.

Ollerenshaw, J. E., Tugarinov, V., and Kay, L. E. (2003). Methyl TROSY: Explanation and experimental verification. *Magnetic Resonance in Chemistry*, 41(10):843–852.

Oroz, J., Chang, B. J., Wysoczanski, P., Lee, C.-T., Pérez-Lara, Á., Chakraborty, P., Hofele, R. V., Baker, J. D., Blair, L. J., Biernat, J., Urlaub, H., Mandelkow, E., Dickey, C. A., and Zweckstetter, M. (2018). Structure and pro-toxic mechanism of the human Hsp90/PPIase/Tau complex. *Nature Communications*, 9(1):4532.

Oroz, J., Kim, J. H., Chang, B. J., and Zweckstetter, M. (2017). Mechanistic basis for the recognition of a misfolded protein by the molecular chaperone Hsp90. *Nature Structural and Molecular Biology*, 24(4):407–413.

Ow, S. Y. and Dunstan, D. E. (2014). A brief overview of amyloids and Alzheimer's disease. *Protein Science*, 23(10):1315–1331.

Park, S. J., Borin, B. N., Martinez-Yamout, M. A., and Dyson, H. J. (2011a). The client protein p53 adopts a molten globule-like state in the presence of Hsp90. *Nature Structural and Molecular Biology*, 18(5):537–541.

Park, S. J., Kostic, M., and Dyson, H. J. (2011b). Dynamic interaction of Hsp90 with its client protein p53. *Journal of Molecular Biology*, 411(1):158–173.

Parsell, D. A., Kowal, A. S., Singer, M. A., and Lindquist, S. (1994). Protein disaggregation mediated by heat-shock protein Hspl04. *Nature*, 372(6505):475–478.

Parsell, D. A. and Sauer, R. T. (1989). The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *Journal of Biological Chemistry*, 264(13):7590–7595.

Pavel, M., Imarisio, S., Menzies, F. M., Jimenez-Sanchez, M., Siddiqi, F. H., Wu, X., Renna, M., O'Kane, C. J., Crowther, D. C., and Rubinsztein, D. C. (2016). CCT complex restricts neuropathogenic protein aggregation via autophagy. *Nature Communications*, 7.

Pederson, K., Chalmers, G. R., Gao, Q., Elnatan, D., Ramelot, T. A., Ma, L. C., Montelione, G. T., Kennedy, M. A., Agard, D. A., and Prestegard, J. H. (2017). NMR characterization of HtpG, the E. coli Hsp90, using sparse labeling with13C-methyl alanine. *Journal of Biomolecular NMR*, 68(3):225–236.

Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences*, 94(23):12366–12371.

Petkova, A. T., Leapman, R. D., Guo, Z., Yau, W. M., Mattson, M. P., and Tycko, R. (2005). Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science*, 307(5707):262–265.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612.

Philo, J. S. and Arakawa, T. (2009). Mechanisms of protein aggregation. *Current pharmaceutical biotechnology*, 10(4):348–51.

Phipps, B. M., Hoffmann, A., Stetter, K. O., and Baumeister, W. (1991). A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. *The EMBO journal*, 10(7):1711–1722.

Pillay, K. and Govender, P. (2013). Amylin uncovered: A review on the polypeptide responsible for type II diabetes. *BioMed Research International*, 2013.

Pockley, A. G. (2003). Heat shock proteins as regulators of the immune response. *Lancet*, 362(9382):469–476.

Preissler, S. and Deuerling, E. (2012). Ribosome-associated chaperones as key players in proteostasis. *Trends in Biochemical Sciences*, 37(7):274–283.

Puig, K. L. and Combs, C. K. (2013). Expression and function of APP and its metabolites outside the central nervous system. *Experimental Gerontology*, 48(7):608–611.

Rajagopal, P., Tse, E., Borst, A. J., Delbecq, S. P., Shi, L., Southworth, D. R., and Klevit, R. E. (2015). A conserved histidine modulates HSPB5 structure to trigger chaperone activity in response to stress related acidosis. *eLife*, 4(MAY):1–21.

Rao, J. N., Jao, C. C., Hegde, B. G., Langen, R., and Ulmer, T. S. (2010). A combinatorial NMR and EPR approach for evaluating the structural ensemble of partially folded proteins. *J Am Chem Soc*, 132(25):8657–8668.

Rapoport, T. A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, 450(7170):663–669.

Rauch, J. N., Zuiderweg, E. R., and Gestwicki, J. E. (2016). Non-canonical Interactions between heat shock cognate protein 70 (Hsc70) and Bcl2-associated anthanogene (BAG) Co-chaperones are important for client release. *Journal of Biological Chemistry*, 291(38):19848–19857.

Reggie, L., Lopez, J. J., Collinson, I., Glaubitz, C., and Lorch, M. (2011). Dynamic Nuclear Polarization-Enhanced Solid-State NMR of a 13 C-Labeled Signal Peptide Bound to Lipid-Reconstituted Sec Translocon. *Journal of the American Chemical Society*, 133(47):19084–19086.

Reiss, A. B., Arain, H. A., Stecker, M. M., Siegart, N. M., and Kasselman, L. J. (2018). Amyloid toxicity in Alzheimer's disease. *Reviews in the Neurosciences*, pages 1–15.

Retzlaff, M., Hagn, F., Mitschke, L., Hessling, M., Gugel, F., Kessler, H., Richter, K., and Buchner, J. (2010). Asymmetric Activation of the Hsp90 Dimer by Its Cochaperone Aha1. *Molecular Cell*, 37(3):344–354.

Revington, M., Holder, T. M., and Zuiderweg, E. R. (2004). NMR study of nucleotide-induced changes in the nucleotide binding domain of Thermus thermophilus Hsp70 chaperone DnaK. Implications for the allosteric mechanism. *Journal of Biological Chemistry*, 279(32):33958–33967.

Revington, M., Zhang, Y., Yip, G. N., Kurochkin, A. V., and Zuiderweg, E. R. (2005). NMR investigations of allosteric processes in a two-domain thermus thermophilus Hsp70 molecular chaperone. *Journal of Molecular Biology*, 349(1):163–183.

Richards, F. M. (1958). on the Enzymic Activity of Subtilisin-Modified Ribonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 44(2):162–166.

Riek, R., Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. (2002). Solution NMR techniques for large molecular and supramolecular structures. *Journal of the American Chemical Society*, 124(41):12144–12153.

Rink, T. J., Beaumont, K., Koda, J., and Young, A. (1993). Structure and biology of amylin. *Trends in Pharmacological Sciences*, 14(4):113–118.

Ritossa, F. M. (1964). Behaviour of RNA and DNA synthesis at the puff level in salivary gland chromosomes of Drosophila. *Experimental Cell Research*, 36(3):515–523.

Rosenzweig, R., Farber, P., Velyvis, A., Rennella, E., Latham, M. P., and Kay, L. E. (2015). ClpB N-terminal domain plays a regulatory role in protein disaggregation. *Proceedings of the National Academy of Sciences*, 112(50):e6872–e6881.

Rosenzweig, R., Moradi, S., Zarrine-Afsar, A., Glover, J. R., and Kay, L. E. (2013). Unraveling the Mechanism of Protein Disaggregation Through a ClpB-DnaK Interaction. *Science*, 339(6123):1080–1083.

Rosenzweig, R., Sekhar, A., Nagesh, J., and Kay, L. E. (2017). Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. *eLife*, 6.

Rudiger, S. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *The EMBO Journal*, 16(7):1501–1507.

Rudiger, S., Freund, S. M. V., Veprintsev, D. B., and Fersht, A. R. (2002). CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17):11085–90.

Rutherford, S. L. and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature*, 396(November):336–342.

Rutz, D. A., Luo, Q., Freiburger, L., Madl, T., Kaila, V. R., Sattler, M., and Buchner, J. (2018). A switch point in the molecular chaperone Hsp90 responding to client interaction. *Nature Communications*, 9(1).

Sabareesan, A. T., Mathew, M., and Udgaonkar, J. B. (2018). Mechanism of aggregation and membrane interactions of mammalian prion protein. *Biochimica et Biophysica Acta (BBA) - Biomembranes*.

Sahlan, M., Zako, T., and Yohda, M. (2018). Prefoldin, a jellyfish-like molecular chaperone: functional cooperation with a group II chaperonin and beyond. *Biophysical Reviews*, 10(2):339–345.

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews Molecular Cell Biology*, 14(10):630–642.

Saio, T., Guan, X., Rossi, P., Economou, A., and Kalodimos, C. G. (2014). Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. *Science*, 344:597.

Sakikawa, C., Taguchi, H., Makino, Y., and Yoshida, M. (1999). On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. *Journal of Biological Chemistry*, 274(30):21251–21256.

Salek, R. M., Williams, M. A., Prodromou, C., Pearl, L. H., and Ladbury, J. E. (2002). Backbone resonance assignments of the 25kD N-terminal ATPase domain from the Hsp90 chaperone. *Journal of Biomolecular NMR*, 23(4):327–328.

Sameshima, T., Iizuka, R., Ueno, T., and Funatsu, T. (2010). Denatured proteins facilitate the formation of the football-shaped GroEL–(GroES) 2 complex. *Biochemical Journal*, 427(2):247–254.

Schanda, P. and Brutscher, B. (2005). Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 127(22):8014–8015.

Schanda, P., Forge, V., and Brutscher, B. (2007). Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc Natl Acad Sci U S A*, 104(27):11257–11262.

Schanda, P., Kupĉe, E., and Brutscher, B. (2005). SOFAST-HMQC experiments for recording two-dimensional deteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR*, 33(4):199–211.

Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006). Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of the American Chemical Society*, 128(28):9042–9043.

Schubert, U., Antón, L. C., Gibbs, J., Norbury, C. C., Yewdell, J. W., and Bennink, J. R. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature*, 404(6779):770–774.

Sekhar, A., Rosenzweig, R., Bouvignies, G., and Kay, L. E. (2015). Mapping the conformation of a client protein through the Hsp70 functional cycle. *Proceedings of the National Academy of Sciences*, 112(33):10395–10400.

Sekhar, A., Rosenzweig, R., Bouvignies, G., and Kay, L. E. (2016). Hsp70 biases the folding pathways of client proteins. *Proceedings of the National Academy of Sciences*, 113(20):E2794–E2801.

Sekhar, A., Velyvis, A., Zoltsman, G., Rosenzweig, R., Bouvignies, G., and Kay, L. E. (2018). Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions. *eLife*, 7.

Serio, T. R., Cashikar, A. G., Kowal, A. S., Sawicki, G. J., Moslehi, J. J., Serpell, L., Arnsdorf, M. F., and Lindquist, S. L. (2000). Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science*, 289(5483):1317–1321.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4):213–223.

Shinde, U. and Inouye, M. (2000). Intramolecular chaperones: Polypeptide extensions that modulate protein folding. *Seminars in Cell and Developmental Biology*, 11(1):35–44.

Shtilerman, M., Lorimer, G. H., and Englander, S. W. (1991). Chaperonin Function : Folding by Forced Unfolding. *Science*, 284(5415):822–825.

Sluzky, V., Tamada, J. A., Klibanovt, A. M., and Langer, R. (1991). Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces (protein stability in solution/degradation pathways/mathematical modeling). *Proc. Natl. Acad. Sci. USA*, 88(November):9377–9381.

Smulders, R. H. P. H., Carver, J. A., Lindner, R. A., Van Boekel, M. A. M., Bloemendal, H., and De Jong, W. W. (1996). Immobilization of the C-terminal extension of bovine αA-crystallin reduces chaperone-like activity. *Journal of Biological Chemistry*, 271(46):29060–29066.

Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D., and Brutscher, B. (2013). BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *Journal of Biomolecular NMR*, 55(4):311–321.

Sörgjerd, K. M., Zako, T., Sakono, M., Stirling, P. C., Leroux, M. R., Saito, T., Nilsson, P., Sekimoto, M., Saido, T. C., and Maeda, M. (2013). Human prefoldin inhibits amyloid-β (Aβ) fibrillation and contributes to formation of nontoxic Aβ aggregates. *Biochemistry*, 52(20):3532–3542.

Sounier, R., Blanchard, L., Wu, Z., and Boisbouvier, J. (2007). High-Accuracy Distance Measurement between Remote Methyls in Specifically Protonated Proteins. *Journal of the American Chemical Society*, 129(3):472–473.

Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997). α-Synuclein in Lewy bodies. *Nature*, 388(6645):839–840.

Sreeramulu, S., Jonker, H. R., Langer, T., Richter, C., Lancaster, C. R. D., and Schwalbe, H. (2009). The human Cdc37·Hsp90 complex studied by heteronuclear NMR spectroscopy. *Journal of Biological Chemistry*, 284(6):3885–3896.

Stejskal, E. O. and Tanner, J. E. (1965). Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics*, 42(1):288–292.

Strauch, A. and Haslbeck, M. (2016). The function of small heat-shock proteins and their implication in proteostasis. *Essays In Biochemistry*, 60(2):163–172.

Street, T. O., Lavery, L. A., and Agard, D. A. (2011). Substrate Binding Drives Large-Scale Conformational Changes in the Hsp90 Molecular Chaperone. *Molecular Cell*, 42(1):96–105.

Street, T. O., Zeng, X., Pellarin, R., Bonomi, M., Sali, A., Kelly, M. J. S., Chu, F., and Agard, D. A. (2014). Elucidating the mechanism of substrate recognition by the bacterial Hsp90 molecular chaperone. *Journal of Molecular Biology*, 426(12):2393–2404.

Stsiapura, V. I., Maskevich, A. A., Tikhomirov, S. A., and Buganov, O. V. (2010). Charge transfer process determines ultrafast excited state deactivation of thioflavin T in low-viscosity solvents. *Journal of Physical Chemistry A*, 114(32):8345–8350.

Sulatskaya, A. I., Lavysh, A. V., Maskevich, A. A., Kuznetsova, I. M., and Turoverov, K. K. (2017). Thioflavin T fluoresces as excimer in highly concentrated aqueous solutions and as monomer being incorporated in amyloid fibrils. *Scientific Reports*, 7(1):1–11.

Swain, J. F., Dinler, G., Sivendran, R., Montgomery, D. L., Stotz, M., and Gierasch, L. M. (2007). Hsp70 Chaperone Ligands Control Domain Association via an Allosteric Mechanism Mediated by the Interdomain Linker. *Molecular Cell*, 26(1):27–39.

Swain, J. F., Schulz, E. G., and Gierasch, L. M. (2006). Direct comparison of a stable isolated Hsp70 substrate-binding domain in the empty and substrate-bound states. *Journal of Biological Chemistry*, 281(3):1605–1611.

Tadeo, X., Castaño, D., and Millet, O. (2007). Anion modulation of the 1H/2H exchange rates in backbone amide protons monitored by NMR spectroscopy. *Protein Science*, 16(12):2733–2740.

Takano, M., Tashiro, E., Kitamura, A., Maita, H., Iguchi-Ariga, S. M., Kinjo, M., and Ariga, H. (2014). Prefoldin prevents aggregation of α-synuclein. *Brain Research*, 1542:186–194.

Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M. A., and Frydman, J. (2009). The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nature Structural and Molecular Biology*, 16(12):1279–1285.

Techtmann, S. M. and Robb, F. T. (2010). Archaeal-like chaperonins in bacteria. *Proceedings of the National Academy of Sciences*, 107(47):20269–20274.

Tharp, W. G. and Sarkar, I. N. (2013). Origins of amyloid-β. *BMC Genomics*, 14(1).

Thoma, J., Burmann, B. M., Hiller, S., and Müller, D. J. (2015). Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. *Nature Structural and Molecular Biology*, 22(10):795–802.

Tissiéres, A., Mitchell, H. K., and Tracy, U. M. (1974). Protein synthesis in salivary glands of Drosophila melanogaster: Relation to chromosome puffs. *Journal of Molecular Biology*, 84(3):389–398.

Todd, M. J., Lorimer, G. H., and Thirumalai, D. (1996). Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proceedings of the National Academy of Sciences*, 93(9):4030–4035.

Toyama, B. H. and Weissman, J. S. (2011). Amyloid Structure: Conformational Diversity and Consequences. *Annual Review of Biochemistry*, 80(1):557–585.

Treweek, T. M., Rekas, A., Lindner, R. A., Walker, M. J., Aquilina, J. A., Robinson, C. V., Horwitz, J., Der Perng, M., Quinlan, R. A., and Carver, J. A. (2005). R120G αB-crystallin promotes the unfolding of reduced α-lactalbumin and is inherently unstable. *FEBS Journal*, 272(3):711–724.

Treweek, T. M., Rekas, A., Walker, M. J., and Carver, J. A. (2010). A quantitative NMR spectroscopic examination of the flexibility of the C-terminal extensions of the molecular chaperones, αA- and αB-crystallin. *Experimental Eye Research*, 91(5):691–699.

Tugarinov, V., Choy, W.-Y., Orekhov, V. Y., and Kay, L. E. (2005). Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proceedings of the National Academy of Sciences*, 102(3):622–627.

Tugarinov, V., Muhandiram, R., Ayed, A., and Kay, L. E. (2002). Four-dimensional NMR spectroscopy of a 723-residue protein: Chemical shift assignments and secondary structure of malate synthase G. *Journal of the American Chemical Society*, 124(34):10025–10035.

Tyagi, N. K., Fenton, W. A., and Horwich, A. L. (2010). ATP-triggered ADP release from the asymmetric chaperonin GroEL/GroES/ADP7 is not the rate-limiting step of the GroEL/GroES reaction cycle. *FEBS Letters*, 584(5):951–953.

Tycko, R. (2014). Physical and structural basis for polymorphism in amyloid fibrils. *Protein Science*, 23(11):1528–1539.

Umehara, K., Hoshikawa, M., Tochio, N., and Tate, S. I. (2018). Substrate binding switches the conformation at the lynchpin site in the substrate-binding domain of human Hsp70 to enable allosteric interdomain communication. *Molecules*, 23(3).

Vainberg, I. E., Lewis, S. A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H. L., and Cowan, N. J. (1998). Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell*, 93(5):863–873.

Vassar, P. S. and Culling, C. F. (1959). Fluorescent stains, with special reference to amyloid and connective tissues. *Archives of pathology*, 68:487–98.

Vickery, H. B. (1950). The origin of the word protein. *The Yale journal of biology and medicine*, 22(5):387–393.

Visa, M., Alcarraz-Viźan, G., Montane, J., Cadavez, L., Castaño, C., Villanueva-Peñacarrillo, M. L., Servitja, J. M., and Novials, A. (2015). Islet amyloid polypeptide exerts a novel autocrine action in b-cell signaling and proliferation. *FASEB Journal*, 29(7):2970–2979.

Volkert, T. L., Baleja, J. D., and Kumamoto, C. A. (1999). A highly mobile C-terminal tail of the Escherichia coli protein export chaperone SecB. *Biochemical and Biophysical Research Communications*, 264(3):949–954.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, 59(4):687–696.

Walsh, K. A. and Brown, J. R. (1962). Internal standards for amino acid analyses. *BBA - Biochimica et Biophysica Acta*, 58(3):596–598.

Wälti, M. A., Schmidt, T., Murray, D. T., Wang, H., Hinshaw, J. E., and Clore, G. M. (2017). Chaperonin GroEL accelerates protofibril formation and decorates fibrils of the Het-s prion protein. *Proceedings of the National Academy of Sciences*, 114(34):9104–9109.

Walton, T. A., Sandoval, C. M., Fowler, C. A., Pardi, A., and Sousa, M. C. (2009). The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *Proceedings of the National Academy of Sciences*, 106(6):1772–1777.

Wang, H., Kurochkin, A. V., Pang, Y., Hu, W., Flynn, G. C., and Zuiderweg, E. R. (1998). NMR solution structure of the 21 kDa chaperone protein DnaK substrate binding domain: A preview of chaperone-protein interaction. *Biochemistry*, 37(22):7929–7940.

Wang, L., jun Hu, Z., ming Luo, Y., wu Huo, Y., Ma, Q., zhi He, Y., ying Zhang, Y., Sun, F., and yang Dong, Z. (2010). Distinct symmetry and limited peptide refolding activity of the thermosomes from the acidothermophilic archaea Acidianus tengchongensis S5T. *Biochemical and Biophysical Research Communications*, 393(2):228–234.

Wang, Z., Feng, H. P., Landry, S. J., Maxwell, J., and Gierasch, L. M. (1999). Basis of substrate binding by the chaperonin GroEL. *Biochemistry*, 38(39):12537–12546.

Waudby, C. A., Ramos, A., Cabrita, L. D., and Christodoulou, J. (2016). Two-Dimensional NMR Lineshape Analysis. *Scientific Reports*, 6.

Weinhäupl, K., Lindau, C., Hessel, A., Wang, Y., Schütze, C., Jores, T., Melchionda, L., Schönfisch, B., Kalbacher, H., Bersch, B., Rapaport, D., Brennich, M., Lindorff-Larsen, K., Wiedemann, N., and Schanda, P. (2018). Structural Basis of Membrane Protein Chaperoning through the Mitochondrial Intermembrane Space. *Cell*, 175(5):1365–1379.e25.

Wentink, A. S. (2014). *Structure and Dynamics of the E. coli Chaperone, Trigger Factor, and its Interaction with Ribosome-Nascent Chain Complexes*. PhD thesis, University College London.

Westermark, P., Engström, U., Johnson, K., Westermark, G., and Betsholtz, C. (1990). Islet amyloid polypeptide: pinpointing amino acid residues linked to amyloid fibril formation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(July):5036–40.

Westermark, P., Wernstedt, C., Wilander, E., and Sletten, K. (1986). A novel peptide in the calcitonin gene related peptide family as an amyloid fibril protein in the endocrine pancreas. *Biochemical and Biophysical Research Communications*, 140(3):827–831.

Wetlaufer, D. B. (1973). Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proceedings of the National Academy of Sciences*, 70(3):697–701.

Wieteska, L., Shahidi, S., and Zhuravleva, A. (2017). Allosteric fine-tuning of the conformational equilibrium poises the chaperone BiP for post-translational regulation. *eLife*, 6:1–20.

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73:1–16.

Wolmarans, A., Lee, B., Spyracopoulos, L., and LaPointe, P. (2016). The Mechanism of Hsp90 ATPase Stimulation by Aha1. *Scientific Reports*, 6(August):1–15.

Wördehoff, M. and Hoyer, W. (2018). α-Synuclein Aggregation Monitored by Thioflavin T Fluorescence Assay. *Bio-Protocol*, 8(14).

Wu, H. (1931). A theory of denaturation. *Chinese Journal of Physiology*, 5(4):321–344.

Xue, C., Lin, T. Y., Chang, D., and Guo, Z. (2017). Thioflavin T as an amyloid dye: Fibril quantification, optimal concentration and effect on aggregation. *Royal Society Open Science*, 4(1).

Yagi-Utsumi, M., Kunihara, T., Nakamura, T., Uekusa, Y., Makabe, K., Kuwajima, K., and Kato, K. (2013). NMR characterization of the interaction of GroEL with amyloid β as a model ligand. *FEBS Letters*, 587(11):1605–1609.

Yoshimura, Y., Lin, Y., Yagi, H., Lee, Y.-H., Kitayama, H., Sakurai, K., So, M., Ogi, H., Naiki, H., and Goto, Y. (2012). Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proceedings of the National Academy of Sciences*, 109(36):14446–14451.

Zahn, R., Buckle, A. M., Perrett, S., Johnson, C. M., Corrales, F. J., Golbik, R., and Fersht, A. R. (1996a). Chaperone activity and structure of monomeric polypeptide binding domains of GroEL. *Proceedings of the National Academy of Sciences*, 93(26):15024–15029.

Zahn, R., Perrett, S., and Fersht, A. R. (1996b). Conformational states bound by the molecular chaperones GroEL and SecB: A hidden unfolding (annealing) activity. *Journal of Molecular Biology*, 261(1):43–61.

Zahn, R., Perrett, S., Stenberg, G., and Fersht, A. R. (1996c). Catalysis of Amide Proton Exchange by the Molecular Chaperones GroEL and SecB. *Science*, 271(5249):642–645.

Zahn, R., Spitzfaden, C., Ottiger, M., Wüthrich, K., and Plückthun, A. (1994). Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature*, 368(6468):261–265.

Zako, T., Sahlan, M., Fujii, S., Yamamoto, Y. Y., Tai, P. T., Sakai, K., Maeda, M., and Yohda, M. (2016). Contribution of the C-Terminal Region of a Group II Chaperonin to its Interaction with Prefoldin and Substrate Transfer. *Journal of Molecular Biology*, 428(11):2405–2417.

Zhang, Y., Roder, H., and Paterson, Y. (1995). Rapid amide proton exchange rates in peptides and proteins measured by solvent quenching and two-dimensional NMR. *Protein Science*, 4(4):804–814.

Zhuravleva, A. and Gierasch, L. M. (2011). Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones. *Proceedings of the National Academy of Sciences*, 108(17):6987–6992.

Zhuravleva, A. and Gierasch, L. M. (2015). Substrate-binding domain conformational dynamics mediate Hsp70 allostery. *Proceedings of the National Academy of Sciences*, 112(22):E2865–E2873.

Zolkiewski, M., Zhang, T., and Nagy, M. (2012). Aggregate reactivation mediated by the Hsp100 chaperones.

Zraika, S., Hull, R. L., Verchere, C. B., Clark, A., Potter, K. J., Fraser, P. E., Raleigh, D. P., and Kahn, S. E. (2010). Toxic oligomers and islet beta cell death: guilty by association or convicted by circumstantial evidence? *Diabetologia*, 53(6):1046–1056.

Zwanzig, R., Szabo, A., and Bagchi, B. (1992). Levinthals Paradox. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 89(1):20–22.

Supplementary figures

A.1 Protein parameters

Protein parameters were calculated the ProtParam swiss tool. For material and methods see

A.1.1 *Pyrococcus horikoshii* strain OT-3 Thermosome

- Number of amino-acids : 549 (1 subunit)
- Molecular weight : 59692.8 Da (1 subunit)
- Formula: C2619H4380N708O835S17
- Extinction coefficient : 10430 $M^{-1}cm^{-1}$ at 280 nm in water (This protein does not contain any Trp residues. This could result in more than 10% error in the computed extinction coefficient)
- Absorbance at 1 mg/ml : 0.175

- Amino-acid composition :

Ala (A) 62	11.3%	Gly (G) 42	7.7%	Pro (P) 15	2.7%
Arg (R) 21	3.8%	His (H) 6	1.1%	Ser (S) 15	2.7%
Asn (N) 11	2.0%	Ile (I) 52	9.5%	Thr (T) 25	4.6%
Asp (D) 38	6.9%	Leu (L) 47	8.6%	Trp (W) 0	0.0%
Cys (C) 1	0.2%	Lys (K) 56	10.2%	Tyr (Y) 7	1.3%
Gln (Q) 17	3.1%	Met (M) 16	2.9%	Val (V) 46	8.4%
Glu (E) 67	12.2%	Phe (F) 5	0.9%		

- Sequence :

```
      10         20         30         40         50
MAQLAGQPIL ILPEGTQRYV GRDAQRMNIL AARIIAETVR TTLGPKGMDK
      60         70         80         90         100
MLVDSLGDIV ITNDGATILD EMDIQHPAAK MMVEVAKTQD KEAGDGTTTA
      110        120        130        140        150
VVIAGELLKK AEELLDQNIH PSIIIKGYTL ASQKAQEILD SIAKEVKPDD
      160        170        180        190        200
EEVLLKAAMT AITGKAAEEE REYLAKLAVE AVKLVAEEKD GKLKVDIDNI
      210        220        230        240        250
KLEKKEGGAV RDTRLIRGVV IDKEVVHPGM PKRIENAKIA LINDALEVKE
      260        270        280        290        300
TETDAEIRIT SPEQLQAFLE QEEKMLKEMV DKIKEVGANV VFVQKGIDDL
      310        320        330        340        350
AQHYLAKYGI LAVRRVKKSD MEKLAKATGA KIVTNIRDLT PEDLGEAELV
      360        370        380        390        400
EERKVAGENM IFVEGCKNPK AVTILIRGGT EHVVDEVERA LEDAIKVVKD
      410        420        430        440        450
ILEDGKIIAG GGASEIELSI KLDEYAKEVG GKEQLAIEAF AEALKVIPRT
      460        470        480        490        500
LAENAGLDPI ETLVKVIAAH KEKGQTIGID VYEGEPADMM ERGVIEPVRV
      510        520        530        540
KKQAIKSASE AAIMILRIDD VIAASKLEKE KEGEKGGGSE EFSGSSDLD
```

A.1.2 *Escherichia coli* Malate Synthase G

- Number of amino acids : 731
- Molecular weight : 81537.6 Da
- Formula: C3590H5683N1035O1079S29
- Extinction coefficient : 93195 $M^{-1}cm^{-1}$ at 280 nm in water when cystines are not reduced

- Absorbance at 1 mg/ml : 1.143

- Amino-acid composition :

Ala (A) 73	10.0%	Gly (G) 51	7.0%	Pro (P) 31	4.2%
Arg (R) 44	6.0%	His (H) 23	3.1%	Ser (S) 34	4.7%
Asn (N) 42	5.7%	Ile (I) 44	6.0%	Thr (T) 33	4.5%
Asp (D) 44	6.0%	Leu (L) 70	9.6%	Trp (W) 12	1.6%
Cys (C) 6	0.8%	Lys (K) 32	4.4%	Tyr (Y) 18	2.5%
Gln (Q) 40	5.5%	Met (M) 23	3.1%	Val (V) 46	6.3%
Glu (E) 46	6.3%	Phe (F) 19	2.6%		

- Sequence :

```
10         20         30         40         50
MSQTITQSRL RIDANFKRFV DEEVLPGTGL DAAAFWRNFD EIVHDLAPEN
      60         70         80         90         100
RQLLAERDRI QAALDEWHRS NPGPVKDKAA YKSFLRELGY LVPQPERVTV
      110        120        130        140        150
ETTGIDSEIT SQAGPQLVVP AMNARYALNA ANARWGSLYD ALYGSDIIPQ
      160        170        180        190        200
EGAMVSGYDP QRGEQVIAWV RRFLDESLPL ENGSYQDVVA FKVVDKQLRI
      210        220        230        240        250
QLKNGKETTL RTPAQFVGYR GDAAAPTCIL LKNNGLHIEL QIDANGRIGK
      260        270        280        290        300
DDPAHINDVI VEAAISTILD CEDSVAAVDA EDKILLYRNL LGLMQGTLQE
      310        320        330        340        350
KMEKNGRQIV RKLNDDRHYT AADGSEISLH GRSLLFIRNV GHLMTIPVIW
      360        370        380        390        400
DSEGNEIPEG ILDGVMTGAI ALYDLKVQKN SRTGSVYIVK PKMHGPQEVA
      410        420        430        440        450
FANKLFTRIE TMLGMAPNTL KMGIMDEERR TSLNLRSCIA QARNRVAFIN
      460        470        480        490        500
TGFLDRTGDE MHSVMEAGPM LRKNQMKSTP WIKAYERNNV LSGLFCGLRG
      510        520        530        540        550
KAQIGKGMWA MPDLMADMYS QKGDQLRAGA NTAWVPSPTA ATLHALHYHQ
      560        570        580        590        600
TNVQSVQANI AQTEFNAEFE PLLDDLLTIP VAENANWSAQ EIQQELDNNV
      610        620        630        640        650
QGILGYVVRW VEQGIGCSKV PDIHNVALME DRATLRISSQ HIANWLRHGI
      660        670        680        690        700
LTKEQVQASL ENMAKVVDQQ NAGDPAYRPM AGNFANSCAF KAASDLIFLG
      710        720        730
VKQPNGYTEP LLHAWRLREK ESHLEHHHHH H
```

A.1.3 *Homo sapiens* Amylin, or IAPP

- Number of amino acids : 37
- Molecular weight : 3906.32 Da
- Formula: C165H262N50O56S2
- Absorbance at 1 mg/ml : 0.413
- Amino-acid composition :

Ala (A) 4	10.8%	Gly (G) 2	5.4%	Pro (P) 0	0.0%
Arg (R) 1	2.7%	His (H) 1	2.7%	Ser (S) 5	13.5%
Asn (N) 6	16.2%	Ile (I) 1	2.7%	Thr (T) 5	13.5%
Asp (D) 0	0.0%	Leu (L) 3	8.1%	Trp (W) 0	0.0%
Cys (C) 2	5.4%	Lys (K) 1	2.7%	Tyr (Y) 1	2.7%
Gln (Q) 1	2.7%	Met (M) 0	0.0%	Val (V) 2	5.4%
Glu (E) 0	0.0%	Phe (F) 2	5.4%		

- Sequence :

```
         10         20         30
KCNTATCATQ RLANFLVHSS NNFGAILSST NVGSNTY
```

A.2 Science Advances article and Supplementary figures

STRUCTURAL BIOLOGY

Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle

Guillaume Mas*, Jia-Ying Guan, Elodie Crublet†, Elisa Colas Debled, Christine Moriscot, Pierre Gans, Guy Schoehn, Pavel Macek†‡, Paul Schanda‡, Jerome Boisbouvier‡

Chaperonins are ubiquitous protein assemblies present in bacteria, eukaryota, and archaea, facilitating the folding of proteins, preventing protein aggregation, and thus participating in maintaining protein homeostasis in the cell. During their functional cycle, they bind unfolded client proteins inside their double ring structure and promote protein folding by closing the ring chamber in an adenosine 5′-triphosphate (ATP)–dependent manner. Although the static structures of fully open and closed forms of chaperonins were solved by x-ray crystallography or electron microscopy, elucidating the mechanisms of such ATP-driven molecular events requires studying the proteins at the structural level under working conditions. We introduce an approach that combines site-specific nuclear magnetic resonance observation of very large proteins, enabled by advanced isotope labeling methods, with an in situ ATP regeneration system. Using this method, we provide functional insight into the 1-MDa large hsp60 chaperonin while processing client proteins and reveal how nucleotide binding, hydrolysis, and release control switching between closed and open states. While the open conformation stabilizes the unfolded state of client proteins, the internalization of the client protein inside the chaperonin cavity speeds up its functional cycle. This approach opens new perspectives to study structures and mechanisms of various ATP-driven biological machineries in the heat of action.

INTRODUCTION

Chaperones, proteins that assist folding and stabilization of other proteins, are essential for cellular homeostasis and viability. Chaperonins, a particular class of chaperones also noted hsp60 in the following, form a giant barrel-like structure composed of two back-to-back stacked rings. While group I chaperonins, such as bacterial GroEL, require a co-chaperonin lid (GroES) to cap the ring chamber, group II chaperonins, present in the eukaryotic cytosol and in archaea, are composed of subunits with a built-in helical protrusion which closes the folding chamber. Biochemical studies have provided limited structural information on key steps of the chaperonin functional cycle and mechanism of protein folding (*1*–*4*). Hsp60 chaperonin undergoes large-scale adenosine 5′-triphosphate (ATP)–dependent conformational changes to promote folding of the client protein within a large chamber. Structures of open and closed states (*5*–*9*) have provided static snapshots of steps along the cycle, but the conformational dynamics of the active ATP-fueled chaperone, as well as the interactions with client proteins being refolded, have remained elusive. Nuclear magnetic resonance (NMR) spectroscopy is the method of choice to study proteins in solution and is, in principle, able to observe conformational changes, interactions, and dynamics under functional conditions. However, the site-specific observation of proteins as large as chaperonin, 1 MDa, remains a significant challenge due to the extensive line broadening of NMR signals in large proteins. Furthermore, retaining the protein in a functional state throughout a structural investigation is complicated by the rapid consumption of ATP, thereby leading to inactivation and inhibition of the ATP-dependent machinery.

To overcome these methodological limitations, we have developed an ATP regeneration system directly inside an NMR tube to keep the ATP concentration constant and to prevent the accumulation of adenosine 5′-diphosphate (ADP), thereby maintaining the chaperonin in a steady working condition powered by ATP hydrolysis. We furthermore elaborated a combined methyl-specific labeling approach to probe the different conformational states of the chaperonin and its client proteins. This combination of methods allows us to probe, on the structural level, the different states involved in the functional cycle of active hsp60 processing client proteins while the chaperonin is being powered by ATP hydrolysis (*10*, *11*). We provide quantitative measures of the thermodynamics and kinetics of the transitions between the open and closed conformations and reveal how the chaperonin affects the folding/unfolding equilibrium of the client protein. Specifically, we find that hsp60 acts as a holdase of unfolded proteins until activation by ATP. ATP binding closes the chaperonin cavity, while the ATP hydrolysis rate controls the timing of the closed state, during which encapsulated proteins can refold for a period of ~17 s. After ATP hydrolysis, the chamber reopens during the ADP residence time (~10 s), which allows release of folded protein and client proteins to be reloaded before the start of a new cycle. Our results reveal how nucleotides and client protein binding regulate the transition between different conformations populated during the hsp60 functional cycle. We anticipate this work to be a starting point to obtain insights into the mechanisms of various chaperones. Furthermore, our approach opens new perspectives for the study of structures and mechanisms of ATP-powered biomolecular machines.

RESULTS

ATP binding closes the chaperonin cavity

We have chosen to study the conformations, nucleotide binding, and client protein interaction of a group II chaperonin from the

University of Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France.
*Present address: Biozentrum University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland.
†Present address: NMR-Bio, F-38000 Grenoble, France.
‡Corresponding author. Email: macek@nmr-bio.com (P.M.); paul.schanda@ibs.fr (P.S.); jerome.boisbouvier@ibs.fr (J.B.)

hyperthermophilic *Pyrococcus horikoshii* (*12*). This hsp60 chaperonin is a homohexadecameric assembly (Fig. 1, A and B), that is, a simplified mechanistic model of its eukaryotic heterohexadecameric counterpart. Furthermore, with an optimal functional temperature above 60°C, where NMR relaxation properties are more favorable, site-specific NMR information can be obtained at the level of individual residues under physiologically relevant conditions. However, studying a 1-MDa protein assembly is challenged by the low sensitivity and resolution associated with rapid NMR spin relaxation (*13*, *14*). By combining selective valine and methionine methyl labeling (*15*, *16*) with methyl–transverse relaxation optimized spectroscopy (TROSY) experiments (*17*, *18*), we have obtained high-quality NMR spectra, allowing us to probe the structure and dynamics at 62 individual sites evenly dispersed throughout the structure of the chaperonin (Fig. 1, B and C, and figs. S1 and S2). In nucleotide-free or ADP-bound states, chaperonins are in an open state allowing the access of client proteins to the cavity (Fig. 1A) (*19*–*22*). However, the event that triggers the open-to-closed transition (Fig. 1B) remains a matter of discussion (*4*, *19*, *23*–*26*). Having highly resolved NMR spectra, we probed the conformational states of chaperonin in the presence of different nucleotide ligands. Chemical shift perturbations induced by the binding of ADP and a nonhydrolyzable ATP analog (App-NHp) revealed that M159, located in the vicinity of the nucleotide binding site, exhibits distinct peak positions in the apo and nucleotide-bound states and, thus, directly reports on the occupancy of the binding site (Fig. 1C). Upon binding of App-NHp, however, the signals of M275 and M279 located in the lid are significantly shifted (fig. S3C), revealing a global conformational change to the closed state (*4*). In contrast, binding of ADP does not affect the signals of these two methionines, confirming that the ADP-bound state, just like the apo state, is in an open conformation (fig. S3B). Negative staining electron microscopy (EM) analysis of the NMR sample (fig. S4) confirmed that App-NHp binding induces the closed conformation. Together, these results show that ATP binding triggers the transition from the open to the closed conformation of the ring, while ADP binding does not affect the closed conformation.

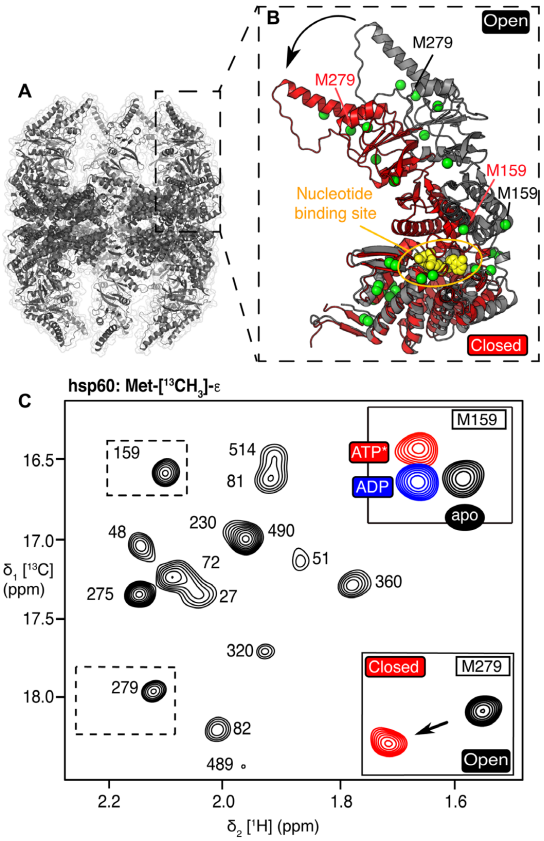


Fig. 1. Methionine-directed methyl NMR allows the identification of the nucleotide binding state and the different chaperonin conformational states. (**A**) Model of the chaperonin structure in the open conformation obtained by homology modeling from Protein Data Bank (PDB 3IZH). (**B**) Magnified view of one hsp60 monomer in open (gray) and closed (red) conformations (obtained by homology modeling from PDB 1Q3R). The methyl groups of methionine residues are represented by green spheres, and the nucleotide in the binding site is represented in yellow. (**C**) 2D ^{1}H-^{13}C methyl-TROSY spectrum of a 12.5 μM U-[^{2}H, ^{15}N], Met-[^{13}CH$_3$] chaperonin sample with the assignment of the methionine residues obtained by combination of a mutagenesis-based approach (fig. S1) with detection of NOEs between methyl groups (fig. S2). The inserts focus on cross peaks of M159 and M279 in the apo state (black) and in the presence of 2 mM ATP analog App-NHp (red, closed state) or ADP (blue, open state).

The open state of chaperonin acts as a holdase of unfolded proteins

Hen egg white lysozyme (HEWL; 14.5 kDa) and malate synthase G (MSG; 82 kDa) were selected as model proteins to observe the interaction of hsp60 with two client proteins with a significant size difference. At 65°C, a temperature close to the optimal conditions for hsp60 activity, HEWL features two sets of NMR resonances corresponding to the unfolded and folded states, which interconvert on a time scale of 350 ms (fig. S5). The addition of chaperonin shifts only the peaks of the unfolded form of HEWL, showing that hsp60 interacts exclusively with the unfolded form (Fig. 2A and fig. S5). Quantitative analysis of the titration data reveals a dissociation constant of 1.6 ± 0.4 μM, with a binding stoichiometry of one unfolded HEWL per chaperonin cavity. NMR-based thermal unfolding studies of HEWL revealed that the presence of hsp60 decreases the melting temperature by 5°C (Fig. 2B and fig. S5), confirming the stabilization of the unfolded state by the chaperonin. The exclusive interaction of hsp60 with the unfolded state is furthermore confirmed by the observation that the translational diffusion coefficients of the unfolded states of HEWL and MSG are shifted to the one of the chaperonin, whereas the folded state of these proteins retain the hydrodynamic properties expected for the globular freely tumbling proteins (Fig. 2C and fig. S6). To determine how these unfolded proteins interact with hsp60, we monitored changes in hsp60 NMR spectra upon the addition of paramagnetically labeled HEWL or MSG (Fig. 2D and fig. S7). Paramagnetic-induced line broadening effects reveal that both unfolded proteins bind to the bottom of the folding chamber nearby the previously proposed sensor loop (Fig. 2E) (*5*). These results demonstrate that hsp60 interacts specifically with unfolded proteins and that the general interaction site for client proteins is located at the base of the chaperonin cavity.

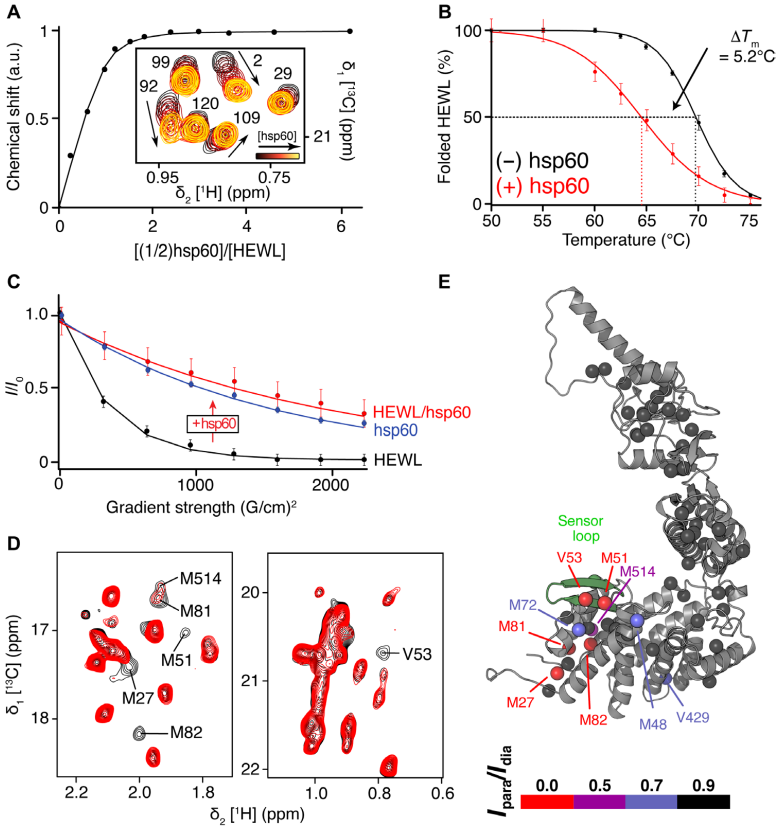


Fig. 2. NMR characterization of the interaction between client protein and hsp60. (**A**) Sum of chemical shift changes of five HEWL methyl signals (V2, I88, V92, V109, and I124) showing a significant shift and no overlap. The inset shows the NMR signal of valine residues of HEWL recorded at 75°C at a constant concentration (31 μM), and an increasing concentration of perdeuterated hsp60 assembly (0 to 97 μM). a.u., arbitrary units. (**B**) Evolution of the folded population of the HEWL between 50° and 75°C without (black) and with (red) hsp60 (two equivalents of HEWL per chaperonin). The melting temperatures (T_m) are indicated by vertical dotted lines. (**C**) Translational diffusion properties of hsp60 alone (blue), unfolded HEWL alone (black), and in the presence of hsp60 (red) characterized by diffusion-ordered NMR spectroscopy. Derived diffusion coefficients at 65°C (hsp60: $5.8 \pm 0.3 \times 10^{-7}$ cm/s^2; unfolded HEWL with hsp60: $4.4 \pm 1.2 \times 10^{-7}$ cm/s^2; unfolded HEWL alone: $30 \pm 4 \times 10^{-7}$ cm/s^2) reveal that only unfolded HEWL is interacting with hsp60. (**D**) Overlay of the 2D ^{1}H-^{13}C NMR spectra of U-[^{2}H, ^{15}N], Val-[$^{13}CH_3$]^{proS}, Met-[$^{13}CH_3$] methyl–labeled sample of hsp60 with two equivalents of diamagnetic (black) or paramagnetic labeled HEWL (red). (**E**) Model of the monomer of hsp60 in the open conformation showing the residues affected by the paramagnetically labeled client proteins. hsp60 sensor loop is colored in green.

Client protein accelerates the functional cycle of hsp60

To study the system in action under steady-state conditions, without accumulation of the competitive inhibitor ADP (*27*), we implemented an enzymatic system directly inside the NMR sample to ensure continuous rapid regeneration of ATP from ADP and phosphoenol pyruvate (PEP; Fig. 3A). This system keeps the ATP concentration constant, prevents ADP accumulation, and keeps the chaperonin catalytically active at a constant rate (Fig. 3, B and C). Moreover, interleaved NMR data collection of this complex reaction mixture allows to simultaneously monitor the state of the chaperonin and client protein, as well as the concentrations of ATP, ADP, and PEP. The distinct sets of methionine NMR resonances report on the open/closed conformational state (M275 and M279) and type of nucleotide bound (M159; Figs. 1 and 3D), while the Ile, Leu, and Val resonances of HEWL report on the equilibrium of folded and unfolded states of the client protein (Fig. 4A and fig. S8). The population of folded HEWL of ca. 50% in the presence of chaperonin alone increased to 75% when powered by ATP and the ATP regeneration system, demonstrating that the active hsp60 is able to enhance the client protein refolding (Fig. 4B). Furthermore, the comparison of ATPase activity of hsp60 with and without HEWL, quantified from the decay of PEP signals, reveals that the presence of client protein enhances the ATPase activity of hsp60 by 40% from 24.7 ± 0.5 to 36.0 ± 1.5 ATP/min per hsp60 (Figs. 3 and 4C and fig. S9). These observations indicate a mutual activation of protein refolding and cycling of the chaperonin between different functional states.

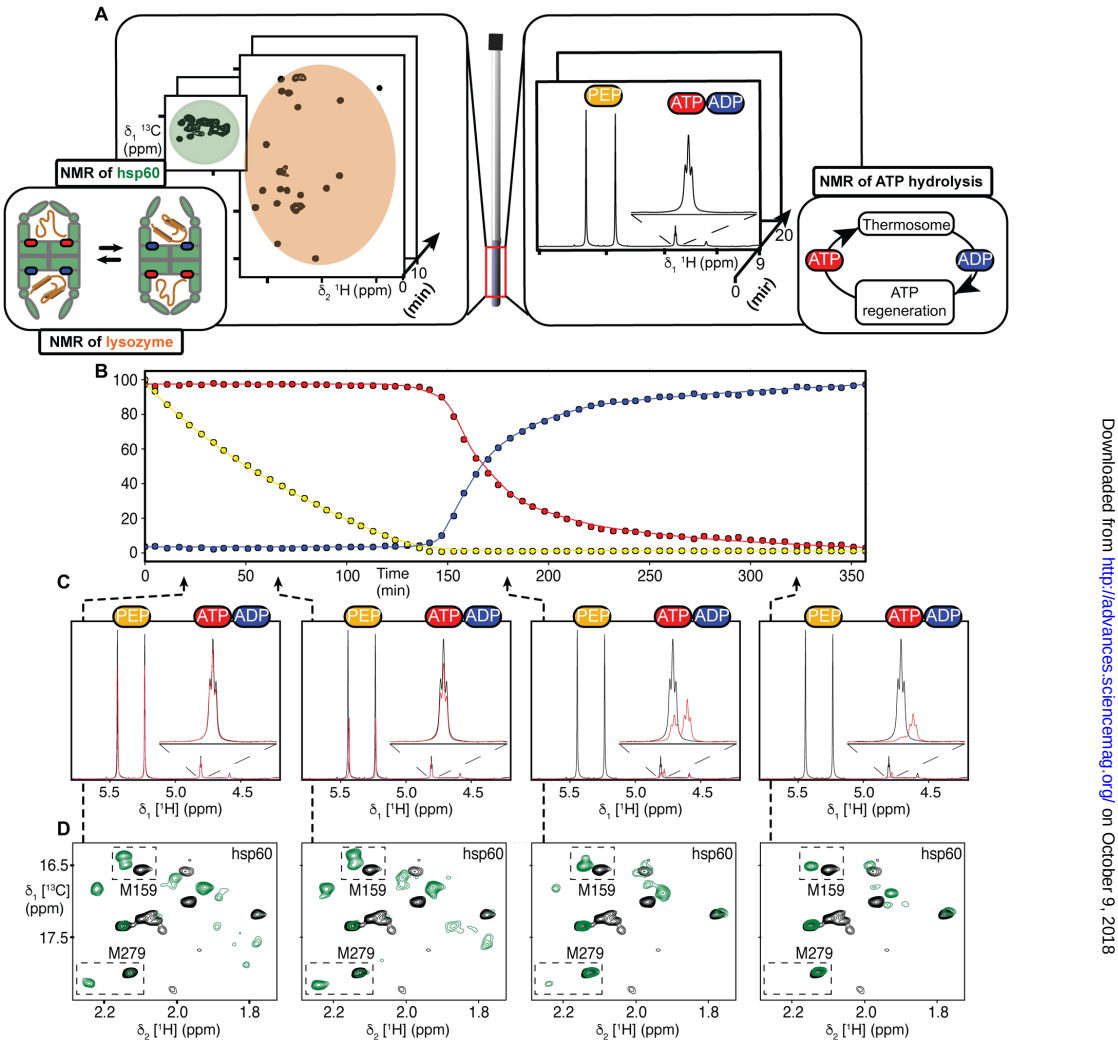


Fig. 3. Real-time NMR study of hsp60 functional cycle in the presence of a client protein. (**A**) Schematic representation of hsp60 functional turnover observed by real-time NMR in the presence of the ATP regeneration system. (**B**) Plot of the PEP, ATP, and ADP concentrations in function of time. Percentage of each species was extracted from the 1D ^{1}H NMR experiment. For PEP, the sum of the initial area of the two signals corresponding to PEP was considered as 100%. For the ATP and ADP, the percentage corresponds to a ratio of the area corresponding to the signal of each molecule. (**C**) 1D ^{1}H NMR spectra recorded at different time points of the experiment [black, reference spectrum at the beginning of the experiment; red, spectrum at the time indicated by an arrow on the graphic (B)]. The zoom on the upper right of each panel corresponds to the signal of ATP and ADP. (**D**) The 2D ^{13}C-^{1}H SOFAST-methyl-TROSY spectra recorded at different times of the experiment corresponding to the hsp60 methionine methyl signals [black, reference spectrum before addition of ATP; green, spectrum at the time indicated by an arrow on the graphic (B)].

Nucleotide binding and hydrolysis control the population of chaperonin conformations

While the chaperonin is being powered by the ATP regeneration system, the NMR reporter of the nucleotide binding state, M159, shows that both the ATP- and ADP-bound states are present in steady state, despite the fact that the concentration of free ADP in solution is negligible (Fig. 4D and fig. S10, A and B). The level of unoccupied binding site is below our detection limit (<3%), showing that the lifetime of nucleotide-free state is negligible compared to the ADP- and ATP-bound states. The reporters of the cavity conformational state (M275 and M279) reveal that there is also a mixture of open and closed states present under these steady-state

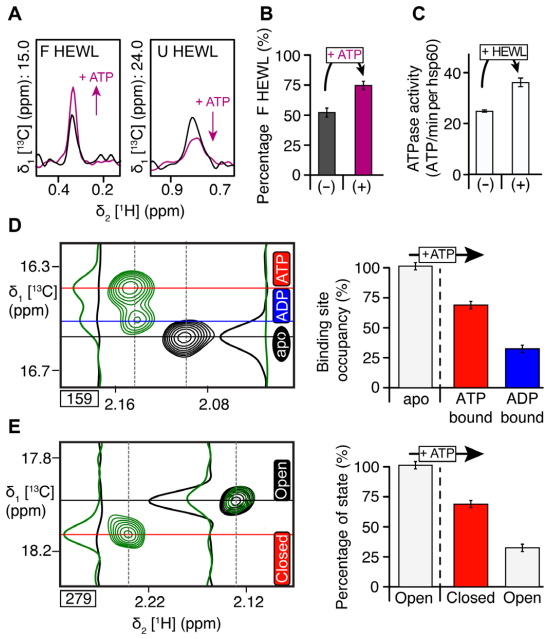


Fig. 4. NMR study of the active hsp60 machinery processing client protein. (**A**) The 1D traces of 2D ^{1}H-^{13}C NMR spectra (fig. S8A), displaying signals of folded (F, left) and unfolded (U, right) states of methyl (Ile-δ1, Val/Leu-proS)–labeled HEWL, recorded before/after the addition of 10 mM ATP. (**B**) Percentage of folded HEWL in the presence of the hsp60 before and after the addition of 10 mM ATP. (**C**) ATPase activity of hsp60 with and without HEWL, calculated from the decrease of the PEP signals (Fig.3 and fig. S9). (**D**) Quantification of the apo, ADP/ATP-bound populations of hsp60 in the presence of HEWL, quantified using intensities of NMR signals of M159 reporter (in black, reference spectrum of apo hsp60; green, spectra after the addition of 10 mM ATP). (**E**) Quantification of the relative populations of open and closed states of chaperonin over time in the presence of HEWL, as derived from the relative intensities of the signals of the two methionines 275 and 279 in the open/closed conformations. Spectra on the left show excerpts of NMR spectra before the addition of ATP (black) and after the addition of ATP (green).

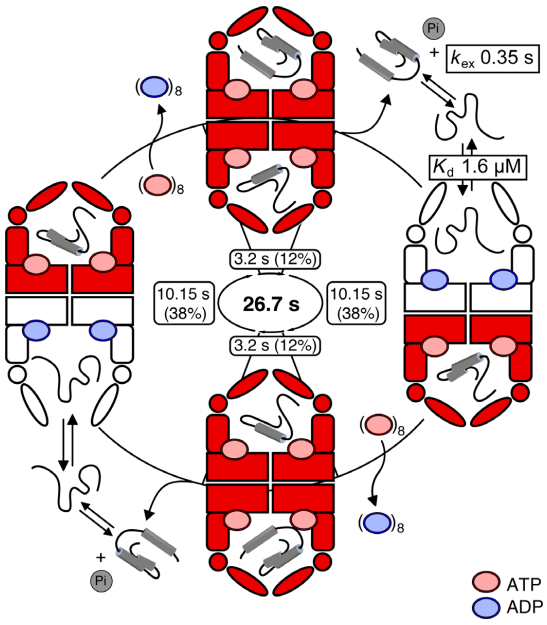


Fig. 5. Model of the functional cycle of chaperonin. Unfolded client protein is stabilized through its interaction with the equatorial domain of the ADP-loaded open hsp60 ring (Fig. 2E and fig. S7). The nucleotide binding sites, initially occupied with ADP, bind ATP without a detectable level of nucleotide-free state. The ATP binding, rather than its hydrolysis, closes the ring, which remains closed until ATP hydrolysis and P_i release. After hydrolysis, the ring reopens but ADP remains bound to the nucleotide binding sites, preventing immediate rebinding of new ATP and chaperonin closing. The chaperonin rings are represented by two symmetric monomers colored in white (inactive ring: ADP bound and open) or red (active ring: ATP bound and closed). The duration for each step of the functional cycle is calculated from the overall ATP hydrolysis rate and the relative populations of open/closed state, as reported in Figs.3 and 4 and fig. S10 and represented on the inside of the circle. K_d, dissociation constant.

conditions (Fig. 4E and fig. S10, C and D). These independent NMR observables allow the determination of population levels of the ATP-bound (62 ± 3%) and ADP-bound states (38 ± 3%), which quantitatively match the ones of the closed and open conformation, respectively (62 ± 3%, 38 ± 3%; Fig. 4, D and E). These relative population levels of open (ADP-bound) and closed (ATP-bound) states of chaperonin are retained also when client protein is present (fig. S10).

DISCUSSION

The chaperonin conformational cycle is driven by ATP; however, its exact role remains a matter of debate (*10*, *11*). A proposed model, in which ATP binding alone is sufficient to close the ring (*4*, *19*, *23*, *24*), was challenged by studies concluding that it is the ATP hydrolysis that drives the ring closing (*3*, *25*, *26*). The quantitative equality of the relative populations of ATP-bound (ADP-bound) states and closed (open) conformations of the chaperonin observed in this study (Fig. 5) implies that ATP binding and chamber closing are intimately linked, and along with the fact that nonhydrolyzable ATP analogs close the hsp60 cavity, these results confirm the view that nucleotide binding, rather than hydrolysis, induces the closure of the ring (Fig. 1B).

This study has also enabled resolution of the cooperativity and conformations of both rings during the adenosine triphosphatase (ATPase) cycle of group II chaperonins. The fact that we observe only two sets of NMR signatures, corresponding to closed/ATP-bound and open/ADP-bound states, leads us to propose that each ring comprising eight subunits acts as a cooperative unit, as proposed earlier (Figs. 3 and 4, D and E, and figs. S9 and S10) (*27*, *28*). This view is corroborated by the absence of mixed conformations within a ring in EM images (fig. S4). Several studies advocated the presence of inter-ring cooperativity in the group II chaperonins, suggesting an alternating mechanism in which one ring is open and the opposite

ring closed (*1*–*3*, *29*, *30*). However, there is also evidence of a nonconcerted mechanism, in particular for the eukaryotic chaperonin CCT (chaperonin containing TCP-1) (*31*, *32*). The extreme model with strict anticooperativity (*3*)—that is, opening of one ring leading to closing of the adjacent ring—predicts the exclusive presence of half-open/half-closed chaperonin particles, that is, populations consisting of 50% open and 50% closed half-rings at each time point. An alternative model based on partial anticooperativity includes doubly closed particles in addition to the half-open/half-closed state (*33*, *34*). The population levels obtained directly from our NMR data of the ATP-powered chaperonin in action allow resolution of these conflicting models. The skewed population levels (62:38 ± 3%) necessarily imply the presence of chaperonin particles that have two closed chambers. Taking into account that nucleotide-free states are absent (Fig. 4D), we propose a modified model (*2*), depicted in Fig. 5, which involves four states along the functional cycle, two of them fully closed and two half-open/half-closed states. The kinetics of ATP processing (Figs. 3 and 4) and the relative populations of states (Fig. 5), detected in our NMR sample, allow determining the average lifetimes of the populated states. With an ATP hydrolysis activity per chaperonin of 36 ATPs/min (Fig. 4C), and assuming that a full cycle consumes 16 ATPs, one catalytic turnover takes 26.7 s. Together with the experimentally observed population levels (Fig. 4, D and E), we can derive that the half-open/half-closed state has a threefold longer lifetime (10.2 s) than the doubly closed state (3.2 s; Fig. 5). The long residence time of the half-open/half-closed state allows the release of the folded proteins and the rebinding of a new unfolded client protein to the chaperonin. In addition, the fact that the ratio of populations in the open and closed conformations equals the ratio of ADP- and ATP-bound states shows that the events following the ATP hydrolysis are subsequently leading to the opening of the chaperonin folding chamber, while the ADP residence time controls the duration of the open state of hsp60 cavity.

Together, our results reveal that nucleotide binding, hydrolysis, and release control the conformational transitions between closed and open states and their relative lifetimes. While the open chaperonin conformation stabilizes the unfolded state of client proteins, its binding inside the cavity speeds up the functional cycle of the chaperonin. The implementation of a fully active chaperoning system, being constantly powered by fresh ATP inside the NMR tube, allows studying biomolecular function in action. This work establishes the feasibility to probe these events at the structural level even in systems as large as 1 MDa. We anticipate our approaches to be applicable to a wide range of protein assemblies, including HSPs (heat shock proteins) involved in cancer (*35*) or in the disaggregation of amyloidogenic proteins (*36*), as well as other large cellular machines in action such as biological motors or polymerases.

MATERIALS AND METHODS

Production and purification of methyl specifically labeled proteins in *Escherichia coli*

Production and purification of specifically methyl-labeled hsp60

E. coli BL21-CodonPlus-(DE3)-RIL cells transformed with a pET-41c plasmid encoding the hsp60 from *P. horikoshii* (also known as thermosome) were progressively adapted in three stages over 24 hours to M9/D_2O medium containing D-glucose-d_7 (2 g/liter; Isotec). As part of our standard labeling protocols, the culture media also contained $^{15}ND_4Cl$ (1 g/liter), but incorporation of ^{15}N in produced samples was not used in this study and these extra ^{15}N spins did not interfere with $^{13}CH_3$-edited NMR experiments reported here. In the final culture, bacteria were grown at 37°C in M9 medium prepared with 99.85% D_2O (Eurisotop). When the optical density (OD) at 600 nm reached 0.6 to 0.7, a solution containing the labeled precursors was added. The precursor solution added for 1 liter of culture medium contained the following: either 100 mg of [αβγ-$^2H_5$, ε-^{13}C]-L-methionine (Cambridge Isotope Laboratories; CIL) for the production of the U-[^2H, ^{15}N], Met-[$^{13}CH_3$] hsp60 sample; or 240 mg of 2-hydroxy-2-[^{13}C]methyl-3-oxo-4,4,4-tri-[^2H]-butanoate (pro-S acetolactate-^{13}C, NMR-Bio), 30 mg of L-leucine-d_{10}, and 100 mg of [α,β,γ-2H5, ε13C]-L-methionine (CIL) for the production of the U-[^2H, ^{15}N], Met-[$^{13}CH_3$], Val-[$^{13}CH_3$]^{proS} hsp60 sample (*15*).

One hour after the addition of the precursors, hsp60 expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The induced culture grew for 4 hours at 37°C before harvesting. Cells were lysed by sonication in buffer containing 25 mM Hepes (pH 7.5), 150 mM NaCl, 5 mM $MgCl_2$, and DNase (deoxyribonuclease) (0.01 mg/ml). After the first heat shock step at 80°C, hsp60 was purified using an anion exchange chromatography step (Resource Q, 6 ml, GE Healthcare) followed by a size exclusion chromatography step (HiLoad 16/60 Superdex 200 pg, GE Healthcare). The final yield reached 20 mg/liter of methyl-specific protonated hsp60. The protein was concentrated to final NMR concentration in D_2O buffer [20 mM MES (pH 6.5 uncorrected), 100 mM NaCl, 150 mM KCl, and 25 mM $MgCl_2$]. After concentration of the sample, the ATPase activity of hsp60 was tested using a malachite green assay (Sigma-Aldrich), and the oligomerization state of the particle was controlled using an analytical size exclusion column (CE Agilent Bio SEC-5A 500 Å) and negative staining EM.

Production and purification of hsp60 methionine and valine mutants

Constructs containing valine-to-alanine or methionine-to-leucine single point mutations were generated by an automated molecular biology platform (RoBioMol, Institut de Biologie Structurale) using an automated polymerase chain reaction–based protocol adapted from the QuikChange site-directed mutagenesis method (*37*). The library of mutants was expressed in parallel in 25 ml of M9/D_2O and labeled on methionine and valine methyl groups following the protocol described above. Purification of hsp60 mutants was done following the protocol described above. The hsp60 mutants were concentrated in D_2O buffer [20 mM MES (pH 6.5 uncorrected), 100 mM NaCl, 150 mM KCl, and 25 mM $MgCl_2$] using centrifugal filter devices (Vivaspin 15, 100,000 MWCO, Vivascience) to a final concentration between 6.25 and 12.5 μM hsp60 (100 to 200 μM monomer).

Production of U-[^2H, ^{15}N], Leu/Val-[$^{13}CH_3$]^{proS}, Ile-[$^{13}CH_3$]^{δ1} HEWL

E. coli BL21-CodonPlus-(DE3) cells transformed with a pET-11a plasmid encoding the HEWL (Addgene plasmid #39233) (*38*) were progressively adapted in three stages over 24 hours to M9/D_2O medium containing $^{15}ND_4Cl$ (1 g/liter) and D-glucose-d_7 (2 g/liter; Isotec). In the final culture, bacteria were grown at 37°C in M9 medium prepared with 99.85 % D_2O (Eurisotop). When the OD at 600 nm reached 0.6 to 0.7, 2-hydroxy-2-[^{13}C]methyl-3-oxo-4,4,4-tri-[^2H]butanoate (proS acetolactate-^{13}C, NMR-Bio) was added at a final concentration of 240 mg/liter 1 hour before induction. Forty minutes later (that is, 20 min before induction), 3,3-[$^2H_2$],4-[^{13}C]-2-ketobutyrate

(NMR-Bio) was added to a final concentration of 60 mg/liter (*15*). One hour after the addition of the first precursors, HEWL expression was induced by the addition of IPTG to a final concentration of 2 mM. The induced culture grew for 3 hours at 37°C before harvesting.

HEWL was purified from inclusion bodies (*38*) and refolded using size exclusion chromatography (*39*). The final yield reached 5 mg/liter of methyl-specific protonated refolded HEWL. The protein was concentrated using centrifugal filter devices (Vivaspin 15, 5000 MWCO, Vivascience) in D_2O buffer [20 mM MES (pH 6.5 uncorrected), 100 mM NaCl, 150 mM KCl, and 25 mM $MgCl_2$].

Production of U-[^{2}H, ^{15}N], Leu/Val-[$^{13}CH_3]^{proS}$, Ile-[$^{13}CH_3]^{\delta 1}$ MSG

E. coli BL21(DE3) cells were transformed by heat shock with the kanamycin-restricted T7lac promoter–controlled expression plasmid pET21b bearing the MSG gene. Cells were progressively adapted to M9/D_2O media containing $^{15}ND_4Cl$ (1 g/liter) and D-glucose-d_7 (2 g/liter; Isotec) in three stages over 24 hours. In the final culture, the bacteria were grown at 37°C in M9 media prepared with 99.85% D_2O (Eurisotop). When the OD at 600 nm reached 0.7, 2-hydroxy-2-[^{13}C]methyl-3-oxo-4,4,4-tri-[^{2}H]butanoate (pro-S acetolactate-^{13}C, NMR-Bio) was added at a final concentration of 240 mg/liter 1 hour prior to IPTG induction and 40 min later (that is, 20 min prior to induction), and 3,3-[$^2H_2$],4-[^{13}C]-2-ketobutyrate (NMR-Bio) was added to a final concentration of 60 mg/liter (*15*). The induction was performed overnight at 20°C. MSG was purified initially by Chelating Sepharose chromatography (GE Healthcare), followed by gel filtration chromatography (Superdex 200 pg, GE Healthcare). The typical final yield after purification was 80 mg/liter of methyl-specific protonated MSG.

Electron microscopy

Sample (0.1 mg/ml) was applied to the clean side of a carbon layer (between a carbon and a mica layer). The carbon was then floated on stain [2% (w/v) ammonium molybdate (pH 7.4)] and covered by a copper grid. Both were fished and air-dried before loading in the electron microscope. Images were taken with a CM12 microscope (FEI) operating at 120 kV and a nominal magnification of 45,000 times under low-dose conditions on an Orius CCD camera (Gatan).

NMR spectroscopy

All NMR experiments were recorded at specified conditions using 700-, 850-, or 950-MHz Bruker Avance III HD spectrometer equipped with a 1.7-mm (850 MHz) or 5-mm cryogenically cooled pulsed-field gradient triple-resonance probes (700, 850, or 950 MHz).

Two-dimensional (2D) SOFAST-methyl-TROSY NMR experiments (*18*) were recorded with an adjusted duration depending on the final concentration of the proteins (experimental time ranging from 10 to 120 min per sample). The angle of the proton excitation pulse was set to 30°, and the recycling delay was optimized to 0.6 s to achieve the highest sensitivity.

The 3D HMQC-NOESY-HMQC experiment was recorded over 90 hours with a 75 μM (470 μM monomer) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$], Val-[$^{13}CH_3]^{proS}$ sample of hsp60 with a nuclear Overhauser effect (NOE) mixing time of 600 ms (optimal NOE mixing time determined from the buildup of the cross-peak intensities in a series of short 2D NOESY spectra). NMR data were recorded on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 950 MHz and equipped with a 5-mm cryogenically cooled pulsed-field gradient triple-resonance probes.

All the 2D (^{13}C,^{1}H)-DOSY NMR experiments were recorded on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 700 MHz and equipped with a 5-mm cryogenically cooled pulsed-field gradient triple-resonance probes. A recycle delay of 1 s was used in all experiments. The diffusion time was set to 100 ms, gradient duration was 2 ms, and recovery delay was set to 0.2 ms, with a total acquisition time of 2 hours for each spectrum. The gradient strength varied from 5 to 42.5 G/cm.

^{13}C-edited 2D Exchange Spectroscopy (EXSY) experiments were acquired to characterize the exchange between folded and unfolded HEWL and to transfer assignment between the two forms. For this purpose, we have modified standard methyl-TROSY experiment (*17*) by adding a first 90° ^{1}H pulse just before proton acquisition to store magnetization along the *z* axis during the EXSY mixing delay. Then, the magnetization was flipped back in the transverse plane by the addition of a second 90° ^{1}H pulse for the detection of ^{1}H signal. All the 2D methyl-TROSY-EXSY experiments were acquired on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 700 MHz equipped with a 5-mm cryogenically cooled pulsed-field gradient triple-resonance probe. A recycle delay of 1 s was used in all experiments. The EXSY mixing times were set to 1, 25, 50, 100, 200, 300, 500, 800, 1000, and 1300 ms, with an average acquisition time of 30 min for each spectrum.

Assignment of methyl group resonances

Assignment of HEWL Ile-δ_1, Val and Leu-proS methyl groups

Folded HEWL assignment was obtained from the published assignment (*40*). Unfolded HEWL assignment was transferred from the folded HEWL using a methyl-TROSY-EXSY experiment recorded at 65°C, where both unfolded and folded populations of HEWL can be observed (fig. S5D).

Assignment of hsp60 Met-ε and Val-γ2 methyl groups

Each mutant sample (40 μl) was loaded in a 1.7-mm tube, and the NMR experiments were recorded on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 850 MHz and equipped with a 1.7-mm TCI MicroCryoProbe. The 2D SOFAST-methyl-TROSY NMR experiments (*18*) were recorded at 75°C with an adjusted duration depending on the final concentration of each mutant (experimental time ranging from 60 to 120 min per sample). For each mutant, spectra (fig. S1) were recorded for the apo state, AppNHp–bound state (1 mM), or ADP-bound state (1 mM). Analysis and comparison of the complete library of mutant spectra allowed the assignment of 33 valines (70%) and 15 methionines (100%). To obtain the missing assignment and confirm the assignment obtained by mutagenesis, a 3D HMQC-NOESY-HMQC experiment was recorded with a 47 μM (0.75 mM monomer) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$], Val-[$^{13}CH_3]^{proS}$ sample of hsp60 (fig. S2). Comparison of the NOE cross peaks with the model of the structure of hsp60 allowed to confirm the mutagenesis assignment and to assign 43 valines (92%) and 15 methionines (100%).

Study of the interaction between client proteins and hsp60

Titration of the interaction between HEWL and hsp60

Spectra were recorded using a different sample for each titration point with a fixed concentration of 31 μM U-[^{2}H, ^{15}N], Val/Leu-[$^{13}CH_3]^{proS}$, Ile-[$^{13}CH_3]^{\delta 1}$ HEWL and a variation of the concentration of U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$] hsp60. Each sample (40 μl) was loaded in a 1.7-mm tube, and the NMR experiments were recorded on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 850 MHz

and equipped with a 1.7-mm TCI MicroCryoProbe. The 2D SOFAST-methyl-TROSY NMR (*18*) experiments were recorded at 75°C.

Thermal unfolding of HEWL with and without hsp60

The reversible thermal (un)folding of HEWL (*41*) was probed using two samples. HEWL reference sample: 31 μM HEWL U-[^{2}H, ^{15}N], Val/Leu-[$^{13}CH_3$]^{proS}, Ile-[$^{13}CH_3$]$^{\delta 1}$. HEWL with hsp60: 15 μM hsp60 (240 μM monomer concentration) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$] sample and 31 μM HEWL U-[^{2}H, ^{15}N], Val/Leu-[$^{13}CH_3$]^{proS}, Ile-[$^{13}CH_3$]$^{\delta 1}$ sample (two equivalents of HEWL per hsp60 particle). Each sample (40 μl) was loaded in a 1.7-mm tube, and the NMR experiments were recorded on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 850 MHz and equipped with a 1.7-mm TCI MicroCryoProbe. The 2D SOFAST-methyl-TROSY NMR experiments (*18*) were recorded from a temperature of 50° to 75°C with steps of 5°C between 50° and 60°C and steps of 2.5°C between 60° and 75°C. Percentage of folded state was extracted from the average ratio between the intensity of the signals in the folded and unfolded states for different residues.

Measurement of the translational diffusion coefficient of the free and hsp60-bound HEWL and MSG

For HEWL, three samples were used to acquire the 2D (^{13}C,^{1}H)-DOSY data sets (*42*) at a temperature of 65°C. Reference sample: 31 μM HEWL U-[^{2}H, ^{15}N], Val/Leu-[$^{13}CH_3$]^{proS}, Ile-[$^{13}CH_3$]$^{\delta 1}$; hsp60 reference sample: 15.5 μM hsp60 (250 μM monomer concentration) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$]; and hsp60 with HEWL: 15,5 μM hsp60 (250 μM monomer concentration) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$] and 31 μM HEWL U-[^{2}H, ^{15}N], Val/Leu-[$^{13}CH_3$]^{proS}, Ile-[$^{13}CH_3$] $^{\delta 1}$ sample (two equivalents of HEWL per hsp60 particle).

For MSG, three samples were used to acquire the 2D (^{13}C,^{1}H)-DOSY data sets. As thermally unfolded MSG is not soluble, the MSG reference spectrum was acquired at 50°C with 31 μM folded MSG U-[^{2}H, ^{15}N], Ile-[$^{13}CH_3$] $^{\delta 1}$. For the sample containing unfolded MSG bound to hsp60, 15.5 μM hsp60 (250 μM monomer concentration) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$] was mixed to 31 μM MSG U-[^{2}H, ^{15}N], Ile-[$^{13}CH_3$] $^{\delta 1}$. MSG was thermally unfolded at 60°C in the presence of hsp60 (two equivalents of MSG per hsp60 particle), and DOSY (Diffusion Ordered SpectroscopY) spectra were acquired at 60°C. The hsp60 reference sample spectrum was acquired at 60°C with a sample containing 15.5 μM hsp60 (250 μM monomer concentration) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$].

Paramagnetic labeling of HEWL and MSG for the detection of intermolecular paramagnetic relaxation enhancement (PRE) effects

Spin labeling of the ε-amino groups of the solvent accessible HEWL (Sigma-Aldrich) or MSG lysine residues with OXYL-1-NHS (1-oxyl-2,2,5,5-tetramethylpyrolline-3-carboxylate-*N*-hydroxysuccimide ester, Toronto Research Chemicals) was carried out following a published protocol (*43*). Protein was dissolved in the labeling buffer [10 mM Na_2CO_3 (pH 9.2)] at a concentration of 40 μM. A stock solution was prepared by dissolving 10 mg of OXYL-1-NHS in 100 μl of dimethyl sulfoxide. After addition of a sixfold molar excess of OXYL-1-NHS over lysine residues to the protein solution, the reaction mixture was incubated for 1 hour at room temperature, followed by 4 hours at 4°C. The excess of spin label was removed by washing the sample with approximately 20 volumes of NMR buffer using centrifugal filter devices (Vivaspin 15, 5000 MWCO, Vivascience). Homogeneity of the lysine labeling was checked by mass spectrometry.

We measured paramagnetic relaxation (PRE) enhancement values, to provide qualitative information on the interactions between HEWL or MSG and hsp60, by comparing peak intensities in 2D SOFAST-methyl-TROSY NMR experiments recorded at 75°C with a 16 μM hsp60 (256 μM monomer concentration) U-[^{2}H] Met-[$^{13}CH_3$], Val-[$^{13}CH_3$]^{proS} in interaction with either 32 μM OXYL-1-NHS–labeled ^{1}H HEWL (I_{ox})/MSG (I_{ox}) or reference diamagnetic HEWL (I_{red})/MSG (I_{red}). The reference sample was obtained by reducing the spin label of the OXYL-1-NHS–labeled sample by incubation with sodium ascorbate (2.2 mM) for 24 hours at 4°C. NMR data were recorded using a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 950 MHz and equipped with a 5-mm cryogenically cooled pulsed-field gradient triple-resonance probes.

NMR study of the hsp60 in action with/without HEWL

Sample conditions

Two samples were used to record the data. Hsp60 reference sample: 26 μM hsp60 (416 μM monomer concentration) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$]. Hsp60 with HEWL: 26 μM hsp60 (416 μM monomer concentration) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$] and 52 μM HEWL U-[^{2}H, ^{15}N], Val/Leu-[$^{13}CH_3$]^{proS}, Ile-[$^{13}CH_3$]$^{\delta 1}$ (two equivalents of HEWL per hsp60 particle). All the samples were prepared in D_2O buffer containing 100 mM MES (pH 6.5 uncorrected), 100 mM NaCl, 150 mM KCl, and 25 mM $MgCl_2$.

Composition of the ATP regeneration system

ATP regeneration system was composed of 3 U of pyruvate kinase from *Bacillus stearothermophilus*, 0.1 mM ribulose-5-phosphate, and 220 mM PEP (Sigma-Aldrich).

Interleaved NMR experiments

The experiment was started by the addition of 10 mM ATP to the reaction mix, and the NMR tube was placed in the spectrometer at a temperature of 65°C. After an equilibration time of 5 min, 2D SOFAST-methyl-TROSY NMR spectra were recorded in interleaved mode with ^{1}H 1D spectra.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/9/eaau4196/DC1

Fig. S1. Examples of mutant spectra used to assign individual methionine and valine correlations.
Fig. S2. Cross-validation of the valine and methionine methyl group assignments.
Fig. S3. Interaction of hsp60 with different nucleotides.
Fig. S4. Conformational changes of hsp60 induced by a nonhydrolyzable ATP analog (App-NHp) investigated by EM and NMR.
Fig. S5. NMR characterization of the HEWL thermal unfolding without/with hsp60.
Fig. S6. Translational diffusion properties characterized by diffusion-ordered NMR spectroscopy.
Fig. S7. Determination of HEWL binding site on hsp60 using PRE experiments.
Fig. S8. Refolding of HEWL by ATP-powered chaperonin.
Fig. S9. Real-time NMR study of hsp60 functional cycle without client protein.
Fig. S10. Population of active hsp60 in ATP/ADP/apo state and closed/open state.
References (*44*, *45*)

REFERENCES AND NOTES

1. G. Kafri, A. Horovitz, Transient kinetic analysis of ATP-induced allosteric transitions in the eukaryotic chaperonin containing TCP-1. *J. Mol. Biol.* **326**, 981–987 (2003).
2. M. G. Bigotti, A. R. Clarke, Cooperativity in the thermosome. *J. Mol. Biol.* **348**, 13–26 (2005).
3. S. Reissmann, C. Parnot, C. R. Booth, W. Chiu, J. Frydman, Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. *Nat. Struct. Mol. Biol.* **14**, 432–440 (2007).
4. A. Nakagawa, K. Moriya, M. Arita, Y. Yamamoto, K. Kitamura, N. Ishiguro, T. Kanzaki, T. Oka, K. Makabe, K. Kuwajima, M. Yohda, Dissection of the ATP-dependent conformational change cycle of a group II chaperonin. *J. Mol. Biol.* **426**, 447–459 (2014).

5. I. G. Muñoz, H. Yébenes, M. Zhou, P. Mesa, M. Serna, A. Y. Park, E. Bragado-Nilsson, A. Beloso, G. de Cárcer, M. Malumbres, C. V. Robinson, J. M. Valpuesta, G. Montoya, Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nat. Struct. Mol. Biol.* **18**, 14–19 (2011).
6. J. H. Pereira, C. Y. Ralston, N. R. Douglas, D. Meyer, K. M. Knee, D. R. Goulet, J. A. King, J. Frydman, P. D. Adams, Crystal structures of a group II chaperonin reveal the open and closed states associated with the protein folding cycle. *J. Biol. Chem.* **285**, 27958–27966 (2010).
7. Y. Huo, Z. Hu, K. Zhang, L. Wang, Y. Zhai, Q. Zhou, G. Lander, J. Zhu, Y. He, X. Pang, W. Xu, M. Bartlam, Z. Dong, F. Sun, Crystal structure of group II chaperonin in the open state. *Structure* **18**, 1270–1279 (2010).
8. L. Ditzel, J. Löwe, D. Stock, K.-O. Stetter, H. Huber, R. Huber, S. Steinbacher, Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93**, 125–138 (1998).
9. Y. Zang, M. Jin, H. Wang, Z. Cui, L. Kong, C. Liu, Y. Cong, Staggered ATP binding mechanism of eukaryotic chaperonin TRiC (CCT) revealed through high-resolution cryo-EM. *Nat. Struct. Mol. Biol.* **23**, 1083–1091 (2016).
10. H. Yébenes, P. Mesa, I. G. Muñoz, G. Montoya, J. M. Valpuesta, Chaperonins: Two rings for folding. *Trends Biochem. Sci.* **36**, 424–432 (2011).
11. T. Lopez, K. Dalton, J. Frydman, The mechanism and function of group II chaperonins. *J. Mol. Biol.* **427**, 2919–2930 (2015).
12. M. Okochi, H. Matsuzaki, T. Nomura, N. Ishii, M. Yohda, Molecular characterization of the group II chaperonin from the hyperthermophilic archaeum *Pyrococcus horikoshii* OT_3. *Extremophiles* **9**, 127–134 (2005).
13. R. Sprangers, L. E. Kay, Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622 (2007).
14. J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, NMR analysis of a 900K GroEL–GroES complex. *Nature* **418**, 207–211 (2002).
15. R. Kerfah, M. J. Plevin, R. Sounier, P. Gans, J. Boisbouvier, Methyl-specific isotopic labeling: A molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **32**, 113–122 (2015).
16. A. M. Ruschak, L. E. Kay, Methyl groups as probes of supra-molecular structure, dynamics and function. *J. Biomol. NMR* **46**, 75–87 (2010).
17. V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, L. E. Kay, Cross-correlated relaxation enhanced $^{1}H–^{13}C$ NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420–10428 (2003).
18. C. Amero, P. Schanda, M. Asunción Durá, I. Ayala, D. Marion, B. Franzetti, B. Brutscher, J. Boisbouvier, Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies. *J. Am. Chem. Soc.* **131**, 3448–3449 (2009).
19. R. Iizuka, T. Yoshida, Y. Shomura, K. Miki, T. Maruyama, M. Odaka, M. Yohda, ATP binding is critical for the conformational change from an open to closed state in archaeal group II chaperonin. *J. Biol. Chem.* **278**, 44959–44965 (2003).
20. J. Zhang, M. L. Baker, G. F. Schröder, N. R. Douglas, S. Reissmann, J. Jakana, M. Dougherty, C. J. Fu, M. Levitt, S. J. Ludtke, J. Frydman, W. Chiu, Mechanism of folding chamber closure in a group II chaperonin. *Nature* **463**, 379–383 (2010).
21. J. Zhang, B. Ma, F. DiMaio, N. R. Douglas, L. A. Joachimiak, D. Baker, J. Frydman, M. Levitt, W. Chiu, Cryo-EM structure of a group II chaperonin in the prehydrolysis ATP-bound state leading to lid closure. *Structure* **19**, 633–639 (2011).
22. C. R. Booth, A. S. Meyer, Y. Cong, M. Topf, A. Sali, S. J. Ludtke, W. Chiu, J. Frydman, Mechanism of lid closure in the eukaryotic chaperonin TRiC/CCT. *Nat. Struct. Mol. Biol.* **15**, 746–753 (2008).
23. O. Llorca, J. Martín-Benito, J. Grantham, M. Ritco-Vonsovici, K. R. Willison, J. L. Carrascosa, J. M. Valpuesta, The "sequential allosteric ring" mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. *EMBO J.* **20**, 4065–4075 (2001).
24. I. Gutsche, J. Holzinger, N. Rauh, W. Baumeister, R. P. May, ATP-induced structural change of the thermosome is temperature-dependent. *J. Struct. Biol.* **135**, 139–146 (2001).
25. A. S. Meyer, J. R. Gillespie, D. Walther, I. S. Millet, S. Doniach, Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* **113**, 369–381 (2003).
26. N. R. Douglas, S. Reissmann, J. Zhang, B. Chen, J. Jakana, R. Kumar, W. Chiu, J. Frydman, Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. *Cell* **144**, 240–252 (2011).
27. H.-y. Chen, X.-l. Tan, J. Lu, C.-x. Zhang, Y. Zhang, S.-l. Yang, Characterization of ATPase activity of class II chaperonin from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biotechnol. Lett.* **31**, 1753–1758 (2009).
28. A. R. Kusmierczyk, J. Martin, Nested cooperativity and salt dependence of the ATPase activity of the archaeal chaperonin Mm-cpn. *FEBS Lett.* **547**, 201–204 (2003).
29. G. Kafri, K. Y. Willison, A. Horovitz, Nested allosteric interactions in the cytoplasmic chaperonin containing TCP-1. *Protein Sci.* **10**, 445–449 (2001).
30. A. Horovitz, Y. Fridmann, G. Kafri, O. Yifrach, Review: Allostery in chaperonins. *J. Struct. Biol.* **135**, 104–114 (2001).
31. M. G. Bigotti, S. R. W. Bellamy, A. R. Clarke, The asymmetric ATPase cycle of the thermosome: Elucidation of the binding, hydrolysis and product-release steps. *J. Mol. Biol.* **362**, 835–843 (2006).
32. P. Lin, F. Sherman, The unique hetero-oligomeric nature of the subunits in the catalytic cooperativity of the yeast Cct chaperonin complex. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10780–10785 (1997).
33. D. Rivenzon-Segal, S. G. Wolf, L. Shimon, K. R. Willison, A. Horovitz, Sequential ATP-induced allosteric transitions of the cytoplasmic chaperonin containing TCP-1 revealed by EM analysis. *Nat. Struct. Mol. Biol.* **12**, 233–237 (2005).
34. A. Horovitz, K. R. Willison, Allosteric regulation of chaperonins. *Curr. Opin. Struct. Biol.* **15**, 646–651 (2005).
35. J. Wu, T. Liu, Z. Rios, Q. Mei, X. Lin, S. Cao, Heat shock proteins and cancer. *Trends Pharmacol. Sci.* **38**, 226–256 (2017).
36. B. Ojha, N. Fukui, K. Hongo, T. Mizobata, Y. Kawata, Suppression of amyloid fibrils using the GroEL apical domain. *Sci. Rep.* **6**, 31041 (2016).
37. C. D. Amero, M. A. Durá, M. Noirclerc-Savoye, A. Perollier, B. Gallet, M. J. Plevin, T. Vernet, B. Franzetti, J. Boisbouvier, A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J. Biomol. NMR* **50**, 229–236 (2011).
38. C. Schlörb, K. Ackermann, C. Richter, J. Wirmer, H. Schwalbe, Heterologous expression of hen egg white lysozyme and resonance assignment of tryptophan side chains in its non-native states. *J. Biomol. NMR* **33**, 95–104 (2005).
39. Z. Gu, Z. Su, J.-C. Janson, Urea gradient size-exclusion chromatography enhanced the yield of lysozyme refolding. *J. Chromatogr. A* **918**, 311–318 (2001).
40. H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, L. J. Smith, A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci.* **10**, 677–688 (2001).
41. F. Meersman, C. Atilgan, A. J. Miles, R. Bader, W. Shang, A. Matagne, B. A. Wallace, M. H. J. Koch, Consistent picture of the reversible thermal unfolding of hen egg-white lysozyme from experiment and molecular dynamics. *Biophys. J.* **99**, 2255–2263 (2010).
42. T. Didenko, R. Boelens, S. G. D. Rüdiger, 3D DOSY-TROSY to determine the translational diffusion coefficient of large protein complexes. *Protein Eng. Des. Sel.* **25**, 319 (2012).
43. J.-J. Lawrence, L. Berne, J. L. Ouvrier-Buffet, L. H. Piette, Spin-label study of histone H1–DNA interaction. *Eur. J. Biochem.* **107**, 263–269 (1980).
44. E. Crublet, R. Kerfah, G. Mas, M. Noirclerc-Savoye, V. Lantez, T. Vernet, J. Boisbouvier, A cost-effective protocol for the parallel production of libraries of $^{13}CH_3$-specifically labeled mutants for NMR studies of high molecular weight proteins. *Methods Mol. Biol.* **1091**, 229–244 (2014).
45. G. Mas, E. Crublet, O. Hamelin, P. Gans, J. Boisbouvier, Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J. Biomol. NMR* **57**, 251–262 (2013).

Acknowledgments: The authors thank B. Brutscher, I. Gutsche, and S. Hiller for the stimulating discussions and comments on the manuscript. **Funding:** This work used the high-field NMR and EM facilities at the Grenoble Instruct-ERIC Center (ISBG; UMS 3518 CNRS CEA-UJF-EMBL) with support from the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INSB-05-02) and the Grenoble Alliance for Integrated Structural Cell Biology (GRAL; ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology. The electron microscope facility is supported by the Rhône-Alpes Region, the Fondation Recherche Medicale (FRM), the Fonds Européen de Développement Régional, and the GIS–Infrastrutures en Biologie Santé et Agronomie (IBISA). This work was supported by the European Research Council (ERC-StG-2012-311318 and ERC-CoG-2010-260887). **Author contributions:** J.B., G.M., P.M., and P.S. designed the experiments, analyzed the results, and wrote the manuscript. G.M., J.-Y.G., E.C., and E.C.D. prepared the samples. C.M. and G.S. performed the EM experiments. G.M., J.-Y.G., E.C., E.C.D., J.B., P.G., P.M., and P.S. performed the NMR experiments. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 8 June 2018
Accepted 1 August 2018
Published 19 September 2018
10.1126/sciadv.aau4196

Citation: G. Mas, J.-Y. Guan, E. Crublet, E. C. Debled, C. Moriscot, P. Gans, G. Schoehn, P. Macek, P. Schanda, J. Boisbouvier, Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Sci. Adv.* **4**, eaau4196 (2018).

Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/4/9/eaau4196/DC1

Supplementary Materials for

Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle

Guillaume Mas, Jia-Ying Guan, Elodie Crublet, Elisa Colas Debled, Christine Moriscot, Pierre Gans, Guy Schoehn, Pavel Macek*, Paul Schanda*, Jerome Boisbouvier*

*Corresponding author. Email: macek@nmr-bio.com (P.M.); paul.schanda@ibs.fr (P.S.); jerome.boisbouvier@ibs.fr (J.B.)

Published 19 September 2018, *Sci. Adv.* **4**, eaau4196 (2018)
DOI: 10.1126/sciadv.aau4196

This PDF file includes:

Fig. S1. Examples of mutant spectra used to assign individual methionine and valine correlations.
Fig. S2. Cross-validation of the valine and methionine methyl group assignments.
Fig. S3. Interaction of hsp60 with different nucleotides.
Fig. S4. Conformational changes of hsp60 induced by a nonhydrolyzable ATP analog (App-NHp) investigated by EM and NMR.
Fig. S5. NMR characterization of the HEWL thermal unfolding without/with hsp60.
Fig. S6. Translational diffusion properties characterized by diffusion-ordered NMR spectroscopy.
Fig. S7. Determination of HEWL binding site on hsp60 using PRE experiments.
Fig. S8. Refolding of HEWL by ATP-powered chaperonin.
Fig. S9. Real-time NMR study of hsp60 functional cycle without client protein.
Fig. S10. Population of active hsp60 in ATP/ADP/apo state and closed/open state.
References (*44*, *45*)

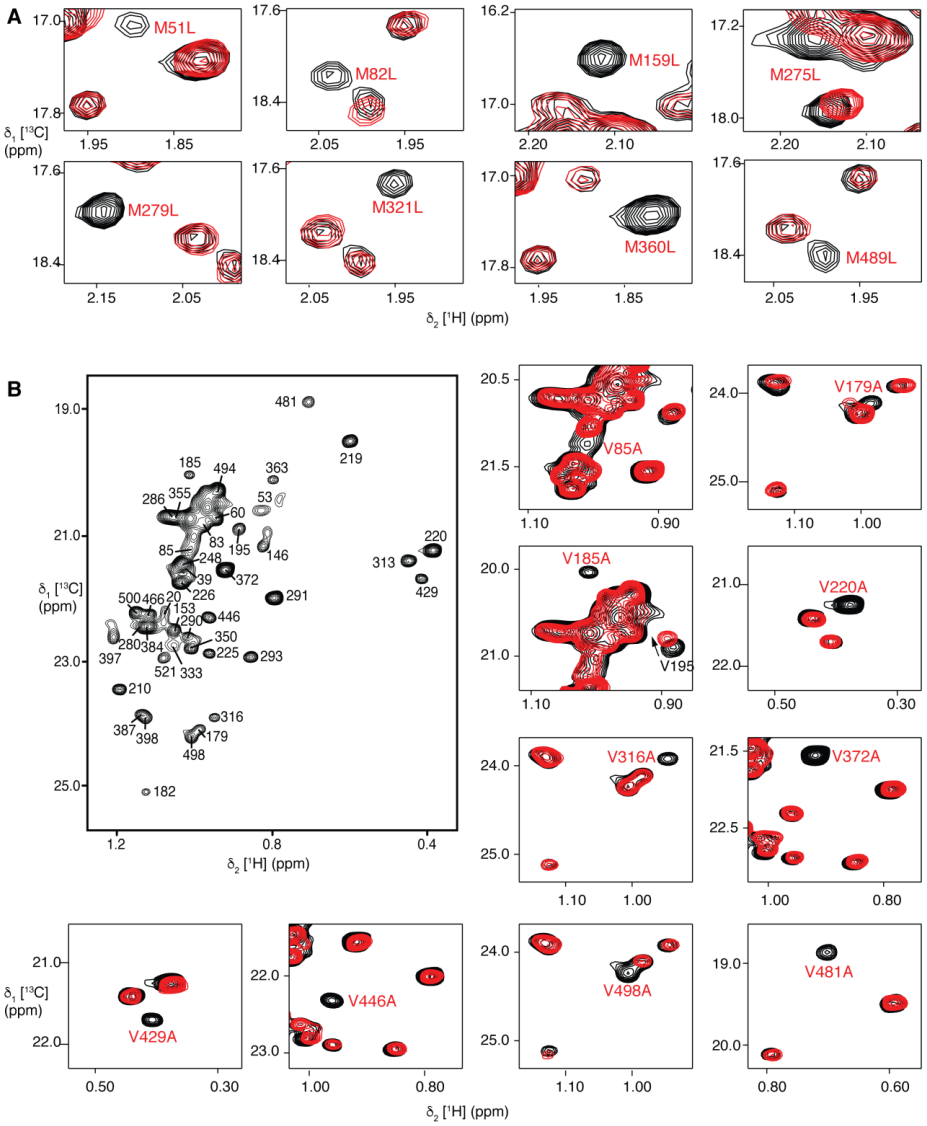


Fig. S1. Examples of mutant spectra used to assign individual methionine and valine correlations. (**A**) Examples of methionine-to-leucine mutants (*37*, *44*, *45*) (residues 51, 82, 159, 275, 279, 321, 360, 489) of hsp60. (**B**) Assignment of the hsp60 Val-[$^{13}CH_3$]^{proS} methyl groups and example of valine-to-alanine mutants (85, 179, 185, 220, 316, 372, 429, 446, 498, 481) of hsp60. 2D ^{13}C-^{1}H SOFAST-methyl-TROSY spectra displayed in this figure were recorded using samples of U-[^{2}H,^{15}N], Val-[$^{13}CH_3$]^{proS} hsp60 (*43*) (A) or U-[^{2}H,^{15}N], Met-[$^{13}CH_3$] hsp60. 2D extracts of spectra for examples of single-point mutant of hsp60 (plotted in red) are overlaid with the reference spectra of the wild-type particle (black). The assignment inferred for the missing resonance in the mutant spectrum is indicated. Spectra of hsp60 mutants were recorded in 1 to 2h (depending on mutant concentration) on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 850 MHz and equipped with a 1.7 mm cryogenic probe.

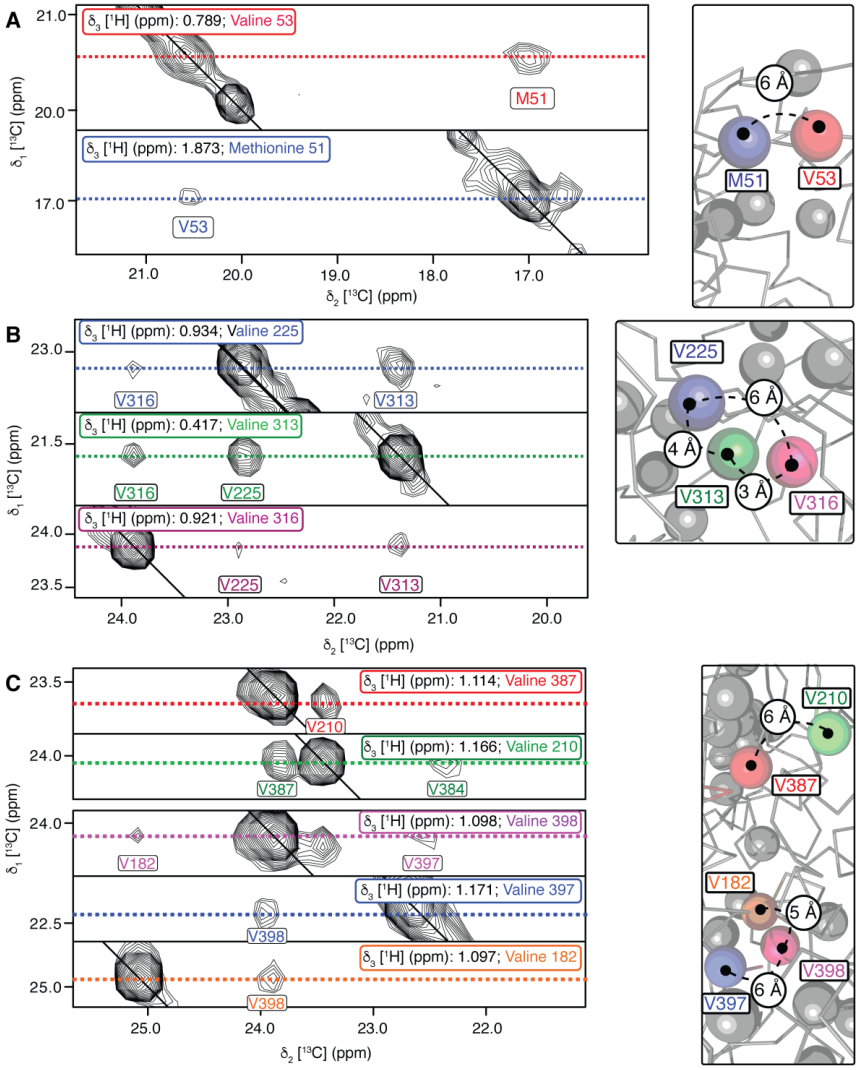


Fig. S2. Cross-validation of the valine and methionine methyl group assignments. 3D HMQC–NOESY-HMQC spectrum was recorded with a sample of 47 µM (750 µM of monomer) U-[^{2}H, ^{15}N], Val-[^{13}CH$_3$]^{proS}, Met-[^{13}CH$_3$] hsp60 sample on a Bruker Avance III HD spectrometer operating at a ^{1}H frequency of 950 MHz and equipped with a 5 mm cryogenic probe. For each example (A, B, C), extracts of the plane at the ^{1}H frequencies for each residue are shown and the ^{1}H-^{1}H distances corresponding to observed NOEs are displayed in the model of the 3D structure of the open hsp60. Lines represent the detected NOE connections between the methyl groups, where each colour is related to the corresponding residues annotated in the depicted 2D extracts.

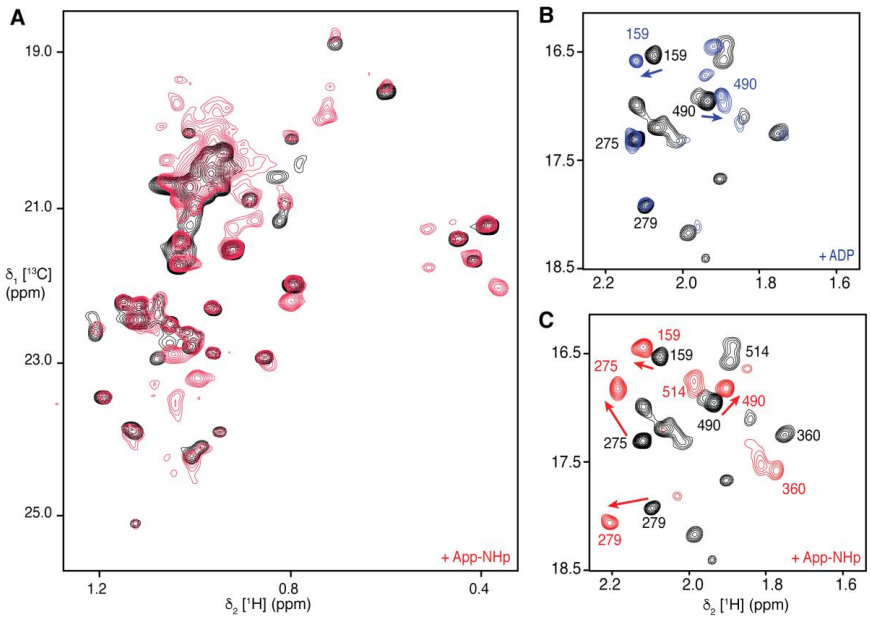


Fig. S3. Interaction of hsp60 with different nucleotides. (**A**) Overlay of the 2D ^{1}H-^{13}C SOFAST-methyl-TROSY spectra of a 12.5 µM hsp60 (200 µM of monomer) U-[^{2}H, ^{15}N], Val-[$^{13}CH_3$]^{proS} sample without (black) and in presence of 2 mM App-NHp (red). (**B** and **C**) Overlay of the 2D ^{13}C^{1}H SOFAST -methyl-TROSY spectra of a 12.5 µM hsp60 (200 µM of monomer) U-[^{2}H, ^{15}N], Met-[$^{13}CH_3$] sample without (black) and with 2 mM App-NHp (red) or 2 mM of ADP (blue). Assignment of the residues in the nucleotide loaded state assigned by mutagenesis is indicated in blue (ADP) or red (App-NHp).

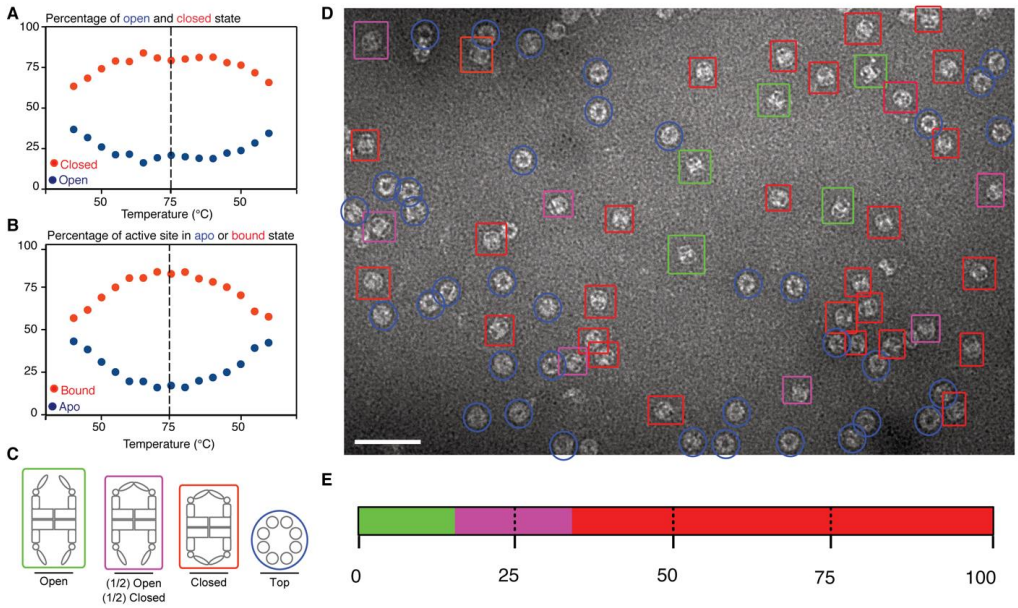


Fig. S4. Conformational changes of hsp60 induced by a nonhydrolyzable ATP analog (App-NHp) investigated by EM and NMR. (**A**) Equilibrium between the open and closed conformations of hsp60 at different temperatures in presence of 2 mM App-NHp. Spectra were recorded at various temperatures increasing from 40°C and up to 75°C and followed by a step by step cool-down from 75°C to 40°C. The percentage of each conformation was extracted from the *ratio* of the intensity of the residues 275 and 279 in the open and closed conformations (Fig. 1C). (**B**) Equilibrium between the nucleotide bound and apo states of the hsp60 at different temperatures in presence of 2 mM App-NHp. The percentage of each state was extracted from the *ratio* of the intensity of the residues 159 and 490 in the bound and apo states (Fig. 1C and fig. S3B). (**C**) Schematic representation of the different possible conformations of hsp60. (**D**) Electron microscopy picture (49000X) with negative staining (ammonium molybdate 2%, pH 7.4) of hsp60 in presence of App-NHp, recorded with a Phillips CM12 electron microscope. Hsp60 sample at a concentration of 1 mg/mL was heated at 75°C for 10 minutes in presence of 2 mM App-NHp just before to be fixed on the carbon grid at room temperature. Colour code used in the picture to show the conformational states of the hsp60 is reported on (C). The scale represents 100 nm (**E**) Statistical analysis of the different conformations corresponding to 10 different EM pictures with 849 particles (47% side view and 53% top view) using the same colour code as in (C).

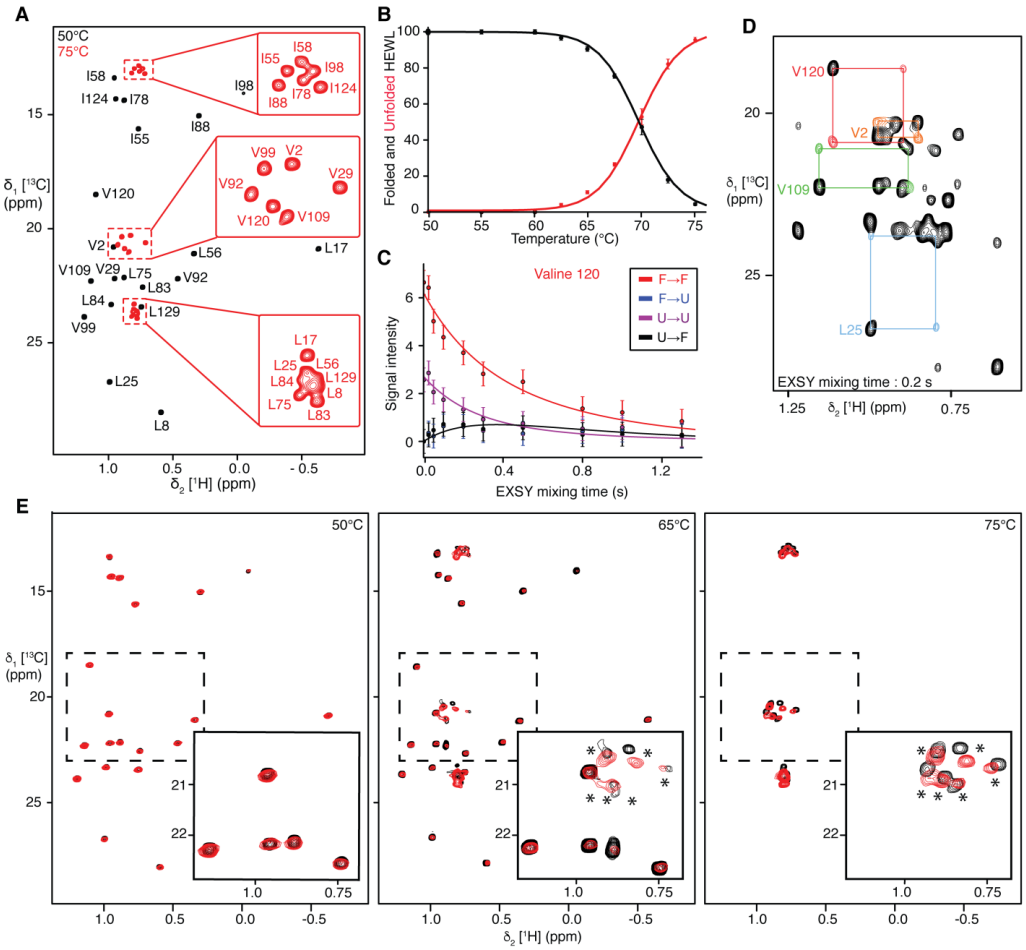


Fig. S5. NMR characterization of the HEWL thermal unfolding without/with hsp60. (**A**) 2D ^{1}H-^{13}C SOFAST-methyl-TROSY spectra of a 31 µM HEWL U-[^{2}H, ^{15}N], Ile-[$^{13}CH_3$]$^{\delta 1}$, Val/Leu-[$^{13}CH_3$]^{proS} sample at 50°C (black) and 75°C (red). Assignment of methyl groups in folded (black) and unfolded (red) state is shown on respective spectra. (**B**) Evolution of the folded and unfolded population of HEWL between 50 and 75°C. Percentage of both states was extracted from the *ratio* between the intensity of the signals in the folded and unfolded states. (**C**) Fit of diagonal and transfer signals detected in 2D methyl-TROSY-EXSY experiments for the valine 120 proS methyl group. The fitted exchange constant between the folded and unfolded states of the HEWL is $k_{ex}=k_{F\rightarrow U}+k_{U\rightarrow F}$ = 2.9 s^{-1}. (**D**) 2D methyl-TROSY-EXSY spectrum recorded on a 61 µM U-[^{2}H, ^{15}N], Ile-[$^{13}CH_3$]$^{\delta 1}$, Val/Leu-[$^{13}CH_3$]^{proS} HEWL sample at 65°C with an exchange delay of 0.2 s used to transfer the assignment from the folded to unfolded state of the lysozyme. Coloured lines represent the connections between the folded and unfolded signals. Corresponding cross-peaks are displayed in the same colour as the residue labels. (**E**) 2D ^{1}H-^{13}C SOFAST-methyl-TROSY spectra of a 31 µM U-[^{2}H, ^{15}N], Ile-[$^{13}CH_3$]$^{\delta 1}$, Val/Leu-[$^{13}CH_3$]^{proS} lysozyme sample without (black) and with (red) 15 µM hsp60 (240 µM of monomer). Spectra were recorded at 50°C (left panel), 65°C (middle panel) or 75°C (right panel). Zoom on a set of signals corresponding to the folded and unfolded states are presented in the inserts. The signals of the unfolded state are highlighted by an asterisk.

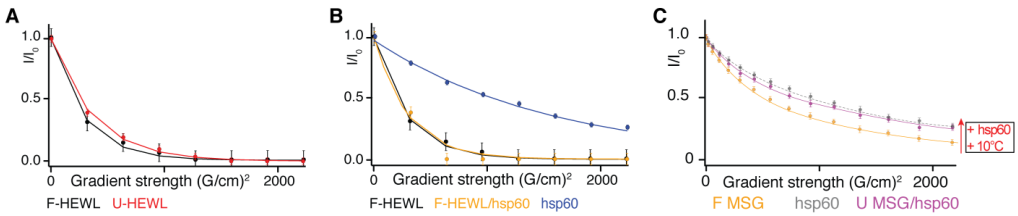


Fig. S6. Translational diffusion properties characterized by diffusion-ordered NMR spectroscopy (*42*) of (**A**) folded HEWL (F-HEWL, black curve) and unfolded HEWL (U-HEWL, red curve) at 65°C without hsp60; (**B**) folded HEWL without hsp60 (F-HEWL, black curve), folded HEWL with hsp60 (F-HEWL, yellow curve) and reference for hsp60 (blue curve) at 65°C; (**C**) Folded MSG at 50°C (F MSG, yellow curve), unfolded MSG at 60°C in presence of hsp60 (U MSG/hsp60, purple curve) and reference for hsp60 alone at 60°C (hsp60, grey curve). Thermally unfolded MSG is not soluble enough to be detected alone using solution NMR. Therefore, diffusion of folded MSG (50°C) was compared with diffusion of thermally unfolded MSG (60°C) stabilized by hsp60. An increase of temperature is expected to increase translational diffusion rate of monomeric protein. Experimentally, the observed diffusion rate of MSG at 60°C, in presence of hsp60, is reduced to the level of hsp60 indicating that unfolded MSG is bound to hsp60 and diffuse as a 1 MDa complex.

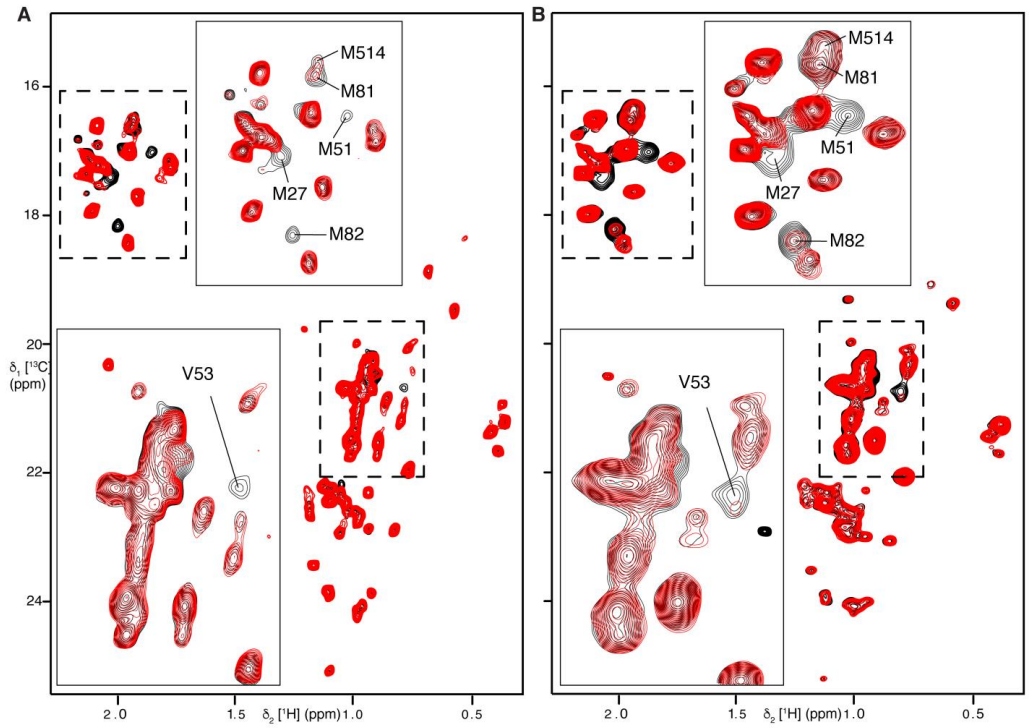


Fig. S7. Determination of HEWL binding site on hsp60 using PRE experiments. Overlay of 2D $^{13}C^{1}H$ SOFAST-methyl-TROSY spectra of a U-[^{2}H, ^{15}N], Val-[$^{13}CH_3$]^{proS}, Met-[$^{13}CH_3$] hsp60 sample, with: (**A**) diamagnetic lysozyme (black) and with paramagnetic lysozyme (red); (**B**) with diamagnetic MSG (black) and with paramagnetic MSG (red). NMR spectra were acquired at 75°C using a 16 µM sample of methyl labelled hsp60 in presence of a two-fold excess of paramagnetically labelled client proteins (HEWL or MSG).

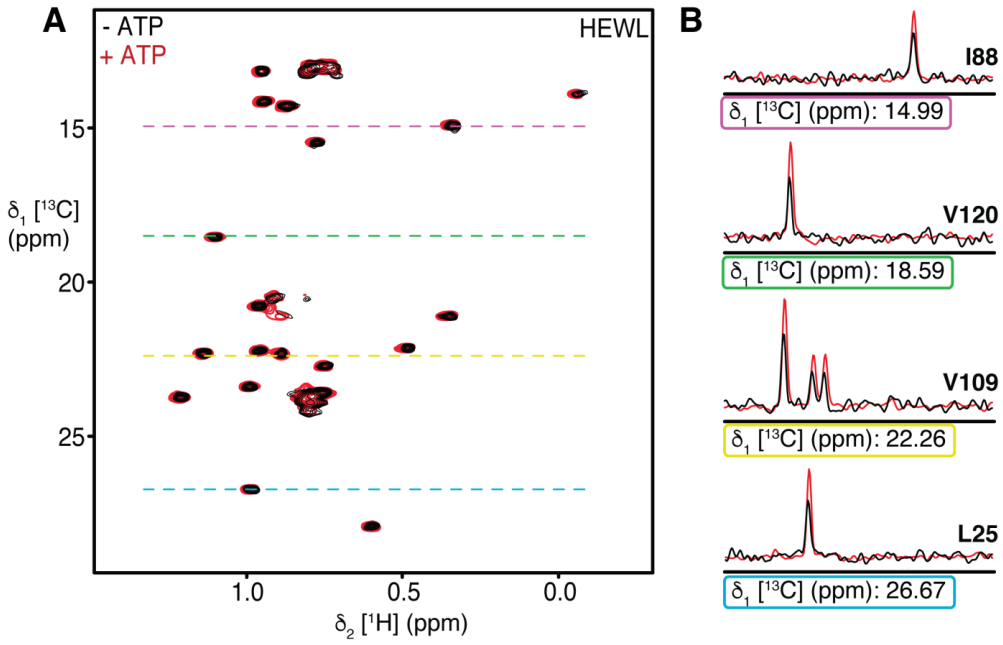


Fig. S8. Refolding of HEWL by ATP-powered chaperonin. (**A**) 2D ^{13}C-^{1}H SOFAST-methyl-TROSY spectra recorded before addition of ATP (black) or 30 minutes after addition of 10 mM of ATP, corresponding to the HEWL Ile-[$^{13}CH_3$]$^{\delta 1}$ and Val/Leu-[$^{13}CH_3$]^{proS} signals. The coloured dotted lines correspond to the 1D extract presented in (B). (**B**) 1D extracts from the 2D ^{13}C-^{1}H SOFAST-methyl-TROSY spectra presented in (A). Before addition of ATP (black) and 30 minutes after addition of ATP (red). For each extract, the chemical shift in the ^{13}C dimension is indicated below the spectrum.

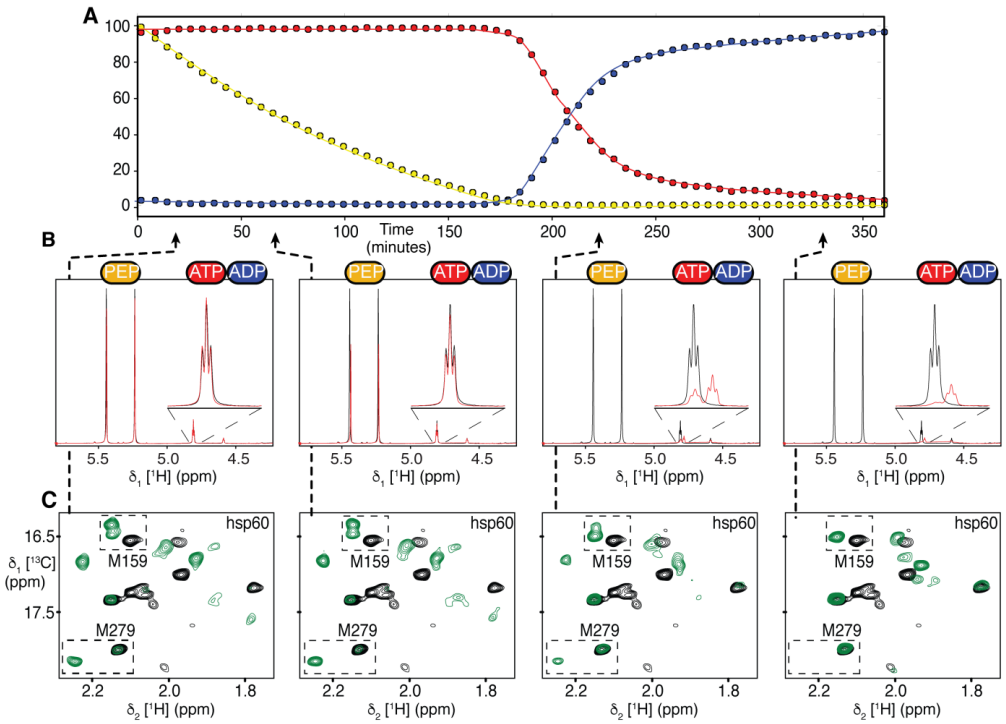


Fig. S9. Real-time NMR study of hsp60 functional cycle without client protein. (**A**) Plot of the evolution of PEP, ATP and ADP concentrations in function of time. Percentage of each species was extracted from 1D ^{1}H NMR experiment as indicated in Fig. 3. (**B**) 1D ^{1}H NMR spectra recorded at different time points of the experiment (black: reference spectrum at the beginning of the experiment; red: spectrum at the time indicated by an arrow on the graphic (A)). The zoom on the upper right of each panel corresponds to the signal of ATP and ADP. (**C**) 2D ^{13}C-^{1}H SOFAST-methyl-TROSY spectra recorded at different time of the experiment corresponding to the hsp60 methionine methyl signals. Black: reference spectrum before addition of ATP; green: spectrum at the time indicated by an arrow on the graphic (A).

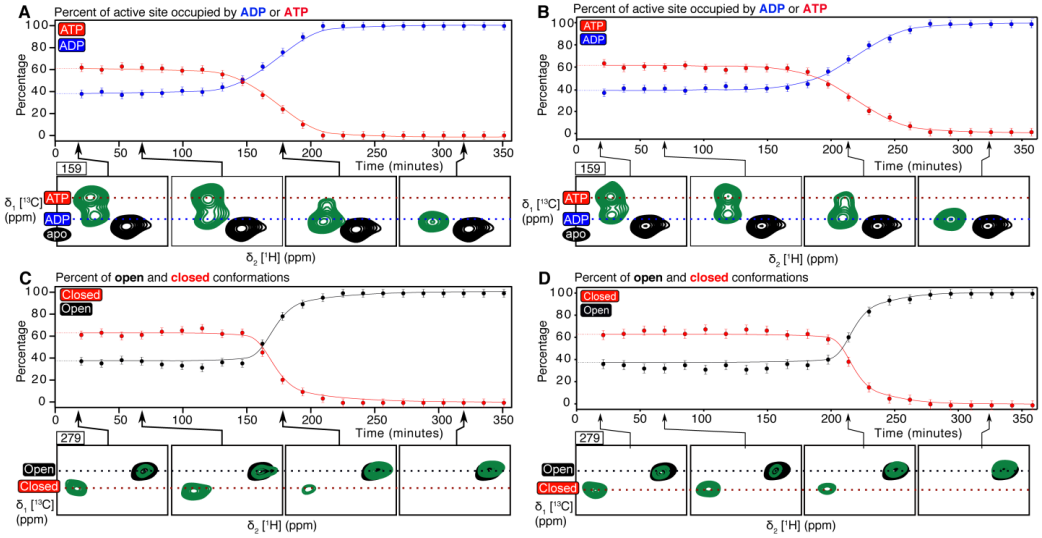


Fig. S10. Population of active hsp60 in ATP/ADP/apo state and closed/open state. The data reported have been acquired in presence of HEWL (left panels A and C) or without HEWL (right panels B and D). Quantification of hsp60 binding site occupied by ATP or ADP. Graphic showing the evolution of ATP and ADP occupied binding site. The percentage of ADP and ATP occupied sites had been extracted from the intensity *ratio* of M159 in both states. Zoom on the 2D (^{13}C-^{1}H)-SOFAST-methyl-TROSY spectra recorded at different times of the experiment corresponding to the methionine 159 signal (black: reference spectrum before addition of 10 mM of ATP; green: spectrum at the time indicate by an arrow on the graphic (A)). (**C** and **D**) Quantification of hsp60 open and closed conformations. Plot showing the evolution of hsp60 open and closed conformations. The percentage of open and closed hsp60 was extracted from the intensities *ratio* of M275 and M279 in both states. Zoom on the 2D ^{1}H-^{13}C SOFAST-methyl-TROSY spectra recorded at different time of the experiment corresponding to the methionine 275 signal (black: reference spectrum before addition of 10 mM of ATP; green: spectrum at the time indicated by an arrow on the graphic (A)).