Interactions between components of the protein homeostasis network

Wiktoria Laura Sadowska

New College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Physical and Theoretical Chemistry

January 2024

Abstract

An intricate set of processes responsible for maintaining the stability of the proteome forms a protein homeostasis network, which is crucial for cell viability and survival. This thesis describes insights into the interactions of the components that belong to that network. We focus on small heat-shock proteins (sHsps), significant members of the chaperone machinery, and glucosidase II (GluII), which is a key enzyme of the ER quality control system. We utilise native mass spectrometry, supported by other biophysical techniques, to broaden the understanding of studied protein-protein and protein-ligand complexes.

First, we characterise the properties of organelle sHsps, a family unique to plants. We show their structural diversification from cytosolic homologs. We find that two of the examined sHsps exhibit polydispersity, never reported in plant sHsps before. Further, we show evidence of dual monomer architecture within higher oligomers among organelle-targeted sHsps.

Then, we demonstrate that organelle-localised sHsps from different evolutionary classes, as well as those from different cellular compartments, co-assemble with diverse modes of interactions. We quantify the energy biases of those interactions using a developed statistical-mechanical model. Studied co-assemblies exhibit a range from a slight to strong preference toward homo-interactions.

We next investigate the subunit exchange between the organelle and cytosolic sHsps. The results indicate that the diversification of the α-crystallin domain led to the incompatibility in the dimerisation interfaces between these two groups. However, more flexible C-terminal and N-terminal regions maintain the assembly of homo-dimer subunits into higher hetero-oligomers.

Lastly, we determine the mode of interactions between GluII and a thiopyridone ligand, a potential host-targeted antiviral drug. Based on the obtained results, we propose a 'dock, lock and latch' molecular mechanism, where the ligand accesses the binding pocket near the dimerisation interface and then binds to the available cysteine through a disulfide bond. This binding induces structural rearrangement followed by dimer dissociation.

Interactions between components of the protein homeostasis network



Wiktoria Laura Sadowska

New College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Physical and Theoretical Chemistry

January 2024

Acknowledgements

First and foremost, my gratitude goes to my supervisor, Professor Justin Benesch, for the invaluable opportunity to pursue a DPhil under his guidance. His mentorship over the past four years has been immensely valuable. Justin has an exceptional ability to perceive whether his students need support and reassurance, or encouragement and a gentle push towards motivation, and he has always been readily available to provide it. He sparked my passion for mass spectrometry and led me through its complexities with immense kindness, ensuring that I learned as much as possible and got all the support I needed.

Although there have been difficult and challenging moments, this journey has been the most incredible adventure of my life. I want to thank with all my heart everyone who contributed to making it such a remarkable experience.

Throughout the years, Justin brought amazing people to our group, many of whom have become my closest friends. Their presence created an exceptional work environment, where I not only received invaluable scientific assistance, but also moral support and shared countless laughs. I would like to extend special thanks to Neel, Konstantin, Josh, and Angela. I am particularly grateful to two special people: Dom, who provided enormous scientific support and help in the lab during the early days, but most importantly over time, became an older brother far away from home; and Ellie, with whom the science was filled with fun and excitement. Time in the lab with her felt like the best way to spend the day. I am thankful for the endless dancing and singing sessions by the mass spectrometers, our unplanned matching outfits, comforting lunch breaks, and the fantastic trips to conferences we shared.

I would like to thank several friends from the lab who were always eager to provide help, share their knowledge, engage in discussions, and enjoy a pint on the rooftop: Andy Dolan, Tarick El-Baba, Corinne Lutomski, Jani Bolla, and Virginia Casablancas.

Dr Georg Hochberg was extremely generous, taking me under his wing during Covid, for which I am very grateful. He invested a lot of time in acquainting me with a completely new field by teaching me phylogenetics. His guidance enabled me to remain productive outside of the lab when it was shut down during the lockdown.

They are people in Oxford who have become my family here and have made it a home. To Almu, Andres, Corinne, and Sam, thank you so much for making

me feel that I am never alone. All the beautiful moments we have created together - weekend trips, morning runs, Sunday brunches, and many more - have always been times filled with happiness, love, and comfort.

My parents taught me by showing an example of what hard work is, but more importantly, they taught me that for every project I decide to pursue, I should do it with passion and love, to always give 100% and never lose the motivation, and to always aim high and think outside of the boxes. Thank you both for making me believe that I can accomplish whatever I want, respecting my decisions, and supporting me in pursuing my dreams.

I am grateful to my brother Oskar, who always has inspired me to be creative, brave, and persistent, and to never be afraid to be myself. I admire you for your amazing ability to turn whatever you do into fun and adventure, which makes everyone around you feel so special.

I also want to thank my granny, who has always been a major role model for me. She has shown me what it means to be a strong, independent woman.

I have been very lucky to have Ola my long-time friend and Polish oasis in the UK. Whenever Oxford felt too overwhelming, I could always escape to London for a refreshing weekend with Ola. Thank you for singing Polish songs with me when I was feeling homesick, for always checking up on me, and for being my YES-man.

I would not have been able to make it through without Adam, who has been my biggest support throughout the whole journey, serving as my enthusiastic cheerleader and my firmest rock on whom I could always count. Thank you for always knowing what to say to put a big smile on my face, for taking care of me, and for being constantly by me side, no matter what. I cannot express enough how grateful I am.

Thank you!

Abstract

An intricate set of processes responsible for maintaining the stability of the proteome forms a protein homeostasis network, which is crucial for cell viability and survival. This thesis describes insights into the interactions of the components that belong to that network. We focus on small heat-shock proteins (sHsps), significant members of the chaperone machinery, and glucosidase II (GluII), which is a key enzyme of the ER quality control system. We utilise native mass spectrometry, supported by other biophysical techniques, to broaden the understanding of studied protein-protein and protein-ligand complexes.

First, we characterise the properties of organelle sHsps, a family unique to plants. We show their structural diversification from cytosolic homologs. We find that two of the examined sHsps exhibit polydispersity, never reported in plant sHsps before. Further, we show evidence of dual monomer architecture within higher oligomers among organelle-targeted sHsps.

Then, we demonstrate that organelle-localised sHsps from different evolutionary classes, as well as those from different cellular compartments, co-assemble with diverse modes of interactions. We quantify the energy biases of those interactions using a developed statistical-mechanical model. Studied co-assemblies exhibit a range from a slight to strong preference toward homo-interactions.

We next investigate the subunit exchange between the organelle and cytosolic sHsps. The results indicate that the diversification of the α-crystallin domain led to the incompatibility in the dimerisation interfaces between these two groups. However, more flexible C-terminal and N-terminal regions maintain the assembly of homo-dimer subunits into higher hetero-oligomers.

Lastly, we determine the mode of interactions between GluII and a thiopyridone ligand, a potential host-targeted antiviral drug. Based on the obtained results, we propose a 'dock, lock and latch' molecular mechanism, where the ligand accesses the binding pocket near the dimerisation interface and then binds to the available cysteine through a disulfide bond. This binding induces structural rearrangement followed by dimer dissociation.

Statement of Originality

This thesis presents most of the research I conducted as a part of the DPhil research program in the Department of Chemistry at the University of Oxford. The work presented here was carried out between October 2019 and December 2023.

Conducted experiments gave rise to four experimental chapters. The last chapter covers my work on a project which is a collaboration with the Zitzmann group (Biochemistry Department, University of Oxford). For each of the presented projects, I received guidance and help from numerous people, for which I am very grateful, but this thesis is a product of my own work, unless stated otherwise in the main text or acknowledgements, included at the end of each chapter.

As many people contributed to the work presented here either indirectly by guiding me and sharing their knowledge that I was able to utilise, or directly by helping with experiments, I have used the pronoun 'we' throughout the thesis.

The figures contain data or schematics produced by myself unless stated otherwise, either in the caption or in the acknowledgement section of the same chapter.

This work has not been submitted for a degree, diploma, or any other academic certification at any other institution, in whole or in part.

Wiktoria Laura Sadowska
January 2024

Contents

List of Figures **xiii**

List of Tables **xvii**

List of Abbreviations **xix**

1 Biological background and biophysical techniques for studying protein assemblies **1**

1.1 Introduction . . . 2
1.2 Protein homeostasis . . . 3
1.2.1 Molecular chaperones . . . 5
1.3 Small heat-shock proteins . . . 6
1.3.1 The structural organisation and broad diversity of sHsps assemblies . . . 7
1.3.2 Biological functions . . . 11
1.3.3 The evolutionary trajectory of sHsps . . . 14
1.4 The ER protein folding machinery . . . 16
1.4.1 The calnexin cycle . . . 17
1.5 Native mass spectrometry in the analysis of protein interactions . . . 19
1.5.1 Nano-electrospray ionisation . . . 23
1.5.2 Ion transmission . . . 25
1.5.3 Quadrupole selection and gas-phase manipulation . . . 27
1.5.4 Mass analysis . . . 30
1.6 Circular dichroism for the study of protein secondary structure . . . 31

2 Functional and structural investigation of organelle-localised sHsps from *A. thaliana* **35**

2.1 Motivation . . . 36
2.2 Introduction . . . 37
2.2.1 Evolution of plant sHsps . . . 37
2.2.2 Functional and structural diversity of organelle sHsps . . . 39
2.3 Results . . . 44
2.3.1 Chaperone activity of mitochondrial plant sHsps . . . 44

2.3.2 CD confirms correct folding of mitochondrial sHsps 48
2.3.3 Organelle sHsps from *A. thaliana* display diversity of oligomeric structures and assembly 51
2.3.4 Two N-terminal environments of organelle sHsps 55
2.4 Discussion 63
2.4.1 Mitochondrial sHsps act according to the canonical holdase activity 63
2.4.2 Structural diversification of organelle sHsps 64
2.4.3 Two orientations of monomers within higher oligomers of mitochondrial sHsps 66
2.5 Acknowledgments 68
2.6 Methods 69
2.6.1 Protein expression and purification 69
2.6.2 Chaperone assays 70
2.6.3 Probing protein secondary structure with CD 71
2.6.4 Native MS of organelle sHsps 71

3 Co-assemblies of organelle sHsps from plants 75
3.1 Motivation 76
3.2 Introduction 77
3.2.1 The diversity of sHsps assemblies 77
3.2.2 Physico-chemical description of protein co-assemblies 79
3.3 Results 83
3.3.1 Diverse co-assembly modes of organelle sHsps from different cellular compartments 87
3.3.2 Almost unbiased co-assembly of mitochondrial sHsps from different evolutionary classes 94
3.4 Discussion 100
3.4.1 Mapping the energy of interactions between organelle sHsps 100
3.4.2 Evolutionary conservation of organelle sHsps 102
3.5 Acknowledgments 105
3.6 Methods 106
3.6.1 Protein expression and purification 106
3.6.2 Mass Spectrometry - subunit exchange of sHsps 107
3.6.3 Description of co-assembly using statistical mechanics 108

4 Incompatibility of dimerisation interfaces in organelle and cytosolic plant sHsps 111
4.1 Motivation . . . 112
4.2 Introduction . . . 113
4.2.1 Oligomerisation interfaces of sHsps . . . 113
4.2.2 Structural differences between cytosolic and organelle sHsps 115
4.3 Results . . . 116
4.3.1 Phylogenetic analysis of CI sHsps shows their strong conservation . . . 116
4.3.2 Mitochondrial(MTII) and cytosolic(CI) sHsps form co-assemblies with favoured even stoichiometries . . . 119
4.3.3 CI sHsps changes dispersity as a result of co-assembly with MTI member . . . 124
4.3.4 Products of reaction between CP and CI sHsps are hetero-oligomers with strictly even stoichiometries . . . 129
4.4 Discussion . . . 131
4.4.1 Dimeric incompatibility in cytosolic and organelle plant sHsps 131
4.4.2 Compatibility of CI and organelle sHsps interfaces responsible for higher oligomer formation . . . 134
4.5 Acknowledgments . . . 137
4.6 Methods . . . 137
4.6.1 Protein expression and purification . . . 137
4.6.2 Characterisation of sHsps complexes by nESI-MS . . . 138
4.6.3 Phylogenetic analysis . . . 138

5 The principles of Endoplasmatic Reticulum α-Glucosidase II inhibition by thiopyridone as an host-targeting antiviral strategy 139
5.1 Motivation . . . 140
5.2 Introduction . . . 141
5.2.1 GluII as an antiviral strategy target . . . 141
5.2.2 Targeting dimerisation interface of GluII . . . 143
5.3 Results . . . 145
5.3.1 In the native state GluII forms a heterodimer with a variety of glycoforms . . . 145
5.3.2 Covalent interactions of GluIIα with thiopyridone . . . 147
5.3.3 Determination of the ligand binding residue . . . 155
5.4 Discussion . . . 160
5.4.1 Disulfide binding of GluII by thiopyridone . . . 160
5.4.2 GluII-thiopyridone interactions led by dock, lock and latch molecular mechanism . . . 161

5.5 Acknowledgments 162
5.6 Methods 164
5.6.1 nESI-MS of untreated and treated GluII 164
5.6.2 Characterisation of the ligand binding by LC-MS 165

6 Concluding remarks **167**

References **173**

List of Figures

1.1 The protein homeostasis network. 4
1.2 Model of the canonical sHsps chaperone activity. 6
1.3 Structural organisation of sHsps. 8
1.4 Structural diversity of sHsps. 10
1.5 Interactions responsible for the chaperone activity of HSP17.6 towards its natural substrate. 12
1.6 The calnexin cycle in the ER quality control system. 17
1.7 Native MS in studying protein assemblies. 21
1.8 Schematic of the QExactive UHMR Orbitrap mass spectrometer. . . 22
1.9 Nano-electrospray ionisation of protein analytes. 24
1.10 Quadrupole mass filter. 28
1.11 Protein secondary structure analysis by circular dichroism. 32

2.1 Evolutionary trajectory of the plant sHsps. 38
2.2 Diagrammatic representation of major sHsp domains based on aligned sequences of studied sHsps. 41
2.3 Multiple sequence alignment of investigated plant sHsps. 42
2.4 *In vitro* chaperone assays of mitochondrial sHsps. 45
2.5 Secondary structure analysis by circular dichroism. 49
2.6 The native mass spectra of organelle-targeted sHsps from *Arabidopsis thaliana*. 52
2.7 CID of HSP21 after previous isolation in the quadrupole. 56
2.8 CID of HSP23.6 shows an untypical manner of dissociation. 57
2.9 Tandem MS experiments of HSP23.5. 59
2.10 CID of HSP26.5 with increasing HCD energy. 61
2.11 Proline effect of HSP26.5 during MS/MS experiments. 62
2.12 Dual orientation of sHsp monomers. 67

3.1 Hypothetical modes of assembly of two monodisperse proteins. . . . 78
3.2 12-mer co-assembly predicted with the statistical-mechanical model. 82
3.3 Locations of the studied sHsps. 84
3.4 Different co-assembly modes of plant sHsps. 86

3.5 Unbiased co-assembly of HSP21(CP) and HSP26.5(MTII) 88
3.6 Self-assembly preference between HSP21(CP) and HSP23.6(MTI) . 90
3.7 HSP21(CP)-HSP23.5(MTI) display preference for self-assembly. . . 92
3.8 Full m/z range mass spectrum of mixed HSP23.6 with HSP21. . . . 93
3.9 Heterogenious co-assemblies of HSP26.5-HSP23.6. 95
3.10 Quantitative description of co-assembly between HSP23.6(MTI) and HSP26.5(MTII). 96
3.11 Quantitative description of unbiased co-assembly between HSP26.5(MTII) and HSP23.5(MTI). 97
3.12 Unbiased co-assembly between HSP17.6(CI) and HSP18.1(CI). . . . 99
3.13 Mapping experimental values of energy bias. 101
3.14 Phylogeny of investigated sHsps. 103

4.1 Oligomeric incompatibility in CI and CII plant sHsps. 113
4.2 Oligomerisation interfaces of sHsps take part in substrate binding . 114
4.3 Comparison of the dimers of *P. sativum* HSP18.1(CI) and *A. thaliana* HSP21(CP) 115
4.4 Phylogenetic tree of CI land plants sHsps. 117
4.5 Gene conversion mechanism and its impact on phylogeny. 118
4.6 Products of subunit exchange between HSP26.5 and HSP17.6. . . . 120
4.7 Abundances of hetero-oligomers of HSP26.5 and HSP17.6, detected before and after equilibration. 122
4.8 Disassembly and reassembly of HSP26.5 and HSP17.6 12-mers . . . 123
4.9 Product of subunit exchange between HSP26.5(MTII) and HSP18.1(CI). 125
4.10 Complexes formed as a result of HSP23.6(MTI) and HSP17.6(CI) co-assembly. 127
4.11 Products of co-assembly between HSP23.6(MTI) and HSP18.1(CI). 128
4.12 Products of subunit exchange between chloroplast HSP21 and cystolic CI members. 130
4.13 Schematic representation of the subunit exchange between cytosolic CI and organelle-localised sHsps. 132
4.14 MSA of tested sHsps with highlighted sequence parts significant in co-assembly. 135

5.1 Crystal structure of trypsin-digested *Mm*GluII. 142
5.2 Structure of C61 and C28 ligands. 143
5.3 Dissociation of C61-treated 1 μM GluII monitored over the course of time by MP measurments. 144
5.4 Native mass spectrum of *hMm*GluII. 146
5.5 Native MS of C61-treated GluII in a range of different concentrations 148

5.6 Collision-induced dissociation of non-covalent complexes formed upon C61 and C28 treatment of GluII. 150
5.7 Tandem-MS of complexes formed after C61 and C28 treatment of GluII with a step-wise increase of the HCD energy. 152
5.8 Intact mass data for *hMm*GluII. 154
5.9 Identification of non-covalent complexes of GluIIα_{C822A} with C61. . 156
5.10 Tandem MS of GluIIα_{C822A}C61$_1$ complexes with a step-wise increase of the HCD energy. 157
5.11 Native MS of untreated and C61-treated GluIIα_{C822A} after removing the ligand excess. 158
5.12 LC-MS data of untreated and C61-treated GluII$_{C822A}$. 159

List of Tables

2.1 Suppression of ADH aggregation by mitochondrial sHsps. 46
2.2 Comparison of experimental and theoretical secondary structure predictions. 50

List of Abbreviations

ACD . . . . . α-crystallin domain

ATP . . . . . adenosine triphosphate

CD . . . . . circular dichroism

CID . . . . . collisional-induced dissociation

CNX . . . . . calnexin

CRT . . . . . calreticulin

CTR . . . . . C-terminal region

CS . . . . . citrate synthase

DTT . . . . . dithiothreitol

EDTA . . . . . ethylenediaminetetraacetic acid

ER . . . . . endoplasmic reticulum

ERAD . . . . . endoplasmic-reticulum-associated protein degradation

ERQC . . . . . endoplasmic reticulum quality control

ESI . . . . . electrospray ionization

GluII . . . . . Glucosidase II

HCD . . . . . high-energy collisional dissociation

HTAs . . . . . host-targeted antivirals

Hsp . . . . . Heat-shock protein

IST . . . . . in-source trapping

IPTG . . . . . Isopropyl β-D-1-thiogalactopyranoside

(k)Da . . . . . (kilo)Dalton

LB . . . . . Luria broth

MP . . . . . mass photometry

MS . . . . . mass spectrometry

MSA . . . . . Multiple sequence alignment

MS/MS . . . . tandem mass spectrometry

nESI . . . . . nano-electrospray ionization

NTR . . . . . N-terminal region

OD . . . . . optical density

PQC . . . . . protein quality control

PAGE . . . . . polyacrylamide gel electrophoresis

PDB . . . . . Protein Data Bank

PBS . . . . . phosphate-buffered saline

SDS . . . . . sodium dodecyl sulfate

SEC . . . . . size exclusion chromatography

SUMO . . . . small ubiquitin-like modifier

sHsp(s) . . . . small Heat-shock protein(s)

TCEP . . . . . (tris(2-carboxyethyl)phosphine)

UV . . . . . ultraviolet

WT . . . . . wild type; unmodified or not mutated

1

Biological background and biophysical techniques for studying protein assemblies

Contents

1.1 Introduction 2
1.2 Protein homeostasis 3
1.2.1 Molecular chaperones 5
1.3 Small heat-shock proteins 6
1.3.1 The structural organisation and broad diversity of sHsps assemblies 7
1.3.2 Biological functions 11
1.3.3 The evolutionary trajectory of sHsps 14
1.4 The ER protein folding machinery 16
1.4.1 The calnexin cycle 17
1.5 Native mass spectrometry in the analysis of protein interactions 19
1.5.1 Nano-electrospray ionisation 23
1.5.2 Ion transmission 25
1.5.3 Quadrupole selection and gas-phase manipulation 27
1.5.4 Mass analysis 30
1.6 Circular dichroism for the study of protein secondary structure 31

1.1 Introduction

Protein homeostasis (proteostasis) is essential for cell viability, development, and growth. It consists of a complex network of components, which all act to affect the levels and stability of proteins in the cell. Loss of proteostasis is implicated in numerous diseases and ageing.

Gaining an understanding of processes responsible for a balanced proteome is significant from a biological perspective, but also targeting those elements raises a potential for the development of therapeutic strategies.

In this thesis, we study interactions between various components of the protein homeostasis network. We focus on small heat shock proteins (sHsps), important members of the chaperone machinery, and glucosidase II (GluII), which is one of the main enzymes of the ER quality control (ERQC) system. Here, we provide a detailed description of those proteostasis components to highlight the need for their better understanding and to familiarise readers with the biological background of projects presented in this thesis and their motivation.

We then introduce biophysical methods, which we utilise to study heterogeneous mixtures, and we discuss the choice of these techniques. We focus primarily on employing native mass spectrometry, which allows a detailed analysis of very polydisperse systems, including high molecular mass complexes.

1.2 Protein homeostasis

The proteome of each cell consists of tens of millions of protein molecules ([1]), which exist in a delicate equilibrium of concentrations and conformations. The processes responsible for the dynamic regulation of this intricate network of functional and structurally stable proteins are part of a so-called "protein homeostasis" ([2]), crucial for cell survival and viability.

Proteins are molecules that exist in marginally thermodynamically stable conformations ([3, 4]). The correct structure is required for protein functionality. Temperature, pH, the presence of chemicals, and other environmental stressors or inherent instability can shift the fragile equilibrium towards a misfolded protein state, which further induces protein unfolding and aggregation ([5]). Disruption of the proteins' native structure means they can no longer carry out their function; moreover, the presence of dysfunctional protein aggregates leads to cytotoxicity and results in a loss of proteostasis.

The misbalancing of protein homeostasis is a hallmark of many diseases such as cancer, Alzheimer's disease, and Parkinson's disease (and other neurodegenerative diseases); additionally, it is also associated with ageing ([2, 6]).

To restore and prevent the loss of equilibrium between components of the proteome, cells have developed a robust homeostasis network, which includes components facilitating protein biosynthesis, correct folding, trafficking, and degradation (Figure 1.1A). This complex system of concerted pathways is also referred to as a protein quality control network (PQC, [7]). The PQC system consists of numerous interconnected components and processes localised across the whole cell in various compartments. In the nucleus, PQC controls correct DNA replication and repair, transcription, and RNA processing. In the cytosol, PQC deals with proteins that

are synthesised in the cytoplasm and remain there to perform their functions. The ER is responsible for the folding and quality control of membrane-bound and secretory proteins. Moreover, the ER-associated degradation (ERAD) pathway identifies misfolded proteins within the ER and targets them for degradation. Mitochondria and chloroplast have their own PQC system for the proteins encoded by both the mitochondrial and nuclear DNA that reside within the organelle. Lastly, peroxisomes and lysosomes contribute to the PQC system by degradation and recycling of various biomolecules, that are damaged beyond repair, further contributing to protein homeostasis.

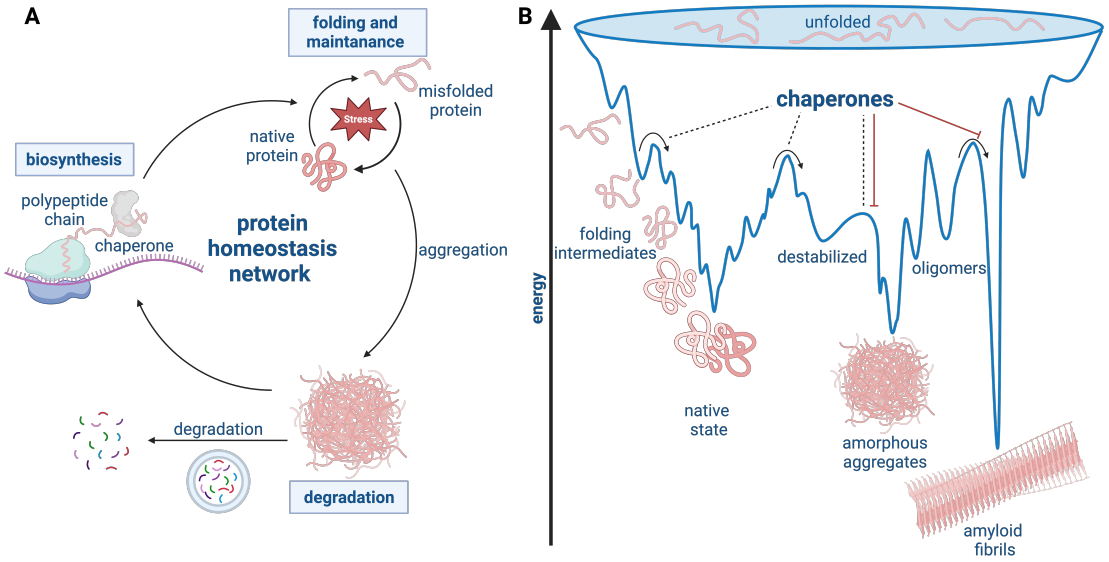


Figure 1.1: The protein homeostasis network. A) Proteostasis is maintained in the cell by numerous concerted actions, which can be divided into three categories: regulation and facilitating of protein biosynthesis, assistance with refolding of unfolding proteins, and removal of misfolded and aggregated proteins. B) A schematic of the protein energy landscape. Chaperones play a significant role in the proteostasis network by modulating the stability of proteins along the conformational pathway. They facilitate the proper folding towards functional, native conformation, and they act to inhibit destabilisation, which can lead to unspecific hydrophobic interactions and result in the formation of amorphous or fibrilar aggregates. Figure adapted from ([4]).

1.2.1 Molecular chaperones

Crucial components of the protein homeostasis machinery are molecular chaperones, that are responsible for conformational maintenance. They fulfil this role by co-translational mediation of the proper, defined, three-dimensional structure of nascent proteins, ensuring functional activity of newly biosynthesised proteins ([8]). If, however, proteins are destabilised, it induces a chaperone response; thereupon, they act to inhibit protein misfolding, restore their already lost native structure, or even disaggregate stress-labile client proteins (Figure 1.1B).

A significant part of the chaperone response are members of the Heat-shock protein (Hsp) family. They were discovered upon upregulated expression in the presence of thermal stress ([9]), which explains the origin of their name, but they are also induced by other types of cellular stressors. Moreover, they are constitutively expressed under normal conditions, with a reported range between 5-10% of the total proteome, for different cell types and organisms([8]), with the precise value varying between organisms but also within the same species.

They are classified based on their molecular weight into HSP100, HSP90, HSP70, HSP60, HSP40, and the small Hsps (sHsps). The large Hsps contribute to the protein homeostasis network by folding, refolding, assembling, trafficking, and degradation of substrate-proteins through an ATP-dependent mechanism. Whereas the sHsps assist them in an ATP-independent manner by acting as holdases and preventing aggregation of stress-sensitive proteins ([10, 11]).

Further, I provide a more comprehensive description of the sHsps family and the structures, activity, and functions their members maintain in the cell.

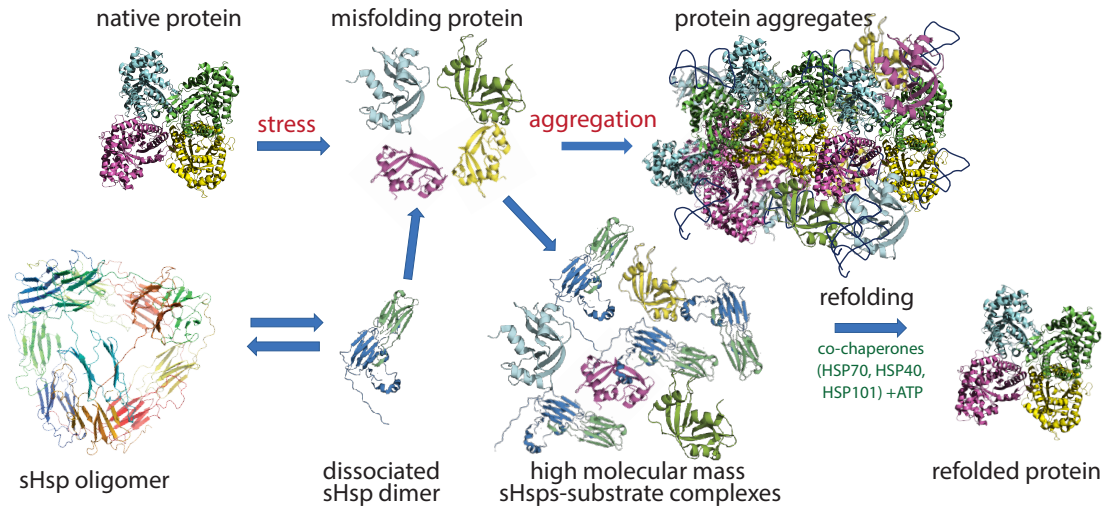


Figure 1.2: Model of the canonical sHsps chaperone activity. A native protein gets unfolded in response to stress, which can further lead to aggregation. This process is prevented by sHsps. Oligomeric structures of sHsps dissociate into smaller species in the presence of stress and capture misfolding client proteins in soluble, high molecular mass, heterogeneous complexes. Next, the ATP-dependent co-chaperones can act to refold the substrates or alternatively degrade them.

1.3 Small heat-shock proteins

The main role of sHsps in the cell is to prevent the aggregation of misfolded proteins, which leads to cytotoxicity and loss of protein functionality.

sHsps are a group of ubiquitous molecular chaperones present across all domains of life, from bacteria to mammals, which highlights their importance in protein homeostasis and suggests they must have diverged very early in the history of evolution. Since then, their evolutionary trajectory has shown intense structural, mechanistic, and functional diversification, which will be discussed further in this chapter.

The most canonical view on how sHsps carry out their function (illustrated in Figure 1.2) is based on their dynamic nature. sHsps in their native state form higher oligomeric structures, which are thought to be a reservoir for smaller, chaperone-active monomers and dimers that constantly exchange subunits with higher-order assemblies, forming an equilibrium ([12–14]). Upon stress, this reaction shifts toward

a higher concentration of smaller oligomers, whose hydrophobic surfaces are available to sense and interact with misfolding stress-labile client proteins. sHsps capture their substrates in a folding-competent state and sequester them into high-molecular-mass complexes ([15, 16]), whereupon they can hand them over to the ATP-dependent chaperone machinery for refolding or degradation ([17]). This mechanism is a source of promiscuity for sHsps, allowing interaction with a broad range of substrates ([18]); hence, chaperones are called the first line of defence against cellular stress.

1.3.1 The structural organisation and broad diversity of sHsps assemblies

The chaperone activity of sHsps can be attributed to their architecture. Their signature organisation of amino acid sequence consists of three regions: α-crystallin domain (ACD), flanked by a flexible N-terminal region (NTR) and a short but dynamic C-terminal region (CTR) (Figure 1.3, top left). Each of them implements a different role in the interactions with substrates.

The ACD is the most evolutionary conserved and ordered region across all sHsps classes ([11, 19, 20]). Its presence is a defining feature of sHsps; however, there are other classes of proteins that have an ACD but are not identified as such. The ACD contains eight to nine β-strands, which forms a β sandwich-like structure (Figure 1.3, top right, [21–23]).

The sHsps monomer subunits dimerise through interactions between the ACDs. In metazoans, sHsps' dimerisation interface involves an extended β6+7-strand, which interacts in an antiparallel orientation with the same strand of the opposite subunit (Figure 1.4, [24, 25]). Non-metazoan ACD-ACD interactions are based on a reciprocal strand exchange of β6, which pairs with β2 of another monomer (Figure 1.4, [22, 23, 26]).

Further, dimers are the building blocks of higher-oligomers (Figure 1.3, bottom right), which assemble through the V/IXI/V motif located in the CTR (called the oligomerisation motif), interacting with the hydrophobic groove between β4 and β8 of the ACD domain ([14, 27, 28]).

The NTR of sHsps contributes the most to their evolutionary diversity by being the least conserved sequence region. The N-terminus, because of its dynamic

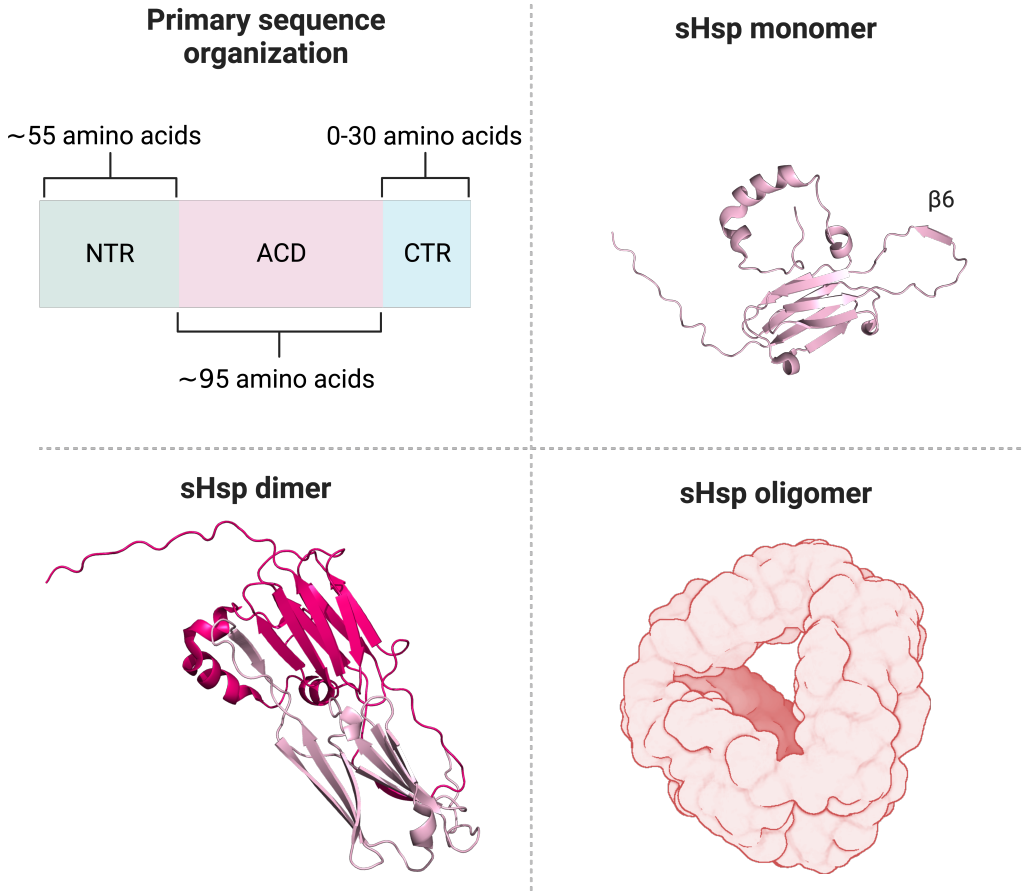


Figure 1.3: Structural organisation of sHsps. Top left: Primary sequence organisation of all sHsps, with the central, most conserved ACD flanked by more flexible and variable NTR and CTR. Top right: HSP16.9 monomer from *Triticum aestivum* illustrating high content of β-sheets, in sHsps secondary structure, which fold into β-sandwich. Bottom left: The monomer further assembles into a dimer through the ACD contacts of two subunits (PDB 1GME). Bottom right: Dimers are building blocks for higher oligomeric structures, here demonstrated by tetrahedral HSP21 12-mer from *Arabidopsis thaliana* (PDB 7ZBW).

character, escapes structural determination, but more ordered parts of this region were identified in the interior cavity of the oligomeric structure, interacting with the ACD of the same or another subunit, further contributing to the stabilisation of the higher-ordered structure ([29–31]). There is evidence suggesting that monomers can display different orientations within higher oligomers, with different architectures of the NTRs ([23, 32]). Data showing high solvent accessibility of the NTR also suggests its outer location. The NTR is also suspected to play a crucial role in the initiation of substrate binding ([33]). Its plasticity and dynamic nature facilitate, rapid sensing of misfolding client proteins, contributing to sHsps activity ([34–36]).

A diversity in amino acid sequences of sHsps is a reason why they display a wide variety of oligomeric structures. Known architectures of yeast, wheat, archaea, and bacterial sHsps exhibit either double disk (HSP16.9 from *Triticum aestivu* [23]) or polyhedral conformations (HSP17 from *Caenorhabditis elegans* [32], HSP16.0 from *Schizosaccharomyces pombe* [37], and HSP16.5 from *Methanococcus jannaschii* [38], Figure 1.4).

Structural investigation of plant sHsps was achievable due to their monodispersity, meaning they adopt only one oligomeric configuration in a native state. So far, all studied plant sHsps have been reported to form one defined oligomeric structure. For this reason, they are a good model system for sHsp investigation. In contrast to some of the mammalian sHsps, which are characterised by high polydispersity - they co-exist in multiple oligomeric stoichiometries - causing a challenge for structural studies ([13, 39–42]). Their assemblies have a form of dynamic distributions, which include monomers and dimers to higher oligomers comprising up to 40 subunits and are characterised by constant subunit exchange.

When sHsps co-localise in the cell, some of them can co-assemble by exchanging subunits between different homologs. They form heterogeneous assemblies of hetero-

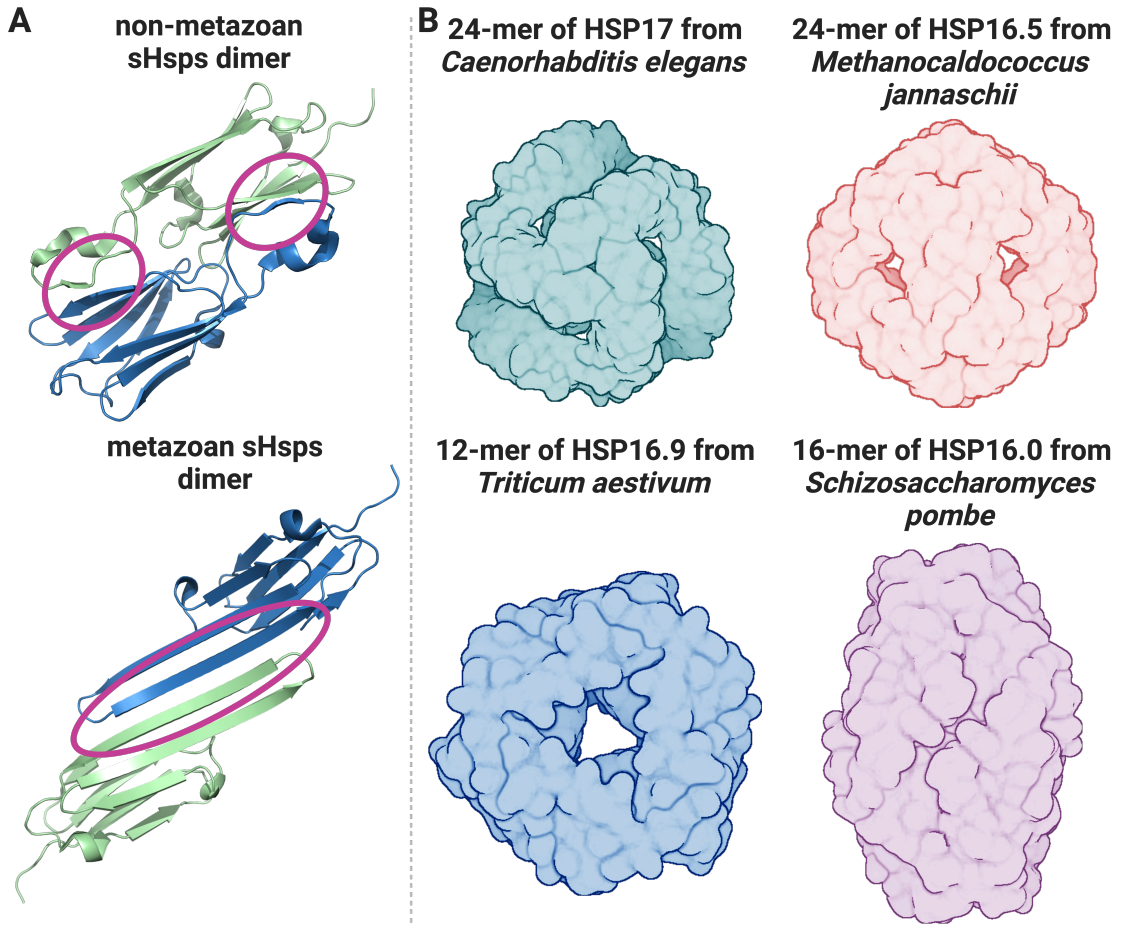


Figure 1.4: Structural diversity of sHsps. A) Comparison of non-metazon and metazoan dimer structure. Top: Dimeric interface of *X. axonopodis* (PDB 3GLA) involves reciprocal strand exchange of β6, interacting with β2 of the opposite subunit. Bottom: human HSPB5 (PDB 4M5S) dimerise through antiparallel binding of fused β6+7 strand. Both interfaces are highlighted in purple. B) sHsps assemble into higher-ordered oligomers with varying sizes and structures. Top left: 24-mer of HSP17 from *C. elegans* with a cubic geometry (PDB 7PE3), top right: 24-mer of HSP16.5 from *M. jannaschii* with octahedral symmetry (PDB 4I88), bottom left: 12-mer of HSP16.9 from *T. astivum* showing double disk shape (PDB 1GME), and 16-mer of HSP16.0 from *S. pombe* with bipyramid architecture (PDB 3W1Z).

oligomers with various stoichiometries ([43–46]). This ability might be a side-effect of the structural conservation of this group, which allows them to maintain their function. It is also possible that heteromerisation is beneficial because it broadens the spectrum of co-existing structures, which may further expand the diversity of substrates sHsps interact with. Which of the two hypotheses is responsible for

keeping sHsp oligomerisation interfaces conserved is still to be determined. The aspect of the ability of sHsps to interact amongst each other is broadly described and studied in Chapter 3 and Chapter 4.

1.3.2 Biological functions

The stress-induced activity of sHsps

The canonical mechanism of sHsps' chaperone activity described above (see Section 1.3 of this chapter, Figure 1.2) was also demonstrated in more detail by Stangel et al. ([47]), by investigating interactions between HSP18.1 (from *Pisum sativum*) and model substrates firefly luciferase (Luc) and citrate synthase (CS). By utilising mass spectrometry, this study revealed that this cytosolic monodisperse plant sHsp, which in its native state forms a 12-mer, becomes polydisperse at elevated temperatures and in the presence of a substrate. A broad range of sHsp-substrate complexes with varying stoichiometries were detected, engaging from one to three substrate subunits and 14 to 36 subunits of sHsps. They further estimated the presence of over 300 sHsps-client coexisting stoichiometries, proving a remarkable heterogeneity of the interactions that provide the holdase activity of sHsps.

Analogous results were observed when we tested interactions between cytosolic HSP17.6 (from *Arabidopsis thaliana*) and a protein identified as its natural substrate: fructose-1,6-bisphosphate aldolase (FBA) from the same plant species. An *in vitro* chaperone assay (see Figure 1.5, left) shows inhibition of aggregation on a level of around 50% at a 1:1 monomeric molar ratio between the two proteins. Complete prevention of aggregation was achieved at a 1:4 molar ratio (FBA:HSP17.6) and hence was chosen for further investigation. Mass photometry measurements for

equivalent experiments reveal the formation of high molecular mass complexes (Figure 1.5, right), which we suspect are responsible for the prevention of aggregation.

Another mechanism of sHsp chaperone activity has been proposed, in which sHsps sequester unfolding substrate proteins into phase-separated regions, as an alternative form of proteome stabilisation ([32]). Those aggregates containing sHsps observed in mammalian cells, roundworms, and yeast, were resolved with the assistance of other chaperones (HSP70, HSP40, HSP100), in an ATP-dependent manner, where the presence of sHsps in these aggregates dictates the substrate reactivation by Hsp70 machinery, followed by client handover([48, 49]). The described aggregase activity of sHsps is concentration and substrate-dependent

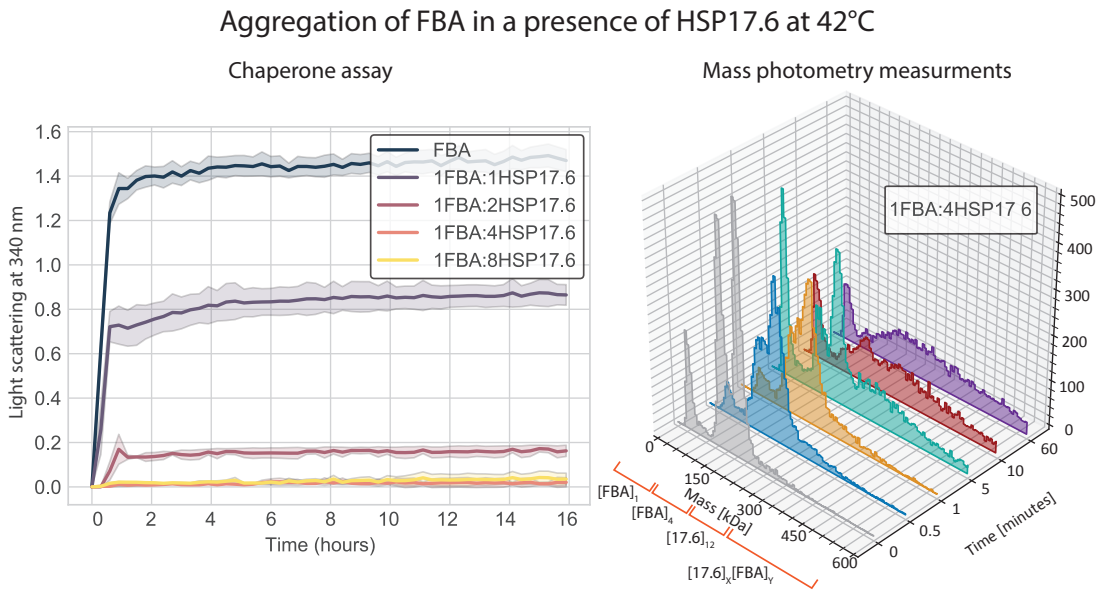


Figure 1.5: Interactions responsible for the chaperone activity of HSP17.6 towards its natural substrate. Left: Chaperone assay of HSP17.6 monitors the formation of aggregates by measuring absorbance at λ=340 nm at 42°C. It illustrates HSP17.6-mediated inhibition of aggregation of its natural target FBA by around 50% already at 1:1 molar ratio and a successive increase in chaperone activity with higher sHsp concentration. Right: Analogous MP experiment, using the same conditions. Measurements were carried out at multiple time points, up to an hour. The obtained data show the appearance of high molecular mass species whose concentration increases with time. They presumably correspond to soluble complexes between FBA and HSP17.6, which contribute to the prevention of aggregation.

and can vary within the same species.

It can be thought of as two different sides of the same coin. sHsps interact with a very wide range of substrates, but the affinity and chaperone activity of a sHsp towards each substrate varies. Some substrates might be chaperoned by a sHsp very efficiently, while others need higher concentrations of sHsp in order for their aggregation to be prevented. There are also cases where inhibition is not possible at all, and that differs for each specific sHsp-substrate pair and depends on concentrations, ratios, and other conditions ([11]). Even though sHsps are very promiscuous and can bind broad range of substrate, which points toward low specificity of their interfaces, in some cases their chaperone activity towards a client is low. This is possibly due to very high interface incompatibility. In this case they are sequestered into aggregates. However, if the holdase activity of a sHsps is high, it can maintain stress-labile proteins in soluble complexes.

The activity of sHsps in non-stress conditions

sHsps are highly expressed as a result of the stress response; however, they are also present and involved in multiple processes in the cells under basal conditions ([50, 51]). For example, there are 10 different sHsps in humans, and only fewer than half are upregulated in the presence of stress ([17]).

HSPB4 and HSPB5 play a crucial role in the eye lens, where they are constitutively expressed and make up to 40% of the total soluble protein cell content ([52]). To avoid crystallisation, in the overall exceptionally high concentration environment, they display high polydispersity. They ensure lens transparency ([29, 40]) and act as chaperones to avoid aggregation of lens proteins due to ageing and constant exposure to UV radiation and oxidative stress ([53, 54]). Mutations in both

HSPB4 and HSPB5 lead to cataracts and other eye disorders, highlighting their importance in proper vision ([55, 56]).

All the mechanisms discussed above involve the activity of sHsps towards a broad range of clients; however, some sHsps exhibit substrate specificity. Several human sHsps are expressed in muscles, and by interacting with cytoskeletal components, they are responsible for their structural integrity ([57, 58]). For example, HSPB1 takes part in mechanosensitive interactions, where it protects filamin C from overtension during contractions, and HSPB5 regulates the passive stiffness of the heart muscle during relaxation by interacting with titin. Identified mutations in HSPBs, which cause cardiomyopathies ([59]), underline the significance of those specific interactions. This illustrates the functional and mechanistic diversity that sHsps fulfil in cells.

Plants, because of their immobility and hence incapability to escape stress, have developed a particularly abundant sHsps network, with some of them having up to 60 sHsps ([60]). Besides upregulation by a variety of environmental stressors, sHsps also play a role in seed maturation in many plant species ([61–63]) and their levels are also enhanced in pollen ([50]), implying their role in the plant development program.

sHsps, despite being categorised as one protein family, had a very early and complex trajectory of evolution that allowed them to diversify. We demonstrate here that they display a variety of structures, functions, and mechanisms they use to pursue their roles; however, all of them act towards maintaining protein homeostasis.

1.3.3 The evolutionary trajectory of sHsps

sHsps are present in nearly all living organisms, in all domains of life from archaea and bacteria to plants and animals, implying that those chaperones essential for

cellular homeostasis diverged incredibly early in the history of life, around 3.8 billion years ago ([64]).

Since then, sHsps have diversified considerably, taking independent evolutionary pathways in each kingdom. This resulted in an impressive variety of diverged subfamilies with distinct structures and functions (described in the previous sections) adapted to different organisms and environments ([42, 65]).

As a consequence of the constant rapid evolution and the significance of sHsps in adaptation to stress exposure, some organisms have developed a wide collection of these proteins. Bacteria have one or two types of sHsps ([66–69]), humans possess 10 sHsps, and plants due to their sessile lifestyle evolved the biggest range of sHsps, with switchgrass having up to 63 copies ([60]). The impressive expansion of plant sHsps makes them an attractive subject for phylogenetic studies.

Another unique feature of plants is that they are the only eukaryotes identified to express a wide diversity of sHsps targeted to membrane-bound cellular compartments, besides *Drosophila melanogaster* and *Toxoplasma gondii*, which both were detected to have a copy of mitochondrial-localised sHsps, in contrast to the rest of the organisms in this group which possess only cytosolic homologs. Organelle-targeted sHsps are a subject of Chapter 2, where they are described and studied broadly.

The ancient origin of sHsps, their significance in proteostasis, and the fact that they developed such a broad heterogeneity of assemblies and functions make them an attractive target for both experimental and theoretical studies. Additionally, the therapeutic potential of chaperone-targeted treatment strategies requires a comprehensive understanding of their interactions. Moreover, unravelling the principles they utilise to evolve and maintain their functions is needed.

We chose to interrogate plant sHsps as an especially suitable model system for studying the evolution, structure, and interactions of those chaperones, as well as

other biophysical properties, and for research aiming to reveal how they fulfil their functions. The reason plant sHps are exceptionally good for the studies mentioned above is their monodispersity (described earlier in Section 1.3.1) and a wide diversity of homologs of plant sHsps targeted to both cytosol and other cellular compartments.

The complexity of sHsps causes experimental challenges. In this work, we employed native mass spectrometry, supported by other biophysical techniques, which allow for the interrogation of very heterogeneous mixtures. In Section 1.5 and Section 1.6, we provide a description of the used methods for readers outside of this field.

1.4 The ER protein folding machinery

Another machinery that is a part of the protein homeostasis network is the endoplasmic reticulum quality control (ERQC) system. It contributes to proteostasis by monitoring and facilitating the correct folding and modification of nascent secretory and membrane proteins. Additionally, it is responsible for the elimination of misfolded polypeptides through ER-associated degradation (ERAD) or autophagy ([7, 70, 71]).

The majority of mammalian secretory and membrane proteins during their translation are transported to the ER, where they are subjected to the activity of ER-localised chaperones and enzymes assisting their structural and conformational maturation, leading to their proper folding. Some of the crucial pathways they undergo in the ERQC include signal-peptide cleavage, N-linked glycosylation, and disulfide bond formation ([70]).

The main mechanism responsible for the accurate folding of N-linked glycoproteins is the calnexin (CNX) cycle, an essential process of the ERQC.

1.4.1 The calnexin cycle

After a protein is imported from the cytosol to the ER, it goes through various processes (Figure 1.6). One of the most basal biosynthetic ER functions is the addition of covalently bound N-linked glycans to proteins, which is present in 90% of glycoproteins ([72]). It involves attachment of a $Glc_3Man_9GlcNAc_2$ to an asparagine side chain of the NXS/T sequence, assisted by oligosaccharyltransferase (OST) enzyme, which takes place on the ER membrane ([73, 74]). Produced N-glycosylated protein is taken over by glucosidase I and II (GluI and GluII), which cleave the two outer glycose residues from the glycan, resulting in monoglucosylated

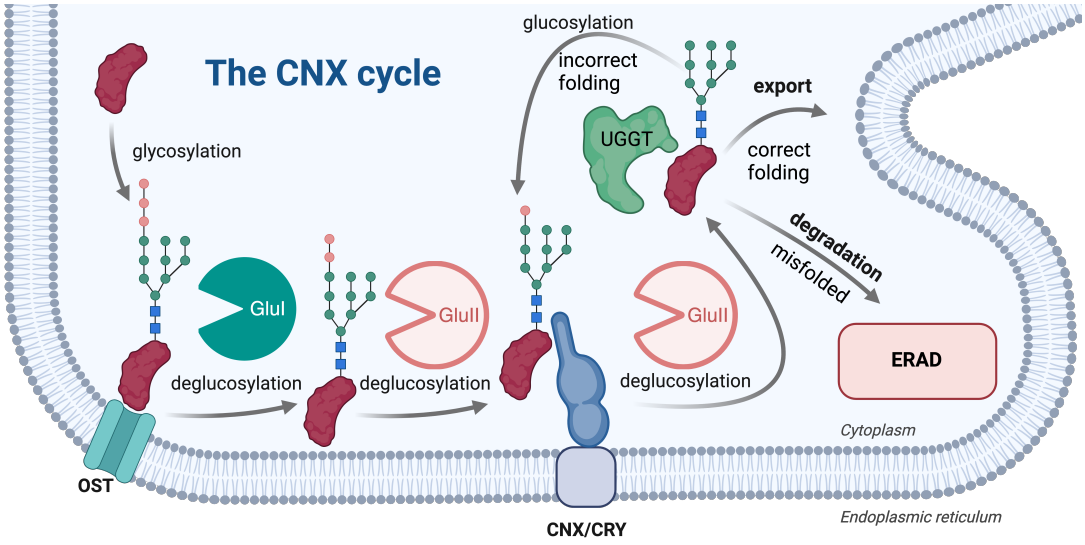


Figure 1.6: The calnexin cycle in the ER quality control system. After a protein is co-translationally imported to the ER, it can get glycosylated by the OST enzyme, which adds $Glc_3Man_9GlcNAc_2$ glycan. GluI and GluII recruit a protein to the ERQC by cleavage of two glucose subunits. This admits the nascent oligosaccharide to the CNX/CRT cycle, where chaperones are responsible for proper folding, preventing aggregation, and early export of their substrates. Further, GluII removes the last glucose residue, and the unglucosylated protein is subjected to a UGGT sensor, which evaluates if the correct folding of glycoprotein is reached. From there, three pathways are possible. If appropriate folding is not reached, UGGT can reattach a glucose subunit, which sends the protein back to the CNX/CRT cycle, giving it another chance to obtain proper conformation. Terminally misfolded proteins are taken over by the ERAD system, which facilitates degradation. If UGGT confirms that the protein is correctly folded, it is exported to its final destination. Created with BioRender.com.

oligosaccharide $GlcMan_9GlcNAc_2$ ([75]).

In the next step, the N-glycan is identified by either both or one of the ER chaperones: CNX and calreticulin (CRT), which facilitate correct folding by prevention of aggregation and premature export to other cellular locations ([76, 77]).

Further, the singly glucosylated glycan is recognised by GluII, which trims the last present glucose residue, causing a release of a glycoprotein from the CNX/CRT, whereupon the UDP-glucose: glycoprotein glucosyl transferase (UGGT) examines the folding of a processed protein, and if its conformation is incorrect, UGGT adds a glucose subunit back to the N-glycan. This leads to a return of the glycoprotein to the CNX/CRT, allowing for another chance for a client protein to reach the proper folding.

Alternatively, terminally misfolded proteins are first subjected to glycan trimming by removing a mannose residue. Produced $Man_8GlcNAc_2$ glycan is recognised and targeted for ER-associated degradation (ERAD).

In another scenario, the native structure of the glycoprotein is reached, which is confirmed by UGGT. The substrate protein is released from the CNX cycle and transported to its destination.

Targeting calnexin cycle as a potential antiviral strategy

A broad spectrum of viruses has an envelope consisting of a lipid bilayer with outer-located glycoproteins embedded in it. Those glycoproteins are crucial in the process of cell infection. They assist in the fusion of the viral membrane with the host cells. Moreover, they contribute to the shielding of virions from the host's immune system, hence, viral infectivity and proliferation rely on those proteins.

Viral-encoded glycoproteins use the host's ERQC and the CNX cycle described above to assist their accurate folding. This gives rise to a potential for host-targeted

broad-spectrum antiviral strategies. Inhibition of ERQC components may lead to incorrect folding of viral glycoproteins and, as a result, limited viral proliferation. In Chapter 5, we describe further and explore this possibility by targeting GluII, which is responsible for the admission and release of glycoproteins to the ERQC. To study the interactions of GluII with a small molecular ligand (190 Da), we utilised mass spectrometry due to its ability to provide high-resolution mass measurements. These measurements effectively distinguish between the number of subunits involved in the complex formation of GluII and the ligand, even in heterogeneous samples.

1.5 Native mass spectrometry in the analysis of protein interactions

Invented at the beginning of the 20th century, mass spectrometry (MS) was first utilised to study small molecules and volatile organic compounds ([78]). Only advances of the past 30 years have broadened its biological application. It started with the investigation of proteins after digestion through peptide analysis, which is still the most commonly used MS application. This further moved in the 1990s to an interrogation of intact protein subunits, then protein-ligand and multisubunit protein-protein complexes. It would not be possible without the discovery of electrospray ionisation (ESI) (described in the later section of this chapter; 1.5.1) and employing it to study biomolecules. It uses gentle ionisation conditions, allowing for the transfer of intact assemblies into the gas phase while maintaining their non-covalent interactions ([79]).

This opened a new door and now MS can be used to study biologically complicated systems to untangle the composition of high molecular mass complexes, their dynamics, and their structure. Moreover, recent developments in the instrumentation and technology, which improved transmission and resolution of mass

spectrometers, make it possible to study increasingly complex and heterogeneous samples and species with masses even up to megadaltons, like ribosomes ([80]), intact viruses ([81, 82]), membrane proteins ([83]), protein aggregates ([84]) or extremely polydisperse proteins like human sHsps ([85]) (described in Section 1.3). We can now interrogate those assemblies in their native-like state to answer questions about their biophysical properties.

Native MS relies on the assumption that the solution-like quaternary structure and conformation are maintained during ionisation, and molecules are transferred from a volatile solvent to the gas phase of the mass spectrometer, with the preservation of hydrogen bonds, salt bridges, and hydrophobic interactions.

After desolvation and ionisation, ions of protein assemblies pass into the vacuum of a mass spectrometer and traverse through it, a process that is called transmission (for more details, see 1.5.2, Figure 1.8, pink). Molecular ions with maintained covalent and non-covalent interactions listed above reach the mass analyser (described in 1.5.4, Figure 1.8, orange), where ions are detected based on their mass to charge (m/z) ratio ([86]). With the current instrumentation, additional gas-phase manipulations are possible, which can provide even more biophysical and structural information (Section 1.5.3, Figure 1.8, green).

Native MS can be used to monitor distributions of protein oligomeric states, the kinetics of their reactions, and the changes they undergo during alteration of conditions ([43, 87]) with a very high mass precision. Figure 1.7 shows an example of a native MS application for studying the assembly of CS from *Synechococcus elongatus*, a cyanobacterium. This protein was found to form Sierpiński fractals, with an 18-mer being the first fractal level. Transmission of complexes even above 1 MDa was achieved. Moreover, while testing samples in different native MS solvents and conditions, an interesting observation was made. The distributions of

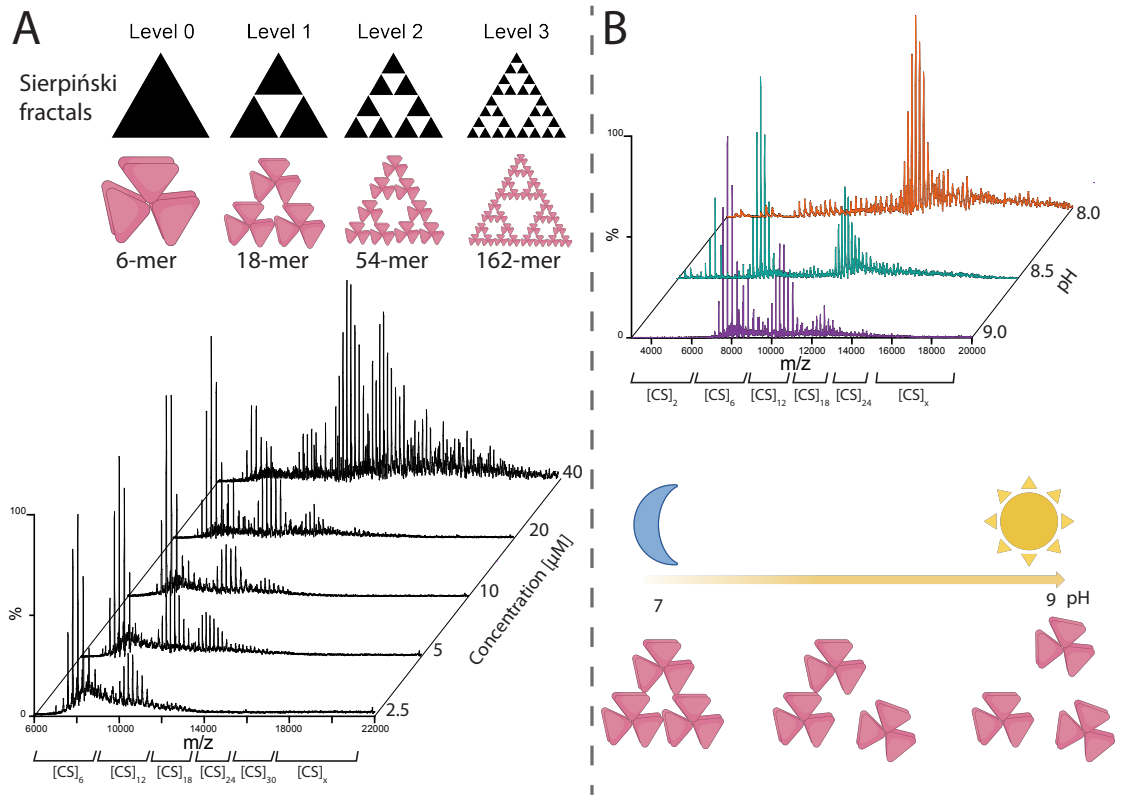


Figure 1.7: Native MS in studying protein assemblies. A) CS from *Synechococcus elongatus* forms Sierpiński triangles fractal structures (top). Bottom: MS data shows that the distribution of CS oligomers is concentration-dependent, with stoichiometries bigger than 18-mer (first fractal level) detected at higher concentrations. Some of the transmitted complexes are above 1 MDa in size. B) During optimisation of suitable conditions for native MS experiments, an interesting dependence was observed of the dissociation of higher oligomers into lower forms with increasing pH, which afterwards was proven to have physiological and functional relevance and act as a CS activation mechanism.

CS are not only concentration-dependent, but also the higher oligomers dissociate with increasing pH. Cyanobacteria's pH is regulated with the circadian rhythm, and so those changes presumably are occurring *in vitro* as well. After testing the activity of CS in different pH values, our collaborators found out that dissociation significantly impacts the activity of this protein (data not shown), presumably by shifting the equilibrium from storage forms (18-mers and higher oligomers) in a neutral pH, to lower, active oligomeric assemblies in higher pH (Figure 1.7B). Native MS data also revealed the mechanism of dissociation, which is based on

dimer subunits leaving the 18-mer.

We showed here that native MS gives the potential to study high molecular mass and heterogeneous protein assemblies. This makes it perfectly suitable for the analysis of protein homeostasis network components. As described in previous sections of this chapter, those proteins are characterised by high heterogeneity of quaternary structures and interactions between themselves and their substrates, which is possible to illustrate with high-resolution mass measurements using native MS ([88]).

Instrumentation setup used in this work: QExactive Orbitrap mass spectrometer

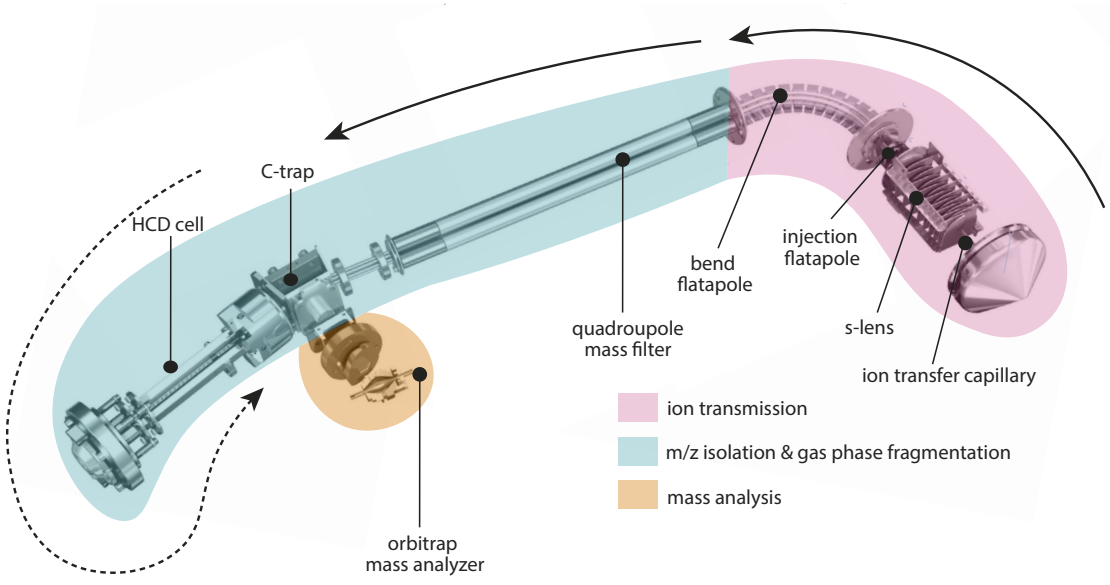


Figure 1.8: Schematic of QExactive UHMR Orbitrap mass spectrometer. Pink: After ionisation by nESI, ions enter the mass spectrometer through the ion transfer capillary of the ion source. The next step is ion transmission via s-lens, injection flatapole, and bent flatapole. Green: The presence of a quadrupole raises the possibility of mass selection and, hence, tandem MS experiments upon isolation of ions from a selected m/z window, which can be fragmented in the HCD cell. Orange: The final step of an nESI-MS experiment is mass analysis, and in this instrument, an Orbitrap mass analyser is used.

All native MS data presented in this thesis were acquired utilising nano-electrospray as an ionisation technique, using a QExactive UHMR Mass Spec-

trometer (ThermoScientific) illustrated in Figure 1.8. After the ionisation step and transferring intact biomolecules into the gas phase of the mass spectrometer through an ion transfer capillary, the ions are focused in the S-lens and then are transmitted through the bent flatapole (Figure 1.8, pink). After that, they approach the quadrupole, which can serve as a mass filter for the selection of the m/z window of interest (Figure 1.8, green). Further, ions can enter the HCD cell, where additional gas-phase manipulation is possible, to finally reach the Orbitrap detector (see Figure 1.8, orange, [89]). Below, we describe each of those steps in more detail.

1.5.1 Nano-electrospray ionisation

The first step of each MS acquisition described above is nano-electrospray ionisation (nESI), which allows for the preservation of native-like conformations of investigated proteins and their non-covalent interactions by applying soft ionisation conditions ([79]). Another benefit of nESI is transferring biomolecules to the mass spectrometer's vacuum straight from the solution.

In comparison to a regular ESI, for a nESI, the sample volume consumption is lower, as well as the required concentration and applied capillary temperature, which is reasoned by the smaller initial droplet size and hence easier disolvation. Additionally, another benefit is a lack of need to use organic solvents, altogether making this ionisation technique suitable for the transfer of very high molecular mass species to the gas phase ([78]).

In this method, a protein sample in a volatile solvent (here ammonium acetate or ammonium bicarbonate) is transferred into a gold-coated glass emitter, with a manually clipped end to provide 1-5 μm diameter ([78, 88]). Analyte-containing charged droplets are ejected from the Taylor cone ([90]), formed at the tip of the emitter by applying gentle backing pressure and a potential difference between

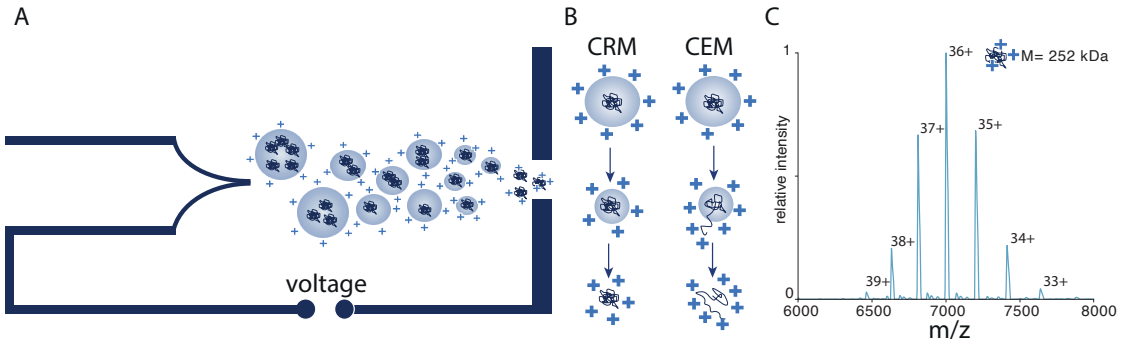


Figure 1.9: Nano-electrospray ionisation of protein analytes. A) The Taylor cone is formed after applying a voltage to the gold-coated emitter containing the protein sample. As a result, charged droplets are formed at the tip of the capillary. Droplets are guided toward the ion source of the mass spectrometer and undergo cycles of evaporation and fission as a consequence of reaching the Rayleigh limit. B) Globular proteins undergo multiple cycles of Rayleigh divisions and evaporation, finally resulting in charged protein, which takes over the droplet charge- Charge Reduction Model (CRM). ESI of disordered molecules can also be explained by the Chain Ejection Model (CEM), in which disordered regions are attracted toward the droplet surface and ejected. C) nESI of analyte results in ions of the same protein carrying variable charge numbers, which is observed as a charge state distribution of each species in the mass spectrum.

the conductive capillary, and inlet of the mass spectrometer (Figure 1.9A). The produced droplets shrink as the solvent evaporates, and consequently, their charge density increases. When the Coulombic repulsions exceed the surface tensions, the Rayleigh limit is reached, and the small droplets undergo fissions. Multiple divisions lead to the final solvent-stripped analyte ion ([91]).

The final ion desorption step is described by two main mechanisms (see Figure 1.9B). The Charge Residue Model (CRM) assumes that evaporation and fission cycles eventually are depleted, initiating charge transfer from the droplet to the analyte ([92]). It is an appropriate description of globular folded proteins' ionisation ([86, 93, 94]).

The other model refers to proteins with unfolded or disordered conformations, and it is called the Chain Ejection Model (CEM). It presumes that the analyte ions get ejected from the droplet before its complete desolvation by attraction and migration of the unfolded regions to the droplet surface (Figure 1.9B, right, [92, 95]).

Proteins studied in this thesis, subjected to the nESI, presumably get ejected according to the CRM model, reasoned by their globular structure. However, we can not exclude the possibility that they undergo the CEM mechanism to some extent because of the high content of disordered regions in their structure.

Produced ions of the same analyte carry various charges (Figure 1.9C), which are observed in the mass spectrum as the charge state distributions of each species ([96, 97]). This allows for the mass assignment of the analyte from the m/z values of the detected peaks.

1.5.2 Ion transmission

After the proteins are ionised and transferred to the gas phase, they enter the ion source through an ion transfer capillary. This capillary can be heated up to 400°C, further helping with desolvation. However, during native MS, the temperature should be optimised to the lowest possible value, which might need to be higher for membrane or glycoproteins. Next, the decreasing pressure gradient guides ions through the neural ion transfer tube to the S-lens region.

S-lens

The S-lens plays the role of a vacuum baffle ([89]) by creating a gradient between the high-pressure ion source and the low-pressure ion optics region. It consists of spaced stacked, flat stainless-steel electrodes and acts as a transmission device. The S-lens focuses the ion beam by applying RF with opposite-phase voltages to neighboring electrodes, guiding the ions to the exit lens. The RF value needs to be optimised for the best transmission, with higher RF amplitude values inducing higher m/z species traversing through the S-lens.

Injection flatapole

The ion beam passes the exit lens and enters an even lower-pressured ion optics region consisting of an injection flatapole and inter-flatapole lens. The injection flatapole is built of four square metal electrodes that are aligned. By applying a DC offset voltage these electrodes are pulsed down to negative voltage values. This creates a negative potential difference with the positive mode of the S-lens.

The injection flatapole is responsible for further focusing of ions and contributes significantly to the desolvation of large protein complexes, also enhancing their transmission substantially by increasing the kinetic energy of ions. It is caused by the pulsed trapping of the ions with negative voltages in the flatapole, which gently shakes off the remaining adducts or removes detergent micelles from membrane proteins while maintaining the non-covalent protein complexes. This process is called in-source trapping (IST) and is crucial for the transmission of high molecular mass protein complexes and experiments involving membrane proteins or interrogation of proteins straight from vesicles.

The inter-flatapole lens is a metal plate with a small hole, kept in a high positive potential to avoid ion leakage. After ion trapping, they are eluted to the bent flatapole, with the DC potential applied to the lens to facilitate further transmission ([89]).

Bent flatapole

The bent flatapole in the QExactive system is also called Advanced Active Beam Guide (AABG). It guides ions through a 90-degree arc, which allows for the disposal of any neutral particles that are not able to follow the curved path when not carrying a charge. It removes neutral gas, solvent droplets, and any inefficiently ionised protein molecules, enhancing overall instrument sensitivity.

The ion beam enters the quadrupole through a flatapole exit lens, which both separates and aligns the two regions.

1.5.3 Quadrupole selection and gas-phase manipulation

Quadrupole mass filter

Initially, quadrupoles were used as mass analysers; however, their low resolving power was the reason for a substitution for more robust alternatives ([86]). Nowadays, quadrupoles primarily serve as both an ion transmission device and a mass filter. It is possible to select from a very narrow to broad m/z window of interest, which gets isolated in the quadrupole, and subject it to further gas-phase manipulation in the higher-energy collisional dissociation (HCD) cell ([92]). The m/z range of the quadrupole in the QExactive mass spectrometer reaches very high values (above 30,000 m/z), which is necessary for the detection of low-charged ions in native MS acquisitions.

The quadrupole is formed of four aligned cylindrical rods (see Figure 1.10). DC and RF voltages applied to the rods are ramped during acquisitions. Opposite rods are connected and display the same voltage sign, which is different from the neighboring rods. The value that determines what m/z range ions have stable trajectories is the RF-to-DC voltage ratio ([98]). Split lens controls the ion injection to the quadrupole, whereupon only selected m/z value ions are maintained and steadily guided by the oscillations to the quad exit lens (Figure 1.10), whereas all the other ions collide with one of the rods, get neutralised and pumped away or alternatively they may be ejected from the rod assembly ([89]).

The described mechanism allows for the interrogation of only selected assembly species of heterogeneous samples in a top-down approach, where intact species

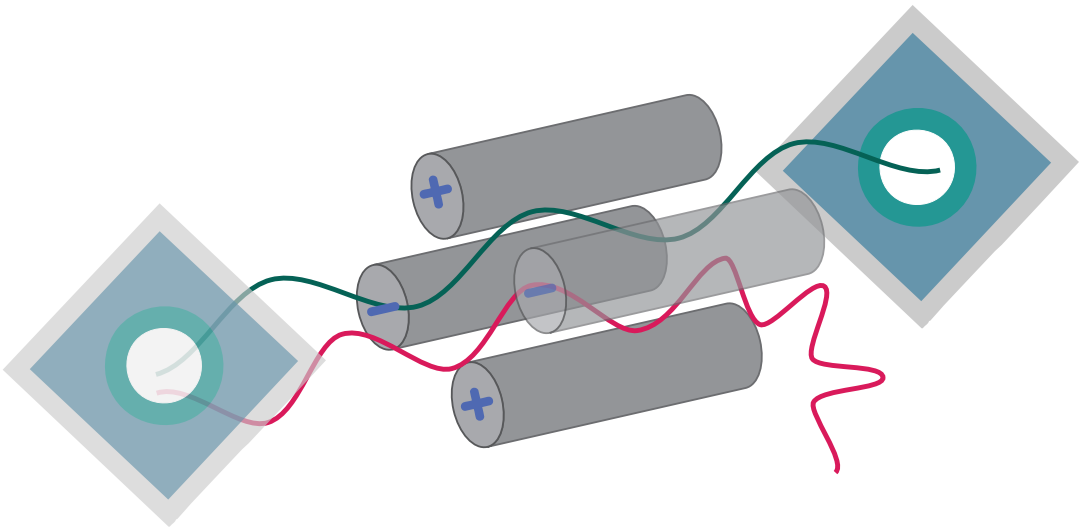


Figure 1.10: Quadrupole mass filter. It comprises an array of four aligned parallel rods. Direct current (DC) and RF potentials are applied onto the rods, with opposite rods being paid and having the same voltage sign. In the QExactive system used in this thesis, the quadrupole serves as a mass filter for isolation of the m/z window of interest, which can further be subjected to fragmentation in the HCD cell (approach called tandem MS). Ions with maintained stable trajectories are defined by the RF-to-DC-voltage ratios.

are subjected to the vacuum of the mass spectrometer, where after m/z isolation of species of interest, the separated ions can be subjected to fragmentation. In Chapter 2, we utilise this mechanism in the tandem MS (MS/MS) experiments of organelle-targeted sHsps, where we describe this process in more detail.

Another interesting utilisation of the quadrupole selection was introduced for soft-landing MS, a relatively new technique that is still under development and combines nESI-MS with cryo-electron microscopy (cryo-EM). Besides having great potential for many applications in structural studies of complex systems, it also shows that protein solution-like conformations are maintained in the gas phase ([99]).

C-trap

C-trap stands for curved linear trap, and it is an ion-collecting and cooling device that is filled with nitrogen gas. Ions entering the C-trap collide with nitrogen particles and, as a result, lose their kinetic energy. Analyte ions of a wide m/z

range get captured in the centerline between four curved rods that build the C-trap, moving freely along the axis. The curved shape of the rods helps with more efficient ion focusing and controlled guiding them to the Orbitrap. One of the electrodes has a gap that extends along its length, which is used for ion ejection. From the C-trap, ions can be guided either to the HCD cell for fragmentation or straight to the detection step in the Orbitrap mass analyser.

Ion fragmentation in the HCD cell

A higher-energy collisional dissociation cell, as the name suggests, is a part of the mass spectrometer where ions can be guided for additional gas-phase manipulation. Quadrupole selected or full m/z range protein assemblies enter the multipole and collide with particles of the gas filling the cell. Usually, nitrogen is used as a collisional gas, but it can be substituted by a heavier option, like Xe or SF_6, to increase the collisional energy.

If enough acceleration voltage is applied, the transmitted ions of noncovalent complexes undergo collisionally induced dissociation (CID, presented and described in more detail in Chapter 2). The collisions induce an increase of the ions' kinetic energy, converted into internal modes, which can cause conformational changes and lead to local unfolding and charge migration to minimise Coulombic repulsions ([100]). This gas-phase reaction may lead to the stripping of one subunit from N-size species, with the leaving monomer taking up a disproportional amount of charge. This results in charge reduction of subunit-stripped N-1 species, which shifts towards higher m/z, and highly charged monomer appearing over lower m/z (see Figure 2.7). With further increase of energy, loss of the second subunit can occur, consequently increasing the charge reduction. The described mechanism can be used to determine the composition of noncovalent complexes and identify interactions between subunits

and the character of those interactions ([88]). It can also be used to improve the peak detection and assignment for very heterogeneous mixtures ([101]) with overlapping peaks. Charge reduction leads to a better separation of charge series, which causes the peak assignment to be less challenging for very polydisperse systems ([39, 86]).

Alterated in the HCD ions are accelerated back to the C-trap, from where they are injected into the Orbitrap mass analyser.

1.5.4 Mass analysis

The Orbitrap plays the role of both a detector and an analyser. It comprises an axially symmetrical electrode system, a central one, which is spindle-shaped, surrounded by two bell-shaped electrodes.

Ions injected from the C-trap enter the Orbitrap and are captured via "electrodynamic squeezing." When trapped, they undergo numerous oscillations around the central electrode. Species with different masses get separated because they oscillate with different frequencies. Oscillating ions are detected by the outer electrode, which measures the image current in the z-axis direction. The mass-to-charge ratio is obtained by Fourier transformation processing ([89]).

The QExactive instrument and included Orbitrap mass analyser offer uniquely high resolution and are well-adapted for transmission of polydisperse and high molecular mass analytes. It is achieved by the reduction of the frequency of RF voltages across the instrument from the injection flatapole to the HCD cell, which enhances the transmission of heavier ions. Additional activation in the injection flatapole and the HCD cell improves the resolution and shakes off the adducts, which results in sharper and better-separated peaks of species close in mass. Moreover, another advantage is the high accuracy and low values of RF-only voltage supply of Orbitrap electrodes and their oscillation frequency-based acquisitions.

Altogether, the instrument employed in this work allowed us to study high-molecular-mass, heterogeneous non-covalent complexes, which are characterised by small mass differences, and some of them by polydispersity or glycosylations. The chosen system allowed us to obtain high-resolution data with an efficient peak separation. Additionally, quadrupole isolation and HCD fragmentation made it possible to determine the composition of some of the studied complexes and the character of interactions between them.

1.6 Circular dichroism for the study of protein secondary structure

Circular dichroism (CD) spectroscopy is an experimental methodology that can be used to study the secondary structure and folding of purified protein samples. CD is defined as differential absorption of left-handed and right-handed circularly polarised light ([102]). A light passing through a solution interacts with chiral molecules to different extents.

Elements that cause unequal absorption of circularly polarised light of the two types (rotating left and right) in protein molecules are chromophores. Measurements of protein samples in the far UV region (180-250 nm) rely on two transitions: n $\rightarrow \pi^*$ of the lone pair of oxygen to an anti-blonding orbital, and $\pi \rightarrow \pi^*$ of the non-bonding π-orbital to the anti-bonding π^*-orbital, within amide groups, and geometries of their polypeptide backbones ([103]). Hence, different structural elements result in the appearance of CD bands, with characteristic CD spectra corresponding to the secondary structure type.

- α-Helix is distinguished by a positive maximum at 193 nm and two negative bands at 222 nm and 208 nm.

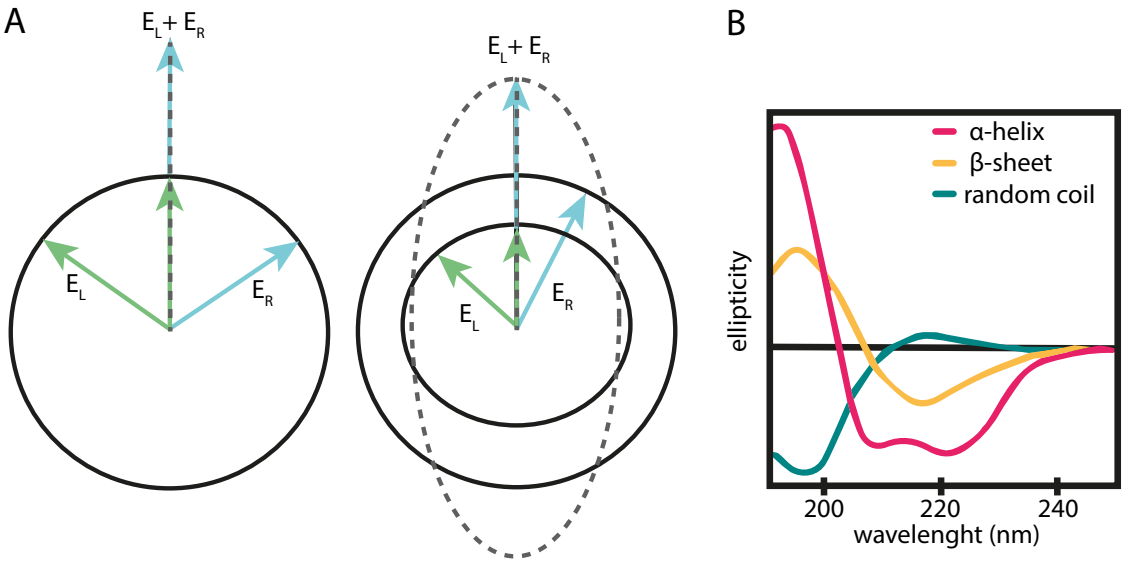


Figure 1.11: Protein secondary structure analysis by circular dichroism. A) CD effect induced by molecules. Left: linearly polarised light generated by left and right-handed polarised components of equal amplitude. Right: Circularly polarised light induced by differential absorbance of the two components. B) Schematic far UV CD spectra illustrate characteristic bands for each secondary structure component. Prediction of secondary structure components for an investigated molecule can be extracted based on data from model samples.

- CD spectra of β-sheets have a characteristic positive band at 195 nm and negative band at 218 nm.
- Disordered proteins show a positive band at 210 nm and a negative near 195 nm.

CD spectra are reported as a difference in absorbance of left and right-handed circularly polarised light or in degrees ellipticity (θ), which is defined as the tangent of the ratio of the minor to the major elliptical axis (see Figure 1.11A, [104]).

Based on those measurements and reference spectra of model proteins with known structures, it is possible to extract contributions from different types of secondary structure components within an investigated sample of analysed molecule, which additionally gives experimental information about the folding of a protein ([103]).

The described set of techniques chosen for this work allowed us to study intricate mixtures of protein homeostasis components to decipher details of their protein-protein and protein-ligand interactions. The lower-resolution CD measurements gave global sample information about the secondary structure and confirmed the proper folding, while mass spectrometry approaches provided high-resolution data with insights into individual subunits.

Here, we introduced the most important facts about the systems we analyse in this thesis. Additionally, each following chapter contains an introduction section, providing more details necessary to gain a complete understanding of the presented further experiments.

2

Functional and structural investigation of organelle-localised sHsps from *A. thaliana*

Contents

2.1 Motivation 36
2.2 Introduction 37
2.2.1 Evolution of plant sHsps 37
2.2.2 Functional and structural diversity of organelle sHsps . 39
2.3 Results 44
2.3.1 Chaperone activity of mitochondrial plant sHsps 44
2.3.2 CD confirms correct folding of mitochondrial sHsps . . . 48
2.3.3 Organelle sHsps from *A. thaliana* display diversity of oligomeric structures and assembly 51
2.3.4 Two N-terminal environments of organelle sHsps 55
2.4 Discussion 63
2.4.1 Mitochondrial sHsps act according to the canonical holdase activity 63
2.4.2 Structural diversification of organelle sHsps 64
2.4.3 Two orientations of monomers within higher oligomers of mitochondrial sHsps 66
2.5 Acknowledgments 68
2.6 Methods 69
2.6.1 Protein expression and purification 69
2.6.2 Chaperone assays 70
2.6.3 Probing protein secondary structure with CD 71
2.6.4 Native MS of organelle sHsps 71

2.1 Motivation

sHsps are highly ubiquitous molecular chaperones, present in almost all living organisms. The vast majority of them are localised in the cytosol; however, plants are unique organisms with a broad range of sHsps targeted to different cellular compartments. Because of their sessile lifestyle and incapability to escape stress, plants evolved the widest variety of sHsps as a coping mechanism in the form of a robust homeostasis machinery. There are as many as eleven evolutionary classes of plant sHsps, and six of them are targeted to cellular organelles ([50, 105]). Since organelle-targeted sHsps classes are absent outside of the plant kingdom, the main focus so far has been on studying the cytosolic sHsps, which is why there are still many questions about organelle sHsps that remain unanswered. What is their biological function? Are they chaperone active? Are they structurally similar to cytosolic sHsps? Do they form higher oligomeric structures? How did they evolve? What are the differences between cytosolic and organelle sHsps? Why are they common in plants but not in other types of organisms?

In the following two chapters, we will try to answer some of those questions by studying four organelle sHsps from *Arabidopsis thaliana* targeted to mitochondria and chloroplasts. We are mainly interested in their biological function, structure, assembly, and evolution. Organelle sHsps, especially the mitochondrial ones, are understudied. Thus, we first investigate if they are chaperone-active by testing their ability to inhibit protein aggregation. Then, we focus on the analysis of their structure - in particular, higher oligomeric species characteristic for most sHsps.

2.2 Introduction

Land plants have unusually high numbers of sHsp homologs, which vary for each plant species: *Arabidopsis thaliana* has 19 sHsps ([106]), *Oryza sativa* has 23, and *Populus trichocarpa* has as many as 36 ([105]). These numbers are significantly higher in comparison to other eukaryotes: 10 for humans and 4 for Drosophila.

As previously mentioned, plants are the only organisms that have a wide spectrum of organelle-localised sHsps, but there are species from outside of the plant kingdom that also possess organelle sHsps. The first such example is *Drosophila melanogaster* ([107]), and the second is *Toxoplasma gondii* ([108]), both targeted to mitochondria. Nevertheless, this is incomparable with a broad diversity of organelle-targeted sHsps found in plants, which can be divided into six evolutionary classes localised in the endoplasmic reticulum, mitochondria, chloroplasts, peroxisomes, and nucleus. It has been hypothesised that extensive diversification of plant sHsps is a coping mechanism of plants against cellular stress that they are frequently exposed to as a result of their immobility, but still, very little is known about why plants need an exceptionally abundant group of sHsps, with different functions and different cellular localisations.

2.2.1 Evolution of plant sHsps

Copies of sHsps homologs have also been identified in bacterial genomes ([48]), but phylogenetic studies show that organelle-targeted sHsps in plants did not evolve via gene transfer from the endosymbiont but rather through a segmental duplication event of the cytosolic sHsps ([109, 110]).

Land plant sHsps have derived from an early event, resulting in a separate evolutionary trajectory from metazoan, archaeal, and eubacterial sHsps ([20, 42,

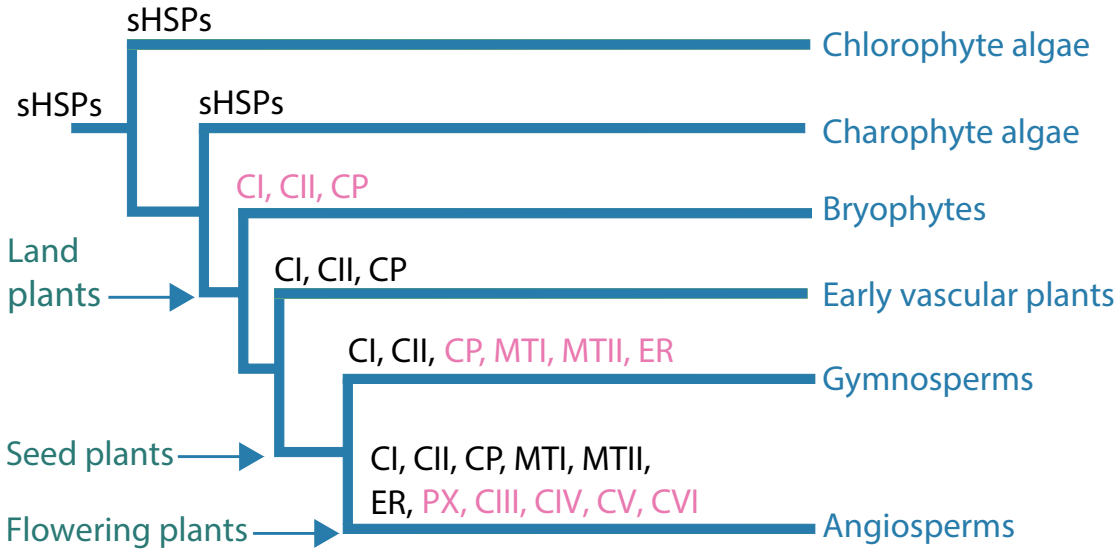


Figure 2.1: Evolutionary trajectory of the plant sHsps. Classes that appeared for the first time in the plant evolution are labelled with light pink on the branches. sHsps are present already in green algae, which means they evolved very early in plant evolution, before the movement of plants into the non-water environment; however, they exhibit low sequence similarity and, thus, are not identified as homologs of any of the land plant classes. Cytosolic CI, CII, and organelle CP appeared in the early land plant lineages as an adaptation to the terrestrial environment. Their wide expansion is observed in seed plants and angiosperms. Adapted from [65].

109–111]). Plant lineage is characterised by a strong selection to maintain gene duplicates, which leads to long lineage-specific expansion caused by the constant birth and death of genes.

Although land plants evolved from algae, none of the algal sHsps lie in the land plant sHsp lineages ([111, 112]), which strongly hints toward the duplication events that happened in land plants, causing the intense expansion of the sHsp family. This expansion did not occur at once 500 million years ago with the appearance of embryophytes, but rather in a step-wise manner. In mosses, we can find homologs of CI and CII ([113, 114]), as well as genes sharing homology with organelle-localised MTI and CP. This indicates that organelle sHsps evolved very early in the land plant evolution. With the appearance of vascular plants, no new classes of sHsps have appeared. The next burst of sHsps expansion came with

seed plants, which broadened their organelle-targeted sHsps collection. The widest expansion of the sHsps classes took place with the appearance of angiosperms, which possess additional CIII, CIV, CV, CVI ER, and PX (see Figure 2.1).

The evolutionary history of land plants sHsps described here suggests that firstly, it was driven by selective pressure connected to the movement of the plants to land and adaptation to a non-water environment (desiccation, temperature changes, heat shock, UV exposure, etc.), and then by their further diversification, natural selection, pressure for maintaining gene duplicates, and their subfunctionalisation ([115]).

2.2.2 Functional and structural diversity of organelle sHsps

Novel roles of organelle sHsps in the new subcellular location

Multiple studies testing the functions of organelle sHsps from outside and inside of the plant kingdom have been reported. Mitochondrial *Dm*HSP22 from *Drosophila melanogaster* (mentioned previously) has been shown to exhibit different expression patterns from sHsps localised in the cytosol of this species, indicating their distinct functions. *Dm*HSP22 is a heat-shock protein with the highest expression levels in ageing flies, suggesting its role in oxidative stress ([116]). HSP22 from tomatoes is also targeted to mitochondria, and is one of the most upregulated proteins during oxidative stress and has been proven to be essential for mitochondrial protection ([117]). Martin De Re et al. showed that another mitochondrial sHsp, the HSP23.8, is directly involved in tomato fruit protection against the consequences of chilling stress ([118]). Both mitochondrial and chloroplastic sHsps in tomatoes are also responsible for the regulation of the ripening of the fruit ([119]).

Extensive research has been performed on chloroplast-targeted HSP21, present in all photosynthesising plants, which is also tested in this chapter. Besides its

unquestionable role in oxidative and heat stress, by the protection of stress-labile protein substrates, HSP21 also regulates chloroplast development during heat stress by preserving transcription ([33, 112]). Some of the reported results suggest it might play a role in maintaining membrane integrity of chloroplast during stress by association with thylakoid membranes at increased temperatures, therefore preventing hyperfluidisation, which might cause leakage ([120]). The precise mechanism of sHsp association with thylakoid membranes is not yet known; however, based on available data, it is suspected that upon dissociation of sHsp oligomers, induced by increased temperature, hydrophobic interfaces are exposed and more prone to interact or penetrate the membrane. This process could be initiated through the NTR of HSP21, which includes an amphipathic α-helix motif. It also has been shown that HSP21 is responsible for thermomemory in *Arabidopsis thaliana* ([121]).

Above, we illustrated that some of the organelle sHsps evolved to pursue new functions in different cellular locations. Multiple studies have tested if it entailed a change in the heterogeneity of interactions with numerous substrates towards more specialised roles. Cytosolic sHsps are known for their promiscuity. They can interact with tens to hundreds of heat-labile proteins ([32, 122, 123]), which makes them responsible for managing protein crowding, in an extremely busy environment. 60 binding partners have been detected for *Dm*HSP22, however, heat shock has not influenced this profile ([124]). Chloroplast HSP22E and F from *Chlamydomonas reinhardtii* have been identified to interact with 39 substrates under heat stress conditions ([125]). These data suggest that promiscuity is a characteristic that has been maintained within organelle-localised sHsps without their activity being restricted to the presence of heat stress.

For cytosolic and organelle sHsps, the highest level of transcriptomes is detected as a response to heat stress. However, some of them are also induced by other

types of stressors like osmotic, oxidative, chilling, wounding, desiccation, and genotoxic ([50]). Their expression profiles vary between different classes but even within the same class, suggesting different roles they might pursue in plants.

It is known that sHsps' expression is also developmentally regulated. Most members of CI and CII are expressed on a higher level in pollen and during seed maturation. Recently, it has been proposed that members of the MTI and PX classes play a role in seed germination ([50, 126]), but in contrast to cytosolic sHsps, none of the organelle-localised sHsps are expressed during pollen maturation. The different expression patterns of organelle-targeted sHsps highlight their diversification from cytosolic sHsps and suggest novel functions that they might fulfil.

Structural changes in organelle sHsps

A common feature for all organelle-targeted sHsps is the presence of a transit peptide. In mitochondrial- and chloroplast-localised sHsps, it has a form of an N-terminal sequence, which gets cleaved once the protein reaches its final destination. As mentioned in Section 1.3.1, the NTR is the most variable region of sHsps. Members

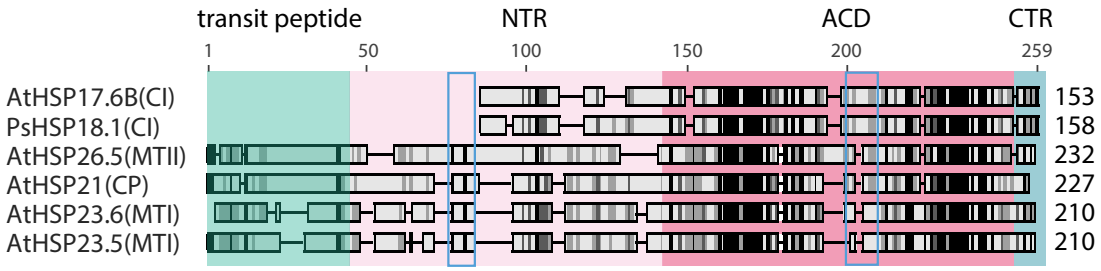


Figure 2.2: Diagrammatic representation of major sHsp domains based on aligned sequences of studied sHsps. The shading represents sequence similarity, with white corresponding to lack of similarity and black means high similarity. As expected, the ACD (light red) is the most conserved region, with high sequence identity preserved across all analysed sequences. It is flanked by the NTR (light pink) and CTR (blue). The NTR of organelle-targeted sHsps is significantly longer and includes conserved α-helical motifs specific only to these sHsps classes, which are absent in cytosolic sequences.

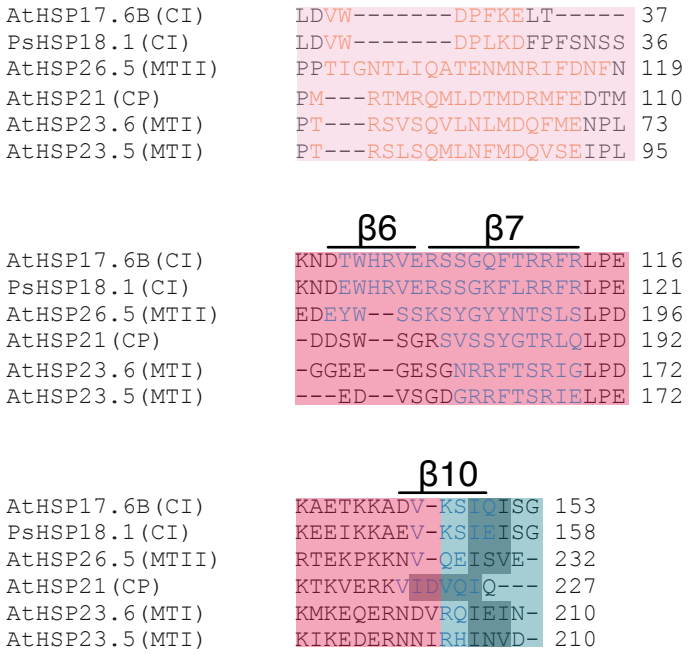


Figure 2.3: Multiple sequence alignment of investigated plant sHsps. A part of the NTR, represented by a light pink colour (top), shows a conserved motif unique to MTI, MTII, and CP classes, exhibiting amphipathic α-helix properties. The dark pink colour illustrates the ACD domain, which for MTI, MTII, and CP classes lacks β6 strand (middle). The CTR (blue) includes the V/IXI/V motif, which is present in all investigated sHsps (bottom), but varies slightly in each case.

of CP, MTI, and MTII classes have significantly longer NTRs (see Figure 2.2) than cytosolic sHsps (other than CV). Moreover, all of them possess the same amphipathic α-helix motif in that region, which is usually deprived of any conserved motifs in other classes (see Figure 2.3).

Another feature unique to MTI, MTII, and CP classes is the lack of β6 strand (Figure 2.3), which might influence their dimerisation and make their structures more prone to dissociation.

All characterised so far plant sHsp monomers assemble into dimers, and the majority of dimers, further into higher oligomeric structures, as described in Section 1.3.1. The NTR has a dual character of an amphipathic helix and an intrinsically disordered protein, causing it to be highly dynamic and flexible, which

is why its precise structural analysis within the oligomeric species remains elusive. It has been suspected that it is localised inside the oligomer. However, multiple recent studies have shown evidence of two different environments for N-termini ([32, 127]). Gundrun Rutsdottir et al. presented data testing for NTR of HSP21(CP) being detected on the outside of a dodecameric structure, suggesting a model in which six arms are located in the interior and the other 6 in the exterior of the oligomer, which may allow interactions with substrates without the need to dissociate ([127]).

For multiple sHsps localised in membrane-bound cellular compartments, the ability to form higher oligomeric structures has been reported ([116, 125, 127]). For example, both HSP22E and F (from *Chlamydomonas reinhardtii*) assemble into larger oligomers, which exist in a dynamic equilibrium with smaller species in the presence of heat stress. These oligomers form high-molecular-mass complexes with substrates during heat stress as a result of their holdase activity, which is a canonical mechanism of sHsps' chaperoning, shown as well for chloroplastic HSP21 ([127]), and peroxisomal HSP15.7 ([50]) from *Arabidopsis thaliana*. The CTR's V/IXI/V motif is crucial for the oligomerisation of sHsps. Sequence alignment shown in Figure 2.3 illustrates the presence of this motif for cytosolic and organelle sHsps investigated in this chapter. This suggests that those proteins should exhibit the ability to form higher oligomers, as was reported for their cytosolic homologs, which we test and describe in Section 2.6.4.

2.3 Results

Here, we present a functional and structural analysis of sHsps targeted to membrane-bound cellular organelles. We focus on sHsps from *Arabidopsis thaliana* localised in mitochondria and chloroplast. This plant has three mitochondrion-targeted sHsps: HSP23.6, HSP23.5, and HSP26.5, and one chloroplastic sHsps: HSP21. HSP21 was studied before by multiple groups, and its chaperone activity has been reported in numerous studies ([33, 127]). For this reason, it was not tested here.

2.3.1 Chaperone activity of mitochondrial plant sHsps

Mitochondrial sHsps from *A. thaliana* were classified as members of the sHsp family based on sequence analysis and expression patterns. To confirm that they are chaperone active, we expressed and purified HSP26.5(MTII), HSP23.6(MTI), and HSP23.5(MTI) and conducted *in vitro* chaperone assays, which rely on monitoring the aggregation of a heat-labile model protein, in the presence and absence of a chaperone, and then evaluating the ability to suppress aggregation. Chaperone activity in those assays can be manifested in two different ways. First, by delaying the increase in the light absorbance, and second, by reaching a lower final level of absorbance coming from created aggregates.

We monitored the absorbance at 340 nm, over 16 hours, for a heat-sensitive model substrate protein, alcohol dehydrogenase (ADH), whose aggregation was induced by 20 mM DTT at 42°C in the absence and presence of sHsps. DTT reduces disulfide bridges, destabilising the secondary structure of ADH, which causes its unfolding and aggregation. The amount and location of reduced disulfide bonds will impact how and where chaperones interact with the protein substrate. Thus, those results might vary from temperature-induced assays involving the same and different

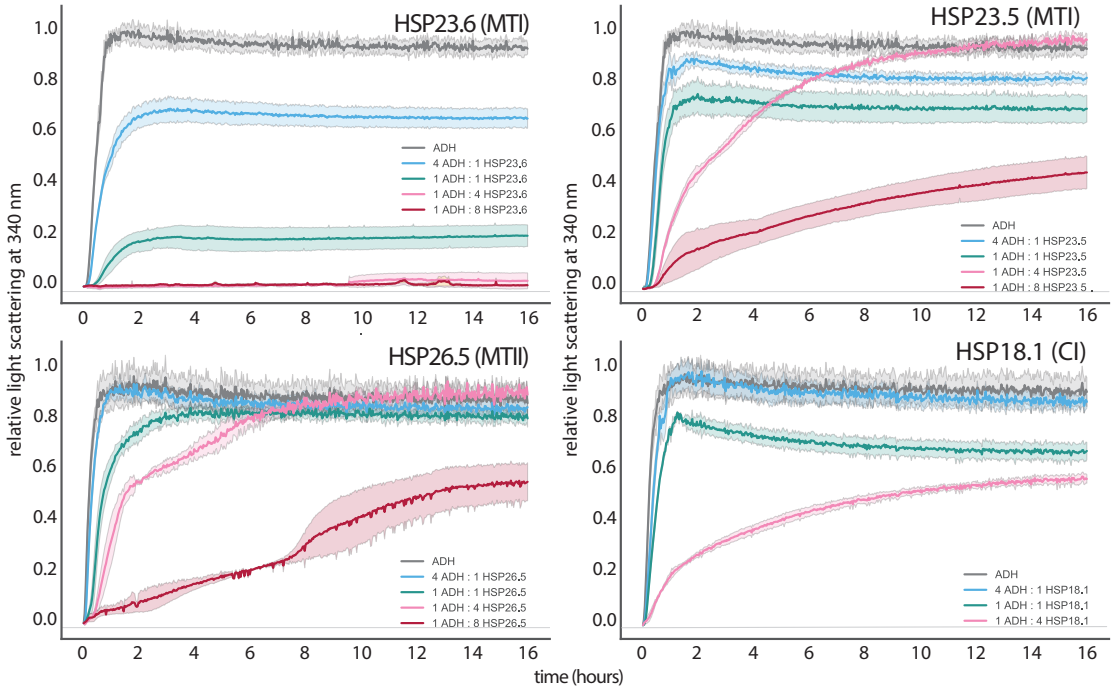


Figure 2.4: *In vitro* chaperone assays of mitochondrial sHsps. Aggregation of a heat-labile protein ADH was DTT-induced at 42°C, and absorbance at 340 nm was recorded over 16 hours in the absence and the presence of mitochondrial sHsps from *Arabidopsis thaliana* (HSP23.6, HSP23.5, and HSP26.5), in a range of molar ratios. Chaperone activity is manifested as a delay and suppression of protein aggregation (increase in light absorbance) of sHsps-protected ADH. Data are averaged over at least three different measurements, and the standard deviations are presented as faded areas. The chaperone activity of a well-studied cytosolic HSP18.1 (from garden pea) was determined for comparison.

substrates. However, the obtained data are sufficient for a general evaluation if tested sHsps are active chaperones. 4 μM ADH was incubated with four different sHsps in a range of varying sHsps concentrations to obtain different molar ratios. We tested the three mitochondrial sHsps mentioned above, and we also added cytosolic CI HSP18.1 (from *Pisum sativum*), which serves here as a comparison. It is a well-studied sHsp, shown before to be an active chaperone ([15, 16, 47, 128]), so additionally, it fulfils a function of a positive control.

Here, as a measure of chaperone activity, we determined a percentage of aggregation (absorbance) level reached after 16 hours of incubation in the presence

of a sHsp, in comparison to the value obtained for ADH after 16 hours in the absence of a chaperone. At least three repeats were recorded. Mean values and standard deviations are collected in Table 2.1.

	% aggregation of ADH (+sHsp) relative to ADH after 16 hours			
ADH : sHsp	HSP23.6	HSP23.5	HSP26.5	HSP18.1
4 : 1	71.0 ± 4.1	88.6 ± 1.4	97.8 ± 4.5	97.0 ± 4.3
1 : 1	22.5 ± 4.6	75.8 ± 5.7	92.4 ± 3.6	74.0 ± 4.1
1 : 4	4.2 ± 3.0	104.6 ± 3.0	105.9 ± 0.8	66.1 ± 1.3
1 : 8	2.0 ± 0.1	50.2 ± 6.3	68.7 ± 4.0	

Table 2.1: Suppression of ADH aggregation by mitochondrial sHsps. Chaperone activity is defined here as the ability to inhibit the aggregation of a stress-labile protein substrate. Percentages of absorbance reached after 16 hours for ADH in the presence of a chaperone relative to an equivalent value in the absence of sHsps, and standard deviations were calculated based on at least three repeats.

As shown in Figure 2.4 (on all four panels), 20 mM DTT-induces aggregation of 4 μM ADH at 42°C. In the absence of sHsps, ADH is denatured in less than an hour. This is represented by the increase in the absorbance caused by the formed protein aggregates, which scatter the light. In contrast, all of the analysed sHsps decrease the aggregation of the substrate, with varying molar ratios for each sHsp required to enable suppression of ADH aggregation to the same level.

For all tested sHsps, results acquired within the first 4 hours show concentration-dependent inhibition of protein aggregation. However, analysing data after that time-point, shows that only HSP23.6 and HSP18.1 are capable of maintaining this correlation (see Fig. 2.4). HSP23.6 is the most chaperone active towards ADH of the sHsp tested here. A 1:4 monomer molar ratio (ADH:HSP23.6) was sufficient to achieve full protection from aggregation, whereas the same ratio of HSP18.1 protected only 33.9% of the fraction from the influence of the DTT-induced unfolding (Table 2.1).

2. Functional and structural investigation of organelle-localised sHsps from A. thaliana

For HSP23.5 and HSP26.5, the results are less trivial to explain, with the first two ratios, 4:1 and 1:1 (ADH:sHsp), protecting either only a very minor fraction of ADH or showing inactivity, 11.4%, 24.2%, and 2.2%, 7.6%, respectively. Interestingly, for a 1:4 molar ratio of client to chaperone, for both HSP23.5 and HSP26.5, a significant delay in aggregation was observed, which is also a sign of chaperone activity. However, the final absorbance value reached at this ratio was slightly higher than for ADH itself (see Fig. 3.9 and Table 2.1), which means that after longer stress exposure time, aggregation could not be prevented. sHsps are stable in higher temperatures and individually do not aggregate; however, they may undergo conformational changes upon dissociation, and they might get incorporated into aggregates of the stress-labile proteins.

At a 1:8 monomeric molar ratio, data show a decrease in aggregation by 49.8% for HSP23.5 and 31.3% for HSP26.5. Both of these datasets have an untypical curve shape, which is an average of three repeats, but it was also observed for other replicates for the HSP26.5. This could indicate two different concentration-dependent mechanisms by which those sHsps inhibit aggregation of stress-labile proteins, which either are not present for the other two tested chaperones or not enough ratios were tested for it to be observed.

We demonstrated that all tested mitochondrial sHsps are capable of suppressing aggregation of a model stress-labile protein, and thus, we can say all of them are chaperone active, with some of them exhibiting even higher efficiency towards ADH than the positive control (cytosolic sHsps). To evaluate which of them is the most chaperone-active, more data involving the examination of a variety of substrates would be needed, which is outside of the scope of this thesis.

To summarise, a finding presented in this section is that interactions between each mitochondrial sHsps from *A. thaliana* and a model client suppress the aggregation

of the latter. It is consistent with the model of sHsps' holdase activity, which states that sHsps capture stress-labile proteins in high-molecular-mass soluble complexes in an ATP-independent manner, whereupon ATP-dependent chaperones can act to refold the target proteins.

2.3.2 CD confirms correct folding of mitochondrial sHsps

Circular dichroism (CD) was employed to ensure the proper folding of the obtained proteins. This technique can also serve as a tool for the investigation of protein secondary structure. Here, we tested the mitochondrial sHsps from *Arabidopsis thaliana* and collated it with the data for HSP18.1 from *Pisum Sativum*, as in the previous section. The experimental results were analysed with Bestsel software ([129]) to calculate secondary structure contents, which we then compared with theoretical predictions based on the amino acid sequence of each of the proteins obtained from Phyre2 ([130]), and we visualise the monomer differences between each species with AlphaFold models predicted for monomers.

The ellipticity measurements in the far-UV for all studied sHsps are presented in Figure 2.5. A global minimum of around 218 nm, characteristic for β-strands, was reached in all tested proteins, additionally, the measurements are in good agreement with the fitted plots. This proves that all tested sHsps are folded correctly. We collected all the experimental and theoretical predictions of secondary structure in Table 2.5, which are in good agreement.

CD-based predictions show a highly similar secondary structure for all tested sHsps, with a dominant proportion of β-sheets and disordered regions. This originates from the high contribution of the ACD domain (see Figure 2.5B), which consists mainly of β-strands and long disordered N-termini, especially for the organelle-targeted sHsps (as shown in Figure 2.2). We would anticipate a lower

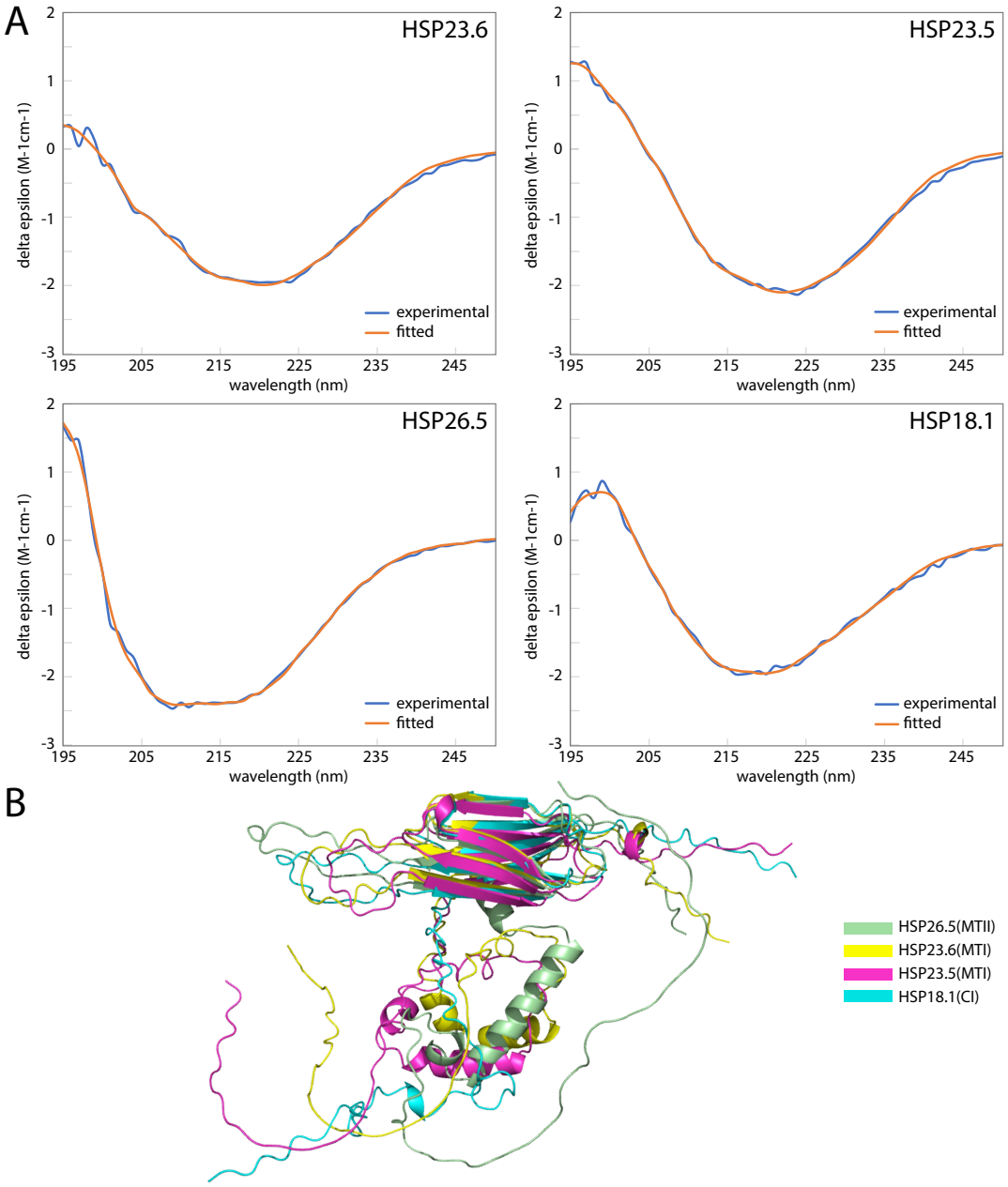


Figure 2.5: Secondary structure analysis by circular dichroism. A) The far-UV CD spectra confirm the correct folding of tested sHsps and allow for secondary structure analysis, which is in agreement with theoretical predictions. A local minimum of around 220 nm reached by all tested sHsps proves the dominant contribution of β-sheets in the secondary structure. It was expected because of the presence of the ACD domain, which is known to have a β-sandwich structure. B) Aligned monomer structure predictions obtained from AlphaFold show a high content of β-sheets in the ACD domain flanked by more disordered regions. The NTRs of organelle sHsps have a higher helical content.

protein		α-helix (%)	β-strand (%)	disordered(%)
HSP23.6	experimental	12.9	40.2	46.8
	predicted	12	36	42
HSP23.5	experimental	11.5	46.1	42.4
	predicted	14	40	39
HSP26.5	experimental	12.1	38.0	50
	predicted	13	36	42
HSP18.1	experimental	11.4	42.2	46.4
	predicted	5	42	35

Table 2.2: Comparison of experimental and theoretical secondary structure predictions. Experimental secondary structure compositions were extracted from BestSel. Phyre 2 was used to obtain theoretical predictions. These values are in agreement. Experimental data are very comparable across all tested sHsps, despite low structure identity, and show the highest contribution of β-sheets and disordered regions, as expected. These structure characteristics are crucial for maintaining the chaperone functions of sHsps.

proportion of α-helices and disordered regions for HSP18.1. This is reasoned by their shorter N-termini than for all organelle sHsps (see Fig. 2.5B, Fig. 2.3), and the fact that CI members do not possess α-helical motif in the N-terminus, however, the obtained values are similar. It might have been caused by a low precision of the secondary structure estimates from experimental data (as they rely on model standard proteins), and the fact that theoretical predictions do not consider that the packing of monomers into higher oligomers might affect their conformation.

Overall, we demonstrated here that sHsps consist mainly of β-sheet structures and disordered regions, which are crucial for maintaining their function (presented in the previous chapter, Section 2.3.1). High levels of disordered regions evince themselves in their flexible and dynamic nature, while β-strand structures facilitate substrate binding and oligomerisation. We also confirmed that the obtained proteins are folded correctly, which was crucial for the experiments described further in this chapter.

2.3.3 Organelle sHsps from *A. thaliana* display diversity of oligomeric structures and assembly

The majority of sHsps form higher oligomeric structures (as described before in Section 1.3.1)). Plant sHsps are favoured for structural studies because so far, all of them have been reported to be monodisperse, meaning they form only one size of oligomer in a native state, in contrast to human sHsps, which exhibit a strong polydispersity with some of them coexisting as a range of different sizes from 12 to 42-mers. The lower heterogeneity of the investigated samples makes plant sHsps more suitable for structural analysis. As shown in Figure 2.1 all classes of organelle-targeted sHsps evolved from cytosolic CI, which forms 12-mers. Recently, Chuanyang Yu et al. reported a tetrahedral structure of a chloroplastic HSP21 12-mer obtained by cryo-EM ([131]).

We used nESI native mass spectrometry to investigate the quaternary structure of all mitochondrial and chloroplastic sHsps from *Arabidopsis thaliana*. In particular, we were interested whether they form higher oligomeric structures, as most chaperones from this family. Moreover, we aimed to find out what oligomeric states and distributions they display. All of the sHsps investigated here have the C-terminal V/IXI/V motif (see Figure 2.3), responsible for oligomerisation, but the motif varies slightly for each species.

Acquired mass spectra for all four studied proteins are shown in Figure 2.6. The top panel represents data recorded for HSP26.5, demonstrating one dominant charge series, which indicates that this protein is monodisperse. After deconvolution, it was assigned as a 12-mer. The barely visible charge series at 9000 m/z comes from an 18-mer. The identical type of distribution was reported for CI sHsps members ([43]) where is serves as a source of polydispersity, reported for a few CI sHsps at elevated temperatures. Even though HSP26.5 does not share this feature, and it

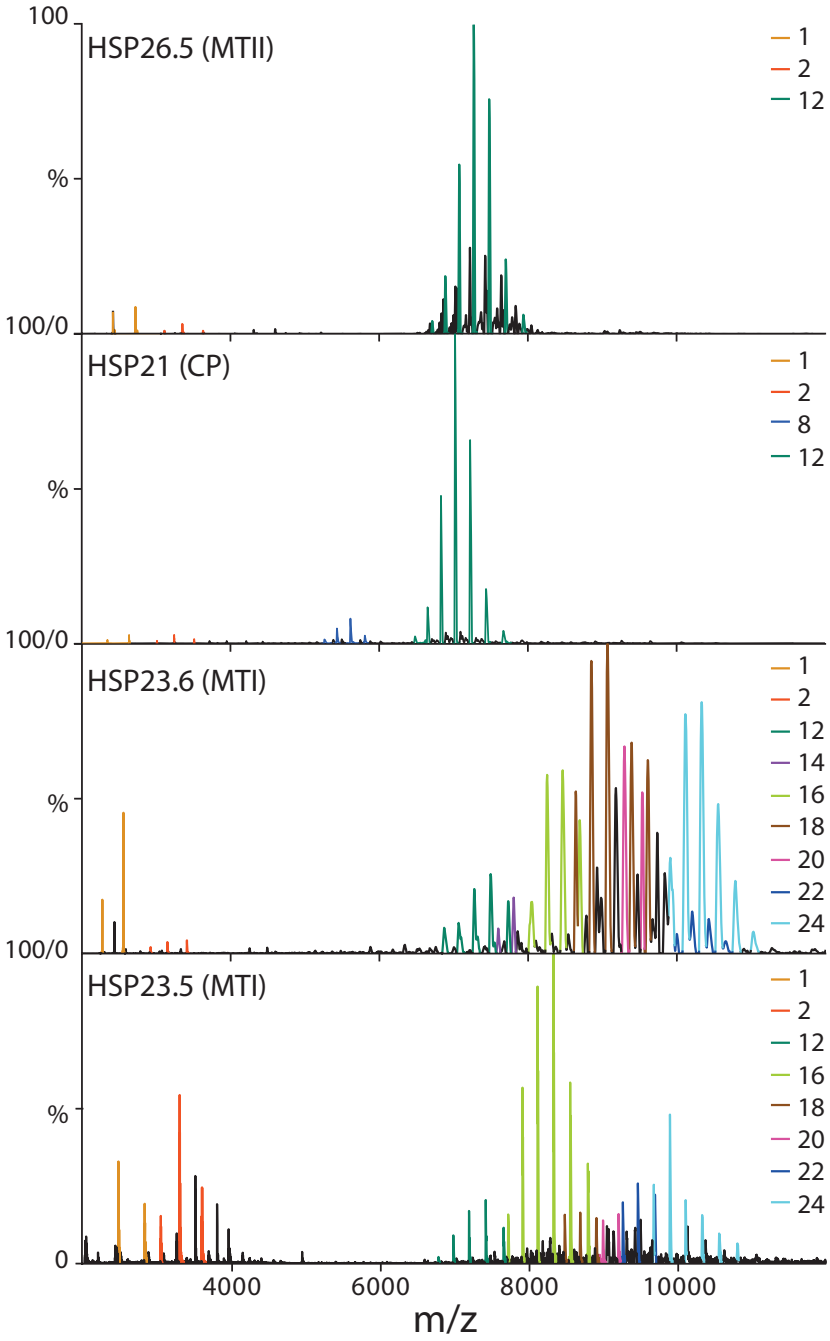


Figure 2.6: The native mass spectra of organelle-targeted sHsps from *Arabidopsis thaliana*. Native MS data show that HSP21 and HSP26.5 are monodisperse like most members of the plant sHsp family. They form predominantly 12-mers, represented by a single charge series around 7,000 m/z (green). Surprisingly, both HSP23.6 and HSP23.5 exhibit polydispersity, forming a range of higher oligomeric structures from 12- to 24-mers, spanning through a broad m/z window from 7,000 to 12,000 (peaks corresponding to one oligomeric size are labelled with the same colour). All spectra show the presence of monomers and dimers over lower m/z, indicating that analysed sHsps exist in the equilibrium between smaller and bigger species.

remains monodisperse even after long incubation at high temperature, it might be an artefact, since MTII evolved from CI over 400 million years ago, which indicates very strong structural conservation of plant sHsps despite low sequence identity (around 30%). Why this particular oligomeric sizes of additional species is observed is not defined; however, since a structure of a double disk consisting of two 6-mers has been reported for a 12-mer of sHsp, we can hypothesise that an 18-mer could include an additional third 6-mer disk, forming a triple disk arrangement.

The HSP21 mass spectrum (Figure 2.6, second panel) shows a low-abundance 8-mer detected at 5500 m/z region, with the vast majority of HSP21 subunits engaged in the formation of 12-mers, which also indicates monodispersity. This data is in agreement with a study mentioned before, showing the tetrahedral structure of HSP21 12-mer. Interestingly, a low intensity 8-mer was also reported for CII members, and extending the concept presented above for the 18-mer structure, an 8-mer could consist of two stacked tetramers, forming a cubic structure or a single ring of four dimers, shaping a square-like arrangement.

Surprisingly, for both MTI class members (Figure 2.6, two bottom panels), we detected distributions of different oligomeric species, ranging from 12- to 24-mers, only with even numbers of subunits. This demonstrates the polydispersity of those proteins, never previously reported for plant sHsps. Intensities for each oligomeric size differ between HSP23.6 and HSP23.5, with distributions centred on 18-mer and 16-mer, respectively. Moreover, 14-mer, the lowest-intensity oligomer for HSP23.6, was not detected for HSP23.5 at all. We might not exclude the possibility that it was not identified because of the overlapping of the peaks or low intensity of peaks corresponding to that oligomer.

Considering the recent finding of HSP21 12-mer showing a structure of each dimer occupying an edge of a tetrahedron, a suggestion of polyhedral structures

being more favoured for sHsp oligomers has been tempting. This could explain why only even numbers of subunits take part in oligomer formation. Moreover, it would be interesting to find out what structure is formed by a 14-mer since a polyhedron with seven edges does not exist, which might be the reason why this oligomer was not detected for HSP23.5 and was the least intense for HSP23.6.

Besides high oligomeric structures detected over higher m/z for all investigated proteins, monomers and dimers were present in the lower m/z region. It is yet another evidence of those assemblies existing in the equilibrium between smaller and bigger species and continuously and dynamically exchanging subunits ([12, 13, 43, 47, 85, 132]).

For all four investigated organelle sHsps, the V/IXI/V motif is present in the CTR (see Figure 2.3). Its importance for oligomerisation has been proved previously in multiple studies. Thus, it is unsurprising that organelle sHsps examined here exhibit the ability to form those assemblies.

To summarise, we confirmed that organelle sHsps can form higher oligomeric structures. Members of CP and MTII classes form dodecameric structures, the same as CI members, from which they evolved. This study reveals for the first time that MTI class members display polydispersity, a feature not previously observed in plant sHsps. It would be interesting to explore the evolutionary processes that transitioned CI from being monodisperse to showing polydispersity in mitochondrial MTI sHsps. This research could determine whether polydispersity enhances their cellular functions or indicates an evolutionary shift towards specialised roles and the emergence of new traits connected to those roles. Since we showed in Section 2.3.1 that mitochondrial sHsps are chaperone active, we speculate that enhancing the diversity of structures they form might improve their ability to interact with a

wider variety of substrates and thus further enhance their promiscuity, which is one of the most significant features of sHsps.

2.3.4 Two N-terminal environments of organelle sHsps

To confirm that oligomers with odd numbers of subunits are not formed by the polydisperse sHsps, we performed collision-induced dissociation (CID). CID allows for better separation of peaks. Thus, the deconvolution of complex spectra of heterogeneous mixtures becomes simpler. It also facilitates the investigation of the building blocks of complexes. The theory behind this gas-phase reaction and how it is performed is described in Section 1.5.3.

CID of chloroplastic HSP21

First, we explored these experiments on a less complex example of HSP21 (Figure 2.7). After obtaining a native MS spectrum, we can choose the m/z window of interest, either a broad range of peaks or a narrow range corresponding to a particular peak, and isolate it in the quadrupole. In the next step, high-collisional energy is applied in the HCD cell only to the separated ions, which results in the stripping of highly charged monomers from the native species. Charge-reduced N-1 species appear over higher m/z region (after subjection to HCD 150 eV). With the increasing value of energy hl(to 300 eV) applied, two subunits can be stripped (N-2) from the parental oligomer (N), which causes further charge reduction.

Figure 2.7 demonstrates the CID mechanism for chloroplastic HSP21. The left panel illustrates the dissociation of a 12-mer upon its previous selection in the quadrupole (green area), which results in the appearance of a new charge series (pink). Knocked-off unfolded monomer, with a high charge uptake, appears in the lower m/z, and charge-reduced 11-mer and 10-mer stripped of one and two subunits,

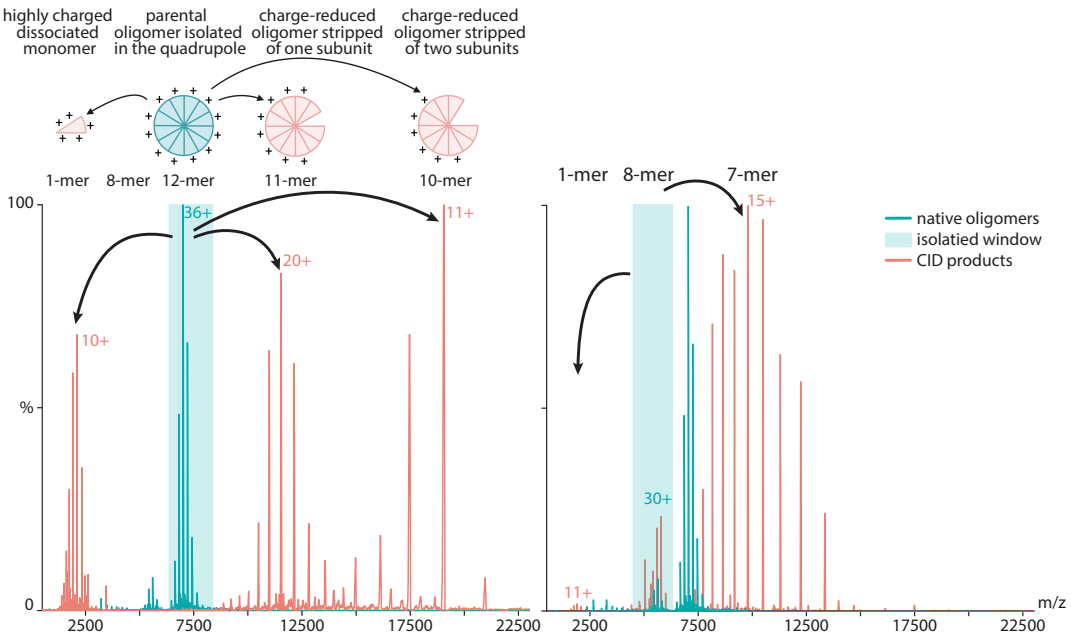


Figure 2.7: CID of HSP21 after previous isolation in the quadrupole. Native MS data (green) represent intact oligomers, further isolated in the quadrupole (green area), upon which they undergo dissociation in the HCD cell (pink). Isolation and CID of the 12-mer (left) lead to the appearance of charge-reduced 11-mer (8,000-15,000 m/z) and 10-mer (15,000-22,000 m/z). The lower m/z region (500-2,500) shows a highly-charged monomer subunit, stripped as a result of subjection to a higher collision energy. Isolation and CID of an 8-mer (right) results in the appearance of a charge series corresponding to a monomer-stripped 7-mer (7,500-15,000 m/z).

respectively, emerge in the higher m/z region. The right panel shows the selection of an 8-mer in the quadrupole, after which CID resulted in the appearance of a 7-mer.

CID of mitochondrial MTI sHsps imply two different conformations of monomers within higher oligomers

We aimed to confirm that oligomeric distributions of MTI members consist exclusively of species with even numbers of subunits. To do so, we first selected a broad m/z window of HSP23.6, and by applying increasing energy in the HCD cell, we performed CID (Figure 2.8). Charge reduction of oligomers caused a shift of the peaks towards higher m/z and their better separation. This allowed for better deconvolution of data of a complex, heterogeneous sample. The product of dissociation resulted in the detection of only odd oligomers (coming from parental

species with even numbers of subunits).

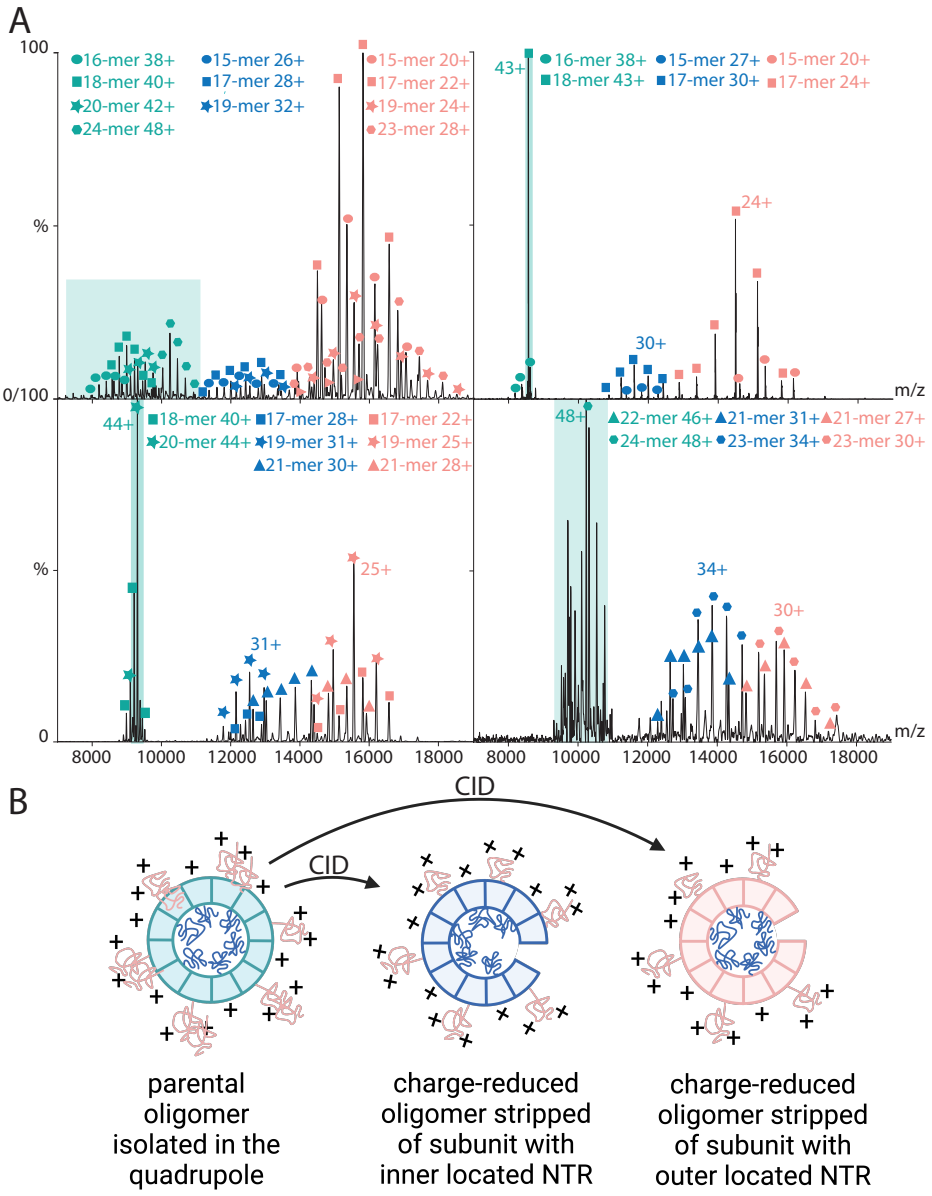


Figure 2.8: CID of HSP23.6 shows an untypical manner of dissociation. A) CID of HSP23.6 leads to the appearance of a charge series with two maxima (called here a double charge series), both corresponding to the same mass of N-1 species. They vary in charge, with lower charged N-1 peaks (pink) stripped of monomers, which took up more charge than in the higher charged N-1 oligomers (blue). Selection windows and peaks corresponding to native oligomers are highlighted in green. The figure represents MS data for different m/z selection windows, corresponding to different oligomeric sizes, which all exhibit this repeating pattern. B) Schematic illustrating a hypothetical explanation of untypical MS data. The double charge series can be explained by two different environments of monomers within the oligomeric structure, which, upon CID, leave with different amounts of charge. We hypothesise that half of the monomers are oriented with their NTRs localised in the inner cavity of the oligomer and the other half on the outside.

Another interesting, untypical phenomenon was observed in the CID spectra of HSP23.6. Dissociation of each oligomeric species resulted in the appearance of two new charge series assigned to the same charge-reduced oligomeric size. To illustrate it better, we performed CID with a narrower m/z selection over different m/z values (Figure 2.8A). The obtained tandem mass spectra of HSP23.6 show the same feature of double charge-series appearing as a result of dissociation of 16, 18, 20, 22, and 24-mer. In order to obtain two different charge series for an ion with the same mass, coming from the dissociation of the same parental oligomer, the dissociating monomer needs to leave in two different ways, taking up different amounts of charge. We speculate that it is caused by two distinct architectures in which monomers are orientated in the oligomer, with half of them having NTRs inside of the polyhedral structure and the other half localised outside (which is schematically presented in Figure 2.8B). We described studies suggesting that not all of the NTRs might fit in the inner cavity of the sHsps' oligomers (Section 2.2.2), considering that the NTRs of organelle sHsps are significantly longer than those of other classes (as shown in Figure 2.2), some of them might be forced to be located outside. This seems like a good explanation of the presented data. The monomers that have NTRs outside could take up more charge in the CID process, and thus, two charge series are created, which differ in charge but correspond to the same mass.

Another explanation for this untypical observation could be two alternative pathways of dissociation of packed monomers from the oligomer; for example, in a sandwich-like structure, more energy would be needed to dissociate monomers in the middle in comparison to outer-located subunits. However, none of the structures reported so far for sHsps showed conformations which would match such a hypothesis, as in both double disk and tetrahedral structures (reported for

12-mers), all subunits have an analogous arrangement and hence need the same amount of energy to remove them.

To check if it occurs for other mitochondrial sHsps, we performed MS/MS experiments for both HSP23.5 and HSP26.5.

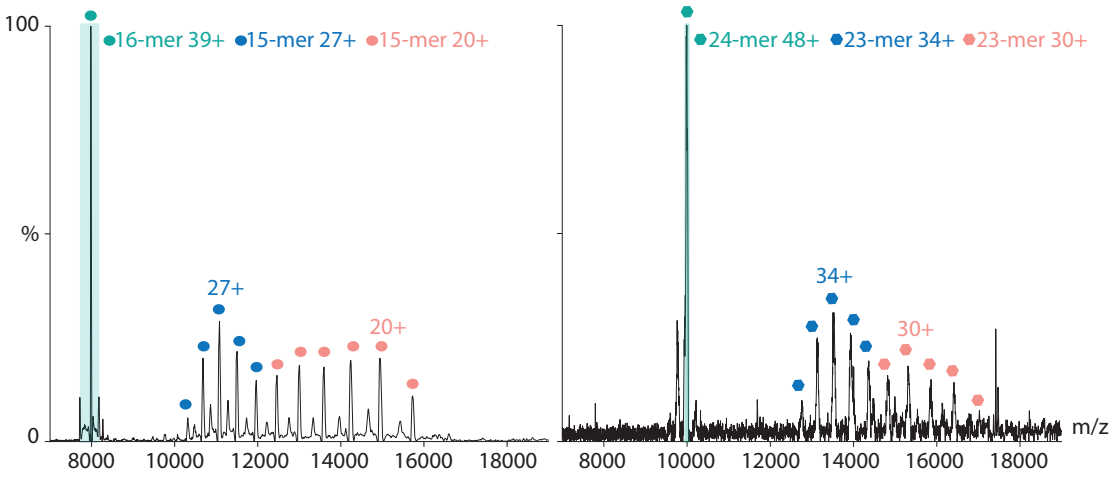


Figure 2.9: Tandem MS experiments of HSP23.5. Green areas represent m/z windows isolated in the quadrupole. Peaks corresponding to native oligomers are marked with green symbols, and N-1 charge series with different maxima are labeled with blue (higher-charged) and pink (lower-charged) symbols. Two different oligomeric sizes of HSP23.5 were subjected to CID. Both 16-mer and 24-mer, as a result of dissociation, form a double charge series, indicating two different orientations of monomers within the oligomeric structures.

We selected 16-mer and 24-mer of HSP23.5 (marked by green area in Figure 2.9) and increased the collisional energy in the HCD cell. Both resulted in the double charge series, with two maxima for different charges.

Proline effect of HSP26.5 (MTII) implies externally located NTR

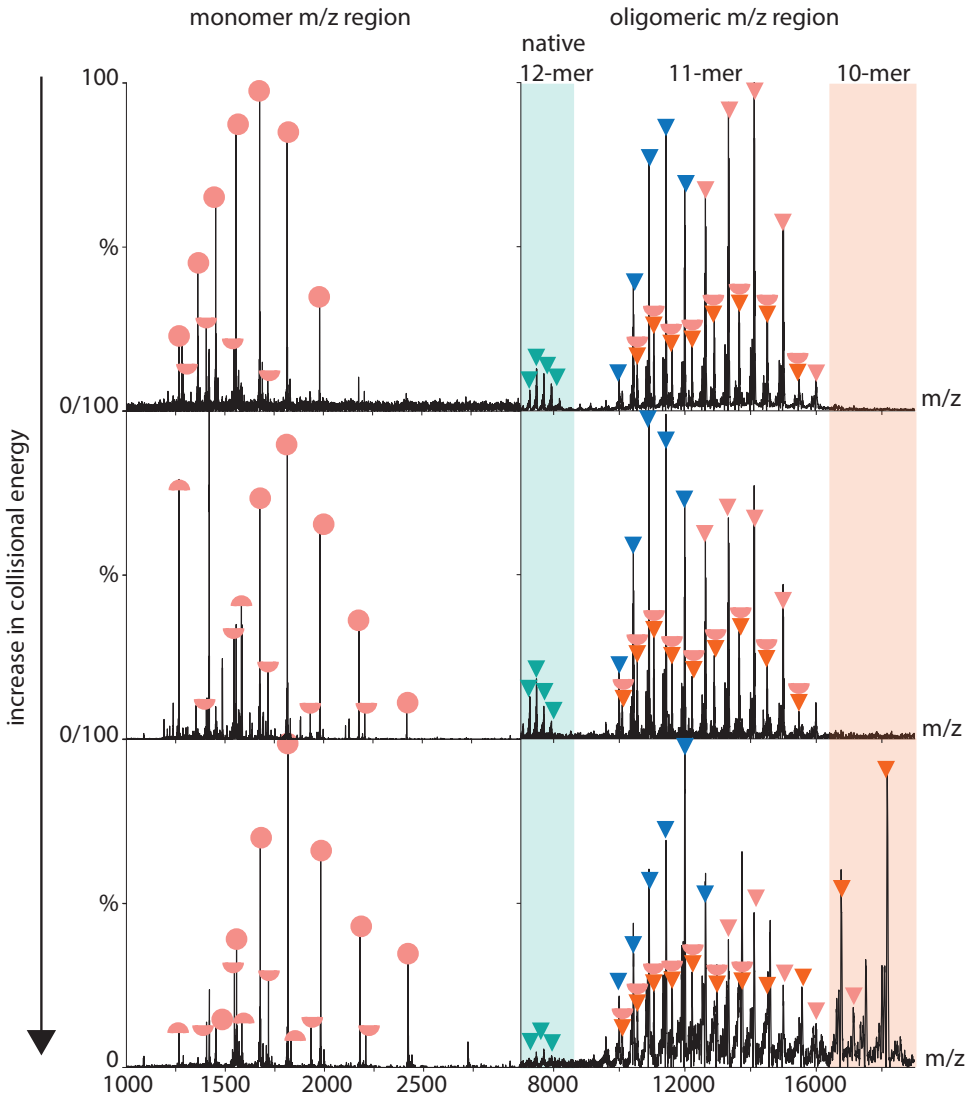
CID of HSP26.5 12-mer resulted in not only a double charge series of 11-mer but also another undefined species, with peaks spanning between the peaks of 11-mer (shown in Figure 2.10). This charge series, after deconvolution, was assigned to a species corresponding to a mass of 233,250 Da, which does not match any different oligomeric size or truncation. We performed denatured MS measurements to check

if contaminations were present in the sample, which led to the identification of the HSP26.5 monomer exclusively (data not shown). The mass difference between the 12-mer and the other detected undefined species is 6,330 Da.

With an increase in the applied HCD energy, we observe not only a shift in the charge distributions towards species with higher charges but also a decrease in the abundance of the 233,250 Da species and the appearance of the N-2 charge series.

In order to determine the origin of the undetermined species, we optimised parameters on the mass spectrometer to focus on the low m/z region while still dissociating a 12-mer of HSP26.5. This is not trivial on the instrument that was used, as it is designed for the transmission of big molecules; however, attempts to perform this experiment on a different mass spectrometer (Synapt G2, Waters) failed to achieve a satisfactory resolution. In all acquired and presented here mass spectra, a highly charged monomer of HSP26.5 is present (see Figure 2.10, left), which is a product of CID. Another charge series detected in all recorded spectra is a low-intensity species assigned to a mass of 15,450 Da (Figure 2.10, left). Summing up the mass of this species (15,450 Da), with a difference between the 11-mer and the undefined species (6,330 Da) identified above, which was detected in the higher m/z mass region, results in a mass of 21,780 Da, which is precisely the mass of HSP26.5 monomer. This suggests that the monomer might undergo dissociation into two parts. In two of the mass spectra, we also managed to spot the 6,329 Da species, corresponding to the smaller fragment.

After investigation of the HSP26.5 amino acid sequence, we defined that this mass corresponds to the N-terminal fragment, 54 amino acids long, with a perfectly matching mass. This fragment ends just before a proline (see Figure 2.10B). We assume that the phenomenon we observe here is described in mass spectrometry proline effect, which is based on the selective cleavage of the amide bond adjacent to



Mass (Da)	Species
21,780	$[HSP26.5]_1$
6,329	$[HSP26.5_{1\text{-}54}]_1$
15,450	$[HSP26.5_{55\text{-}190}]_1$

Mass (Da)	Stoichiometry
261,810	$[HSP26.5]_{12}$
239,840	$[HSP26.5]_{11}$
239,840	$[HSP26.5]_{11}$
232,040	$[HSP26.5]_{10}[HSP26.5_{55\text{-}190}]_1$
217,900	$[HSP26.5]_{10}$

Figure 2.10: CID of HSP26.5 with increasing HCD energy. High and low m/z acquisitions correspond to the same HCD energy applied. Right: MS data representing intact 12-mer (green), N-1 (white area), and N-2 m/z region (orange area). Besides the double charge series described above for other organelle sHsps (labelled with blue and pink triangles), an unexpected charge series (pink-orange symbol) spans between N-1 peaks. It was assigned to 232,040 Da, which is 6,329 Da smaller than 11-mer. Sequence analysis allowed us to determine it as an 11-mer lacking 54-amino acids long NTR peptide, which ends just before a proline. Left: Optimising parameters on the mass spectrometer to focus on the lower m/z region allowed the detection of both parts of the cleaved monomer: 6,329 Da and 15,450 Da, which make up the exact monomer mass after summing up.

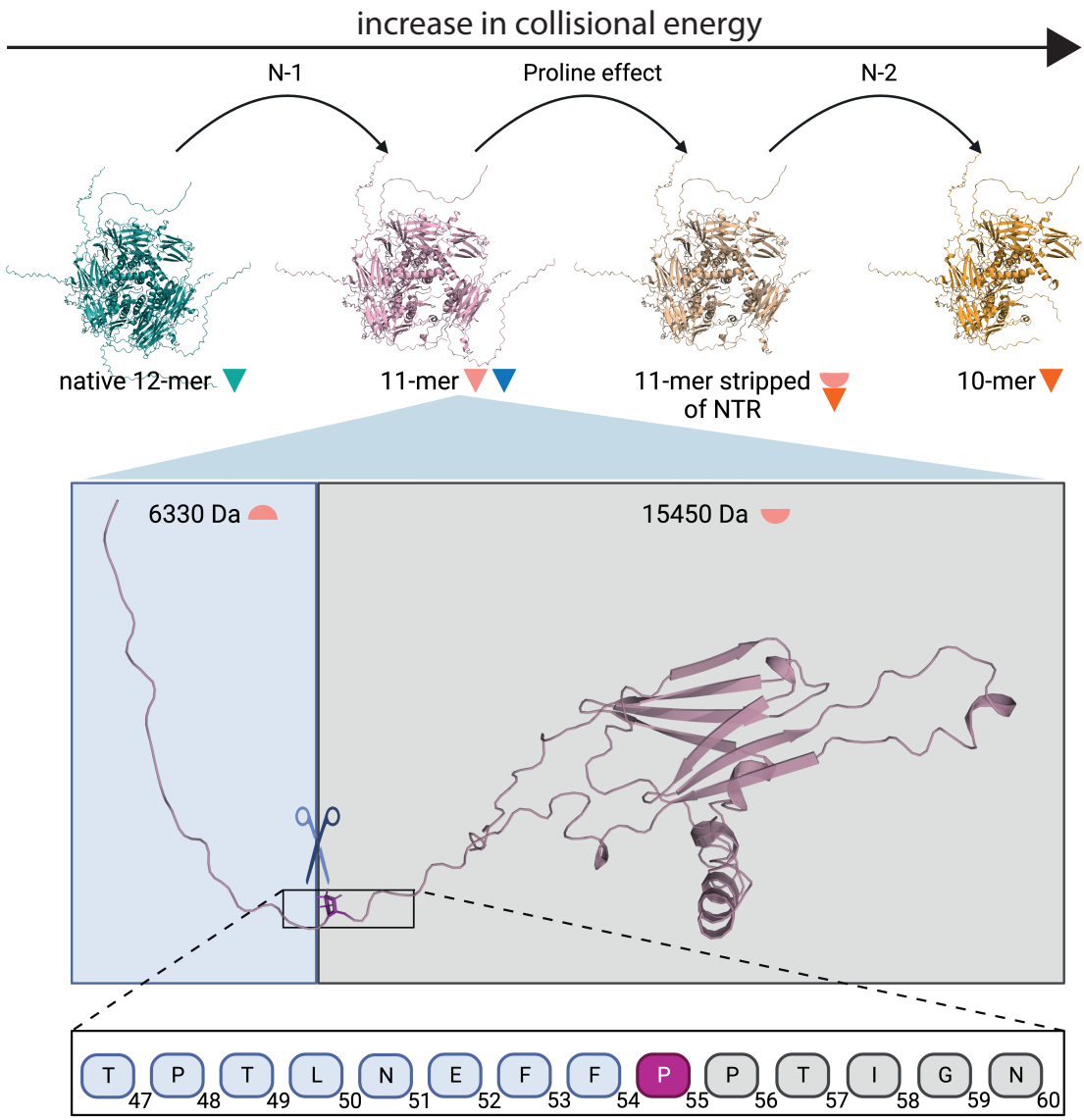


Figure 2.11: Proline effect of HSP26.5 during MS/MS experiments. Schematic explaining the proposed model of subunit dissociation. With an increase in the HCD energy, HSP26.5 first loses a monomer (either higher charged with the outer NTR or lower charged with the NTR on the inside). Further increase of collisional energy causes a proline effect, cleavage of the N-terminal peptide (most likely located on the outside) from the 11-mer. The last step is a dissociation of the remaining part of the broken monomer, and hence, a 10-mer is being formed. The bottom panel shows a close-up of the monomer from the AlphaFold predicted oligomeric structure and highlights where exactly the proline effect takes place and the amino-acid sequence it corresponds to.

an N-terminal proline ([133]). The mechanism of this phenomenon was studied quite broadly, but in a context of digestion-generated peptides, cleaved during tandem MS, which allows for improvement of peptide identification. Raufs et al. described the

proline effect based on multiple variations of proline-including peptides: AAPAA, AVPLG, AAAPA, AGPGA, and APPAA. They reported a selective N-terminal cleavage of the amide bond adjacent to the proline residue.

The data describe very well a series of events (schematically illustrated in Figure 2.11) where the native 12-mer subjected to CID, dissociates not only as a highly charged monomer but also as a 54-amino acids long N-terminal peptide, as a result of the proline effect. Upon further increase of the HCD energy, we observe the release of the rest of this subunit and the production of the N-2 charge series.

It is another proof of the two different orientations of the NTRs. Being localised on the surface of the structure might be a good reason to undergo a proline effect. Interestingly, the AlphaFold-generated structure (Figure 2.11) also presents the NTRs on the outside of the oligomer, with the proline in question being exactly on the edge of the surface of a tetrahedron.

Additional unlabelled low-intensity peaks in Figure 2.11, which closely neighbour the 11-mer charge series, also correspond to N-1 series but with HSP26.5 subunits lacking C-terminal fragment of either 9 or 39 amino acids. Those peptides may have been cleaved post-translationally, but subunits lacking these CTRs also got incorporated into higher oligomers.

2.4 Discussion

2.4.1 Mitochondrial sHsps act according to the canonical holdase activity

sHsps act towards stress-labile proteins by inhibition of their aggregation, which was first determined for the cytosolic CI plant sHsps, and it is thought of as a canonical chaperone activity of sHsps. Since then, it has been described for several members of different evolutionary classes, including cytosolic CII, CIV, mitochondrial CP,

and peroxisomal Po ([19, 50]). However, some of the proteins qualified as sHsps show a lack of ability to prevent aggregation. One such example is HSP21.7, a member of the cytosolic CV. Perhaps it diverged to pursue a different function in the cell or evolved to interact in a more specialised manner. Taking up novel functions, which contribute to protein homeostasis in a different way, would not be surprising, considering the higher number of homologs of plant sHsps, influenced by the pressure for subfunctionalisation.

Despite their different subcellular locations, we demonstrated that mitochondrial sHsps from *A. thaliana* can act canonically as efficient chaperones by inhibiting the aggregation of a heat-labile model substrate.

The precise mechanism they utilise to fulfil this function within the cell remains unclear, with various theories proposed. Most studies suggest that higher oligomeric structures serve as a reservoir for chaperone active forms, keeping their hydrophobic interfaces occupied in a non-stress environment until the moment they can be exploited by a protein in need. It also allows the avoidance of uncontrolled interactions and limits the crowdedness of an already busy cellular environment.

2.4.2 Structural diversification of organelle sHsps

The primary structure of the studied organelle sHsps shows significant changes. Their long NTR includes α-helical motif unique to these classes. Moreover, all of them lack the $\beta 6$ strand in the ACD, which is crucial for dimerisation. It entails the possibility of more prone dissociation of their higher oligomers.

All plant sHsps studied so far have been reported to display monodispersity, with a dodecameric structure reported for CI and CII sHsps members. Similarly, native MS data presented in this chapter show that both chloroplastic HSP21

and mitochondrial HSP26.5 form monodisperse 12-mers. Interestingly, a low-intensity 8-mer was reported for CII *Ta*HSP17.8, which is also visible in the HSP21 spectrum (Fig. 2.6). Chloroplastic sHsps are evolutionarily the oldest organelle-localised class of plant sHsps. They diverged from cytosolic sHsps over 450 million years ago, which explains those similarities. However, it also highlights a strong conservation of those structures.

Surprisingly, the members of mitochondrial class MTII (HSP23.6 and HSP23.5) exhibit polydispersity, which has never been reported for plant sHsps before. They form higher oligomers ranging from 12- to 24-mers, with only even stoichiometries, forming an equilibrium with monomers and dimers present over lower m/z (Figure 2.6).

Polydispersity is common amongst mammalian sHsps, especially those localised in the eye lens. They maintain high heterogeneity of co-existing forms due to an exceptionally crowded protein environment. It serves to facilitate the solubility of a broad client protein spectrum.

It remains unknown whether mitochondrial sHsps evolved polydispersity for the same reason, as the number of sHsp homologs targeted to mitochondria is significantly lower than to the cytosol, or perhaps they evolved to form different structures as a consequence of distinct functions they might pursue in the new subcellular location. Alteratively, it is possible that the change of the dispersity happened as a result of thermodynamic relaxation, not connected to the change of functions of mitochondrial sHsps.

2.4.3 Two orientations of monomers within higher oligomers of mitochondrial sHsps

The NTR of sHsps, because of its flexible nature, escapes structural determination. The x-ray structure reported for cytosolic HSP16.9 ([23]) described more ordered regions of the NTR to be localised in the inner cavity of the 12-mer. In the previous section of this chapter (Section 2.2.1), we mentioned studies that suggested dual orientations of N-terminal regions.

Firstly, we presented CD data, which confirmed the correct folding of the obtained mitochondrial sHsps. Further, the MS/MS results showed an atypical behaviour of oligomeric structures. After dissociation in the gas phase of the mass spectrometer, each of the mitochondrial sHsp oligomers leads to the appearance of two charge series corresponding to the same mass of oligomer-stripped species, which differ in charge. It is evidence of two different environments of the monomers, which could be explained by their different orientations within the oligomeric structure (Figure 2.12). We hypothesise that some of the sHsp structures examined here have monomers with the NTRs oriented on the outside of the oligomer; hence, during the CID, they take up more charge and lead to N-1 charge series with a lower charge. The other monomer species with the NTRs localised in the inner cavity of oligomers are characterised by lower charge and need higher HCD values to be stripped.

Another potential experiment, which could also reveal the dual architecture of NTRs in the structures of mitochondrial sHsps, involves limited proteolysis MS. After the digestion process, we anticipate identifying two distinct NTR fragment profiles: one featuring partially cleaved NTRs and the other with entirely intact NTRs. Such findings would support our hypothesis, demonstrating that some NTRs are positioned on the exterior of the oligomer while the remainder are situated within the inner cavity.

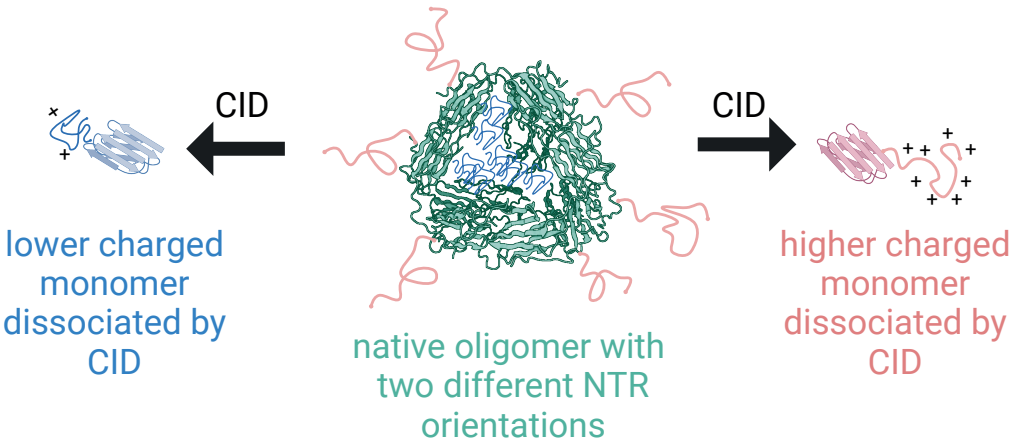


Figure 2.12: Dual orientation of sHsp monomers. Schematic illustrating two different conformations of monomers within the higher oligomeric structure, with half of them locating their NTRs on the surface of the assembly and the other half in the inner cavity of a tetrahedral oligomer. This model explains the mass spectrometry data presented in this chapter.

This hypothesis is supported by yet another interesting observation of the proline effect, which HSP26.5 12-mer undergoes during the CID. Stripping off only the N-terminal peptide of the monomer, which is involved in the formation of a 12-mer, suggests its outer orientation. Proline effect is observed in the top-down MS for identified peptides in the lower m/z, however losing part of a subunit from an intact complex is a rather untypical observation, which further suggests the outer location of the cleaved NTRs. The proline effect is not observed for the cytosolic sHsps, which suggests that the exceptionally long NTR of the organelle sHsps is the reason for their dual architecture.

To summarise, we showed here that some of the organelle sHsps diverged significantly. Possibly, they evolved on their way to subfunctionalisation in a different cellular location. In the next chapter, we aim to find out how that affected their co-assembly and, hence, the structure of interfaces responsible for oligomerisation.

2.5 Acknowledgments

I want to thank Dr Zihao Wang, who trained me when I started my PhD. He familiarised me with the laboratory and instruments and helped with protein expression and purification at the very beginning. I am very grateful to Dr Dominik Saman, who showed me how to perform chaperone assays, introduced me to Circular Dichroism, and helped with those experiments. I am very grateful to Navoneel Sen, who generated the AlphaFold structures presented here. Two Part II students, Lauren Parsons and Georgie Mullan, joined for some parts of this project. It was a pleasure to work with them, transfer my knowledge, and have company during experiments, as well as partners for discussions. Lauren purified HSP18.1 and conducted chaperone assays, and Georgie helped with HSP26.5 purification. HSP21 samples were provided by Prof. Cecilia Emanuelsson, for which I am very grateful, and Prof. Elizabeth Vierling provided plasmids for untagged HSP26.5, HSP23.5, and HSP23.6. I also want to thank Patrick Treffon from the Vierling lab for providing an empty SUMO plasmid and guiding me on how to insert a gene of interest.

2.6 Methods

2.6.1 Protein expression and purification

Predicted mature sequences (after the removal of transit peptides) of HSP23.6 (AT4G25200), HSP23.5 (AT5G51440), and HSP26.5 (At1G52560) were encoded in a pET-23b vector (kindly provided by the Vierling Lab, UMass Amherst). BL21(DE3)pLysS competent cells (Novagen) were transformed with each of the vectors and plated on LB-agar plates containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. Single colonies were picked and grown overnight at 37°C in 20 mL of Luria Broth (LB), containing the appropriate antibiotics, shaking at 180 rpm. Starter cultures were inoculated in 1 L of LB media (containing appropriate antibiotics) and grown at 37°C in an incubator shaking at 180 rpm until optical density (OD) reached values between 0.6 and 0.8. Expression was induced by adding IPTG to a final concentration of 0.4 mM, after which cells were incubated overnight at 20°C. Cells were harvested by centrifugation at 4000 xg for 10 minutes, flash-frozen in liquid nitrogen, and stored at -80°C until purification.

The cells were resuspended in 20 mL of lysis buffer (25 mM Tris, 1 mM EDTA, pH: 8.0), and two protease inhibitor cocktail tablets (Roche) were added, after which the cells were lysed by microfluidising. The cell lysate was centrifuged at 20,000xg for 20 min to separate the soluble fraction, which was next filtered and loaded onto two connected 5-mL HiTrap Q Sepharose Fast Flow columns (GE Healthcare), which were washed with the lysis buffer, before eluting the protein with a gradient of a lysis buffer containing 1 M NaCl. Collected fractions were analysed by 13.5% SDS-PAGE gel. Fractions containing protein of interest were collected, buffer exchanged back to the lysis buffer, and loaded onto two connected 5-mL HiTrap Q Sepharose High Performance (GE Healthcare) columns, and the process of anion

exchange chromatography was repeated as described above. The purest fractions were collected, buffer exchanged into PBS buffer (pH 7.4) with 1 mM DTT, and concentrated to 2 mL, using centrifuge concentrators with appropriate molecular weight cut-off, upon which they were injected onto the Superdex 200/20 column, and run at 0.5 mL/min for 1 column volume. Collected fractions were analysed with SDS-PAGE, and the purest fractions were concentrated. The concentration was determined using NanoDrop by measuring absorbance at 280 nm and calculated using the extinction coefficient obtained from the UniProt ProtParam tool (based on the monomer sequence). Samples were aliquoted, snap-frozen, and stored at -80°C.

2.6.2 Chaperone assays

Chaperone assays were conducted on a FLUOstar® Omega (BMG LabTechTM) microplate reader. All experiments were held in a PBS buffer. Protein samples were buffer-exchanged or solubilised first if needed. Alcohol dehydrogenase (ADH) from *Saccharomyces cerevisiae* (Sigma-Aldrich) was used as a model substrate. Varying sHsps-ADH molar ratios were tested, keeping the ADH concentration constant across all measurements (4 µM). First, the 96-well plates were loaded with sHsps samples, equilibrated at 37 °C for 10 minutes, and then the substrate was added. Just before starting the acquisitions, 20 mM DTT was added to initiate the aggregation. The experiments included measurements of positive and negative controls, as well as background acquisition. Each sample contained at least three repeats. Chaperone assays were held at 42 °C, and the absorbance was measured at 340 nm over a 16-hour window, with orbital shaking after each cycle.

2.6.3 Probing protein secondary structure with CD

Protein samples were buffer exchanged into a CD buffer (10 mM Potassium Phosphate, 50 mM Sodium Sulphate; pH 7.2), diluted to 10-20 µM monomer concentration to a final volume of at least 400 µL, or for the folding tests of the MS samples (in ammonium acetate) they were diluted with the CD buffer right before the measurements. For all CD acquisitions, we used a 1 mm quartz cuvette (cleaned with 70% nitric acid beforehand) and recorded data with a Chirascan^{TM} V100 Circular Dichroism spectrometer (Photophysics). The temperature was controlled by a Series 800 Thermoelectric Cooler Controller (Alpha Omega Instruments) and a Julabo AWC100 circulating water bath (JULABO GmbH, Seelbach, Germany). Background spectra were acquired using 400 µL of a CD buffer, after which the background subtraction was enabled for the protein sample measurements, held at 25°C, and scanned in the range of 180-250 nm, with at least three repeats obtained for each sample. BeStSel software ([129]) was used for data analysis and extracting the secondary structure predictions based on experimental results, which were compared to theoretical data, predicted based on amino acid sequence using Phyre2 ([130]).

2.6.4 Native MS of organelle sHsps

Protein samples were thawed on the day of experiments and buffer-exchanged into 200 mM ammonium acetate (pH 6.9) by running them through Bio-Spin columns (BioRad) twice. The concentration of proteins was determined as described above (Section 2.6.1). The samples were diluted to 20 µM monomer concentration if needed.

All native MS measurements were obtained on a Q Exactive™ Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Scientific™). Protein samples were nano-electrosprayed in a positive-ion mode, using in-house prepared gold-coated glass capillaries, by application of modest backing pressure (around 0.5 mbar) and with

a capillary voltage of 1.1-1.4 kV. To improve desolvation and limit the presence of adducts, the source temperature was set between 50-150 °C, and desolvation voltage was optimised for each sample and was set between -50 V and -100 V. Nitrogen was introduced into the HCD cell as a collision gas, with the trapping pressure set to 7.5. HCD energy was set between 0-10 V for the measurement of intact oligomers. Applying low values in the HCD cell helps with the transmission of high-mass species while not dissociating any other oligomers. For the activation measurements, dissociation was induced in the HCD cell by applying collision energy (CE) between 150-300 eV upon selection of the m/z window of interest in the quadrupole (as stated in the main text). The mass spectra of N-1 charge series were acquired at HCD=150 eV, and ejection of the second subunit was recorded at HCD=300 eV, however both of these reactions were observed already at lower collisional energy values, with lower intensity of dissociation products.

Ion transfer optics, optimised for the best transmission of big oligomeric species, was set as follows: injection flatapole 10 V, inteflatapole lens 9 V, bent flatapole 8 V, transfer multipole 4 V, C-trap entrance lens 3 V. The instrument was operated with "Low Detector Optimization" at a resolution of 17,500 at 200 m/z. Spectra were acquired with averaging of 10 microscans, injection time of 100 ms, and AGC target set to 10^6. Mass spectra were recorded in the m/z range from 500 to 20,000, with the noise threshold parameter being set to 3. Data were deconvolved using UniDec software ([134]).

Data were input with no smoothing, binning, or subtraction, but for some of the data, non-linear sampling was used (set between 10-100). FWHM was fitted for each data set manually. The data were deconvolved in the m/z range of 1,000-20,000 m/z and scanned through a charge range of 10-50$^+$. Data deconvolution was performed in various mass ranges, adjusted for each sample. All other UniDec

parameters were left at their default values. The oligomer and mass tool was used to assign oligomeric states.

3

Co-assemblies of organelle sHsps from plants

Contents

3.1 Motivation 76
3.2 Introduction 77
3.2.1 The diversity of sHsps assemblies 77
3.2.2 Physico-chemical description of protein co-assemblies . . 79
3.3 Results 83
3.3.1 Diverse co-assembly modes of organelle sHsps from different cellular compartments 87
3.3.2 Almost unbiased co-assembly of mitochondrial sHsps from different evolutionary classes 94
3.4 Discussion 100
3.4.1 Mapping the energy of interactions between organelle sHsps 100
3.4.2 Evolutionary conservation of organelle sHsps 102
3.5 Acknowledgments 105
3.6 Methods 106
3.6.1 Protein expression and purification 106
3.6.2 Mass Spectrometry - subunit exchange of sHsps 107
3.6.3 Description of co-assembly using statistical mechanics . 108

3.1 Motivation

sHsps, when co-localised in the cell, can exchange subunits forming hetero-oligomers. The diverse assembly nature of those chaperones has been a topic of many studies. These were however limited to the investigation of co-assemblies between cytosolic sHsps. For plant sHsps, it has been reported that members of the same cytosolic class can exchange subunits ([43, 87]), forming hetero-oligomers. On the contrary, members of different evolutionary classes exhibit specific self-assembly as a result of subfunctionalisation. The gain of new functions caused the structural diversification of those sHsps and, hence, the loss of the ability to exchange subunits ([26]).

The fact that sHsps exhibit various interaction modes is still not well understood. Whether the co-assembly influences their activity and how it is regulated by evolution is not clear. Here, we test organelle sHsps from different evolutionary classes, localised in either the same or different cellular compartments. We examine their ability to co-assemble to establish how evolution affects their interfaces. Moreover, we test how sHsps evolve when they are physically separated, and neither entropic bias for heteromerisation nor evolutionary pressure for subfunctionalisation is present.

In the previous chapter, we presented data that show the structural diversification of organelle-localised sHsps. As functionally distinct proteins require separate structures, we use the subunit-exchange ability of sHsps to estimate if those sHsps diverged significantly to pursue novel roles in different subcellular locations and, consequently, lost the ability to co-assemble, as has been described for cytosolic sHsps ([26]).

Here, we present a broad range of modes of interactions between different sHsps, which were observed in the experimental MS data. Those interactions

were further quantified using a statistical-mechanical model to verify the assembly preference of each studied pair.

3.2 Introduction

3.2.1 The diversity of sHsps assemblies

As mentioned in Section 1.3.1, most sHsps assemble into large oligomers, then in Section 2.3.3, we showed that it is also true for organelle-targeted plant sHsps from *A. thaliana*. Oligomerisation of proteins plays a crucial role in their function, regulation, and, thus, in cellular homeostasis ([135]).

Gene duplication is responsible for the enlarging of the proteome. If we consider an oligomeric protein right after gene duplication, both daughter genes encode hetero-oligomerising proteins, which co-assemble in an entropy-favoured manner, exhibiting binomial distribution (Fig. 3.1, middle). After that, the daughter genes can diverge, but the co-assembly can be maintained and beneficial if formed protein complexes remain active and functional or the sub-function is present within the same complex (Fig. 3.1, right). In another scenario, daughter genes evolve new functions, and because proteins with separate roles require structurally different proteins, their interfaces become incompatible, which keeps them from co-assembling (Fig. 3.1, left).

sHsps are known for a wide diversity of their assembly manners. They show all of the above-described modes, from specific self-assembly to obligate co-assembly (Fig. 3.1). It is still unclear how relevant assembly is for their function, how it influences their activity, and how those interactions are regulated by evolution.

Some of the co-assembling sHsps form a broad spectrum of hetero-oligomers carrying variable numbers of subunits of different sHsps (Fig. 3.1, middle), which has been hypothesised to widen the already broad range of substrates that those chaperones interact with ([78]). For example, human HspB1 and HspB5

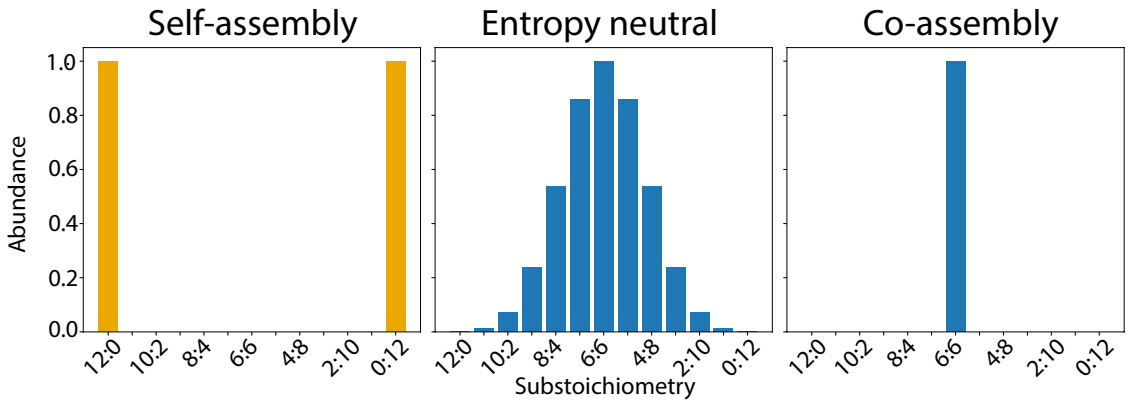


Figure 3.1: Hypothetical modes of assembly of two monodisperse proteins. Left: Two proteins do not exchange subunits and form only homo-oligomers. Middle: 'Unbiased' co-assembly, showing all possible stoichiometries of hetero-oligomers in a binomial fashion, which is determined by the maximal entropy of the system and where there is no energy preference towards homo- nor hetero-interactions. Right: Obligate n:n co-assembly (represented by a 6:6 hetero-12-mer), where subunit exchange between the two proteins is preferential. Yellow bars represent homo-mers and blue bars illustrate hetero-mers.

form a diversity of hetero-oligomers within different oligomeric states, which was demonstrated to stimulate their chaperone activity of parental species, created as a result of subunit exchange between the two ([136]). Another example is mixed oligomers of HspB5 and HspB4, which form heterogeneous complexes ranging from 1:1 to a 1:3 molar ratio. They are present in the eye lens and are responsible for proper vision by inhibiting the aggregation of insoluble proteins ([55]). HspB2 and HspB3 form heteromers of various sizes, with a well-defined and unusual subunit ratio of 3:1, which was shown to change the biochemical properties of those complexes in comparison to their homo-oligomers ([44]).

Organisms frequently exposed to temperature changes, such as plants, have evolved a variety of sHsp homologs, which are co-expressed during stress. As a result of recurring gene duplications, each plant species usually has multiple members of each evolutionary class (more details are included in Chapter 2). Previously, it has been described that cytosolic plant sHsps from the same classes

can co-assemble, forming hetero-oligomers even if it concerns two proteins from species that are evolutionarily distant, their ability to co-assemble is conserved. However, members of different evolutionary families (from the same or different species) lose this ability, as their interfaces become incompatible and they exhibit obligate self-assembly ([19, 26]).

Hochberg et al. ([26]) described how two cytosolic sHsps, HSP18.1 and HSP17.7, from the same plant (garden pea), but different evolutionary classes (CI and CII, respectively), which evolved through a gene duplication event over 450 million years ago, lost their ability to hetero-oligomerise on their way to subfunctionalisation. Next, the authors developed a statistical-mechanical model describing the energy difference needed for two monodisperse oligomers to exhibit specific self-assembly.

Here, we study interactions between organelle-localised sHsps from the same plant species but different evolutionary classes targeted to the same or different subcellular compartments, by analysing modes of their assembly. We intend to evaluate whether their ability to exchange subunits was lost due to them gaining novel functions, as has been reported for cytosolic sHsps. Moreover, we want to establish how sHsps interfaces evolve when there is no pressure for subfunctionalisation nor thermodynamic bias to co-assemble. To quantify interactions between them, we use statistical mechanics. Utilising our developed model allows us to determine if each investigated pair shows a preference toward self- or co-assembly based on experimental native MS data.

3.2.2 Physico-chemical description of protein co-assemblies

Statistical mechanics is a field that describes macroscopic entities and assemblies by applying statistical and probabilistic frameworks to describe the properties of these assemblies.

Here, we use a statistical-mechanical model based on the Boltzmann equation to quantitatively describe heterogeneous co-assemblies between pairs of oligomeric sHsps. This model was developed to evaluate, based on experimental data, the energy bias of interactions of a pair of sHsps. It allows us to compare MS results of numerous sHsps pairs, quantify heteromerisation between them, and evaluate their assembly preference.

This model uses equilibrium statistical mechanics where we assume that all oligomeric species present in the system are in equilibrium. Hence, the experimental data were recorded after equilibration at 37°C for at least an hour, which was verified to be enough time to reach the equilibrium by comparing measurements after 1 hour to even up to 48 hours of incubation.

To study the assembly of two proteins, A and B, the model considers three types of interactions: A-A, B-B and A-B, which are grouped according to their energies into homo-interaction E(A-A) or E(B-B) and hetero-interaction E(A-B) energy. In this work, we use a simplification, which sets the interactions between the same subunits to zero: E(A-A)=E(B-B)=0, which allows us to analyse the system considering only the hetero-interaction energy.

In this work, we use a model, which is an expansion of what was developed by Hochberg et al. ([26]). In the first step, we use statistical mechanics to study the abundance of all the possible stoichiometries of oligomers of the same size formed between proteins A and B. As a result, we obtain a spectrum of distributions, from specific self-assembly to obligate, n:n co-assembly, for a range of energy differences (Fig. 3.2). This is further used to extract energy differences for experimental data by fitting the energy bias to the experimentally extracted abundances.

Statistical-mechanical description of co-assembly

It is entropically favoured for the subunits to co-assemble in a number of different ways. On the other hand, it needs to be considered that either A-A and B-B or A-B contacts may be energetically favoured. Consequently, it may come with a penalty and depends on the number of unfavourable contacts formed within the oligomer. The model needs to assume a specific geometry of the oligomer; for simplicity in this case, a ring arrangement with an oligomer size bigger than or equal to a 12-mer was chosen, even though we know now that some sHsps have a tetrahedral quaternary structure. This will lead to a conservative underestimation, as the tetrahedral arrangement has four vertices and, therefore, allows for more inter-subunit contacts, but with the same number of them in each case. We use the Boltzmann distribution equation to extract relative populations of analysed assemblies:

$$x_i = \frac{\sum_{k=1}^{\binom{N}{i}} e^{-\Delta\Delta G_k^i / k_B T}}{\sum_{j=0}^{N} \sum_{k=1}^{\binom{N}{j}} e^{-\Delta\Delta G_k^j / k_B T}} \tag{3.1}$$

Where x_i is the mole fraction of the N-size oligomer, which contains i numbers of protein A subunits obtained experimentally. We have developed an algorithm to count all the different subunit arrangements (k) and symmetries ([45]). $\Delta\Delta G_k^i$ is the energy difference between homo and hetero-interactions of subunits A and B in an N-sized oligomer, containing i number of subunits A in a k arrangement. Proteins A and B, with $\Delta\Delta G_k^i = 0$ corresponding to entropy-favoured unbiased co-assembly, exhibit binomial distribution (Figure 3.2A) at an equal molar mixing ratio.

Further modifications of the model take into account that it is challenging to mix two species at exactly a 1:1 molar ratio; moreover, we test here at least two different

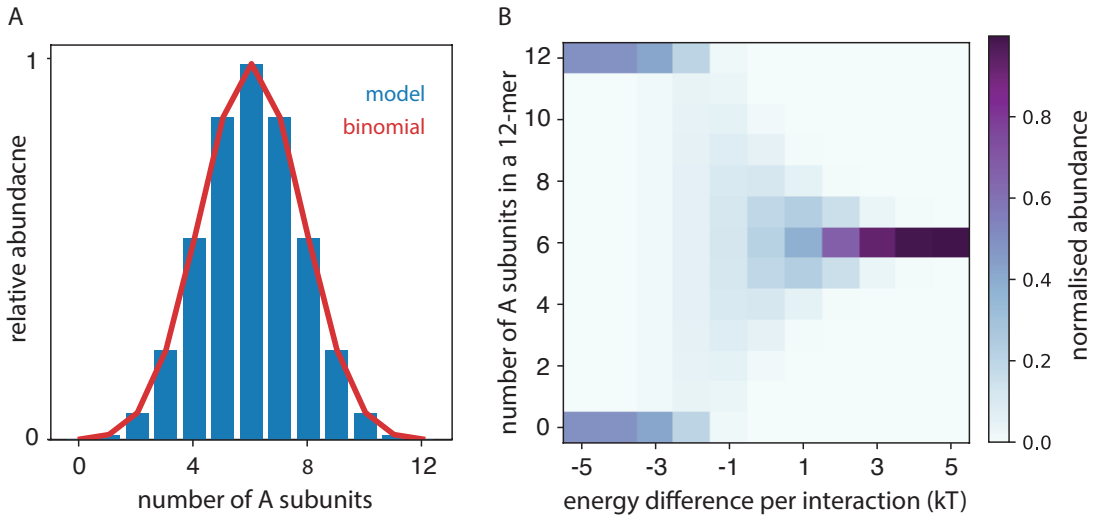


Figure 3.2: 12-mer co-assembly predicted with the statistical-mechanical model. A) Prediction of dodecameric abundances for zero energy difference between homo- and hetero-oligomerisation, which recovers the binomial distribution. B) Predicted distributions for 12-mers as a function of energy bias. This shows that the energy barrier to evolve from entropy-neutral co-assembly to either obligate co-assembly or specific self-assembly is on the level of one hydrogen bond (4 kT). Figure used with the permission from ([45]).

mixing molar ratios for each pair. First, using the Boltzmann equation, the method calculates x_i based on finding a number of subunits of A and B in sub-stoichiometries, so the obtained values are not influenced by imprecisely identified molar ratios.

We scan through energy biases from -5 to 5 kT (Fig. 3.2B) to get the abundances of each of the sub-stoichiometries for each energy bias within this range. The resulting complete spectrum of distributions for 12-mer as a function of energy difference is presented in Figure 3.2B. This statistical-mechanical description method shows that the energy barrier from unbiased co-assembly to obligate self-assembly or obligate n:n co-assembly is 4 kT, which is on the level of one hydrogen bond.

It would not be possible to obtain experimental values of abundances for each sub-stoichiometry without recent advances in mass spectrometry, which now allows us to study extremely heterogeneous systems with very high mass resolution. Here,

we use native MS experimental data to extract the energy biases for each investigated pair. First, we assign all identified charge series in the MS spectrum, then deconvolve them to mass (kDa), calculate what stoichiometry each identified mass corresponds to, and apply the described statistical-mechanical model on fitted abundances. For each pair, we get a value of energy bias, which describes if a pair of interests favours co- or self-assembly and what is the energy bias for that.

3.3 Results

As mentioned above, here we analyse the ability of sHsps targeted to membrane-bound cellular compartments to co-assemble. We aim to gain insights into how their interfaces evolved while diverging to the new cellular locations (Figure 3.3).

The evolution of cytosolic sHsps is influenced by two factors that are absent for the organelle ones. First is the evolutionary pressure for evolving new functions by redundant genes. This is affected by the second factor, which acts as its evolutionary constraint: the fact that heteromerisation is entropically favoured, navigating different subunits to form co-assemblies with various stoichiometries. It has been proposed that relieving this constraint is a key step in the evolutionary trajectories of paralogs toward evolving new functions ([26]). Because the protein pairs of interest are physically separated, this constraint does not apply. It allows us to study how the absence of that constraint directed the evolution of organelle-localised sHsps.

To do this, we measured native MS spectra for mixed pairs of sHsps targeted to membrane-bound cellular compartments. We use the mass spectra to extract abundances of sub-stoichiometries formed between the various sHsps pairs. Then, we apply the model described above (Section 3.2.2) to determine if each of the investigated pairs favours self- or co-assembly and extract the energy bias of these

assemblies. This allows us to compare interactions between different sHsps and put it in the context of evolution.

We examined each possible pair between the following sHsps from different evolutionary classes: HSP21(CP), HSP26.5(MTII), HP23.6(MTI), and HSP23.5(MTI), all from the same plant species *Arabidopsis thaliana*. Each of the tested sHsp pairs was mixed at multiple molar ratios and incubated at 37°C for at least an hour. This was tested beforehand to be a sufficient amount of time to reach the equilibrium by comparing results obtained after 1 hour to data collected after 24 hours of incubation. MS measurements were performed before and after incubation. The Unidec software was used to deconvolve the MS data, with binning applied if necessary to limit the impact of salt adducts and truncations. Knowing the monomer masses of each protein, we could assign the appropriate sub-stoichiometries to each of the obtained masses. Further, for each molar ratio, the experimentally extracted abundances were fitted using the statistical-mechanical model, which calculates the energy bias of the co-assembly in question.

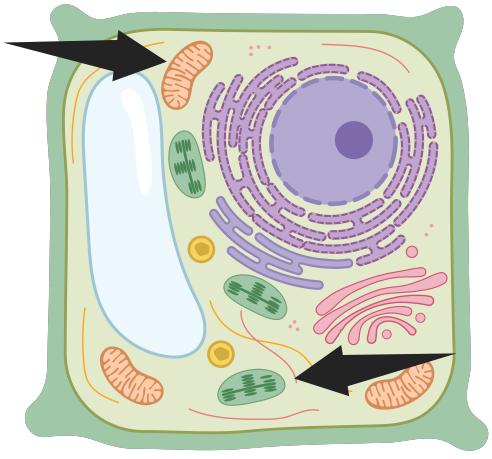


Figure 3.3: Locations of interrogated sHsps. Schematic illustration of a plant cell, which highlights the locations of tested sHsps. The physical separation of mitochondrial and chloroplastic proteins results in a lack of thermodynamic pressure for heteromerisation, normally acting on oligomeric paralogs as an evolutionary constraint on the functional divergence. Created with BioRender.com.

3. Co-assemblies of organelle sHsps from plants

Lastly, we analysed cytosolic sHsps from the same evolutionary class but different plant species as a reference point. For that purpose, interactions between two CI sHsps were studied, HSP17.6 and HSP18.1 from *A. thaliana* and *P.sativum*, respectively.

In this chapter, we focus mainly on the 12-mer mass region of assemblies, as those are the most abundant species in every analysed sample. Even in the systems including polydisperse sHsps, the most abundant species are homo- or hetero-12-mers. Another reason for this is that the used model applies to a single oligomeric state.

As previously mentioned, sHsps are well known for their diversity of assembly. Figure 3.4 shows three representative examples of different modes of interactions between plant sHsps, data acquired before and after incubation. Measurements presented on the left (around 0 min) were conducted immediately after mixing two protein samples, which were kept on ice beforehand, in comparison to at least an hour of incubation at increased temperature (right). The three representations illustrate contrasting manners of sHsps assembly.

The top left panel (Figure 3.4) presents two charge series originating from two homo-oligomers (left). Incubation of mixed proteins for more than an hour at 37°C results in homo-12-mers still being the most abundant species (right). However, eleven very low-intensity, new charge series are detected. The low-intensity peaks come from appearing oligomers containing subunits of both proteins in various stoichiometries. This pair displays a strong preference for self-assembly. The proteins are capable of subunit exchange, but the most intense peaks correspond to homomers.

The middle left panel (in Figure 3.4) shows two charge series centred at 7200 m/z corresponding to monodisperse homo-12-mers (consisting strictly of subunits of either HSP21 or HSP26.5). After an hour of incubation, a shift in the mass spectrum is observed. Multiple additional peaks are detected, which are assigned as eleven

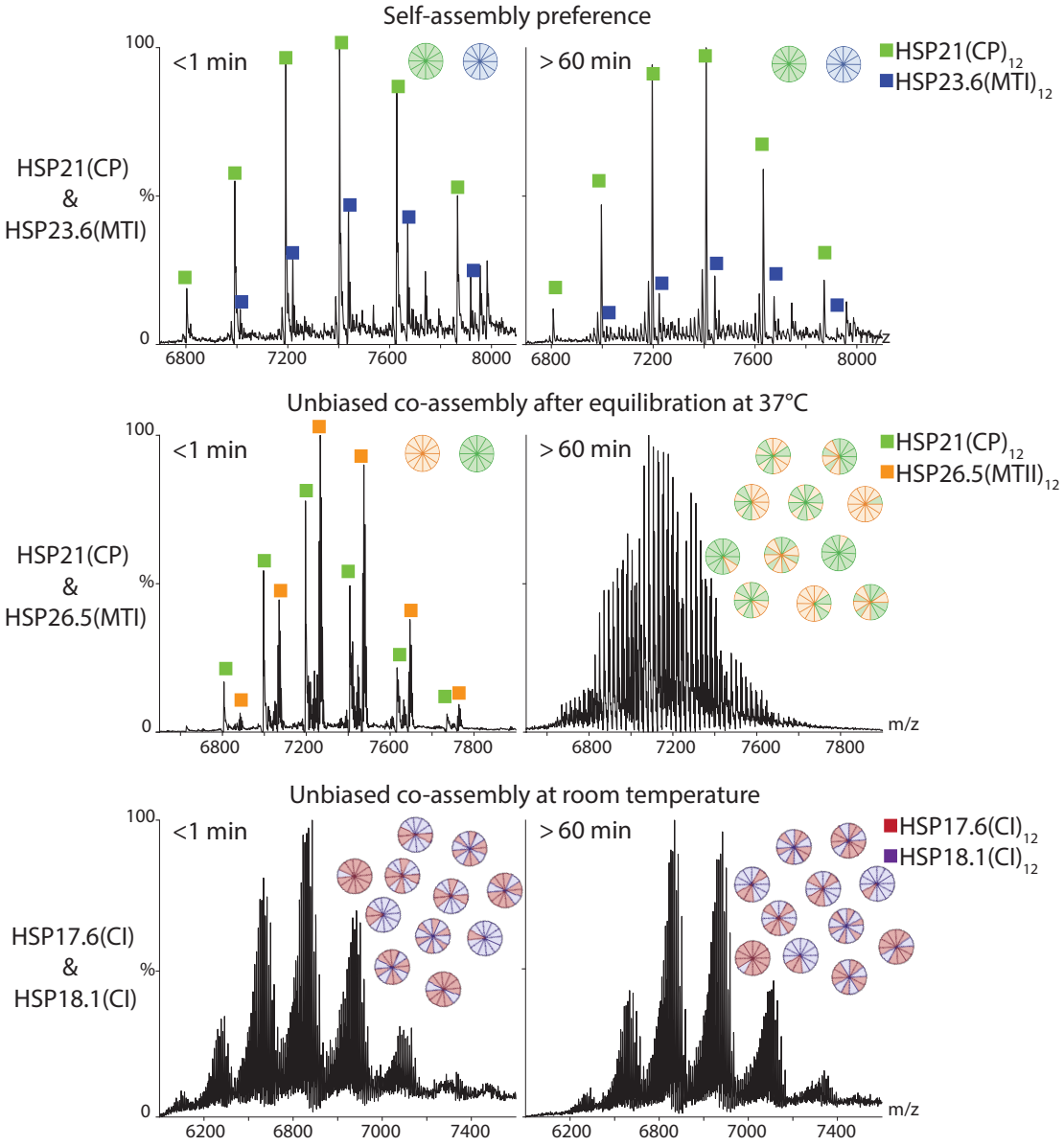


Figure 3.4: Different co-assembly modes of plant sHsps. Only peaks corresponding to homo-oligomers are labelled. Left: Native mass spectra acquired prior to incubation (two appropriate samples, kept on ice after buffer exchange, measurements recorded immediately after mixing). Right: Native MS acquisitions after at least an hour of incubation at 37°C. Top: Before incubation, two charge series correspond to homo-12-mers, which are still the most abundant after incubation, with the heat-shock-induced appearance of new, very low-intensity charge series corresponding to hetero-oligomers, suggesting a preference for self-assembly. Middle: Homo-12-mers (represented by two charge series) before incubation are substituted by new peaks corresponding to 11 new charged series of formed hetero-oligomers. Bottom: Multiple charge series are visible before incubation, meaning hetero-oligomerisation starts prior to temperature induction.

new charge series. This is a sign of undergoing hetero-oligomerisation, with peaks corresponding to different stoichiometries appearing in the mass spectrum.

The last example shows a mixture of two sHsps, for which multiple charge series are detected right after mixing (Figure 3.4, bottom), immediate co-assembly (at room temperature) without the need to be incubated at a higher temperature proves their high interface compatibility. Incubation causes the charge state distribution to become narrower, reaching the equilibrium in which the preference for co-assembly is stronger.

Here, we have shown different manners by which sHsps interact. In the next part of this chapter, we will go into more detail of those interactions, their description, and quantification. Finally, we will compare all the obtained energy difference values for every pair and put them into the context of phylogeny.

3.3.1 Diverse co-assembly modes of organelle sHsps from different cellular compartments

Unbiased co-assembly of HSP21(CP) and HSP26.5(MTII)

First, we investigated interactions between two organelle-localised, monodisperse, dodecameric plant sHsps, HSP21 and HSP26.5, targeted to chloroplasts and mitochondria, respectively. This means that neither pressure for subfunctionalisation nor entropy bias could have influenced their evolution. These sHsps originated from a gene duplication event around 450 million years ago from cytosolic sHsps ([109]).

Figure 3.5 (left) shows native MS spectra of those two proteins, recorded at three molar ratios: 2:1, 1:1, and 1:2 (HSP21:HSP26.5), after equilibration at 37°C. Multiple charge series were detected in each spectrum. Distribution of charge series shifts with changing ratio, moving towards higher m/z with increasing concentration of sHsps higher in mass, with the most intense charge series centring at 6800 m/z for 2:1, then

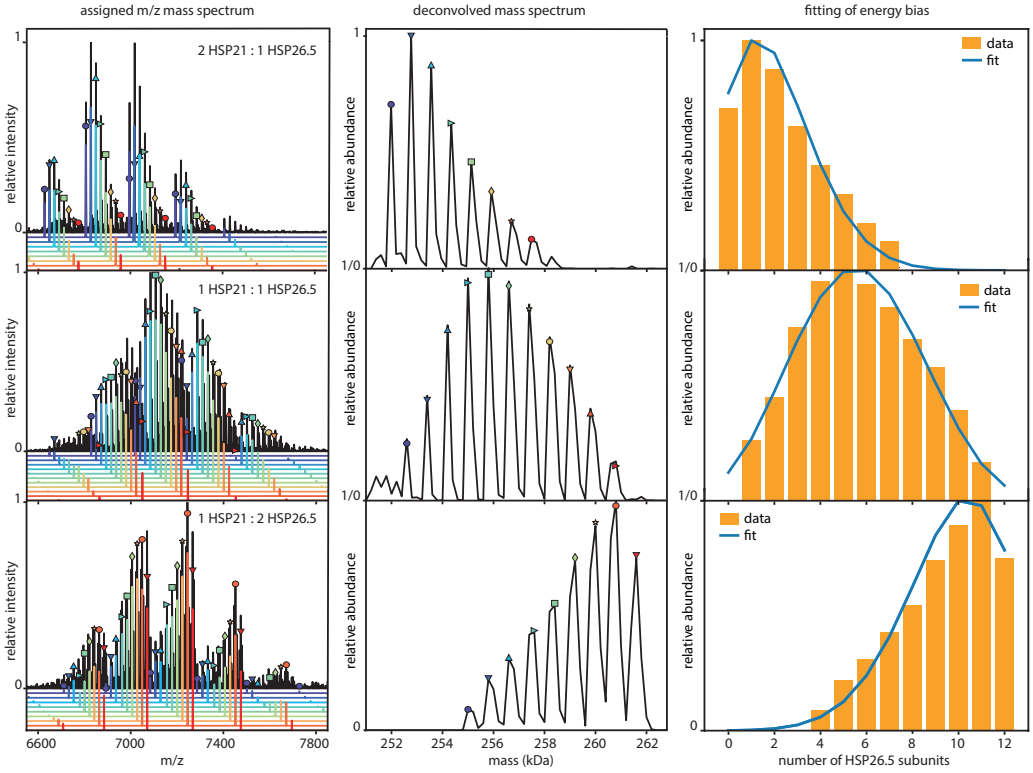


Figure 3.5: Unbiased co-assembly of HSP21(CP) and HSP26.5(MTII). Left: Native MS of subunit exchange products between two sHsps from different membrane-bound cellular compartments: HSP21 and HSP26.5, at molar ratios of 2:1, 1:1, and 1:2 (HSP21:HSP26.5), where different charge series, coming from different masses, are assigned. Middle: Deconvolved masses corresponding to appropriate charge series (marked with the same symbol). We use extracted masses of complexes to calculate which stoichiometry they correspond to. Right: Experimental abundances (orange bars) for each stoichiometry and the fitted distribution using a statistical-mechanical model (blue). An extracted energy bias of -0.80 kT of interactions means that the homo-interactions are slightly more favoured for this pair, but the difference is on the order of the strength of a weak hydrogen bond. HSP21-HSP26.5 shows almost unbiased co-assembly.

7000 m/z for 1:1 and 7200 m/z for 1:2. This indicates statistical incorporation of subunits into complexes. For an equal molar mixture expected to co-assemble in an unbiased manner, a 6:6 stoichiometry would be the most intense complex. However, observed distributions are slightly shifted in the HSP21&HSP26.5 mass spectrum. This discrepancy may be attributed to the low accuracy of Nanodrop absorbance measurements, leading to imprecise protein ratios. Such variations impact all mixing experiments described in this and the following chapter. Nevertheless, the model

employed for quantifying energy biases is unaffected by the initial ratio inputs, as it calculates the fractions independently, thus mitigating this issue. Additionally, the observed distributions may be influenced by the distinct co-assembly preferences of each sHsp in the mixture and their varying tendencies to form heteromers, considering that each may have a different preference for heteromerization.

MS data were deconvolved to obtain masses represented by each detected charge series (Fig. 3.5, middle). Since the monomer mass for each subunit is known, we were able to calculate sub-stoichiometries corresponding to each mass. Figure 3.5 (middle) demonstrates that eleven charge series were detected at a 1:1 molar ratio. They correspond to every possible stoichiometry of hetero-12-mer between HSP21 and HSP26.5, representing an almost binomial distribution, which is entropy-favoured and characteristic for the assembly of two proteins just after gene duplication, having identical amino acid sequences (Fig. 3.1, middle). This means that those proteins co-assemble in an almost unbiased manner. At each of the molar ratios, the most abundant species is heteromer, and its stoichiometry depends on the molar ratio.

To obtain the exact value of the energy bias, we use the statistical-mechanical model, which allows us to extract energy preference between hetero- and homo-interactions, by fitting the energy difference to the abundances obtained from MS data, for all three molar ratios (Fig. 3.5, right). The extracted energy difference is -0.80 kT of interactions, which is on a level of weak hydrogen bond, and around 1/6 of an average hydrogen bond energy, and 1/15 of a strong one([137]). This value of extracted energy means that the investigated pair slightly disfavours hetero-interactions compared to homo-interactions. It is surprising that two sHsps localised in different membrane-bound cellular organelles still co-assemble without a significant preference for either self- or co-assembly. This proves a high compatibility of their

interfaces. Even though those proteins have been physically separated for hundreds of millions of years, their interfaces do not show biophysical signs of incompatibility.

Preferential self-assembly of CP and MTI sHsps

The next two investigated pairs are also between sHsps targeted to two different cellular compartments. Again, monodisperse chloroplastic HSP21 and two polydisperse proteins belonging to MTI, HSP23.6, and HSP23.5 (shown in the previous chapter, Section 2.6.4). MTII (shown above) class is evolutionarily more distant from CP than MTI (shown below). As mentioned previously, we will focus on the quantification of their assembly within the 12-mer m/z region, as restricted by the employed model.

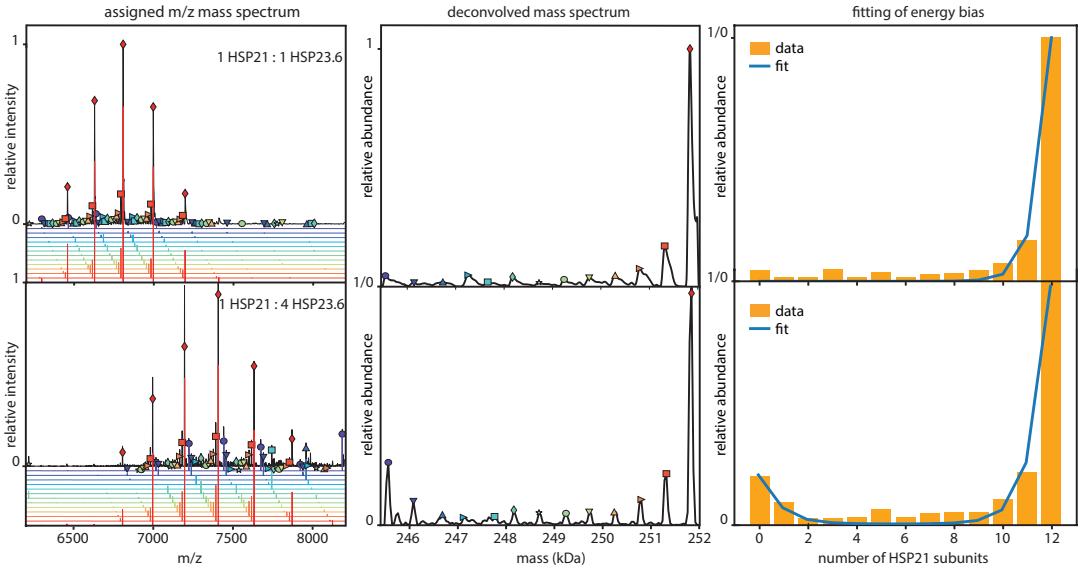


Figure 3.6: Self-assembly preference between HSP21(CP) and HSP23.6(MTI). Left: Native MS data obtained at two molar ratios between two sHsps with different dispersities, localised to different membrane-bound cellular compartments. Middle: Extracted abundances show the highest intensity for homomers, with very low-intensity heteromers being formed. Right: Applying the model on extracted abundances (blue) results in an energy bias value of -1.72 kT. Homo-interactions are more favoured than hetero-interactions, this pair shows a strong preference for self-assembly.

For a pair of HSP21 and HSP23.6 (Figure 3.6), we present quantification of interactions based on native MS mixing experiments at two different ratios. More experiments were performed; however, the polydispersity of one of the proteins made

the data hard to assign correctly, so data at only two molar ratios were analysed. The polydispersity of MTII is also the reason why in the mass spectra, homo-12-mer of HSP21 shows as the most abundant species; in fact, the concentrations of MTII were either equal to or higher than HSP21, but the intensities of 12-mers for MTII members are low. Optimisation to obtain higher 12-mer intensity led to an increase in intensities of peaks corresponding to other oligomeric sizes and, hence, higher heterogeneity and peak overlap, which did not allow for correct peak assignment.

Moreover, to ensure the equilibrium was reached within an hour, we repeated the acquisition with a sample incubation at a higher temperature for 24 hours, which confirmed the equilibrated state of the mixture. This data is presented on the bottom panel of Figure 3.6.

Data at three molar ratios were obtained for mixtures of HSP21 with HSP23.5 (Figure 3.7). For all of the experiments between CP and MTI sHsps, native MS spectra vary significantly in comparison to the first tested pair. Here, the data show one dominant charge series in all the m/z spectra, which corresponds to a homo-12-mer, accompanied by other low-intensity species assigned to different stoichiometries of heteromers. This is evidence of a strong preference towards self-assembly for those pairs. Most subunits are involved in the formation of homo-oligomers, but the other low-intensity charge series, assigned as every possible stoichiometry of hetero-12-mer, shows that those proteins are capable of co-assembly.

After applying the statistical-mechanical model to the abundances obtained from MS experiments, acquired at two different molar ratios of the HSP21 and HSP23.6 pair and at three molar ratios of the HSP21 and HSP23.5, we extracted the energy bias values of -1.72 kT, and -1.41 kT, respectively. Both pairs interact in the same manner, favouring homo-interactions more than hetero.

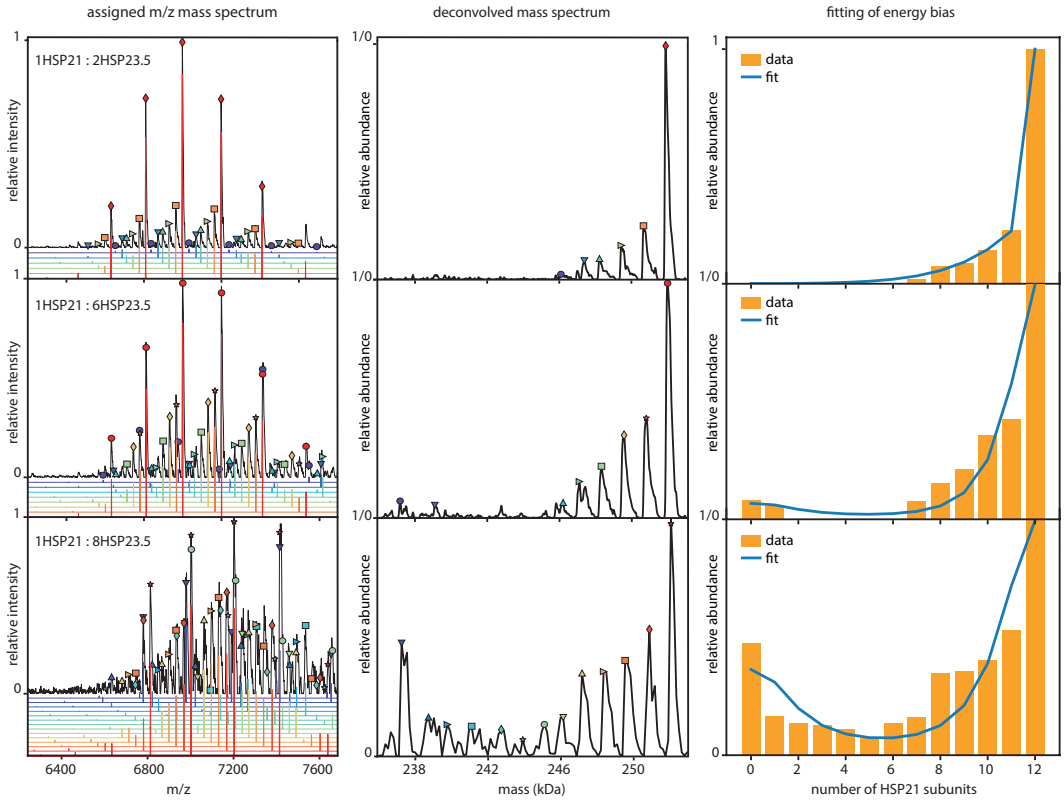


Figure 3.7: HSP21(CP)-HSP23.5(MTI) display preference for self-assembly. Left: Native MS data at three molar ratios between two sHsps with different dispersity from different membrane-bound cellular compartments led to the detection of multiple charge series. Middle: Each charge series corresponds to a mass marked with the same symbol on the middle panel of the deconvolved MS data. Right: After the assignment of stoichiometries, it is observed again that homo-oligomers are the most dominant species, accompanied by multiple stoichiometries of hetero-oligomers. Extracted energy bias of -1.41 kT means that the HSP21-HSP23.5 pair is biased towards self-assembly, and the homo-interactions are favoured.

Even though we are not applying the model outside of the 12-mer mass region, we analysed MS data correlating to other sizes of oligomers to evaluate their ability to co-assemble within other oligomeric states. We used 1:4 molar ratio MS data of HSP21&HSP23.6, as an example. Figure 3.8 demonstrates the full m/z range. It shows that 12-mer is still the most dominant species in the sample. We used the data acquired at the 1:4 molar ratio mixture to analyse the 16-mer m/z region, as only these data provided sufficient intensity for that species to be analysed. Only HSP23.6 homo-16-mer was detected, confirming the lack of hetero-oligomerisation in higher

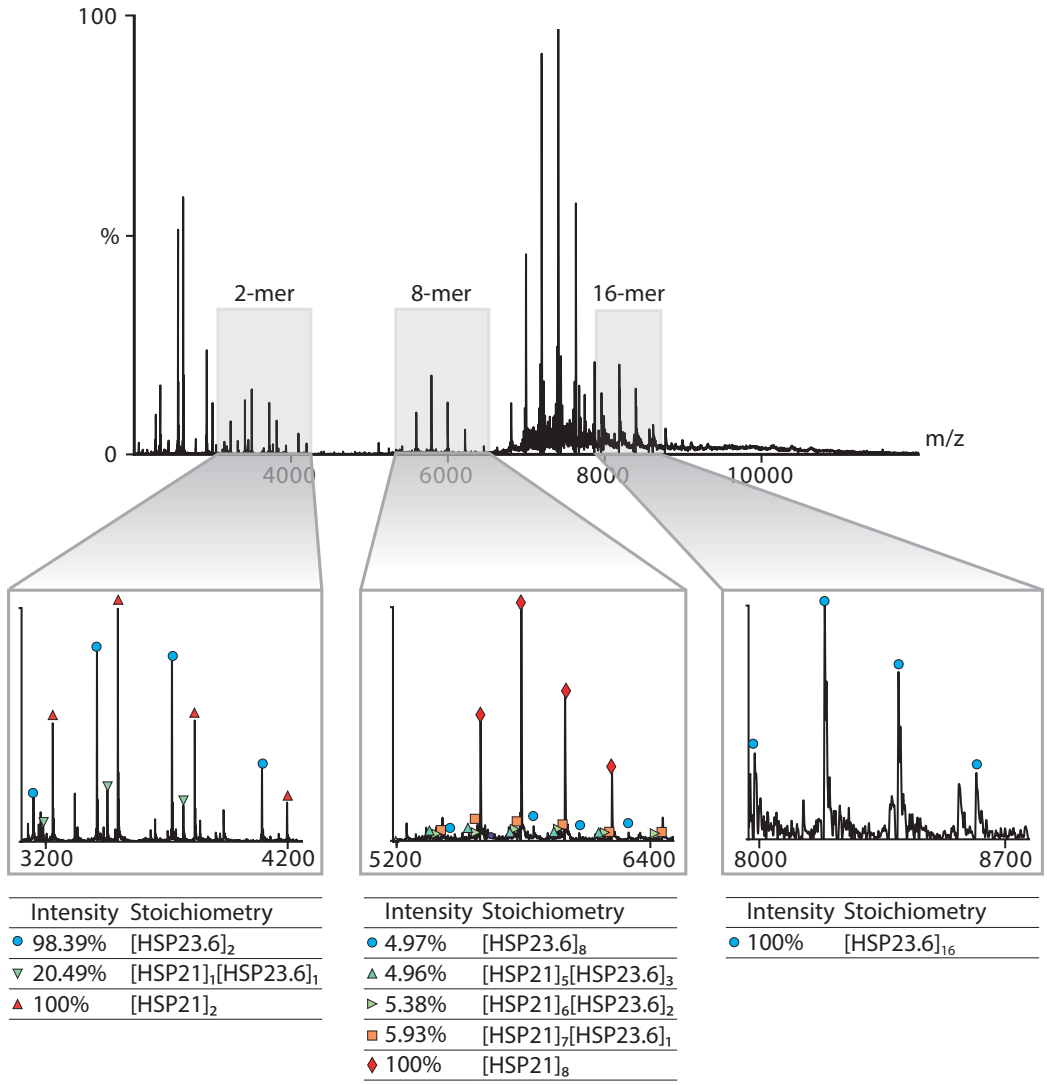


Figure 3.8: Full m/z range mass spectrum of mixed HSP23.6 with HSP21. Top: Native mass spectrum at 1:4 molar ratio. 12-mer is the most abundant species in the sample. Bottom close-ups of different oligomeric sizes m/z regions correspond to dimers (left), 8-mers (middle), and 16-mers (right). For species smaller than a 12-mer, subunit exchange was detected on a very low level. For species higher than a 12-mer, no hetero-oligomerisation was identified.

species. As we showed in Chapter 2, HSP21 forms octamers, even if not to a high extent. We checked if HSP21 is capable of incorporating HSP23.6 into octameric structures. Homo-8-mers are the most abundant species within this oligomeric size; however, other charge series have been detected, which, after deconvolution, were assigned to very low-intensity heteromers in various stoichiometries (Figure 3.8, bottom middle), surprisingly also homo-8-mer of HSP23.6 was detected.

Deconvolution of the dimer region led to the detection of three species (Figure 3.8, left), both homodimers of HSP21 and HSP23.6, and a heterodimer containing subunits of both. Again, the pair shows a similar manner of interactions as in the 12-mer region, with homo-oligomers having the highest intensities. The calculated relative ratio between dimeric species is 1:0.2:1 ($HSP23.6_2$:$HSP21_1HSP23.6_1$:$HSP21_2$), which makes only 14,7% of all HSP23.6 subunits involved in dimerisation to be involved in hetero-interactions.

Altogether, the results clearly show that CP sHsps interact with MTI members, forming hetero-oligomers of different sizes; however, they show a strong preference for self-assembly.

3.3.2 Almost unbiased co-assembly of mitochondrial sHsps from different evolutionary classes

A common generalisation reported in the literature is that plant sHsps from different evolutionary classes do not co-assemble ([138]). However, this hypothesis was based on limited combinations of only cytosolic sHsps. Above, we verified that it is not true for sHsps from different cellular compartments. Here, we look at pairs between sHsps from the same organelle (mitochondria) but different evolutionary classes: MTI and MTII. Those two classes diverged around 300 million years ago in seed plants. Even though members of MTI are polydisperse (as shown in Chapter 2), once again, we quantify energy biases only for the data from the 12-mer region. The statistical-mechanical model can be applied to only one oligomeric size. However, a whole m/z range of native MS spectra acquired at three molar ratios of HSP26.5 and HSP23.6 shows that 12-mers are the most abundant oligomeric species within the tested samples (see Figure 3.9, left).

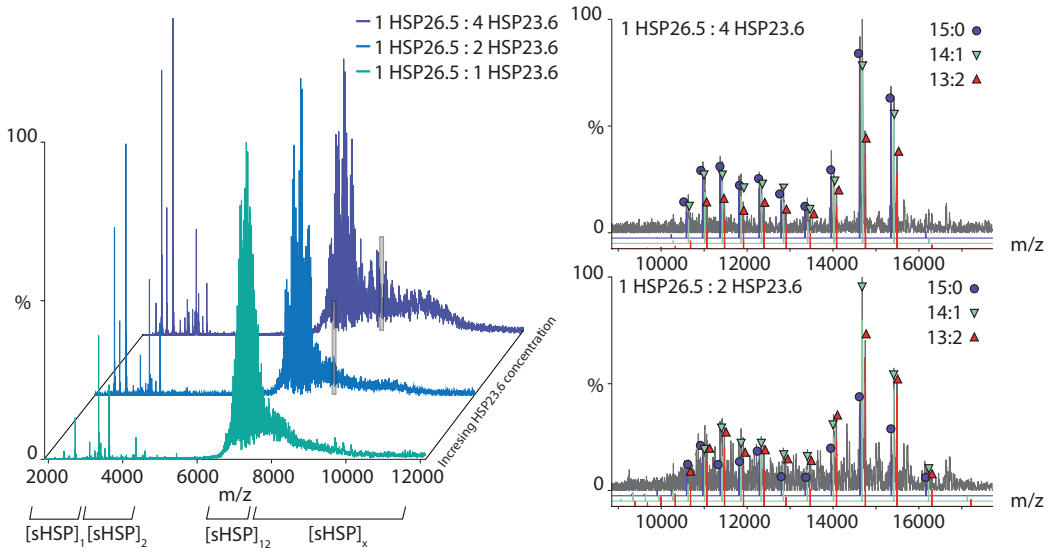


Figure 3.9: Heterogeneous co-assemblies of HSP26.5-HSP23.6. Left: The full m/z range of MS data at three molar ratios shows that after reaching the equilibrium, 12-mers are the most dominant species in the samples. The grey squares indicate the m/z windows isolated and subjected to CID. Right: CID-produced N-1 species detected at both ratios, 1:4 and 1:2 (HSP26.5:HSP23.6), include both homo- and hetero-15-mers. This means that even though HSP26.5 doesn't form 16-mers, it gets incorporated by HSP23.6 into higher oligomeric species.

For both pairs, three mixing molar ratios were tested. We chose data recorded at different molar ratios for each pair, as only the spectra that were possible to deconvolve correctly were used for further analysis, which was restricted by the heterogeneity of samples.

Multiple charge series were detected in each spectrum. After the deconvolution of MS data and assigning sub-stoichiometries to extracted masses, we can see that both of the investigated pairs co-assemble (Figure 3.10 and Figure 3.11). The most intense species in each MS spectrum is a hetero-oligomer, and its stoichiometry depends on the mixing ratio as their distributions shift, depending on the concentrations. In both pairs, most of the subunits of both proteins are involved in hetero-interactions. Members of MTI and MTII interact in a very similar manner to those presented earlier in this chapter, a pair of HSP21(CP) with HSP26.5(MTII).

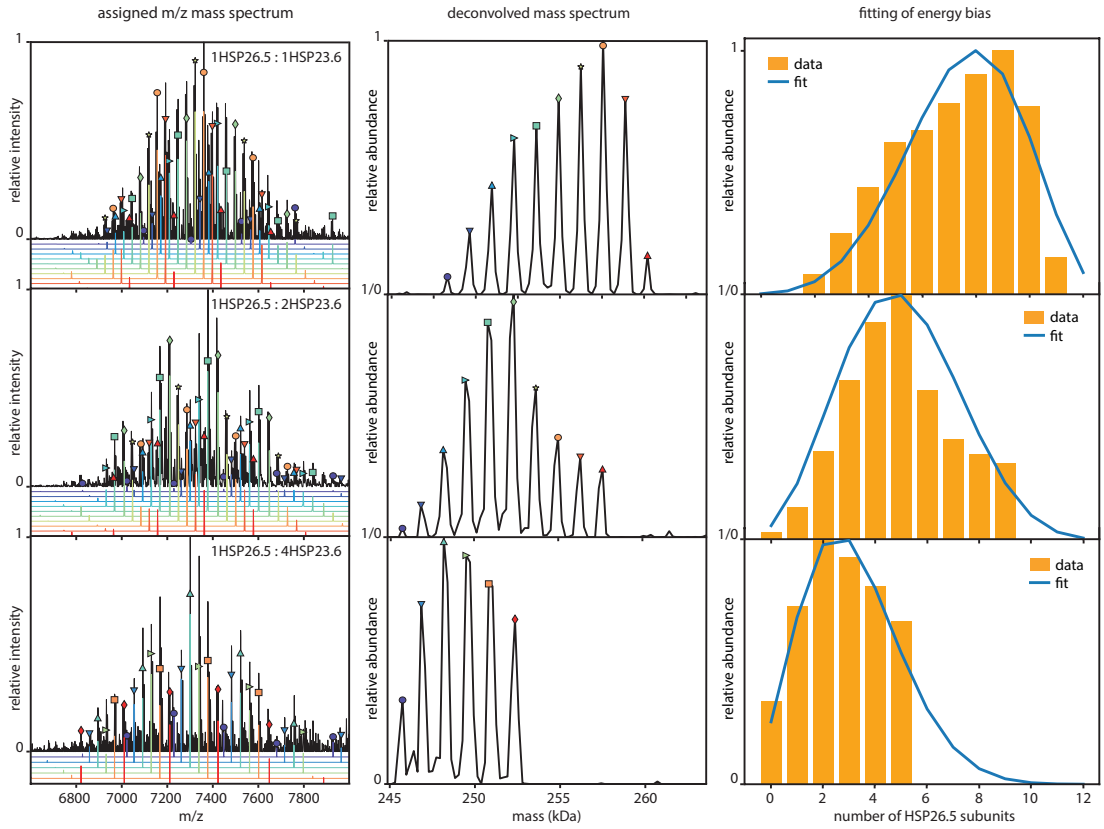


Figure 3.10: Quantitative description of co-assembly between HSP23.6(MTI) and HSP26.5(MTII). Left: Native MS at three ratios between two mitochondrial sHsps, with different dispersities. Middle: Deconvolved MS data resulted in the identification of oligomeric distributions. We identified which sub-stoichiometries the masses correspond to. At each molar ratio, the most abundant species was a hetero-oligomer. Right: By fitting the distributions using the statistical-mechanical model, we determined the energy bias of -0.64 kT, which means that for the HSP23.6-HSP26.5 pair, the homo-interactions are more stable, and there is a slight bias towards self-assembly.

We fitted the energy difference to abundances extracted from the data at three molar ratios for each pair, using the statistical-mechanic model (Fig. 3.10 and Fig. 3.11). The extracted values of energy bias between hetero- and homo-interactions are -0.64 kT for HSP26.5&HSP23.6, and -0.59 kT for HSP26.5&HSP23.5 pair. This means that for both of them, there is a slight preference for homo-interactions in these mixtures. This energy difference is very similar to the one extracted for the HSP21&HSP26.5, a mixture of two sHsps localised to different membrane-bound cellular compartments. Again, it is on the scale of a weak hydrogen

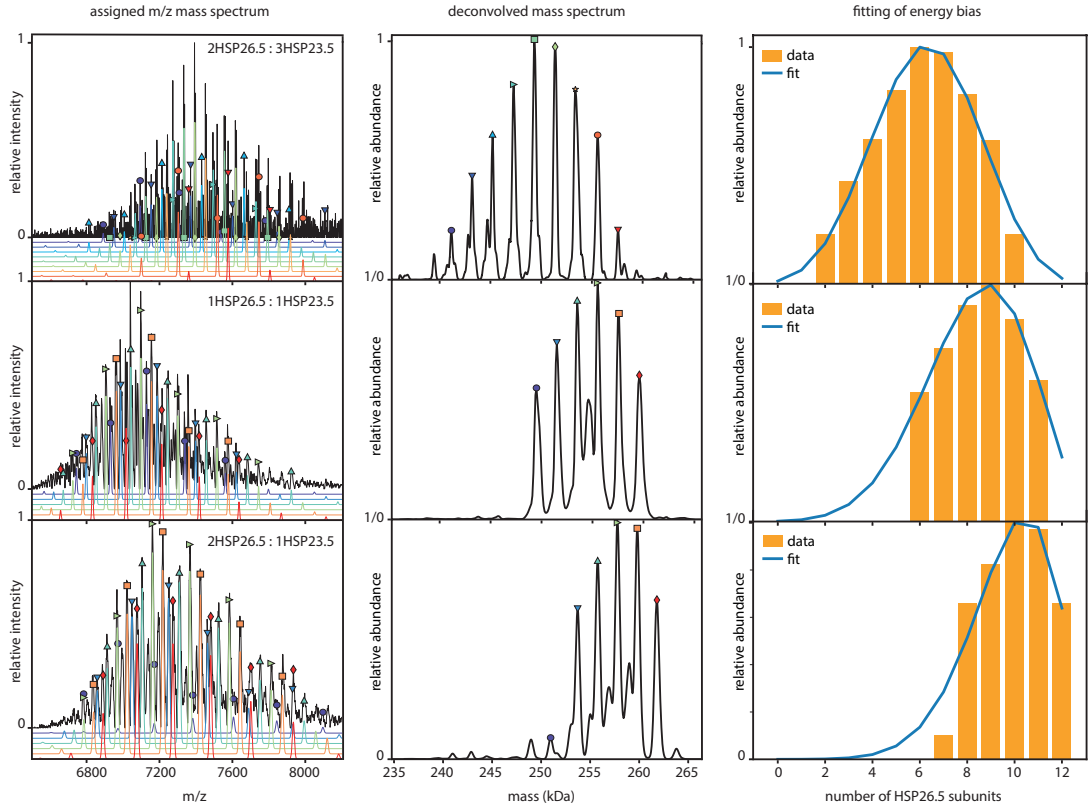


Figure 3.11: Quantitative description of unbiased co-assembly between HSP26.5(MTII) and HSP23.5(MTI). Left: Native MS data at three molar ratios between MTII and MTI sHsps. Multiple charge series were detected in each mass spectrum. Middle: Deconvolced mass distributions. The top plot (data at 2HSP26.5:3HSP23.5 molar ratio) exhibits almost binomial distribution, which is characteristic of energy-neutral co-assembly. Right: After applying the model, we extracted the energy bias of -0.59 kT. HSP26.5-HSP23.5 is another pair showing a slight preference for self-assembly, which is on an energy level of a weak hydrogen bond.

bond. Overall, the two discussed pairs interact in an almost unbiased manner with a slight preference towards self-assembly.

We then tried to analyse the MS data in a higher m/z region, corresponding to 16-mer for the HSP26.5 and HSP23.6 pair, however, because of its heterogeneity, the assignment was challenging. To overcome this problem, we performed MS/MS experiments, where we selected the window of 100 m/z, corresponding to the 16-mer m/z range (highlighted in grey in Figure 3.9), and subjected it to CID. Resulting charge-reduced N-1 species include both homo- and hetero-15-mers in

both analysed mass spectra (Fig. 3.9, right), with their distributions depending on the molar ratio. At the 1:4 ratio, the most abundant product of dissociation was complex with stoichiometry 14:1, which means that it was produced by dissociation of 14:2 and 15:1. HSP26.5, individually does not form oligomers higher than 12-mer (Chapter 2), however in the presence of polydisperse HSP23.6 it gets incorporated into higher oligomers (as shown in Figure 3.9, right), which is a proof of co-assembly affecting the sHsp's structure, which may impact biophysical properties and hence cellular processes.

Previously, based on the data for cytosolic sHsps, it was thought that sHsps from different evolutionary classes do not co-assemble. We have shown here that this is not the case for organelle sHsps, and this generalisation was misleading. CI and CII families diverged 450 million years ago, which was enough time for them to evolve their interfaces to the degree that they are not compatible anymore, which keeps them from co-assembly. In comparison, MTI and MTII are closer evolutionary relatives, which could be a reason why their interactions are still maintained. More studies should be conducted to determine the exact principles regulating co-assemblies of sHsps. The data presented shows that sHsps mutually affect each other's structures and dispersity, which may have an impact on cellular processes when they co-localise in the cell.

Quantifying energy biases of subunit exchange between two CI sHsps

It has been shown before that two cytosolic sHsps from the same evolutionary class can co-assemble. Here, we want to quantify those interactions, extract the energy bias using the developed statistical-mechanical model, and compare it to the data presented in this chapter. We use HSP18.1 from *Pisum sativum* and HSP17.6 from *Arabidopsis thaliana*. Both of them are monodisperse, dodecameric proteins.

Obtained proteins were truncated, which made the data analysis more challenging; we applied binning to limit the impact of sample heterogeneity. Painter et al. ([87]) described co-assembly between Class I sHsps from the same plant species as exchanging only dimeric-subunits and as a result forming only even-stoichiometries; however, another study revealed co-assemblies between CI *Pisum sativum* HSP18.1 and *Triticum aestivum* HSP16.9, which form hetero-dodecamers with both odd and even numbers of subunits of each sHsp ([43]).

Two CI sHsps, HSP18.1 and HSP17.6 from different plant species, exchange

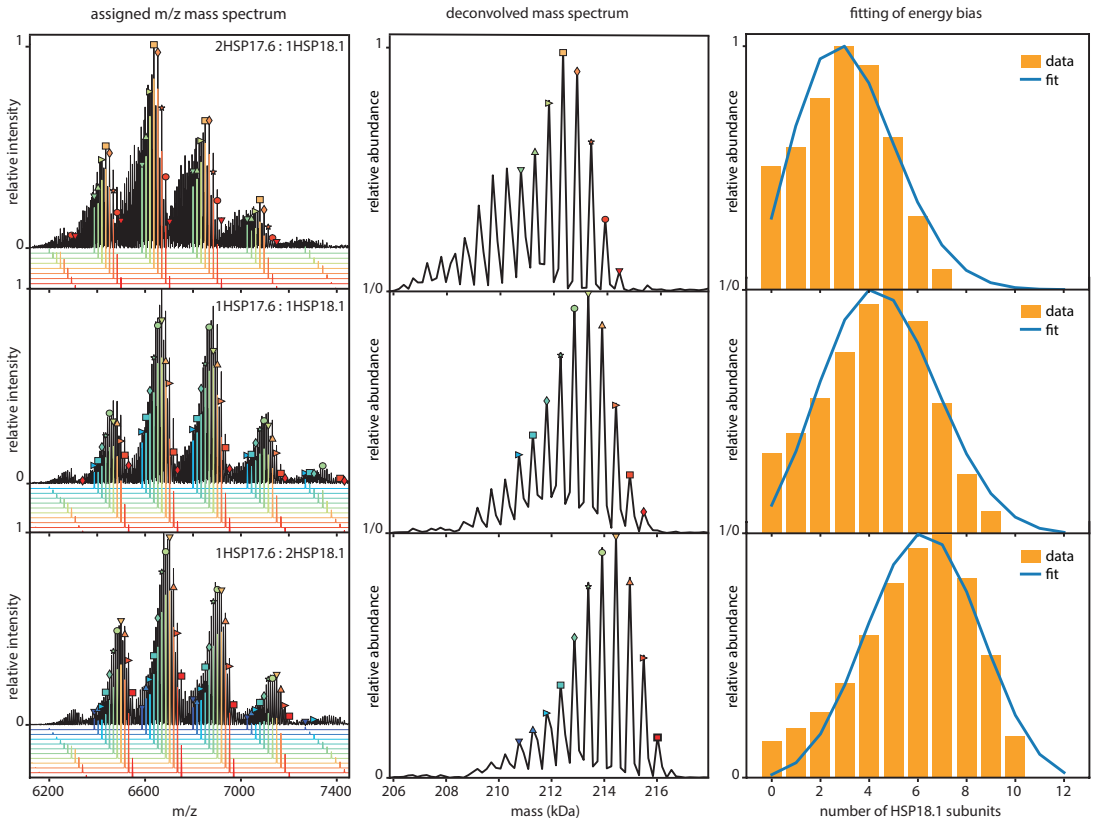


Figure 3.12: Unbiased co-assembly between HSP17.6(CI) and HSP18.1(CI). Left: Native MS data at three molar ratios between two monodisperse cytosolic sHsps from the same evolutionary class (CI). Middle: Detected masses based on MS acquisitions, assigned to multiple stoichiometries, exhibit an almost binomial distribution, which is a sign of entropy-neutral co-assembly. Unlabelled peaks correspond to complexes with truncated subunits. Right: Based on the fitting of the 'counting model', extracted -0.52 kT energy bias shows that for the HSP17.6-HSP18.1 pair, there is a slight preference towards self-assembly.

subunits already at room temperature (Fig. 3.4), which means that their interfaces are highly compatible. We examined products of subunit exchange at three different molar ratios. Multiple charge series were detected in each spectrum, all of them assigned as hetero-oligomers (Fig. 3.12). The resulting hetero-oligomers of HSP18.1 and HSP17.6 exhibit almost binomial distribution, which shifts with a change in ratio. This indicates unbiased co-assembly. The extracted energy preference difference of -0.52 kT means a slight preference toward self-assembly.

3.4 Discussion

3.4.1 Mapping the energy of interactions between organelle sHsps

In this chapter, we investigated assemblies between organelle sHsps from different evolutionary classes from either the same or different cellular organelles to determine how their interfaces evolved.

We demonstrated that all organelle sHsps from *A. thaliana* we studied here can hetero-oligomerise. They exhibit various modes of interactions, from a strong preference for self-assembly (between CP and MTI) to almost unbiased co-assembly (between MTI & MTII, and CP & MTII).

We showed that sHsps, not only from different evolutionary classes but also from different cellular organelles, are capable of subunit exchange within higher oligomeric structures. The most noticeable example is HSP21 and HSP26.5, which are targeted to different cellular compartments and co-assemble in an almost unbiased manner. This is surprising, taking into consideration that those proteins have been physically separated for hundreds of millions of years. Interfaces of those proteins have barely changed, even though they had over 400 million years to do so. For this pair, the evolutionary constraint of thermodynamic bias towards

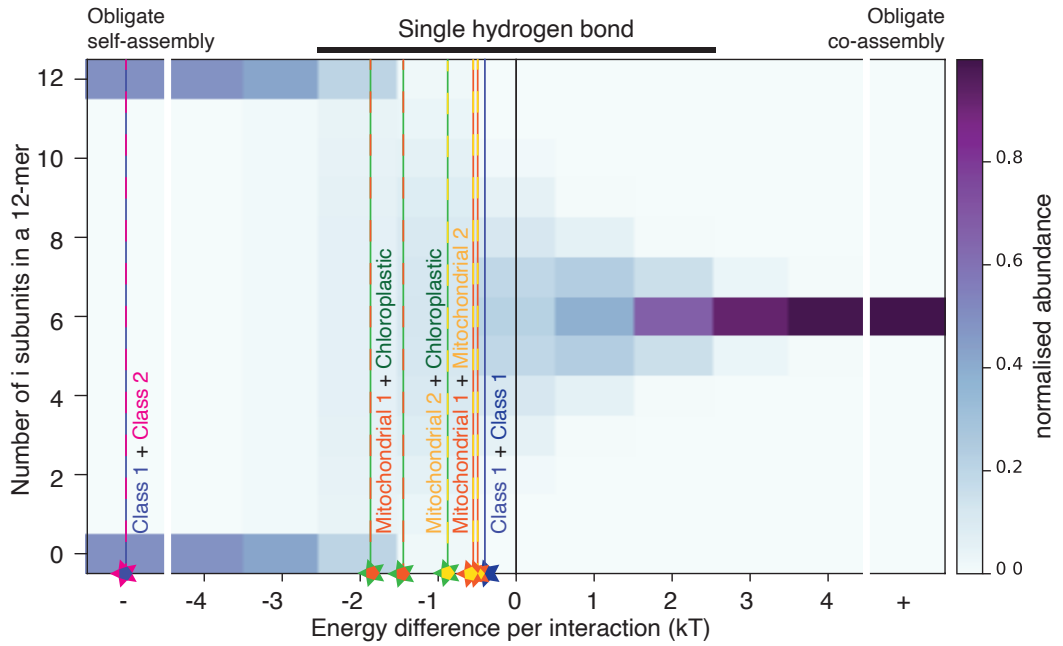


Figure 3.13: Mapping experimental values of energy bias. The 12-mer sub-stoichiometries are shown as a function of energy bias. Each line corresponds to obtained experimental value of energy bias, based on MS results. All of the investigated pairs show some level of preference towards self-assembly; however, none of them overcame an energy barrier of 4 kT, which is on a scale of two hydrogen bonds, which would keep them from interacting, as for the CI&CII pair.

co-assembly is not present. This implies that it is crucial to maintain interfaces responsible for those interactions and suggests that promiscuity and assembly into higher oligomers keep sHsps conserved.

Interestingly, for one of the MTI&MTII pairs, it was detected that HSP23.6 incorporates subunits of HSP26.5 into hetero-16-mers, which individually it does not form. We showed that co-assembly induces biophysical changes in dispersity and protein structure. This might impact the activity and function those sHsps pursue in the mitochondria.

Further, we quantified the energy bias for each investigated pair by applying a developed statistical-mechanical model to the MS data. This allows us to compare analysed pairs. As a reference point, we use the energy bias extracted for CI&CI

mixtures and a value obtained by Hochberg et al. ([26]) for CI&CII pair. All obtained values are mapped in Figure 3.13.

Three of the studied pairs exhibit almost unbiased co-assembly with a slight preference towards homo-interactions (on a scale of a weak hydrogen bond). Extracted energy bias towards self-assembly for MTI&MTII and MTII&CP is comparable but slightly higher than for CI&CI. MTI&CP is the only pair that displays a strong preference for homo-interactions. This shows that those sHsps slowly evolve towards obligate self-assembly; however, none of them changed interfaces to the degree that would prevent their interactions completely.

CI&CII evolved towards self-assembly on their way to subfunctionalisation and have the highest value of energy bias. This may suggest that pressure for subfunctionalisation for proteins from the same cellular compartment is very strong, and when it is not present, sHsps evolve much slower.

The developed model shows that the energy needed to keep two dodecameric proteins from interacting is less than 4 kT, which is on the order of the strength of two weak hydrogen bonds. Results presented here indicate that none of the analysed pairs overcame this relatively low barrier, even though they had hundreds of millions of years to do so. This means that the evolution of organelle sHsps is either very slow or it is preferred for them to maintain those interactions. Even in the absence of thermodynamic pressure to hetero-oligomerise, they have not evolved towards complete self-assembly, so their interfaces are still compatible.

3.4.2 Evolutionary conservation of organelle sHsps

To put these results into an evolutionary context, we place experimentally obtained values on a phylogenetic tree created from sequences of studied proteins (Figure 3.14). It illustrates that energy bias towards self-assembly between classes CP&MTII and

MTII&MTII are comparable to interactions between two CI members, proving strong conservation of oligomerisation interfaces. A noticeable change in the mode of assembly observed between CP&MTI means that on the branch which led to MTII diversification, correspond to substantial changes in sequence, which modified the compatibility of interfaces and, hence, influenced the way sHsps from those two classes interact.

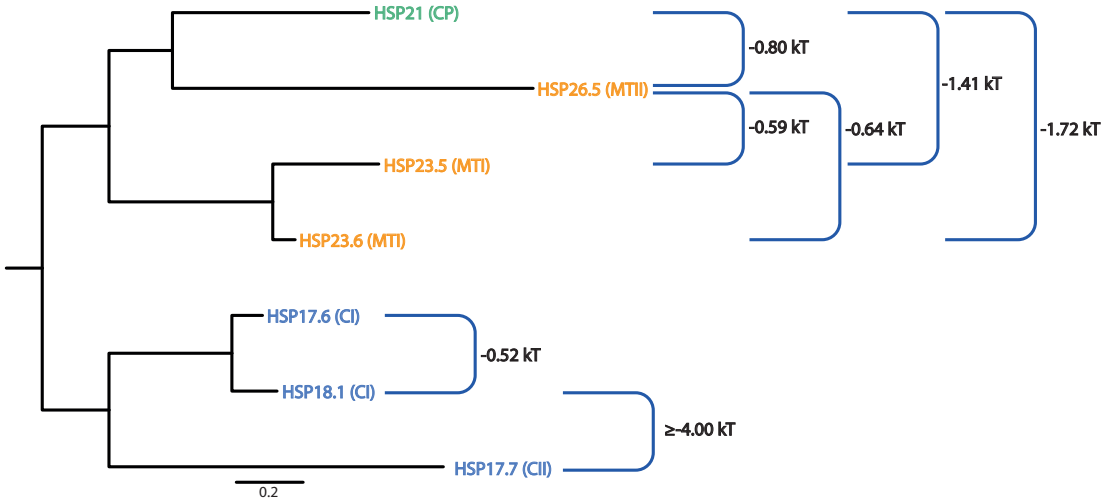


Figure 3.14: Phylogeny of investigated sHsps. Experimental values of energy biases mapped on a phylogenetic tree constructed only based on sequences of investigated proteins. It shows evolutionary distance affecting their interactions with an additive correlation between them and their energy preference towards self-assembly.

Furthermore, an additive property of the obtained values was observed. For example, summing up values obtained for HSP21&HSP26.5; -0.80 kT, with -0.59 kT extracted for HSP26.5&HSP23.5, gives -1.39 kT, which is close to the energy bias of HSP21&HSP23.5 pair (see Figure 3.14). This shows that the evolutionary distance of organelle sHsps resembles their co-assembly.

Comparing those values to energy biases of cytosolic sHps, while taking into account their evolutionary distances, it seems the values for organelle sHsps are smaller, implying stronger conservation of oligomerisation interfaces within sHsps targeted to membrane-bound cellular compartments.

Here, we have shown that when the pressure for subfunctionalisation and thermodynamic bias to heteromerise is not present, the evolution of plant sHsps is slowed down substantially. The reason for this may be the promiscuity of sHsps and broadly developed chaperone networks, to which those interfaces contribute significantly.

3.5 Acknowledgments

I want to thank Dr Dominik Saman, who developed the statistical-mechanical model described and used in this chapter. He invested a significant amount of time to help me understand and use it for the analysis of the MS data. Dr Saman also helped prepare some of the figures included here and helped describe the principles of how the model works, which I am hugely grateful for. Prof. Cecilia Emanuelsson very kindly provided the HSP21 samples. I also want to thank Dr Patrick Treffon from the Vierling lab for providing an empty SUMO plasmid and guiding me on how to insert a gene of interest. I highly appreciate Dr Georg Hochberg's kindness, who took me in during the first lockdown and taught me phylogenetics. I joined his group at the MPI for two months, where I learned how to use bioinformatic tools for protein evolutionary studies. Dr Hochberg supervised me throughout those two months and familiarised me with a new area of science, which was very useful for my projects and allowed me to be productive during a hard time of lockdown. I am also very grateful to Prof. Elizabeth Vierling for her many discussions and guidance.

3.6 Methods

3.6.1 Protein expression and purification

The full-length cytosolic (HSP17.6B, HSP18.1) and organelle (HSP23.5, HSP23.6, HSP26.5 - not containing transit peptides) sHsps were encoded in a modified pET23b vector containing a Ulp1 protease cleavable N-terminal His-SUMO tag (empty vector provided by the Vierling Lab, UMass Amherst).

BL21(DE3)pLysS competent cells (Novagen) were transformed with vectors. Single colonies were picked and grown overnight at 37°C in 20 mL of Luria Broth (LB) media containing 100 µg/mL of ampicillin and 34 µg/mL of chloramphenicol. Starter cultures were inoculated in 1 L of LB media (containing appropriate antibiotics) and grown at 37°C in an incubator shaking at 180 rpm until optical density (OD) reached values between 0.6 and 0.8. Expression was induced by adding IPTG to a final concentration of 0.4 mM, after which cells were incubated overnight at 20°C. Cells were harvested by centrifugation at 4000 xg for 10 minutes, flash-frozen in liquid nitrogen, and stored at -80°C until purification.

The cells were resuspended in 20 mL of lysis buffer (50 mM phosphate, 250 mM NaCl, 12.5 mM imidazole, 5 mM BME, pH 8), and protease inhibitor cocktail tablets (Roche) were added, after which the cell suspensions were lysed by microfluidising. The soluble fraction of lysate was separated by centrifugation at 20000xg for 20 minutes. The supernatant was filtered (with 0.21 μM filters) and loaded onto two connected 5-mL HisTrap columns (GE Healthcare). Proteins were eluted by running a gradient of buffer containing 500 mM imidazole. Collected fractions were analysed by SDS-PAGE gel, and the purest ones were transferred into the dialysis cassette with an appropriate molecular weight cutoff. Overnight dialysis against low imidazole buffer was set in the presence of Ulp1 hydrolase, cleaving the His-SUMO tag, which

was expressed and purified in-house. The cleaved protein samples were loaded again onto the HisTrap columns, the flow-through was collected, buffer-exchanged into PBS buffer (pH 7.4) with 1 mM DTT, and concentrated to 2 mL. The protein samples were injected into the Superdex 200/20 column and run at 0.5 mL/min for 1 column volume. The purest, pooled fractions were concentrated, after which their concentrations were determined by measuring absorbance at 280 nm with Nanodrop and using the extinction coefficient calculated based on protein sequence on the ProtParam website. The samples were aliquoted, snap-frozen, and stored at -80°C.

3.6.2 Mass Spectrometry - subunit exchange of sHsps

Stored at -80°C, protein samples were thawed and buffer-exchanged into 200 mM ammonium acetate (pH 6.9) using Bio-Spin columns (BioRad). Concentrations of samples were determined as described in Section 3.6.1 and diluted to 20 µM monomer concentration. Proteins were mixed in different volumes to obtain the desired molar ratios. MS measurements were conducted prior to incubation and after incubating the mixtures for at least an hour at 37°C, using a shaking heat block.

Native MS experiments were conducted on a Q Exactive™ Quadrupole-Orbitrap Mass Spectrometer (Thermofisher). Nanoelectrospray ionisation was carried out in positive-ion mode, using gold-coated glass capillaries prepared in-house, by applying modest backing pressure (around 0.5 mbar) and a capillary voltage of 1.4 kV. The source temperature was set between 50-150°C, and in-source trapping was set to -100 V in order to improve desolvation. Nitrogen was used as the collision gas in the HCD cell, with a trapping pressure set to 7.5.

Ion transfer optics, optimised for the best transmission, were set as follows: injection flatapole 10V, inteflatapole lens 9V, bent flatapole 8V, transfer multipole 4V, C-trap entrance lens 3V.

The instrument was operated in "Low Detector Optimization" with a resolution of 17,500 (at 200 m/z). Spectra were acquired with averaging of 10 microscans, injection time of 10 ms, and AGC target set to 10^6.

Mass spectra were recorded in the m/z range from 1,000 to 20,000, with the noise threshold parameter being set to 3, and deconvolved using UniDec software ([134]).

Data were input with no smoothing and no binning, but for some of the data, non-linear subtraction was used (set between 10-100). During the data processing, binning was applied, and the value was optimised for each data set, which allowed the impact of truncations and the presence of salt adducts to be minimised. FWHM was fitted for each data set manually. Data deconvolution was performed in various mass ranges, adjusted for each mixing pair. All other UniDec parameters were left at their default values. The oligomer and mass tool was used to assign stoichiometries of each detected heteromer, and then their intensities were extracted.

3.6.3 Description of co-assembly using statistical mechanics

Our method involves a detailed computational approach to understand protein oligomerisation. We developed a recursive algorithm for generating all stoichiometric combinations within protein oligomers. This process begins with the simplest subunit configuration and iteratively builds up more complex stoichiometries, ensuring complete coverage of all geometric combinations.

For experimental validation, we applied this model to a variety of protein pairs, analysing them at different mixing ratios through native mass spectrometry. The interaction energies for homo-oligomerization (E_{aa} and E_{bb}) were set to zero, focusing our analysis on hetero-interactions (E_{ab}). Each protein pair and mixing ratio was fitted individually with the model, allowing E_{ab} and the stoichiometric ratio to vary. These iterative fittings were crucial to accurately match our model

with experimental observations. The final energy values reported in our study are averages of these fittings, providing a comprehensive understanding of the energetics in protein interactions across different conditions.

4

Incompatibility of dimerisation interfaces in organelle and cytosolic plant sHsps

Contents

4.1 Motivation 112
4.2 Introduction 113
4.2.1 Oligomerisation interfaces of sHsps 113
4.2.2 Structural differences between cytosolic and organelle sHsps 115
4.3 Results 116
4.3.1 Phylogenetic analysis of CI sHsps shows their strong conservation 116
4.3.2 Mitochondrial(MTII) and cytosolic(CI) sHsps form co-assemblies with favoured even stoichiometries 119
4.3.3 CI sHsps changes dispersity as a result of co-assembly with MTI member 124
4.3.4 Products of reaction between CP and CI sHsps are hetero-oligomers with strictly even stoichiometries 129
4.4 Discussion 131
4.4.1 Dimeric incompatibility in cytosolic and organelle plant sHsps 131
4.4.2 Compatibility of CI and organelle sHsps interfaces responsible for higher oligomer formation 134
4.5 Acknowledgments 137
4.6 Methods 137
4.6.1 Protein expression and purification 137
4.6.2 Characterisation of sHsps complexes by nESI-MS 138
4.6.3 Phylogenetic analysis 138

4.1 Motivation

In the previous chapter, we revealed that organelle sHsps interact with each other, even if they are targeted to different cellular compartments. Here, we investigate the ability of organelle sHsps to interact with even more evolutionary distant homologs, cytosolic CI members. In the Section 3.3.2, we showed that CI members co-assemble in an almost unbiased manner amongst each other, but they do not interact with CII sHsps, as a result of subfunctionalisation of the latter, which was reported previously ([14]).

CI evolved from algae, and evolutionarily, it is the most ancient plant sHsps class. It is also the most conserved and numerous plant subfamily of these chaperones. We performed a phylogenetic analysis to assess what evolutionary mechanism keeps those proteins interacting and is responsible for their high abundance.

All organelle sHsps from plants diverged from CI over 450 million years ago. That is a significant amount of time, sufficient to structurally differentiate. We aim to determine if organelle sHsps, on their way to gaining new functions in different subcellular locations, evolved by changing their interfaces from cytosolic CI sHsps, which would influence their co-assembly. The goal is to evaluate if organelle sHsps escaped the strong evolutionary mechanism, which maintains a high conservation of CI.

Moreover, we analyse regions of amino acid sequences of investigated sHsps, which are responsible for those interactions, to evaluate how changes between them that took place over those hundreds of millions of years impact experimental observations of co-assembly between cytosolic CI and sHsps targeted to membrane-bound cellular compartments.

4.2 Introduction

4.2.1 Oligomerisation interfaces of sHsps

As we already discussed in the previous chapter, sHsps exhibit various modes of co- and self-assembly. All results in Chapter 3 illustrate the co-assembly of organelle sHsps. On the other hand, Hochberg et al., by studying CI and CII sHsps, which diverged more than 450 million years ago ([65, 109]), demonstrated how differences in regions of their ACD domains, not directly involved in the dimerisation interface, affected the structure of dissociating plant sHsps monomers and consequently cause lack of ability to hetero-oligomerise between the members of those two classes (see Figure 4.1). This hints towards the selectivity of the ACD, regulating heterodimerisation abilities between sHsps.

Moreover, this study presented how another selective contact between the CTR

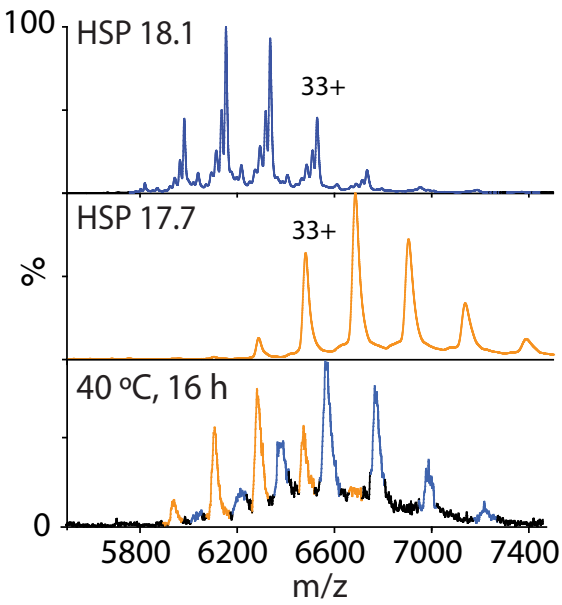


Figure 4.1: Oligomeric incompatibility in CI and CII plant sHsps. Native MS of CI HSP18.1 (top), CII HSP17.7 (middle), and after mixing the two and equilibrating at a higher temperature (bottom), shows detection of only homo-12-mers, which means lack of subunit exchange between those sHsps. In other words, CI and CII exhibit specific self-assembly. Used with permission from ([26]).

and the hydrophobic groove of the ACD (between β4 and β8 strands) regulates further the assembly of dimers into higher oligomeric structures.

The last type of interaction influencing sHsps assembly involves the NTRs. Those contacts are promiscuous, and occur at the last step of assembly, and are responsible for the stabilisation of higher-oligomeric structures ([26]), as well as for controlling their kinetic stability.

In Chapter 3, we demonstrated that it is difficult to evolve interfaces of sHsps that take part in oligomerisation interactions and, hence, diverge towards self-assembly. Perhaps developing selectivity against co-assembly would also disfavour self-assembly. The ability to form higher oligomeric structures is conserved across the sHsps family. The same contacts that take part in co-assembly are crucial for interacting with substrates (see Figure 4.2). This is demonstrated by cross-linking experiments (data acquired by Dr K. Ballard), in which two cytosolic CI sHsps

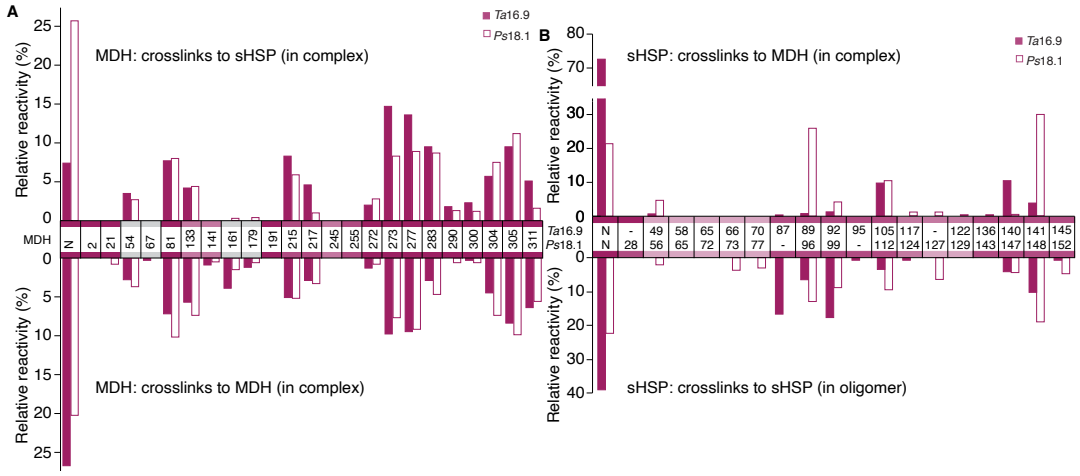


Figure 4.2: Oligomerisation interfaces of sHsps take part in substrate binding. A) A model substrate of sHsps (MDH) crosslinked with itself in complexes, shows that the same interfaces are then bound by sHsps. B) The same is true for sHsps; cross-linking MS data shows that oligomeric interfaces take part in interactions with aggregation-prone substrates, implying competition of the interfaces. This suggests a mechanism in which a stress-labile protein exposes hydrophobic interfaces upon unfolding, which are bound by dissociating sHsps through oligomerisation interfaces. Data and the figure produced by Dr K. Ballard (data not published).

were crosslinked either to themselves in a complex or with a model substrate; MDH, using 7 Å crosslinker bis(sulfosuccinimidyl)glutarate (BS2G). Participation of these sequence regions in client binding means that changing those interfaces could result in the loss of the sHsps function they serve in the cell and highlights the significance of those interfaces in protein homeostasis.

4.2.2 Structural differences between cytosolic and organelle sHsps

Recently, a cryo-electron microscopy structure of *Arabidopsis thaliana* plastid HSP21 (PDB 7BZW) has been reported ([131]). We compare it with *Pisum sativum* HSP18.1 dimer structure (PDB 5DS2) obtained by X-ray crystallography (Figure 4.3). The co-assembly ability of these two exact proteins is tested and described later in this chapter; hence, we can use those models to justify experimental observations.

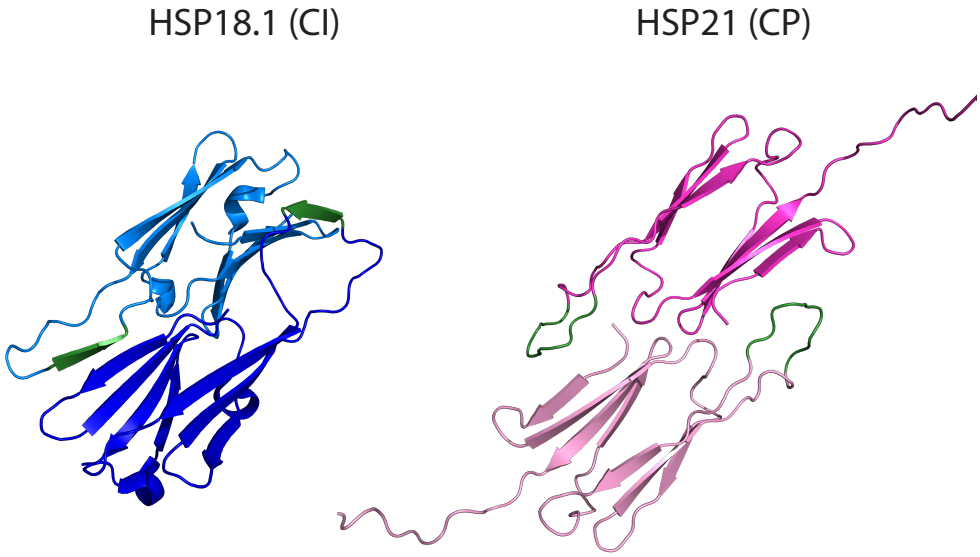


Figure 4.3: Comparison of the dimers of *P. sativum* HSP18.1(CI) and *A. thaliana* HSP21(CP). Left: Monomers of HSP18.1 interact through reciprocal β6-strand (green) exchange, which creates contacts with β2-strand (PDB 5DS2). Right: The dimerisation interface of HSP21 involves β5-7 loop (green) interaction with β2-strand (PDB 7BZW).

The crucial dimerisation interface of CI sHsps involves the interaction of β6

with β2-strand forming a reciprocal swapped dimers. However, HSP21 does not have a β6-strand; instead, it interacts with β2 through a flexible β5-β7 loop.

This unique mechanism of dimerisation may result in a less stable dimeric interface of HSP21 and hence may foster the dissociation of HSP21 into monomers and impact its ability to co-assemble with cytosolic sHsps.

4.3 Results

4.3.1 Phylogenetic analysis of CI sHsps shows their strong conservation

Cytosolic CI has a very complex evolutionary history, which started more than 450 million years ago. In all plant organisms, the CI subfamily is always the largest, most diverse ([65]) and highly expressed as a result of heat response.

In the previous chapter, we showed that two sHsps belonging to the CI subfamily interact in an almost unbiased manner, which implies their strong conservation. We conducted a phylogenetic analysis to determine what evolutionary mechanism is responsible for the intense expansions and conservation of that sHsp class, which maintains its co-assembly. In the next sections of this chapter, we experimentally evaluate whether organelle sHsps are still influenced by this mechanism or if they managed to escape it throughout those 450 million years, as observed for Class II members.

Around 200 sequences from all land plant clades were identified as CI homologs by BLASTP. Based on sequence search results and their MSA, a phylogenetic tree of the CI subfamily was generated, which is presented in Figure 4.4.

Obtained evolutionary data present an atypical pattern, repeating across the whole phylogenetic tree. Genes coding sHsps from the same species are the closest on the tree; in other words, the closest evolutionary relatives are coming from

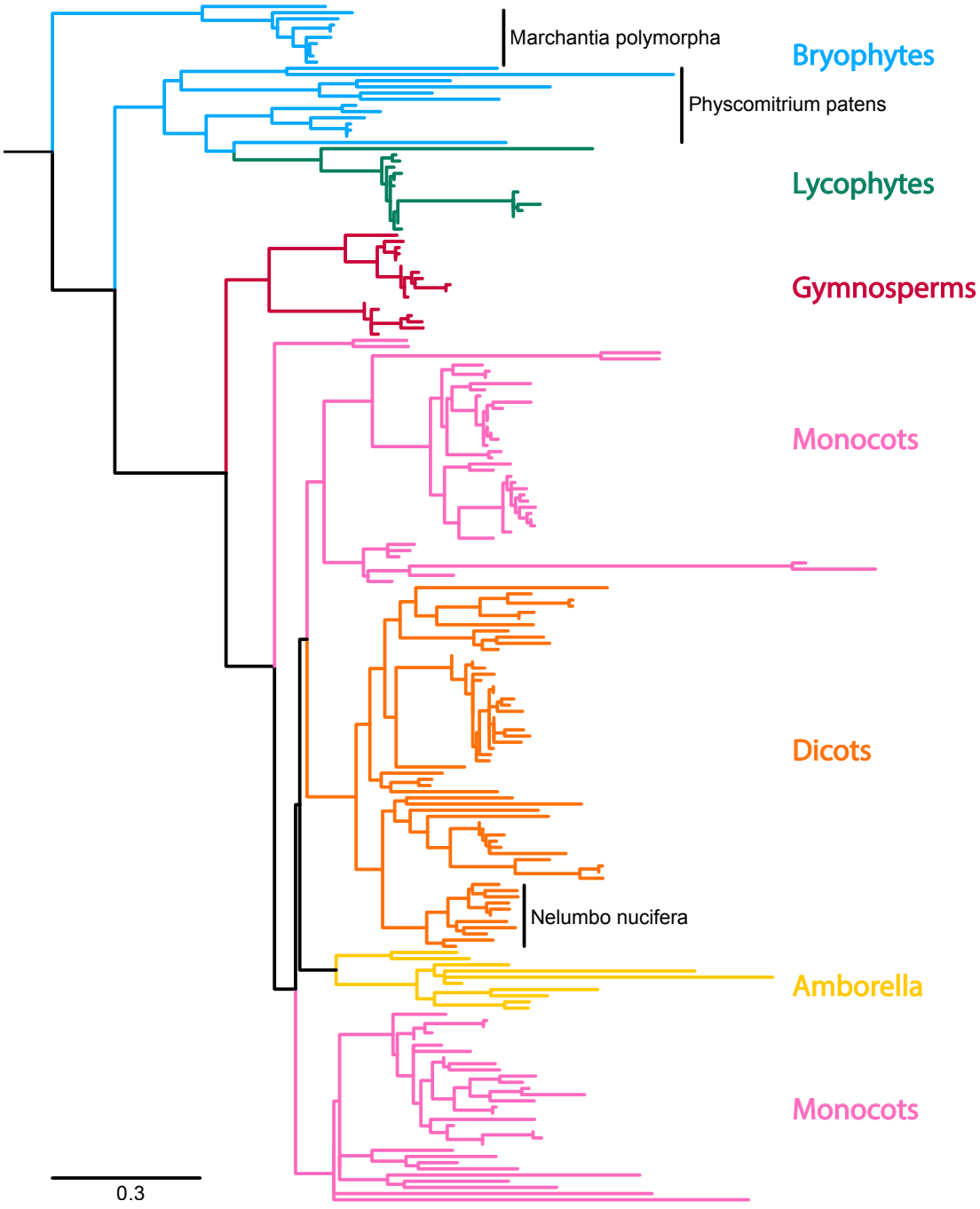


Figure 4.4: Phylogenetic tree of CI sHsps. Each plant clade is labelled with a different colour. Evolutionary analysis of land plant CI sHsps shows that genes coding homologs from the same species are the closest to each other on the tree (illustrated with three labelled examples), which is a sign of reoccurring gene conversion, keeping CI sHsps conserved by overwriting sequences, and keeping them similar, and hence, maintaining their coassembly. A high number of neighbouring paralogous genes is a sign of individual duplication events, which are responsible for the expansion of the CI class.

the same plant species. In Figure 4.4, we highlighted three examples to illustrate the reoccurring observation.

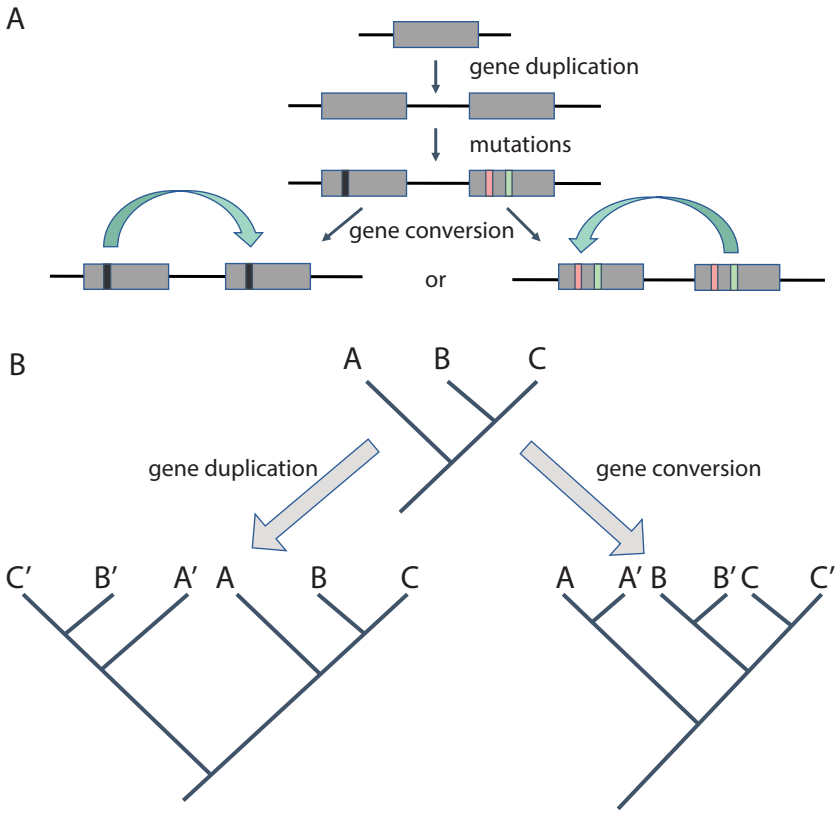


Figure 4.5: Gene conversion mechanism and its impact on phylogeny. A) After gene duplication, sequences undergo mutations, which cause their diversification. Gene conversion can act upon genes that are in close proximity only early after the gene duplication event, when their similarity is high. That causes homogenisation of the genes. B) Expansion of phylogeny by gene duplication results in neighbouring branches corresponding to homologs from different species, whereas gene conversion causes the branches of homologs from the same species to stay closely related.

Usually, as a result of gene duplication, orthologs (genes coming from different species) are evolutionarily closer than paralogs (genes from the same species), schematically demonstrated in Figure 4.5B. However, this is not the case for CI sHsps, which show not only a high number of homologs maintained in each species but also their closely maintained relationship. This suggests recurring gene conversion, a mechanism responsible for keeping genes of paralogs from diversification. It involves overwriting of a sequence by its neighboring homolog (see Figure 4.5A), which is

only possible if gene similarity is high. The effect it has on phylogeny (illustrated in Figure4.5B) is that neighbouring branches originate from proteins of the same species, leading to gene homogenisation, visible in the generated CI phylogenetic tree.

Gene location analysis showed that more than 50% of sequences included in the phylogenetic tree involve directly neighboring sHsps. This information not only proves that gene conversion was possible to occur but also shows a very high number of individual gene duplicate events, which cannot be explained by whole genome duplication, leading to plant polyploidy.

These results imply that gene conversion maintains the interaction of CI sHsps by overwriting sequences and keeping them similar. When they escape this mechanism, they preserve their conservation through constant birth-and-death evolution.

4.3.2 Mitochondrial(MTII) and cytosolic(CI) sHsps form co-assemblies with favoured even stoichiometries

Subunit exchange between HSP26.5(MTII) and HSP17.6(CI)

Both expressed CI sHsps (HSP17.6 from *A. thaliana* and HSP18.1 from *P. sativum*) were obtained as a full-length version with a small fraction (around 10%) lacking the N-terminal peptide, which presumably got post-translationally cleaved in between two phenylalanines for both CI sHsps. Even though the full-length proteins were around nine times more abundant than their shorter variants, this caused heterogeneity in the mass spectrum corresponding to the 12-mer region because of the incorporation of the truncated subunits into oligomers. This gave rise to additional peaks in the higher m/z region, hence making the investigation of subunit exchange with organelle sHsps more challenging. However, distinguishing appropriate subunits with MS was possible, so we proceeded using those samples.

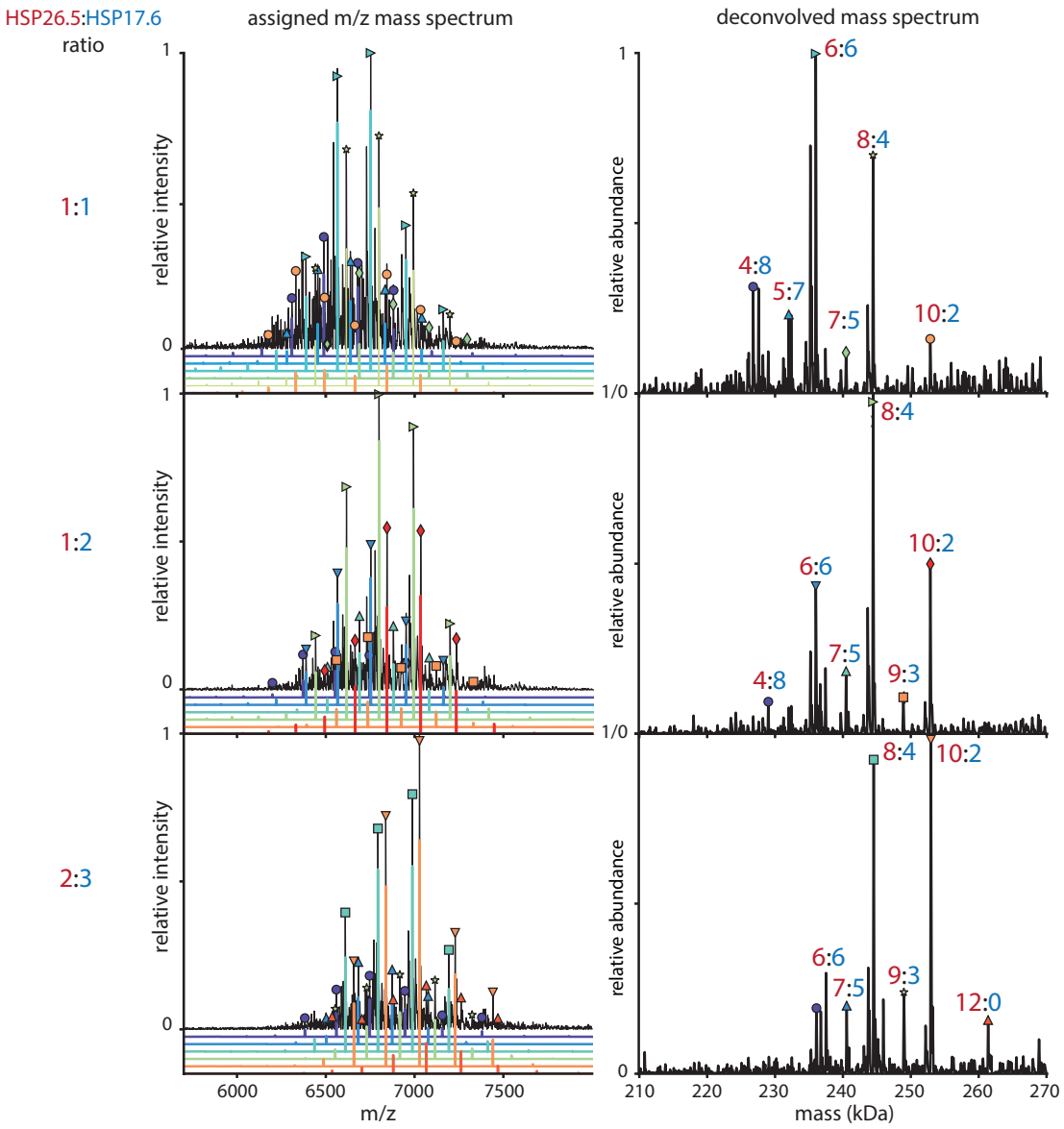


Figure 4.6: Products of subunit exchange between HSP26.5 and HSP17.6. Unlabeled species arose from hetero-dodecamers containing truncated HSP17.6 subunits. At all three investigated molar ratios, complexes comprised of even numbers of each type of subunit are much more abundant than products containing odd numbers of each sHsps. With an increase of HSP26.5 concentration, the distribution of hetero-oligomers shifts towards higher m/z, which after mass deconvolution can be assigned to a higher content of hetero-oligomers with a bigger number of HSP26.5 subunits contributing to complex formation.

First, we tested subunit exchange between dodecameric HSP26.5(MTII) and HSP17.6(CI) at three different molar ratios. MS acquisitions were recorded before, and after at least an hour of equilibration at 37°C. All obtained mass spectra (see

Figure 4.6) show multiple peaks that are not detected for individual sHsps. After the deconvolution of the assigned charge series, complexes containing subunits of both sHsps were identified. A repeating pattern was observed at all mixing ratios of charge series corresponding to hetero-12-mers with even stoichiometries, being more abundant than complexes containing odd numbers of each subunit. For example at a 1:1 molar ratio a species with the highest intensity, $[HSP26.5]_6[HSP17.6]_6$, is followed by $[HSP26.5]_8[HSP17.6]_4$ and $[HSP26.5]_4[HSP17.6]_8$ complexes, which are more intense than peaks corresponding to $[HSP26.5]_5[HSP17.6]_7$ and $[HSP26.5]_7[HSP17.6]_5$.

For the studied mixtures, the stoichiometry of the most abundant complex varied at each molar ratio, which means that distributions of hetero-oligomers are dependent on relative subunit availability in the solution. This suggests a statistical incorporation of dimer subunits into dodecamers.

Another surprising observation for this sHsps pair was perceived in spectra recorded prior to incubation, which showed the presence of the hetero-oligomers (Figure 4.7) even before incubation at a higher temperature. The detected distributions did not change significantly before and after equilibration, suggesting that the equilibrium of those mixtures was reached immediately after mixing (within less than a minute) of the solutions, which were kept at around 4-8°C (before mixing). This extremely rapid subunit exchange reaction may be attributed to flexible NTRs and CTRs, which assemble homodimers into 12-mers.

The data presented in Figure 4.7 showing a slight intensity increase of odd stoichiometry complexes raised a concern whether the equilibrium for those assemblies was reached within an hour, as it looks like single subunit exchange is also taking place but in a much slower manner. To confirm that the equilibrium was reached and that the observed effect is thermodynamic rather than kinetic, we performed an experiment in which the goal was to disassemble the homo-12-mers

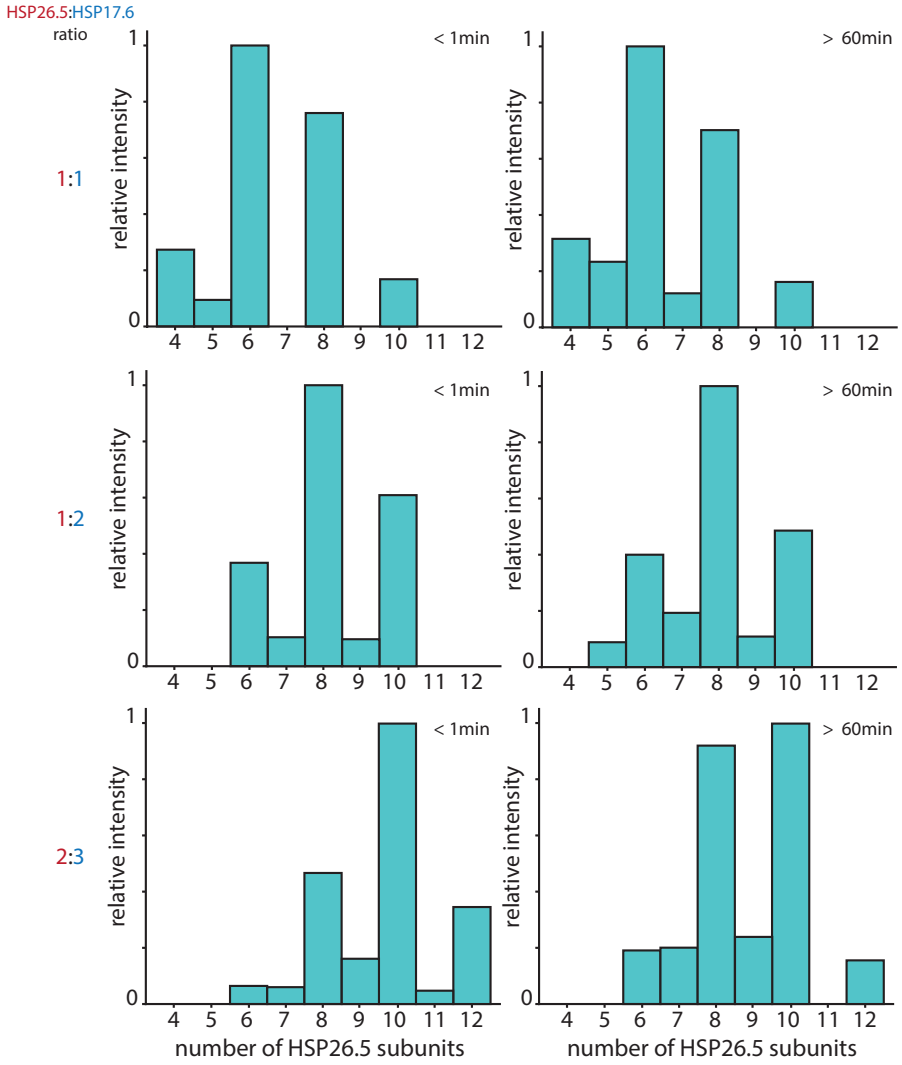


Figure 4.7: Abundances of hetero-oligomers of HSP26.5 and HSP17.6, detected before and after equilibration. Left: Measurements carried out immediately after mixing cold (4-8°C) samples show that hetero-oligomers were formed instantly, before equilibration at elevated temperature. Right: Comparing those intensities with data obtained after at least an hour of equilibration at 37°C indicates no significant difference. It implies that the monitored reaction is very rapid, and equilibration is reached without the need to incubate the mixture at a higher temperature.

before mixing, which was conducted by dilution of individual sHsps samples in 30% isopropanol. Then, the two proteins in disrupted conformations were mixed, after which we allowed for their re-assembly by exchanging solutions to ammonium acetate and testing the products of subunit exchange after re-assembly.

Figure 4.8 presents the superposition of MS data obtained for HSP17.6 and

HSP26.5 individually, showing two homo-12-mers (top left). Those sHsps after solution disruption in 30% isopropanol form mostly monomers and dimers (top right), which reassemble into hetero-12-mers with strictly even-stoichiometries after buffer-exchange back to ammonium acetate (bottom). This confirms that one

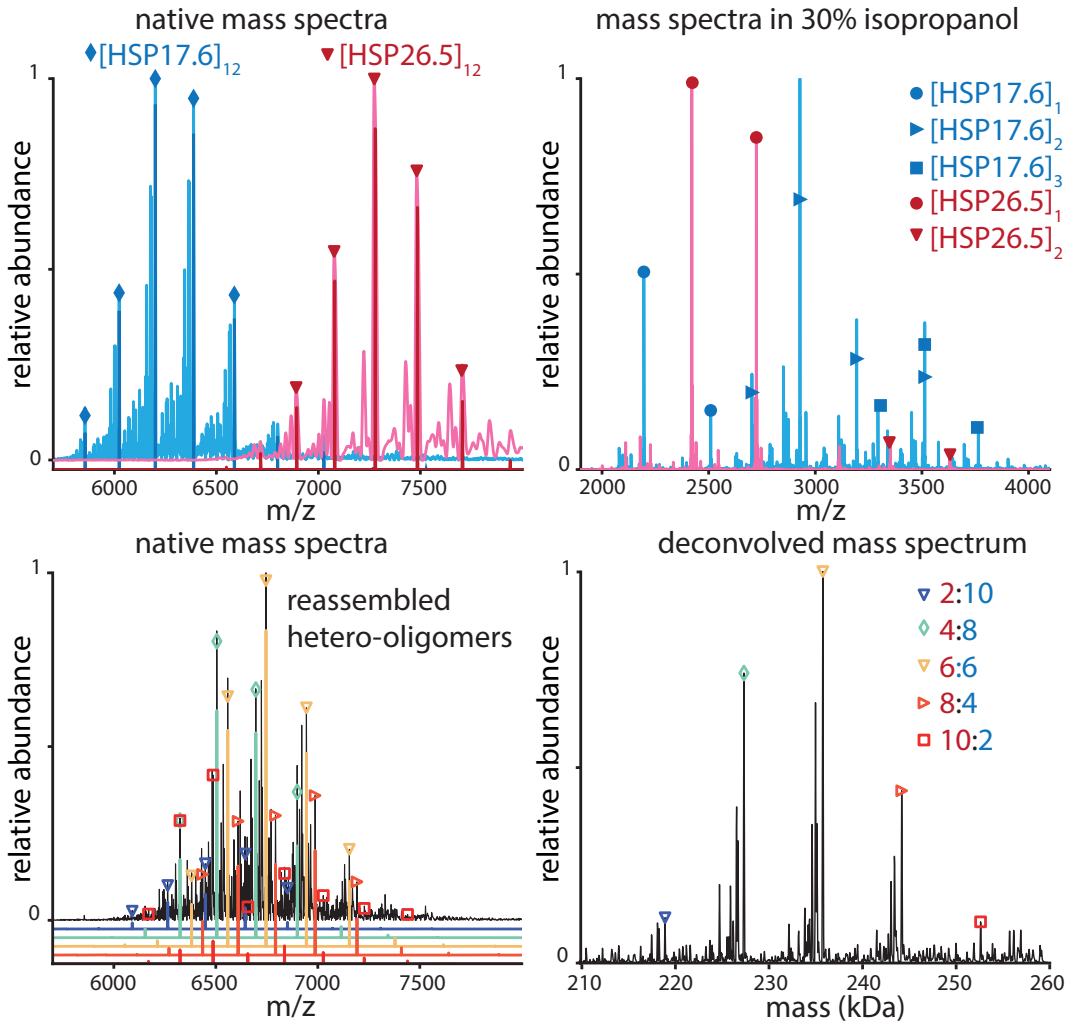


Figure 4.8: Disassembly and reassembly of HSP26.5 and HSP17.6 12-mers. Superposition of native MS presenting homo-12-mers of HSP17.6 (blue) and HSP26.5 (pink) sprayed individually (top left). Solution disruption with 30% isopropanol causes intense protein precipitation and disassembles higher oligomers into mostly monomers for HSP26.5 (blue) and HSP17.6 (pink) into monomers and dimers (top right). After mixing both proteins in the disassembled state at a 1:1 molar ratio and buffer exchanging the sample into ammonium acetate, recorded native MS spectrum (bottom left) shows that only complexes comprised of even numbers of each subunit type are formed, with the distribution centred on 6:6 stoichiometry, which is visible after deconvolution (bottom right).

hour of incubation at a higher temperature is enough time for this system to reach equilibrium, and that for these mixtures 12-mers containing even numbers of each subunit are favoured.

Moreover, the spectra obtained in 30% isopropanol for individual sHsps (Fig. 4.8, top right) show that HSP26.5(MTII) forms mostly a monomer, whereas in a HSP17.6(CI) sample dimer is the most abundant species. This may suggest a lower dimer stability of the mitochondrial sHsps.

Products of subunit exchange between HSP26.5(MTII) and HSP18.1(CI)

The second investigated pair of cytosolic and organelle sHps was between two dodecameric species: mitochondrial HSP26.5(MTII) from *A. thaliana* and cytosolic HSP18.1(CI) from *P. sativum*. Native MS data recorded at three molar ratios (Figure 4.9) show a similar manner of co-assembly as mixtures studied above. Again, peaks coming from even stoichiometries of hetero-oligomers are more intense. Unlabeled peaks (in Figure 4.9) correspond to complexes that include truncated HSP18.1 subunits. Similar to the previously discussed mixtures, we also observe the shift of distributions of hetero-oligomers in correlation to their molar ratio, suggesting statistical dimer incorporation.

These results demonstrate that the manner of subunit exchange described in this chapter is conserved between the members of CI and MTII classes regardless of whether sHsps come from the same or different species.

4.3.3 CI sHsps changes dispersity as a result of co-assembly with MTI member

The next two interrogated mixtures are between MTII class member HSP23.6, which, as described in Chapter 2, exhibits polydispersity, and two monodisperse

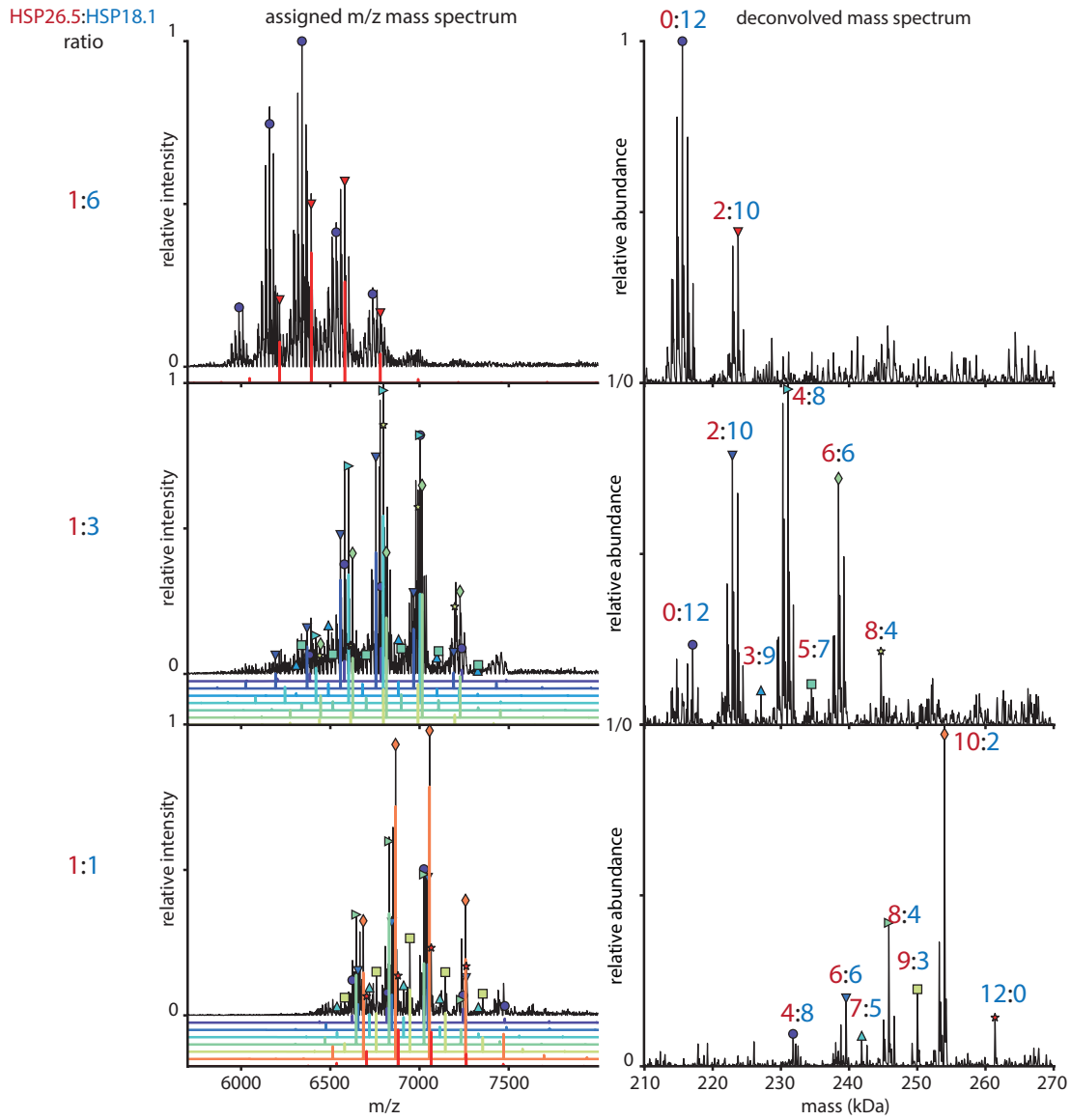


Figure 4.9: Product of subunit exchange between HSP26.5(MTII) and HSP18.1(CI). Co-assembly between HSP26.5 and HSP18.1 at molar ratios of 1:6 (top), 1:3 (middle), and 1:1 (bottom) shows that even stoichiometries of hetero-oligomers are favoured. Unlabelled peaks correspond to complexes consisting of truncated HSP18.1 subunits.

CI members HSP17.6 and HSP18.1. Both pairs were tested at a 1:1 molar ratio. Different m/z regions were analysed, to determine their subunit exchange at different oligomeric states. The high polydispersity of the obtained samples originating from the polydispersity of the mitochondrial protein and truncated subunits of cytosolic

sHsps resulted in a necessity to apply a background subtraction, which allowed data analysis and correct peak assignments.

Subunit exchange between HSP23.6(MTII) and HSP17.6(CI)

Spectrum recorded after equilibrating an equal molar mixture of HSP23.6 and HSP17.6 (Figure 4.10) shows that they exchange subunits within various oligomeric-size structures. Hetero-oligomers were detected within dimers, 12-mers, 14-mers, and 16-mers, but we can not exclude the possibility that co-assembly also takes place within other oligomers, which were not detected, because of the polydispersity of the sample. Only complexes with even stoichiometries were identified within higher oligomers, similar to the results for mixtures between MTII and CI presented in the previous section.

A hetero-dimer consisting of both HSP17.6 and HSP23.6 subunits was detected in the mass spectrum (see Figure4.10, bottom left). However, it was less abundant than both homo-dimers, suggesting low compatibility of HSP17.6 and HSP23.6 dimerisation interfaces. This hetero-dimer was not detected to be involved in the formation of higher oligomers, implying low stability of the formed hetero-dimeric species comprised of both of the investigated sHsps.

The presented results also demonstrate that HSP17.6, which is a monodisperse dodecameric protein, in the presence of HSP23.6, gets incorporated into higher oligomers, which individually it does not form. It indicates that the co-assembly of the analysed pair influences the biophysical properties of sHsps, in this case, structure, which is known to impact the functions proteins pursue in the cell.

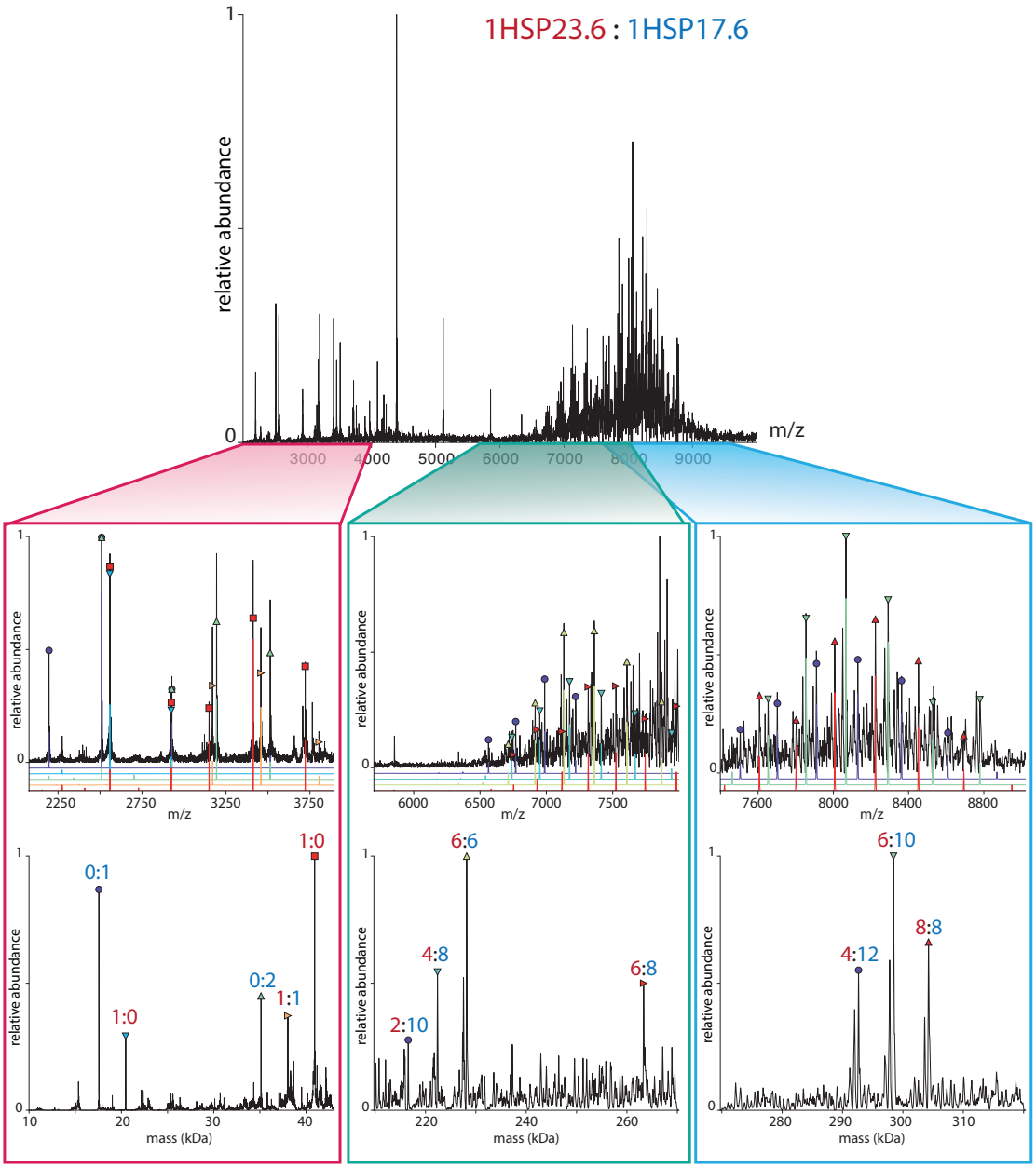


Figure 4.10: Complexes formed as a result of HSP23.6(MTI) and HSP17.6(CI) co-assembly. Top: Full m/z native spectrum measured at a 1:1 molar ratio after more than an hour of equilibration. Bottom: Zoomed-in views of m/z regions corresponding to different oligomeric sizes. Monomers and both hetero- and homo-dimers are present in the mass spectrum, with a higher abundance of species containing only one type of subunits (left). HSP23.6 with HSP17.6 co-assemble into hetero-12-mers and hetero-14-mers comprised of only even numbers of each type of subunit (middle). Similarly, the 16-mer m/z region shows a preference to co-assemble strictly into even-stoichiometry complexes (right). This implies a co-assembly reaction based on homo-dimeric subunit exchange.

Subunit exchange between HSP23.6(MTII) and HSP18.1(CI)

The second pair of CI and MTII was also studied at an equal molar ratio. Obtained MS data for HSP23.6 and HSP18.1 pair (Figure 4.11) are analogous to the above-

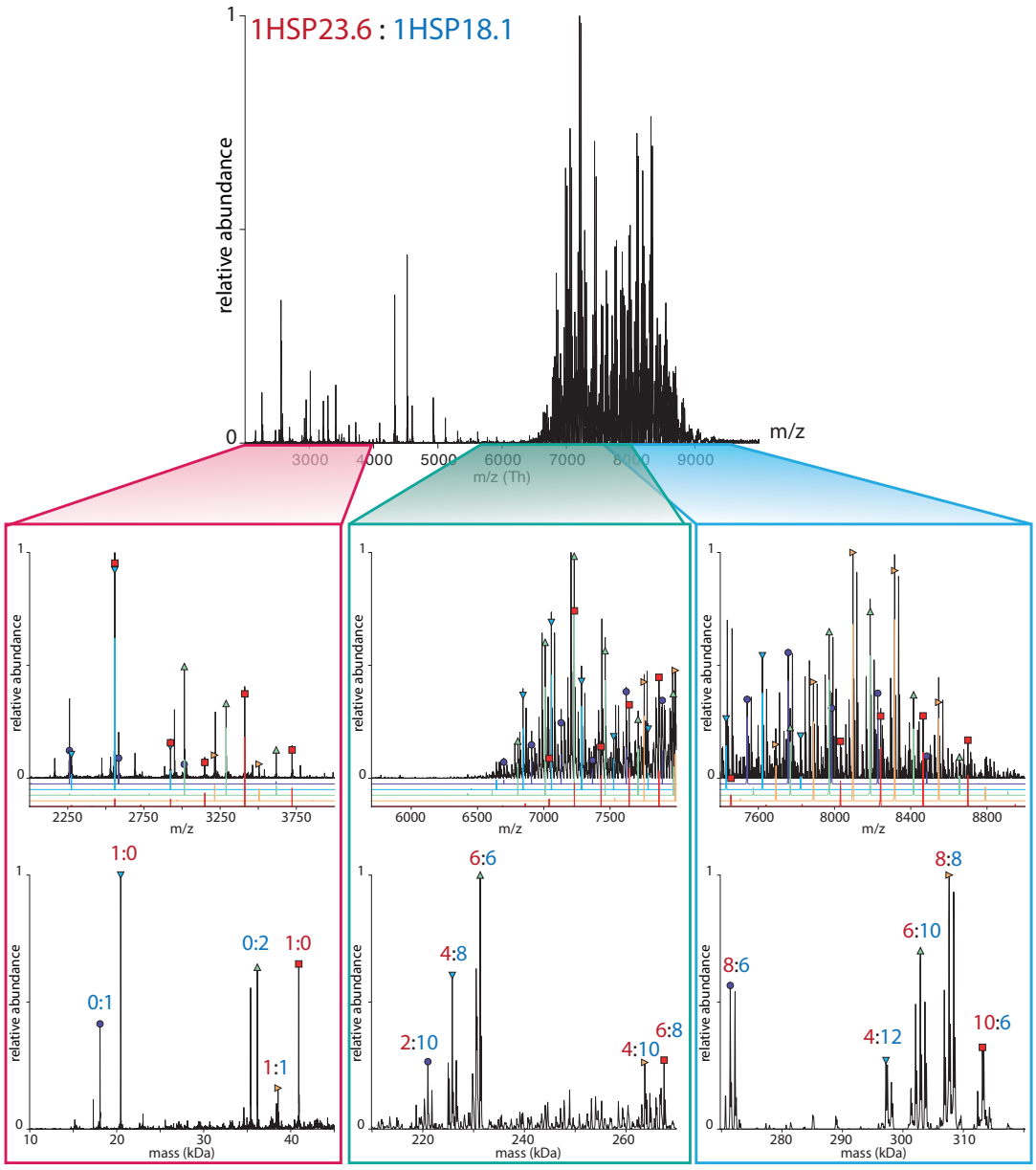


Figure 4.11: Products of co-assembly between HSP23.6(MTI) and HSP18.1(CI). Full m/z native spectrum measured at a 1:1 molar ratio after more than an hour of equilibration (top). Zoomed-in views of m/z regions corresponding to different oligomeric sizes (bottom). Monomers of both sHsps are visible in the spectrum, which also demonstrates a much higher abundance of homo-dimers in comparison to hetero-dimers (left). Co-assembly between HSP23.6 and HSP18.1 results in hetero-12-mers and hetero-14-mers consisting of only even numbers of each subunit type (middle), same as for the 16-mer m/z region (right).

discussed results. Again, only complexes with even stoichiometries were detected within higher oligomers, which suggests dimer-based subunit exchange. The presence of HSP18.1 subunits was identified in hetero-oligomers higher than a 12-mer, which HSP18.1 mainly forms within those experimental conditions. This suggests a change of HSP18.1 dispersity influenced by the presence of HSP23.6. The numerous detected peaks, which are unlabeled, correspond to hetero-oligomers consisting of truncated HSP18.1 subunits.

A hetero-dimer consisting of the two investigated sHsps was detected; nevertheless, its abundance was very low. This demonstrates a strong preference for homo-dimerisation and, thus, implies incompatibility of dimer interfaces, and as a consequence, supports the hypothesis that higher oligomers of those sHsps exchange subunits consisting of homo-dimers resulting in favoured even stoichiometries of detected hetero-oligomers.

4.3.4 Products of reaction between CP and CI sHsps are hetero-oligomers with strictly even stoichiometries

The last two investigated combinations are co-assemblies between CP and CI sHsps. Products of reactions between HSP21 and HSP18.1 (Figure 4.12, top) at a molar ratio of 1:3, and HSP21 mixed with HSP17.6 (Figure 4.12, middle), in an equal molar solution, were detected by native MS. Only hetero-12-mers containing even numbers of each subunit were identified in both mixtures, indicating dimeric subunit exchange, like for the rest of the samples discussed in this chapter. In this case, no hetero-oligomerisation resulting in odd stoichiometries was identified.

Moreover, we performed measurements before equilibration to determine if the co-assembly is as rapid as for the described reaction between HSP26.5 and HSP17.6 (Figure 4.7), which reached equilibrium immediately after mixing, prior

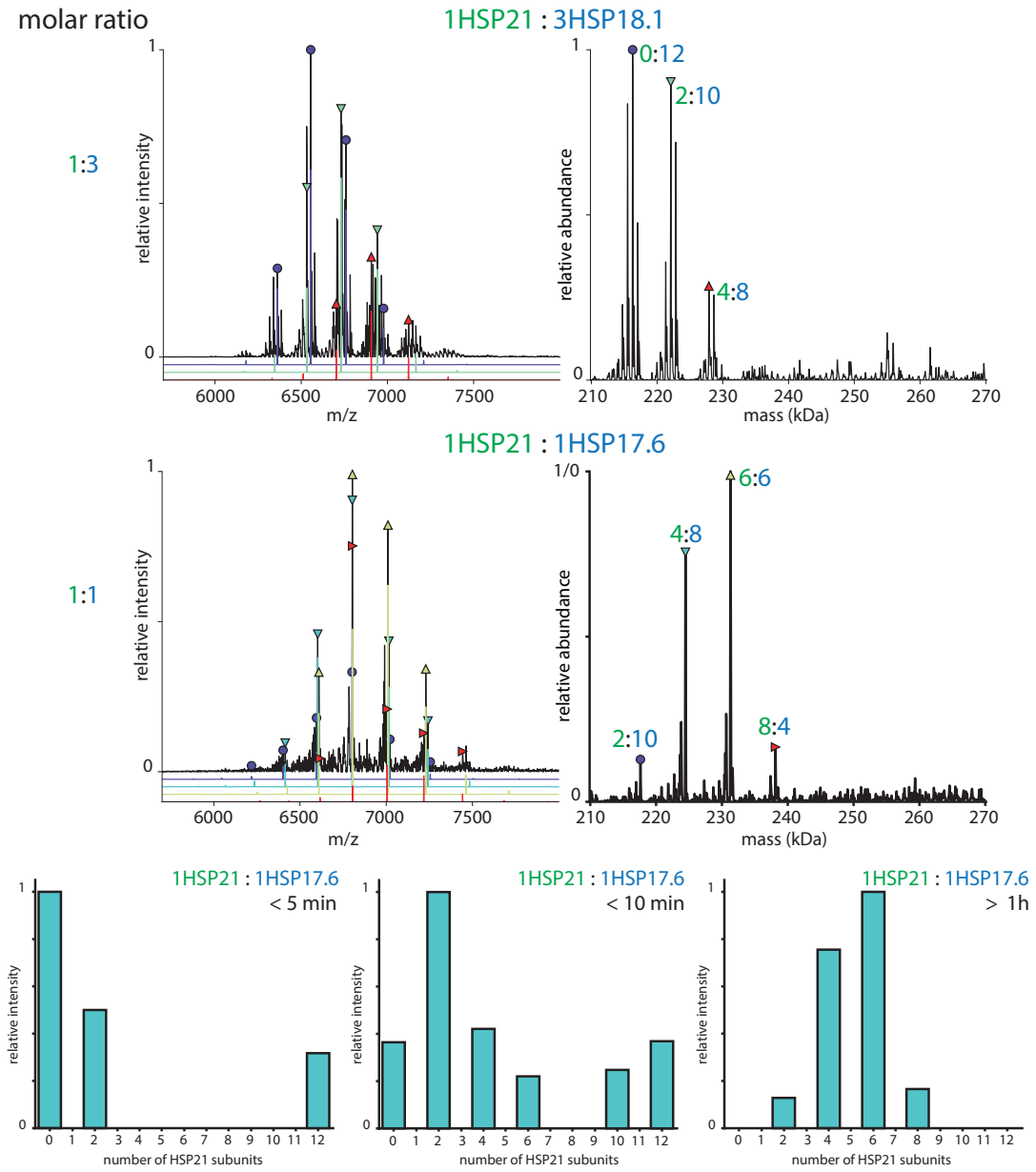


Figure 4.12: Products of subunit exchange between chloroplast HSP21 and cystolic CI members. Both native MS spectra show that co-assembly between HSP21 (CP) and both CI members, HSP18.1 (top) and HSP17.6 (middle), form hetero-oligomers comprised of only even numbers of each subunit type. Bottom: Extracted from native MS data, abundances of hetero-oligomers between HSP21 and HSP17.6 were detected immediately after mixing of cold (4-8°C) samples (left), after around 10 minutes at room temperature (middle) and after more than an hour of equilibration at 37°C (right). The presented data illustrate that firstly, HSP21 homo-dimer gets incorporated into HSP17.6 12-mer. This might be explained by a more rapid dissociation of HSP21 and, hence, a higher availability of their smaller oligomeric species.

to incubation at a higher temperature. Relative intensities extracted from native MS data acquired instantly after mixing, 10 minutes after (room temperature), and after more than an hour of equilibration at 37°C (Figure 4.12, bottom) reveal that this pair needs more time to reach the equilibrium.

These results also demonstrate the mechanism by which subunit exchange occurred (Figure 4.12, bottom). Firstly, two homo-12-mers were detected, corresponding to oligomers consisting strictly of either HSP21 or HSP17.6 subunits, accompanied by $[HSP21]_2[HSP17.6]_{10}$ complex, suggesting incorporation of HSP21 dimer into hetero-12-mer. This could mean that the availability of smaller species of HSP21 is higher than the cytosolic sHsp, which indicates a more dynamic dissociation of HSP21. After 10 minutes at room temperature, the $[HSP21]_2[HSP17.6]_{10}$ species was the most abundant in the sample, accompanied by the appearance of three additional hetero-oligomers $[HSP21]_4[HSP17.6]_8$, $[HSP21]_6[HSP17.6]_6$, and $[HSP21]_{10}[HSP17.6]_2$, while two homo-oligomers were still present but in lower intensity. This followed the final measurement after reaching equilibrium, which shows a dominant complex of $[HSP21]_6[HSP17.6]_6$, with three more even-stoichiometries of hetero-oligomers present in the sample.

4.4 Discussion

4.4.1 Dimeric incompatibility in cytosolic and organelle plant sHsps

Here, we presented an in-depth investigation of the subunit exchange between organelle sHsps from *A. thaliana* and cytosolic CI members (HSP18.1 from *P. sativum* and HSP17.6 from *A. thaliana*). All plant organelle sHsps derived from cytosolic CI over 450 million years ago (described in more detail in Chapter 2.2.1). As we showed in Chapter 3, interactions between organelle sHsps are maintained,

even between those localised in different cellular compartments. We wanted to examine if this ability to interact is also maintained with even more evolutionarily distant homologs, cytosolic sHsps from CI, or if it was lost over this time as a consequence of gaining new functions in the new sub-cellular locations.

All tested pairs show an ability to co-assemble into hetero-oligomers with even stoichiometries, which suggests dimer-based subunit exchange, leading to the formation of complexes comprised of different homo-dimers (see Figure 4.13).

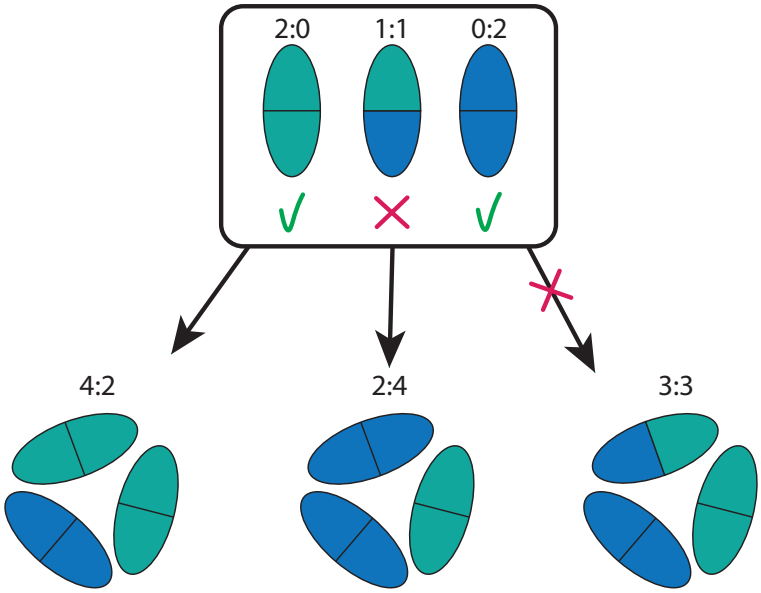


Figure 4.13: Schematic representation of the subunit exchange between cytosolic CI and organelle-localised sHsps.. The incompatibility of the dimerisation interface between CI and organelle sHsps leads to unfavoured hetero-dimeric structures. This causes predominant homo-dimeric subunit exchange, which leads to even stoichiometries of hetero-oligomers.

Results obtained for the MTI&CI mixtures demonstrate that heterodimers were detected at low intensities, and their engagement in higher oligomeric structures was absent. This implies that even though those subunits are capable of heterodimerisation, formed structures are characterised by low stability, hence their lack of participation in higher-ordered structure assembly.

4. Incompatibility of dimerisation interfaces in organelle and cytosolic plant sHsps

The mode of co-assembly between CI and organelle sHsps presented here, with the preference for even numbered subunits, is not a kinetic effect but rather a thermodynamic endpoint, as the preference persists after longer incubation time at a higher temperature (37°C), and this preference is also observed after disassembly of higher oligomers and reassembling back into 12-mers. This may be explained by the diversification of dimer interfaces, which caused their incompatibility.

We compared the available structures of CI HSP18.1 (from *P. sativum*, PDB 5DS2) and organelle HSP21 (from *A. thaliana*, PDB 7BZW), which show significant differences in dimerisation interfaces (see Figure 4.3). If assumed that all dimer structures within CI and within organelle sHsps are conserved, which is justified by sequence alignment (Figure 4.14), theoretical predictions, and experimental results, we can explain the co-assembly modes they exhibit.

The ACD is responsible for dimerisation across all sHsps. In CI, sHsps dimer stabilisation is based on a reciprocal strand exchange between β6 and β2, which interact through the pi system on a histidine (located on β6) with a tryptophan (on β2), which is conserved across CI members. However, those residues are absent in amino acid sequences of all tested organelle sHsps (see Figure 4.14, green).

Moreover, in all organelle sHsps, β6 strand is absent (shown on sequence alignment - Figure 4.14), which is a source of a structural difference of their dimers (see Figure 4.2.2, [131]). Instead, organelle-targeted sHsps have a flexible β5-7 loop interacting with β2 stand through limited hydrogen bonding of a tryptophan (on β5-7 loop) with an aspartic acid (on β2) (both highlighted on the sequence alignment-Figure 4.14, yellow, [131]).

The crucial changes described above in the dimerisation interface are the reasons for their incompatibility and either a lack of dimerisation or low intensity and stability of the created structures. Both subunits need to participate in strand

donation in order to stabilise the dimer efficiently; hence, it could not be pursued by the sHsps in question.

Presumably, the lack of β6 in organelle sHsps might cause their dimers to be less stable and more prone to dissociation. This is justified by available structural information and supported by our data (Section 4.3.4, Figure 4.12), showing that HSP21 dimer gets incorporated into hetero-12-mer as a first step of observed subunit exchange, indicating its higher availability in solution. Moreover, the comparison of the data after the disassembly of 12-mers in 30% isopropanol (Section 4.3.2, Figure 4.8) shows that mitochondrial HSP26.5 dissociated mostly into monomers, whereas HSP17.6 disassembled mainly into dimers after subjecting to the presence of an organic solvent, which disrupts the hydrophobic interactions and may lead to unfolding. This indicates that even after introducing very harsh conditions, the CI sHsps dimer remains stable. This consequently means much stronger stabilisation by the β6 strand exchange compared to the β5-7 loop contacts.

4.4.2 Compatibility of CI and organelle sHsps interfaces responsible for higher oligomer formation

Moreover, we analysed the less selective and more rapid interactions responsible for the assembly of dimers into higher oligomers, which were maintained in all investigated pairs. Two oligomerisation interfaces are considered. The first, more specific interaction of CTR that contains conserved V/IXI/V motif, which interacts with a hydrophobic groove between β4 and β8 (Figure 4.14, red). The second interface involves a heterogenous NTR, with a flexible and dynamic structure, which further stabilises higher-ordered oligomers ([29–31]).

The ACD-CTR is the first contact in the two-step assembly reaction of dimer into higher oligomers, and these interactions are rate limiting and more selective ([26]).

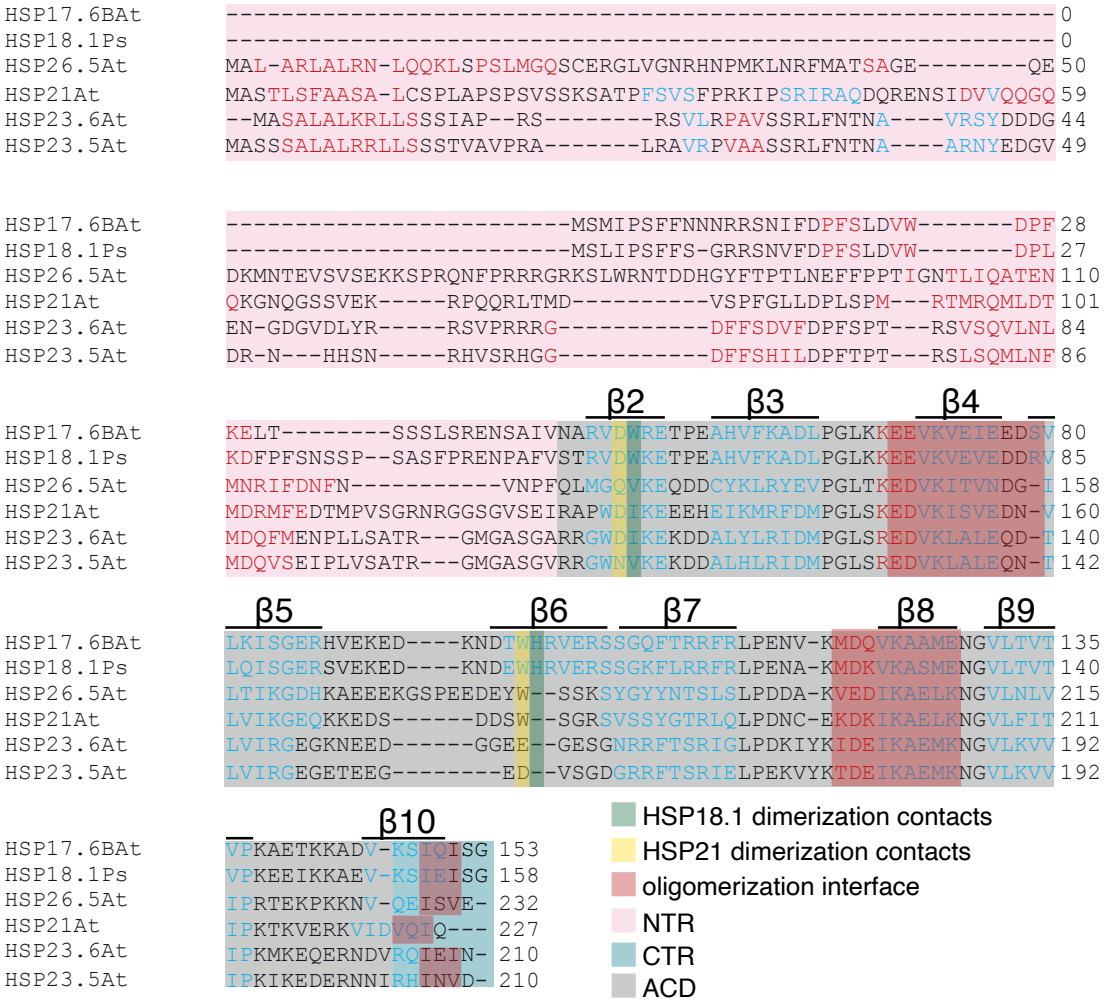


Figure 4.14: MSA of tested sHsps with highlighted sequence parts significant in co-assembly. All organelle sHsps lack β6 strand, which is a crucial dimerisation interface in CI, relying on the pi system between histidine and tryptophan (green). Instead, the organelle-targeted sHsps interact through tryptophan localised on a flexible β5-7 loop with aspartic acid on β2 strand (yellow). Contact of V/IXI/V motif and a hydrophobic groove of β4 and β8 (red) is responsible for the assembly of dimeric subunits into higher oligomers. Upon the formation of those contacts, NTR contributed to further stabilisation (pink). Red font represents helical structures, and blue font illustrates the contribution of β-strands of the amino acid sequence.

The hydrophobic groove between conserved strands β4 and β8 and the V/IXI/V motif are present in all interrogated proteins; thus, in all pairs, each subunit is able to both donate and receive the CTR. This allows for the incorporation of homo-dimers into hetero-oligomeric structures (as shown in Figure 4.13).

A siginifact difference was noticed in the equilibration time between HSP26.5&CI

(Figure 4.7, Section 4.3.2) and HSP21&CI (Figure 4.12, Section 4.3.4) mixtures. It may be explained by the higher similarity of the HSP26.5 V/IXI/V motif to that of CI, whereas the C-terminus of HSP21 is shorter, and its V/IXI/V motif is less compatible with the rest of the analysed protein sequences (Figure 4.14, red).

An exceptionally long NTR that is unique to organelle sHsps (Figure 4.14, pink) and has a higher helical content than the NTR of CI members, might contribute to the further stabilisation of the structure. Those NTR interactions were previously reported to be more promiscuous and responsible for the kinetic stability of oligomers ([26]).

We demonstrated that cytosolic CI is a highly abundant and conserved class of sHsps. A large number of CI paralogs can be explained by frequent gene duplication events, followed by gene conversion (concerted evolution) and birth-and-death evolution, led by pressure to maintain many copies of those proteins. All organelle sHsps diverged from CI over 450 million years ago ([65, 109]). We showed here that throughout this time, organelle sHsps escaped this mechanism and differentiated themselves from cytosolic sHsps, which led to structural incompatibility of their dimerisation interfaces, presumably as a consequence of gaining new functions in their different sub-cellular locations. It is remarkable that organelle sHsps diverged from CI hundreds of millions of years ago, but interactions between them are still maintained to some extent. We showed here that the ability to co-assemble is maintained by more flexible and less conserved CTR and NTR regions interacting with conserved ACD. Co-assembly led by regions with more plasticity is observed to be more rapid. Overall, this proves a very slow evolution of sHsps interfaces hence suggesting their importance in cellular proteostasis.

Additionally, some sHsps studied here can participate in subunit exchange with oligomeric species bigger than they individually form. This suggests that they undergo conformational changes upon hetero-oligomerisation; moreover, it proves the high plasticity of those subunits, where the ACD domain takes part in dimerisation, the CTR is responsible for the assembly of higher-oligomers, further stabilised by exceptionally long and disordered NTRs. Presented examples analysed above illustrate that co-assembly can affect protein structure, which is crucial for the functions those molecules pursue in the cellular environment.

4.5 Acknowledgments

The phylogenetic analysis presented in this chapter was conducted with Dr Georg Hochberg's supervision. I am very grateful for all his guidance and input on this project. Two proteins used in this chapter were expressed and purified with the help of part II students: Lauren Parsons and Georgina Mullan. I also want to thank Prof. Elizabeth Vierling and Prof. Elizabeth Waters for their advice, engaging discussions, and shared enthusiasm for this project. Thanks to Dr Dominik Saman, who showed me how to use SUMO contracts and most efficiently express and purify SUMO-tagged proteins, which I utilised to obtain proteins engaged in experiments presented here.

4.6 Methods

4.6.1 Protein expression and purification

Cytosolic HSP17.6, HSP18.1, and organelle HSP26.5, HSP23.6 were expressed and purified as described in Section 3.6.1. HSP21 was kindly provided by Prof. Emanuelsson.

4.6.2 Characterisation of sHsps complexes by nESI-MS

Protein samples were prepared and sprayed on a mass spectrometer, as described in Section 3.6.2. Protein samples for solution disruption experiments, with an initial concentration of around 180 μM, were diluted with 100% isopropanol to a final isopropanol concentration of 30%. This caused intense protein precipitation. After filtering out the white precipitate using Spin-X Centrifuge Tube Filters, disassembled proteins were sprayed on a mass spectrometer, using parameters described in Section 3.6.2. Samples of solution-disrupted HSP17.6 and HSP26.5 were mixed in a 1:1 molar ratio. Mixed sHsps in 30% isopropanol were buffer-exchanged to 200 mM ammonium acetate using Amicon Ultra Centrifugal Filters with 10 MWCO. The protein sample in re-assembled conditions (200 mM ammonium acetate) was nanoelectrosprayed using the same MS parameters.

4.6.3 Phylogenetic analysis

Class I sHsps homologs were searched with BLASTP using amino acid sequences of known Class I members. Top hits of sequences from each plant kingdom were identified, downloaded, and aligned using muscle3.8.31. Multiple sequence alignment (MSA) was viewed, reviewed, and edited in AliView software ([139]). Profile alignment was generated, it included each of the MSAs for individual plant kingdoms.

The initial approximately-maximum-likelihood phylogenetic trees were calculated using FastTree with default settings and visualised in FigTree v1.4.4. The final tree was produced with RAxML, based on 194 identified CI sHps full-length amino acid sequences. The tree was rooted on CII sHsps plant paralogs.

Chromosomal locations were analysed using data available on the EnsemblPlants database by manually searching the locations of each gene, coding the sequence included in the phylogenetic tree, and analysing the identity of neighbouring genes.

5

The principles of Endoplasmatic Reticulum α-Glucosidase II inhibition by thiopyridone as an host-targeting antiviral strategy

Contents

5.1 Motivation 140
5.2 Introduction 141
5.2.1 GluII as an antiviral strategy target 141
5.2.2 Targeting dimerisation interface of GluII 143
5.3 Results 145
5.3.1 In the native state GluII forms a heterodimer with a variety of glycoforms 145
5.3.2 Covalent interactions of GluIIα with thiopyridone . . . 147
5.3.3 Determination of the ligand binding residue 155
5.4 Discussion 160
5.4.1 Disulfide binding of GluII by thiopyridone 160
5.4.2 GluII-thiopyridone interactions led by dock, lock and latch molecular mechanism 161
5.5 Acknowledgments 162
5.6 Methods 164
5.6.1 nESI-MS of untreated and treated GluII 164
5.6.2 Characterisation of the ligand binding by LC-MS 165

5.1 Motivation

ERQC, described in more detail in Chapter 1, is responsible for the proper folding and modification of membrane and secretory proteins, as well as the degradation of misfolded polypeptides ([70]) by ER-associated degradation (ERAD), and consequently plays a significant role in proteostasis. This machinery is used by viruses, which take advantage of their host's ERQC system for the correct folding of their surface proteins ([140]). Currently, most therapeutics target viral proteins, but increasing resistance to these antivirals created a need for new drugs employing different strategies ([141]). Recent studies focus on targeting host-viral interactions that are necessary for viral life-cycle, biosynthesis, and infectivity ([142, 143]).

Here, intending to develop a host-based antiviral strategy targeting a broad range of viruses, we demonstrate a new approach based on the disruption of the dimerisation interface of GluII, a protein of ERQC, responsible for the admission and release of folding glycoproteins. Inhibition of this enzyme, used by a wide range of enveloped viruses, could cause misfolding of viral glycoproteins and, as a result, decrease viral proliferation and infectivity. Our main aim in this project is to characterise the binding between GluIIα and C61 - its potential inhibitor identified in a virtual screening that causes GluII heterodimer dissociation. We also attempted to determine which residues take part in those interactions. We employed mass spectrometry techniques and supported the obtained results with data from previous studies carried out by the Zitzmann group and Evotec.

5.2 Introduction

5.2.1 GluII as an antiviral strategy target

GluII is an ER-localised enzyme that plays a central role in the ERQC and ERAD (Section 1.4). It is responsible for the processing of nascent glycoproteins by cleaving glucose residues of N-linked oligosaccharides ([75]), whereupon they are recognised by co-chaperones, which ensure the correct folding of client proteins. A broad spectrum of enveloped viruses (such as lentiviruses, herpesviruses, and flaviviruses) take advantage of that mechanism. The proper folding of their envelope glycoproteins relies on it, allowing for viral proliferation and further infectivity. It is a reason why the ER protein-folding machinery is a promising target for novel host-based antiviral strategies.

GluII is a soluble heterodimeric glycoprotein consisting of a 108 kDa α-subunit, playing a catalytic role, and a 58 kDa β-subunit, which is responsible for the retention of GluII in the ER through its C-terminal peptide. Both of these subunits are needed for protein activity and solubility ([144]). They are also characterised by high conservation in eukaryotes, with a 92% and an 87% sequence identity between human and murine homologs of GIIα and GIIβ, respectively. Crystal structure reported by Caputo et al. (PDB: 5F0E, see Figure 5.1, [145]) was obtained for trypsinised murine GluII, lacking most of the β-subunit, as the crystallisation of a full-length protein caused difficulties. The N-terminal region of GluIIβ was preserved during trypsin digestion. This region includes most of the dimerisation interface; thus, the structure illustrates which residues may take part in dimer assembly. It is crucial for the work presented in this chapter, as we develop a novel host-targeted antiviral (HTA) strategy based on heterodimer dissociation.

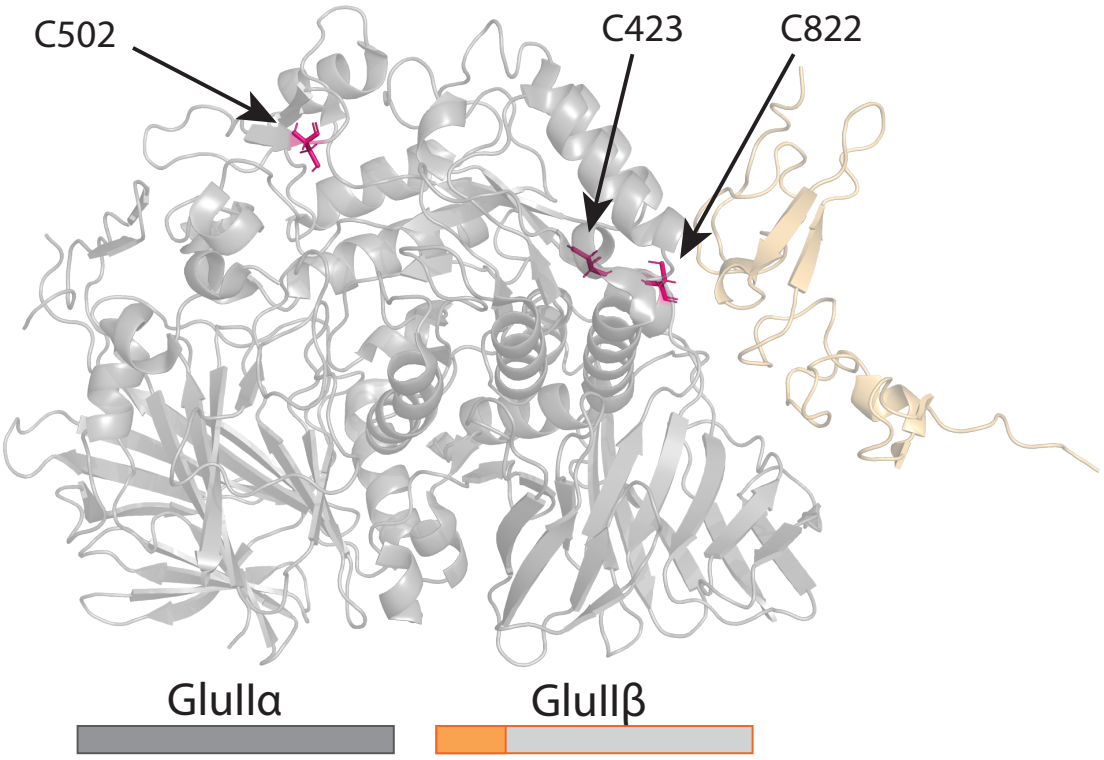


Figure 5.1: Crystal structure of trypsin-digested *Mm*GluII. Cartoon representation shows α-subunit (dark grey) and the preserved N-terminal part of β-subunit (orange). The bottom schematic illustrates the loss of subunit β regions due to trypsinisation (light grey). Free cysteine residues are labelled in pink, two of which (C423 and C822) are in close proximity to the interface between GluIIα and GluIIβ.

The α/β interface disruption is caused by interactions of small-molecular ligands with GluIIα, leading to loss of GluII activity.

So far, developing strategies targeting GluII have focused on designing small molecular glucose mimetics, binding in the active site of the enzymatically active α-subunit. Derivatives of iminosugars have been the most studied inhibitors, showing promising results. However, the widespread spread of glucose-binding enzymes is the reason for their low specificity, causing numerous side effects, which has created a need for a different strategy. While the work on more specific antiviral drugs binding to GluII active sites continues, another approach targeting the same protein has been taken, targeting the interface between subunit α and β of the heterodimer.

Interactions between the two monomers, which keep them together, are formed by salt bridges involving basic residues of α-subunits and acidic residues of β, resulting in a positive charge on the former and a negative on the latter. Disruption of those interactions could lead to decreased activity of GluIIα, which requires the MRH domain of GluIIβ for its maximal efficiency ([146, 147]). Additionally, loss of β-subunit and, as a consequence, loss of retention peptide could cause translocation of GluII out of the ER.

5.2.2 Targeting dimerisation interface of GluII

In order to induce the dissociation of the GluII heterodimer as a novel broad-spectrum antiviral strategy, a collaboration between the Zitzmann group and Evotec began a search for a suitable ligand - a small molecule that would cause electrostatic or steric interruptions in the dimerisation interface. The virtual screen of the region of interest was carried out, and to optimise interactions with either GluIIα or β, computer-aided drug design was performed, resulting in 82 potential compounds. These compounds were further tested with a time-resolved (TR) fluorescence resonance energy transfer (FRET) assay, which illustrates ligand activity as an α/β interface disruption. Results pointed at a 190 Da compound (called here C61) as a promising HTA targeting GluII dimerisation interface (see Figure 5.2, [148]).

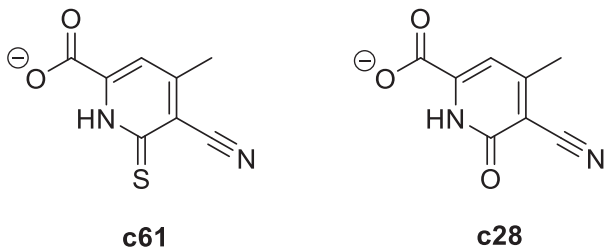


Figure 5.2: Structure of C61 and C28 ligands. C61 - 190 Da thiopyridone - is a potential GluII inhibitor, identified in virtual screening, and proved to cause dissociation of GluII. Conversely, C28, which differs only by one atom - sulfur substituted with oxygen - shows a lack of activity toward GluII.

C61 is a pyridone with a thiol group, which could interact with one of the free cystines of GluII (labelled in Figure 5.1), forming a disulfide bond. Several analogues of C61 were synthesised, where the sulfur was either substituted or blocked (called here C28, illustrated in Figure 5.2), none of which have been shown to be active in an RT-FRET assay, suggesting the significance of the thiol group in α/β interface disruption. MS experiments performed by Evotec also suggest binding C61 to a peptide corresponding to a fragment containing C822 (one of the free cysteines, illustrated in Figure 5.1), which is in close proximity to salt bridges responsible for dimerisation. Moreover, fluorescence N_3-TAMRA assays developed by Evotec indicate C61 incorporation with subunit-α ([148]). In extensive experiments, Dr Furgus Bremner (previously from the Zitzmann group) showed dissociation of GluII caused by C61 using MP (Figure 5.3 presents a summary of some of these experiments). Here, we use mass spectrometry techniques to determine the character of interactions between C61 and GluIIα responsible for heterodimer dissociation.

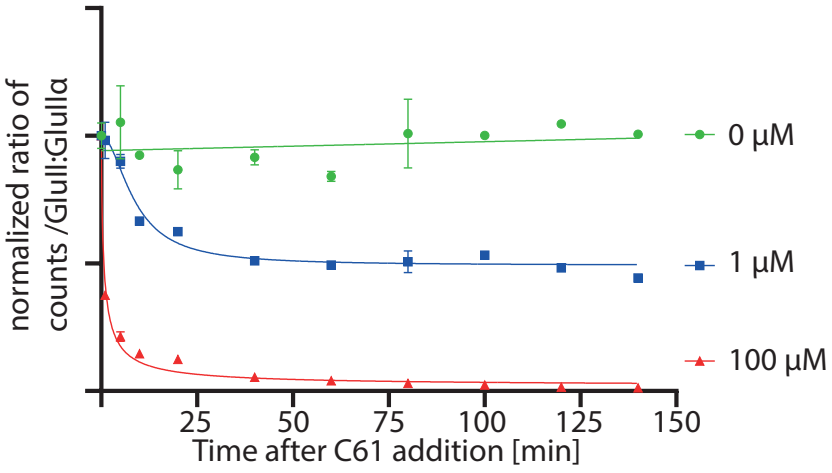


Figure 5.3: Dissociation of C61-treated 1 μM GluII monitored over the course of time by MP. Experiments were carried out for two different C61 concentrations: 1 μM, 100 μM, and in the ligand absence. Results show a depletion of dimer relative to α-subunit as a function of time in response to the addition of C61. It serves here as a measure of GluII dissociation. The α/β interface remains intact for untreated samples throughout the time course of experiments. Used with permission from ([148]).

5.3 Results

5.3.1 In the native state GluII forms a heterodimer with a variety of glycoforms

First, we wanted to confirm the assembly nature of full-length GluII using native mass spectrometry and evaluate if it is a good technique for studying interactions of that protein with small molecular ligands. A crystal structure of $Mm\text{GluII}_{Tryp}$ (PDB:5F0E) (Figure 5.1) demonstrates that it forms a heterodimer ([145]). The structure was obtained for a protein lacking most of the β-subunit, but full-length GluII tested by MP confirmed its heterodimeric assembly ([148]). Considering GluII is a glycoprotein, which is a challenge for native MS because of its heterogeneity and decreased hydrophobicity, high source temperature was used to transmit higher m/z species. A resulting mass spectrum shows the presence of three different species: GluII subunit β, GluII subunit α, and a heterodimer consisting of the two (see Figure 5.4).

The most abundant charge series in the mass spectrum represents subunit α of GluII. It occurs in two charge series, which we assume originate from two different glycoforms, the first carrying one glycan and the second with two occupied glycosylation sites (Figure 5.4, bottom middle panel). That explanation is in agreement with glycomics analysis ([148]), which has confirmed two glycosylation sites of GluIIα: N75 and either T205 or T215, with the first being more occupied than the second. N75 is also presented as a glycosylation site on the crystal structure (PDB: 5HJO) and was identified as $Man_2GlcNAc_2$ (with functional relevance), which matches with the MS data, showing the detection of 108,900 Da species (theoretical mass of GluIIα carrying $Man_2GlcNAc_2$ is 108,809 Da). The second, low-intensity species presumably corresponds to a less occupied glycosylation at site T205/T215,

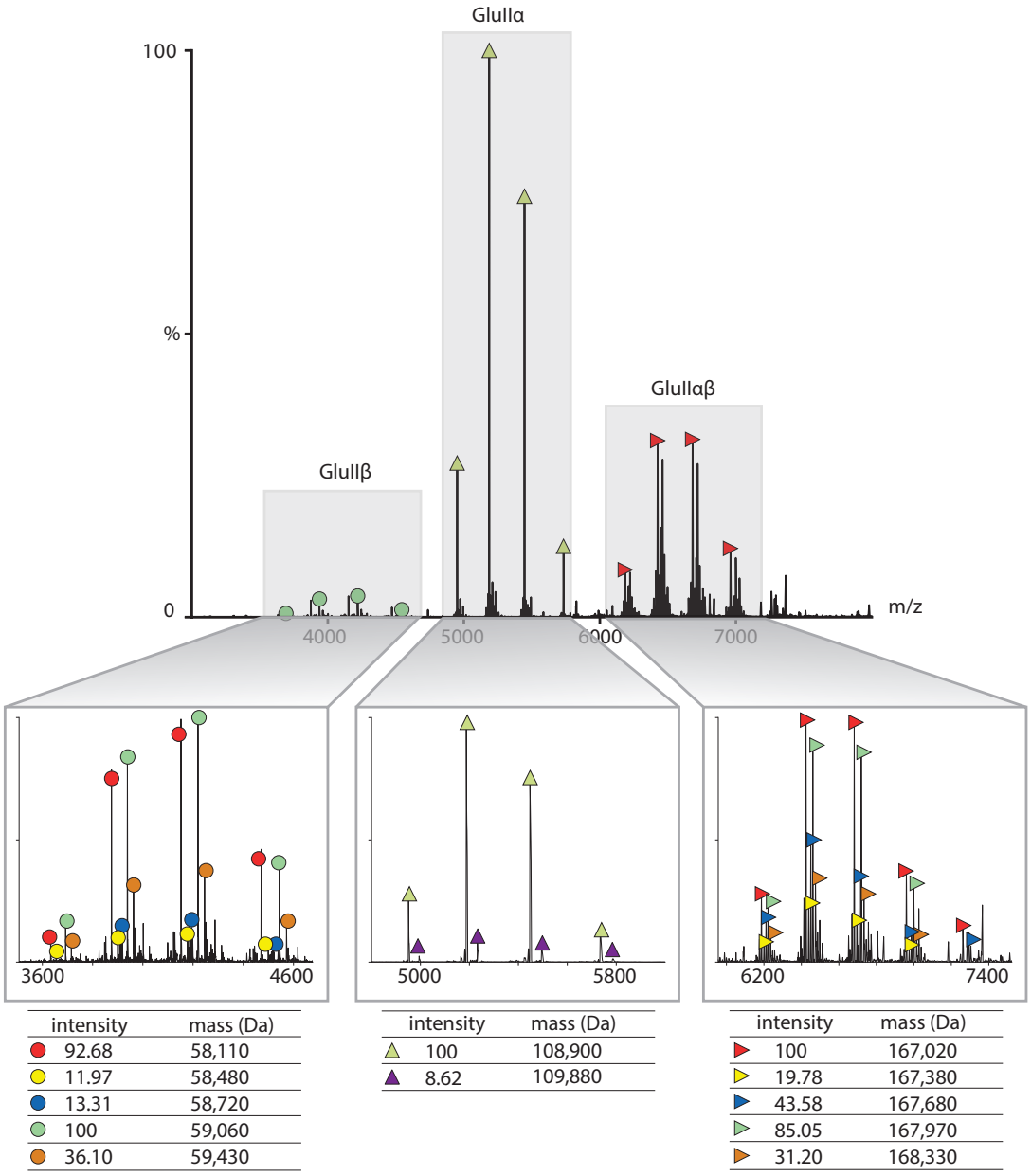


Figure 5.4: Native mass spectrum of *hMm*GluII. Three oligomeric states are visible in the mass spectrum: low-intensity GluIIβ in the lower m/z region, the most intense charge series representing GluIIα spanning through m/z between 5,000-6,000, and a heterodimer consisting of the two with peaks localised between 6,000-7,000 m/z. The bottom panels present close-ups of each of these regions. High heterogeneity of subunit-β originates from glycosylation and proteolysis, as detected masses are lower and higher than the predicted mass based on the amino acid sequence - 58,574 Da. GluIIα occurs in two glycoforms, predominantly carrying N-linked $Man_2GlcNAc_2$ at residue N75 and less occupied O-linked $GalNAc_1NeuAc_2$ at T205/T215. The heterogeneity of heterodimer can be attributed to GluIIβ. It was confirmed by summing up masses of the more intense GluIIα with each of the listed masses of GluIIβ (bottom right), which gives rise to each dimer mass detected (labelled with the same colour).

indicated before as an O-glycosylated $GalNAc_1NeuAc_2$ (with a theoretical mass of 109,750 Da), which agrees with identified species assigned to a mass of 109,880 Da.

The β-subunit located over lower m/z exhibits much higher heterogeneity (Figure 5.4, bottom left). The glycomic analysis identified only one glycosylation site for that subunit, but with a bigger diversity of glycans occupying it ([148]). However, that is not the only explanation for such a broad range of detected species. The extracted masses assigned to each charge series are both lower and higher than the theoretical mass of 58,574 Da predicted based only on the amino acid sequence of GluIIβ (see Figure 5.4 bottom left). This observation could be explained by proteolysis of that subunit, which was also previously suggested ([148]), and could also rationalise its low intensity and the fact that crystallisation of full-length GluII remains elusive.

The high heterogeneity of heterodimer stems from the heterogeneity of GluIIβ (Figure 5.4, bottom right). It was confirmed by summing up various experimentally obtained masses of β-subunit with the more intense glycoform of GluIIα, which perfectly matches detected heterodimer masses (Figure 5.4, bottom).

5.3.2 Covalent interactions of GluIIα with thiopyridone

After verifying with native MS experiments that GluII forms heterodimer in the native state, additionally, we ensured we could use that technique for further experiments to determine the character of interaction between the α-subunit of GluII and a potential antiviral drug (C61) - 190 Da thiopyridone ligand (Figure 5.2). Too high heterogeneity of GluIIα subunit, or its poor transmission in the mass spectrometer, would disqualify native MS as a suitable approach. Monomeric GluIIα carrying one glycan, however, is the most intense species in the mass spectrum, and further, we focus on the m/z region corresponding to that subunit.

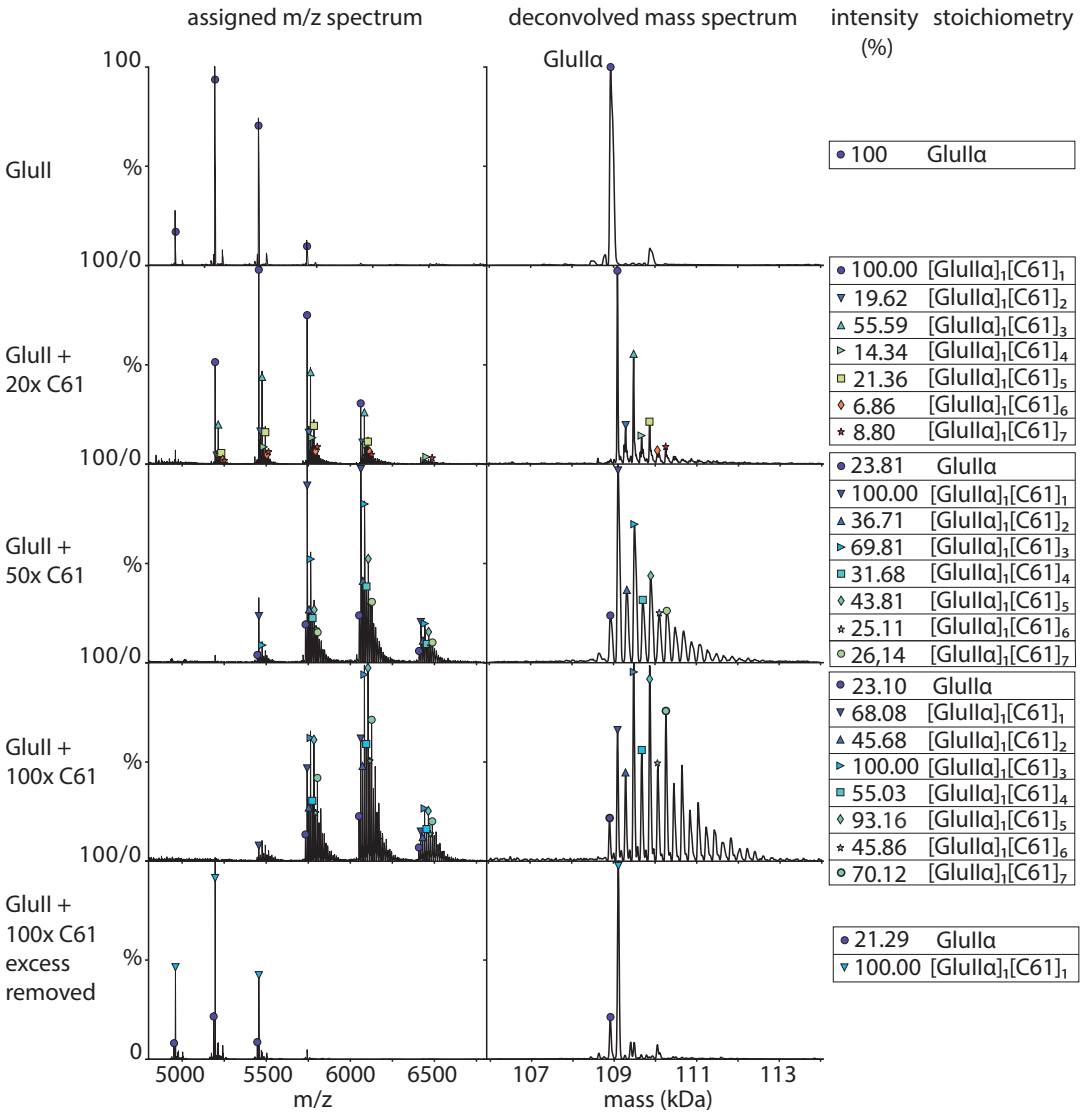


Figure 5.5: Native MS of C61-treated GluII in a range of different concentrations. Close-ups to GluIIα m/z region of native MS spectra of untreated and treated GluII(left), deconvolved mass spectra (middle), and assigned complex stoichiometries of the eight most abundant species (right). The top row represents untreated GluIIα, which shows unbound GluIIα, as expected. The next three data sets correspond to the treated sample with increasing molar excess of the ligand (20, 50, and 100), for all of which numerous complexes were detected, with varying numbers of C61 subunits. An interesting pattern was observed of stoichiometries with odd numbers of C61 being more intense than with even, suggesting dimerisation of the ligand. The last row illustrates data of treated GluII, for which the excess of C61 was removed upon 20 minutes of incubation, showing only one subunit of the ligand interacting with GluIIα.

5. The principles of Endoplasmatic Reticulum α-Glucosidase II inhibition by thiopyridone as an host-targeting antiviral strategy

The strategy of C61 activity relies on disrupting the interface of the heterodimer by binding to GluIIα and, as a consequence, causing dissociation. We treated GluII with C61 in a range of ligand concentrations and with varying incubation times (Figure 5.5).

Numerous complexes consisting of GluIIα and C61 with various stoichiometries were detected, with their distributions depending on C61 molar excess and incubation time (see Figure 5.5). For GluII with a 20-fold excess of the ligands, incubated for 20 minutes, GluIIα_1C61$_1$ is the most abundant complex, followed by six more complexes with up to six more C61-subunits. In the sample with 50-fold excess incubated for a shorter time, the ratio of unreacted GluIIα is higher. The most intense is again the GluIIα_1C61$_1$ species, but it is accompanied by more of the higher-mass complexes, with a greater number of the ligands interacting with the α-subunit. GluII treated with a 100-fold molar excess of C61 immediately after ligand addition resulted in the highest peak corresponding to GluIIα_1C61$_3$ species.

Interestingly, an unexpected pattern was observed across all samples; complexes including odd ligand numbers being more abundant than even numbers. We suspect that it is caused by the C61 dimerisation, whereupon it might non-covalently interact with the protein, which is interrogated later in this chapter.

The fact that we can see an increasing number of ligands bound to GluIIα with its increasing concentration made us argue that those interactions are non-specific. To confirm that, we carried out two experiments. In the first, we removed the excess of the drug after incubation by buffer-exchanging the sample, which led to the detection of GluIIα and C61 complex in 1:1 stoichiometry with both native and intact mass MS (Fig. 5.5, bottom and Fig. 5.8, third panel). The second experiment consisted of applying 180 eV in the HCD cell using the sample with a 100-fold ligand excess (Fig. 5.6), which caused depletion of all the complexes

besides the GluIIα_1C61$_1$ (see Figure 5.6, first two panels). Altogether, it proves that one subunit of C61 binds to GluIIα in a different, much stronger manner than all the other non-covalently interacting molecules, which are most probably representations of non-specific interactions.

The structure of C61 could allow the formation of a disulfide bond with one of the free cysteines of the GluIIα, presumably a residue located close to the dimerisation

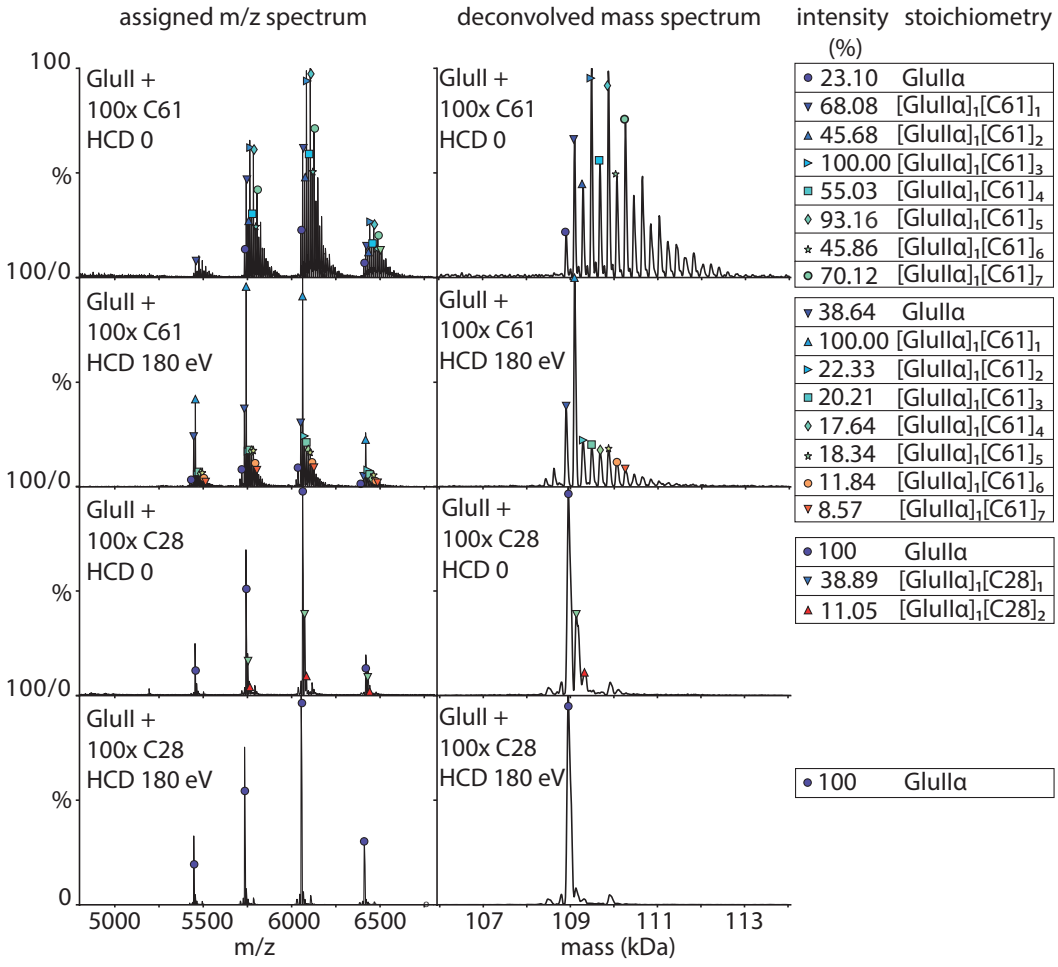


Figure 5.6: Collision-induced dissociation of non-covalent complexes formed upon C61 and C28 treatment of GluII. A 100-fold excess of C61 resulted in the formation of numerous complexes (top row). After subjecting the sample to 180 eV in the HCD, depletion of all complexes with more than one C61-subunit was observed. The same molar excess of C28-ligand resulted in the detection of only three species, with the most abundant being unbound GluIIα, followed by complexes containing one and two C28-subunits, respectively, which dissociate upon applying HCD energy of 180 eV.

interface (see Figure 5.1). The data we have presented so far suggests it could have a covalent character, as it is not disrupted by high collisional energy nor is it disrupted after the excess thiopyridone is removed. To test this, we carried out equivalent measurements with the excess of the ligand and introduced high collisional energy, but in the presence of a different small molecule, which differs from C61 by only one atom, sulfur is substituted with oxygen (called here C28), and thus is unable to form a disulfide bond (see C28 structure in Figure 5.2). After incubation in the presence of a 100-fold excess of C28, we identified unbound GluIIα as the most intense species in the sample, followed by two lower-intensity peaks corresponding to GluIIα_1C28$_1$ and GluIIα_1C28$_2$, which undergo dissociation after being subjected to HCD of 180 eV (see Fig. 5.6, two bottom panels). The substitution of sulfur with oxygen in the ligand structure deprived it of the ability to form a bond with GluIIα, which can survive harsh collisional conditions applied in the mass spectrometer. Additionally, C28 is characterised by a much lower affinity for non-specific interactions, which can be attributed to a higher electronegativity of oxygen in comparison to sulfur, leading to a stronger attraction of π bond electrons to the oxygen, effectively making it less prone to non-covalent interactions. On the other hand, electrons of the C=S bond in C61 are less strongly pulled towards the sulfur atom due to its lower electronegativity. Moreover, sulfur's larger size compared to oxygen allows for better orbital overlap, which could facilitate electron delocalisation.

Tandem MS allowed us to follow the above-described reactions in more detail. Narrow m/z windows were selected to isolate single peaks corresponding to GluIIα_1C28$_1$, GluIIα_1C61$_1$, and GluIIα_1C61$_3$. Mass spectra were recorded for a step-wise increase of the HCD energy (Fig. 5.7).

As suspected, the data demonstrates almost complete dissociation of C28 after reaching 180 eV, in contrast to the isolated GluIIα_1C61$_1$ peak, which remained stable

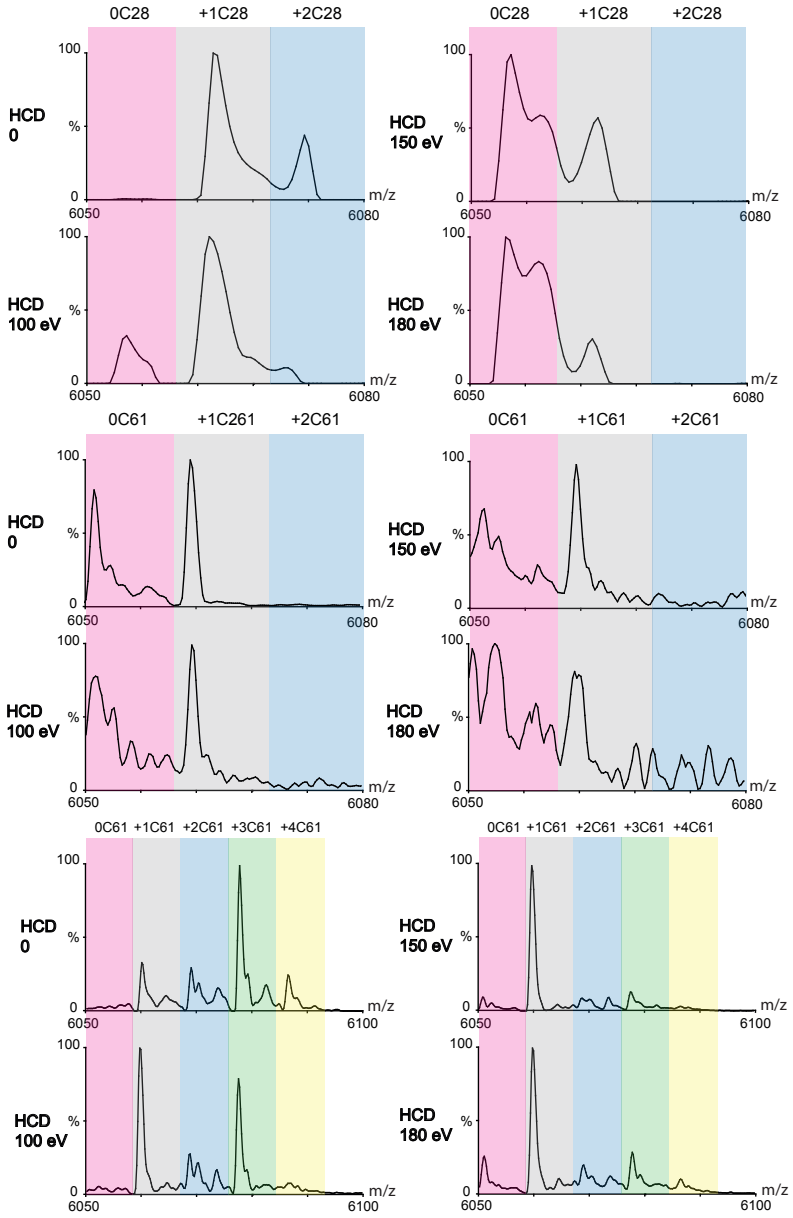


Figure 5.7: Tandem-MS of complexes formed after C61 and C28 treatment of GluII with a step-wise increase of the HCD energy. Isolation of a narrow-range m/z window, targeted to specific peaks of interest, and then monitoring their behavior in the presence of high-collisional energy allowed to illustrate the comparison between GluIIα_1C28$_1$ and GluIIα_1C61$_1$ complexes. The first one loses the ligand subunit with the increase of the energy and the peak shifting towards lower m/z corresponding to the unbound form. On the contrary, the GluIIα_1C61$_1$ peak survives harsh collisional conditions and remains stable throughout the experiment. It means that this complex does not dissociate, thereby suggesting a different character of binding of those two ligands. The bottom panel represents the isolation of a peak corresponding to the GluIIα_1C61$_3$ complex. With an increase in HCD energy, the intensity of that peak decreases, and instead, GluIIα_1C61$_1$ appears, which attests to the loss of dimerised C61.

throughout all incremented energies, which suggests that C61 binds covalently with GluII, whereas C28 does not. Tracking of the GluIIα_1C61$_3$ complex in the presence of increasing HCD energy shows a loss of two C61 subunits, and GluIIα_1C61$_1$ appears as a product of dissociation, without the intermediate of GluII bound with two C61 subunits. This confirms our previous hypothesis, relating to C61 dimerisation and non-covalent interactions of the formed dimers with the target protein (Figure 5.7, bottom).

Finally, we used LC-MS to establish the character of binding of C61 to GluIIα, which introduces denatured conditions before mass spectrometry acquisitions (Figure 5.8). For a sample examined with an excess of the small molecule, a mass corresponding to a complex with three C61 subunits was detected. This can be explained by the exposure of all cysteines during unfolding, and if the ligand is in excess, it can occupy all of them that are free. The WT *hMm*GluIIα possesses three unoccupied cystines, which is in agreement with the LC-MS results (Figure 5.7, second panel).

When the excess of C61 was removed before denaturation, a complex with stoichiometry 1:1 was identified, which disappeared after adding a reducing agent (TCEP), and a species identified as unbound GluIIα appeared instead (Figure 5.7, two bottom panels).

In conclusion, the binding of one subunit of C61 with GluIIα can be maintained throughout buffer exchange, subjection to high collisional energy, and denaturing conditions, but does not survive the addition of TCEP, proving that disulfide bonding is responsible for the formation of that complex.

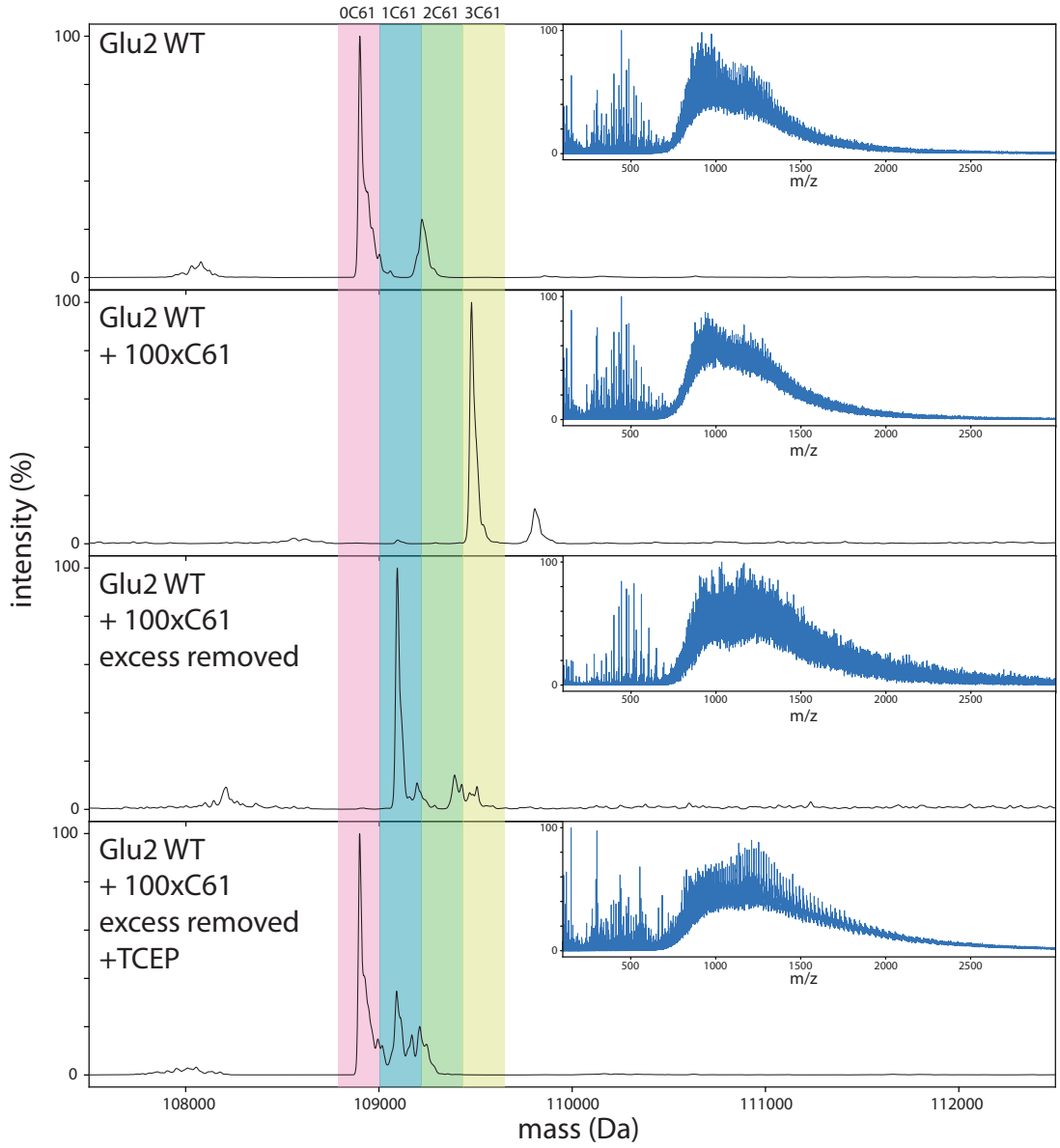


Figure 5.8: Intact mass data for *hMm*GluII. Deconvolved mass spectra of denatured protein samples, with corresponding m/z spectra (in the top right corner). The top panel illustrates untreated GluIIα (with a mass corresponding to no C61-subunits bound. The lower intensity peak coming from the second glycoform is visible. The second spectrum represents a C61-treat GluII sample denatured in the presence of a 100-fold molar excess of the ligand. Extracted mass matches with GluIIα_1C61$_3$, indicating that in denaturing conditions and excess of the ligand, all free cysteines of GluIIα reacted with the C61. When the excess of the ligand was removed prior to protein denaturation, intact mass results show only one subunit of small molecule specifically bound to the target protein (third panel), which was removed after adding a reducing agent (bottom panel), suggesting a crucial role of disulfide binding in the formation of complexes between GluIIα and C61. Presented groups of peaks: unbound GluII (pink), bound to one C61 subunit (blue), two C61 subunits (green) and three C61 subunits (yellow).

5.3.3 Determination of the ligand binding residue

The next aim was to determine which cysteine is involved in the binding characterised in this chapter. GluIIα has seven cysteine residues, four of which are involved in intramolecular disulfide bonding, stabilising protein structure. From the remaining free cysteines, we focus on C822, which was chosen based on a previous screening ([148]) and supported by the fact that according to the crystal structure, it is a part of the dimerisation interface. To verify if replacing cysteine at position 822 causes the protein to lose the ability to form a disulfide bond and, in consequence, make it responsible for the binding of the C61, it was replaced with alanine.

GluIIα_{C822A} was obtained in two different batches, which resulted in purified proteins with what we assume are different glycoforms (see top panel in Fig. 5.9 and Fig. 5.11). We can also observe truncations in samples from both purifications. Whilst it caused an increase in heterogeneity of mass spectra, the mass of the most abundant species matches the expected one and can be distinguished from the other species. Consequently, it is possible to determine interactions of GluIIα_{C822A} with C61.

Native MS measurements demonstrate that incubation of GluIIα_{C822A} with a 100-fold excess of C61 results in the detection of numerous complexes with varying numbers of C61 subunits (Figure 5.9, middle). However, only complexes with 1:1 stoichiometry can survive applying HCD of 180 eV (Figure 5.9, bottom), as was the case regarding the data presented earlier for the WT protein.

Next, we again carried out tandem MS experiments to ensure a lack of dissociation for two different peaks, both corresponding to (GluIIα_{C822A})$_1$C61$_1$. Monitoring single peaks of interest confirmed that they remain stable after subjection to high collisional energy (Figure 5.10).

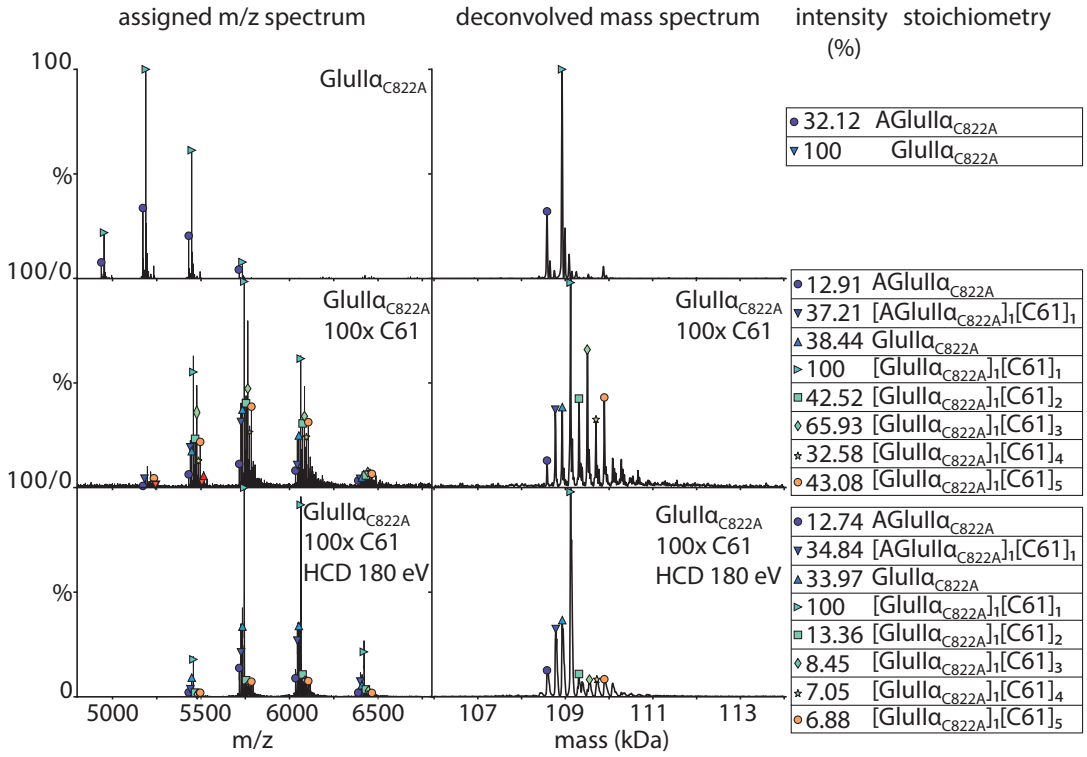


Figure 5.9: Identification of non-covalent complexes of GluIIα_{C822A} with C61. MS data (left), corresponding deconvolved mass spectra (middle), and intensities of assigned stoichiometries (right) of untreated and C61-treated GluIIα_{C822A}. The native mass spectrum of untreated GluIIα_{C822A} shows two species corresponding to the mutant. The less intense charge series was called AGluIIα_{C822A} to allow for further data analysis. The same sample was incubated with a 100-fold excess of C61 (middle row). Multiple complexes were detected, all identified as GluIIα_{C822A}-C61, with an increasing number of C61 subunits. MS spectrum shows an untypical pattern of stoichiometries with odd numbers of the ligand being more intense than with even numbers. Further, subjecting it to 180 eV in the HCD cell (bottom row) caused the depletion of complexes besides (GluIIα_{C822A})$_1$C61$_1$, suggesting a covalent character of binding.

Native mass spectra recorded after removing ligand excess from GluIIα_{C822A} samples reveal that it binds one subunit of C61 (Figure 5.11, bottom). The recorded mass spectrum presents eight charge series which correspond to less-intense unbound and more intense bound GluIIα_{C822A} (in 1:1 stoichiometries) for each of its (glycosylated and truncated) forms, which is analogous to WT GluIIα.

LC-MS was carried out for GluIIα_{C822A} from two different expression batches (Figure 5.12, first and third panel). One of them was subjected to denatured

conditions in the presence of 100-fold ligand excess (second panel). The detected mass corresponds to subunit α bound with two C61 molecules, which is one molecule less than was identified for the WT protein. This is in agreement with expectations, as this variant is deprived of one cysteine and, therefore, has only two free cysteine residues, which after protein unfolding can bind present in the excess ligand. When the sample was buffer-exchanged prior to the intact-mass

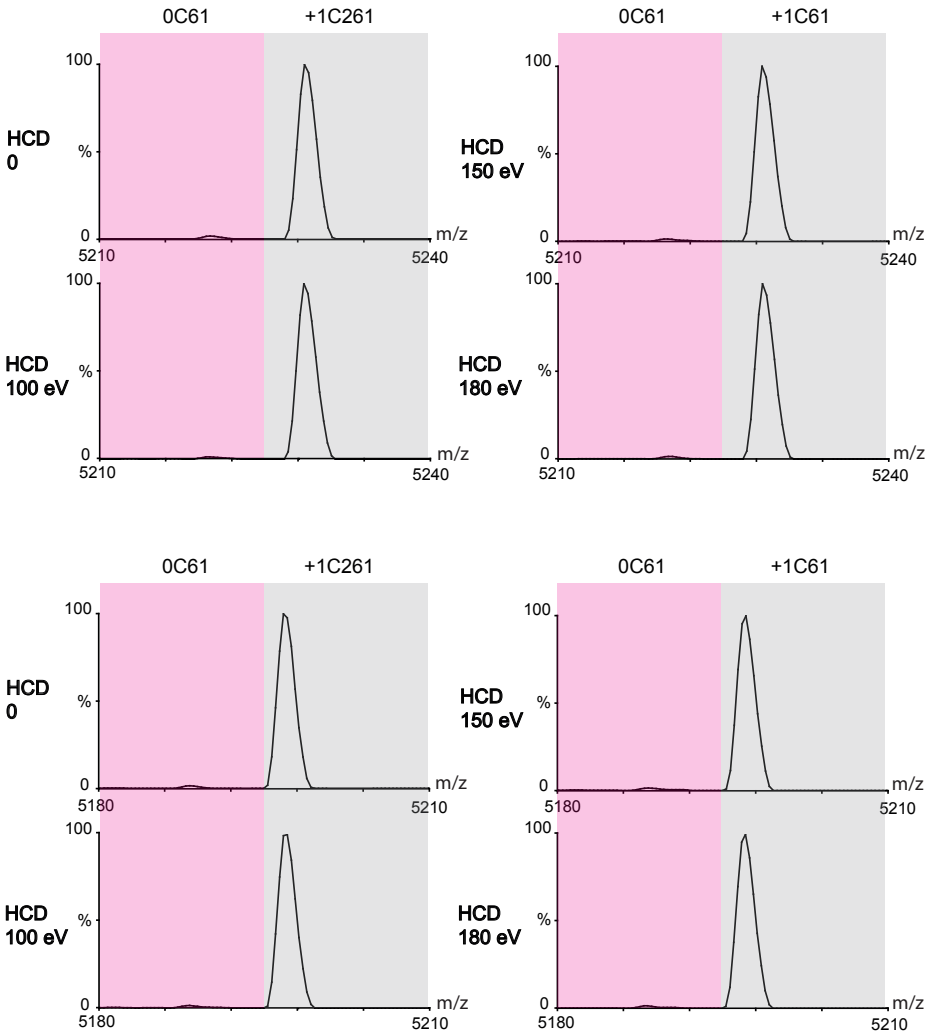


Figure 5.10: Tandem MS of GluIIα_{C822A}C61$_1$ complexes with a step-wise increase of the HCD energy. Monitoring of two different quadrupole-isolated peaks, both corresponding to the (GluIIα_{C822A})$_1$C61$_1$ complex, shows that both remain intact throughout the experiment, which confirms the lack of dissociation and points towards covalent bonding being responsible for the formation of those complexes.

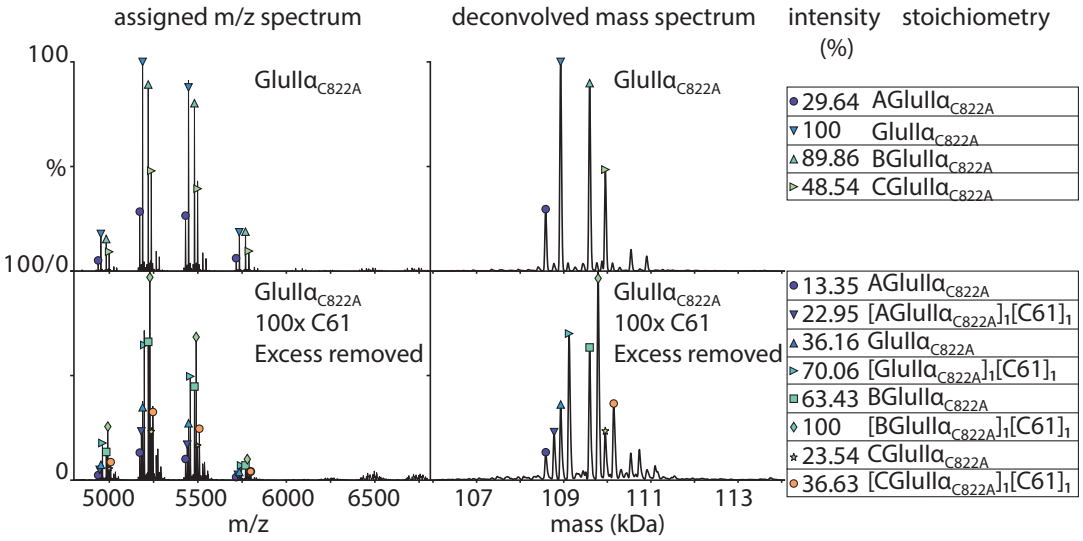


Figure 5.11: Native MS of untreated and C61-treated GluIIα_{C822A} after removing the ligand excess. The figure presents m/z spectra (left), deconvolved data (middle), and intensities of assigned species (right). GluIIα_{C822A}, purified in a second batch (top row), resulted in a much different peak distribution, which can be explained by proteolysis and multiple glycoforms of GluIIα_{C822A}. The most intense charge series is the closest to a predicted mass. To simplify data analysis of the treated sample, we called all the other species AGluIIα_{C822A}, BGluIIα_{C822A}, and CGluIIα_{C822A}. After incubation of the protein sample with a 100-fold molar excess of C61 and removing the excess before the native MS measurement by four cycles of buffer exchange, complexes containing only one C61 subunit were detected for all forms of GluIIα_{C822A} (including A, B, and C).

experiments (GluIIα_{C822A})$_1$C61$_1$ was identified, as with the WT GluII.

This data reveals that GluIIα_{C822A} also forms a covalent bond with C61, suggesting that either C822 is not responsible for the binding of that ligand or that in its absence C61 can be bound to the neighbouring cysteine in position C423, which also lies on α/β interface and can cause its disruption. To verify this possibility a double mutant with substituted cysteine residues in both positions 822 and 423 should be examined. The results presented here are inconclusive, and we can not unambiguously state which cysteine is responsible for the covalent binding of the ligand in question.

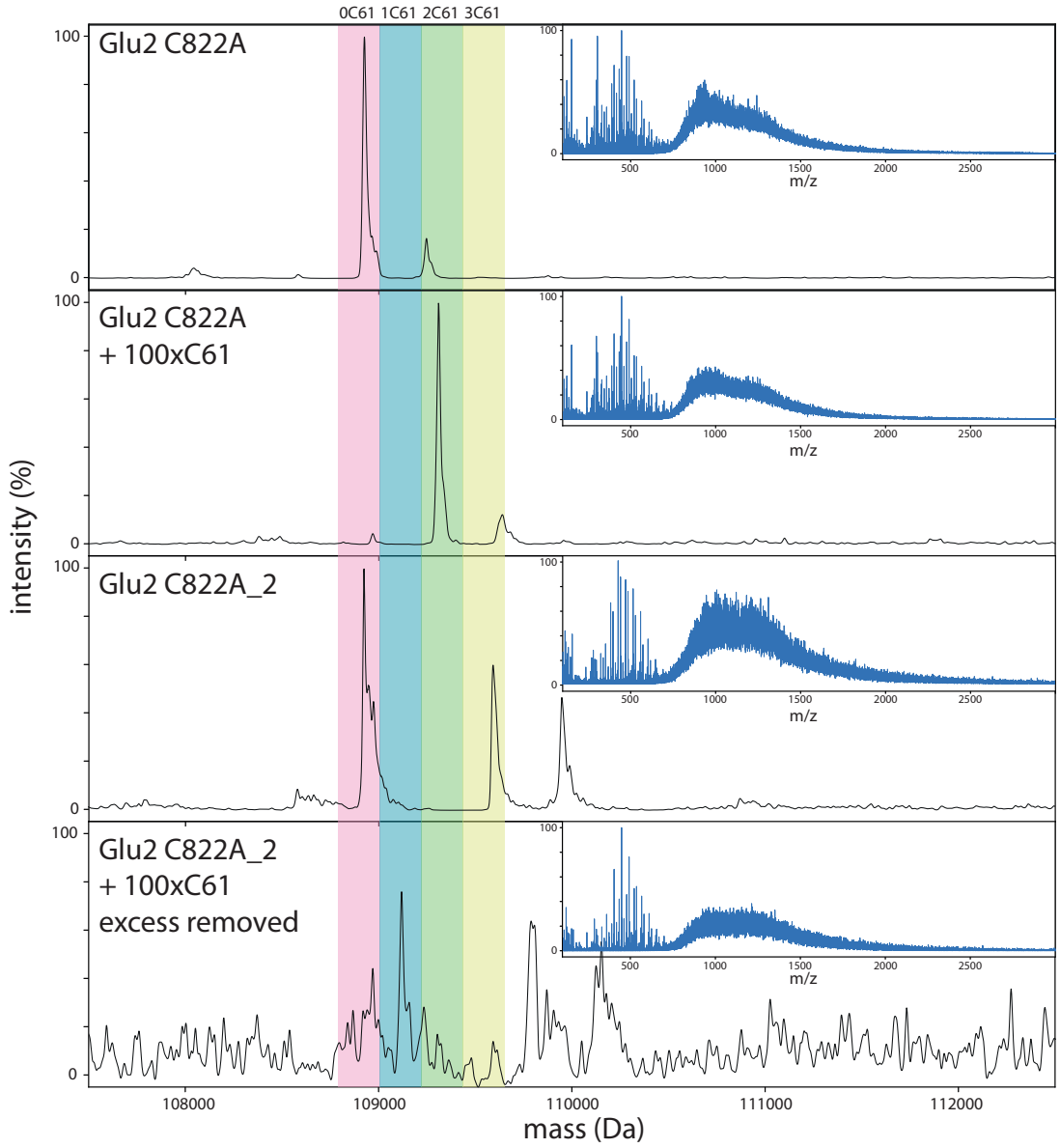


Figure 5.12: LC-MS data of untreated and C61-treated GluII$_{C822A}$. Deconvolved mass spectra of denatured protein samples, with corresponding m/z spectrum (in the right-top corner). The top panel illustrates untreated GluIIα_{C822A} purified in the first batch (with a mass corresponding to no C61 subunits bound), and an additional lower intensity peak coming from the second glycoform. The second spectrum represents the C61-treated GluIIα_{C822A} sample denatured in the presence of a 100-fold molar excess of the ligand. Extracted mass matches with (GluIIα_{C822A})$_1$C61$_2$, indicating that in denatured conditions and ligand excess, all the available cysteines of GluIIα_{C822A} bind the ligand. The third data set shows again untreated GluIIα_{C822A}, purified in the second batch, which resulted in a more heterogeneous spectrum, presumably coming from a greater variety of glycoforms. The bottom spectrum illustrates results after C61 treatment and removing the excess of the ligand before measurements. Complexes corresponding to (GluIIα_{C822A})$_1$C61$_1$ were identified, suggesting the specific binding of one C61 subunit by a covalent bond in native conditions. Presented groups of peaks: unbound GluII (pink), bound to one C61 subunit (blue), two C61 subunits (green) and three C61 subunits (yellow).

5.4 Discussion

5.4.1 Disulfide binding of GluII by thiopyridone

The goal of this project is to develop a new strategy, which could serve as host-targeted antivirals. Our scope was to determine the character of binding between a small molecular ligand (C61) and α-subunit of GluII, shown by our collaborators to cause its dissociation and hence inhibition.

We have confirmed that full-length GluII forms a heterodimer in native conditions, which was previously illustrated by MP data and a crystal structure of trypsin-digested GluII. Native MS data revealed high heterogeneity of that protein, coming from a variety of glycosylations and indicating the proteolysis that causes protein truncations.

Then, we have demonstrated the high affinity of the C61 ligand to its target when it is present in high concentration. The data suggest that C61 also undergoes dimerisation and can interact non-covalently with a target protein in both monomeric and dimeric states. Further, we proved that only one subunit of the drug binds in a much stronger manner. The fact that this binding survives subjection to high collisional energy and denaturing conditions suggests its covalent character. Substitution of sulfur with oxygen in the ligand caused a loss of the ability to form these bonds between GluIIα and the ligand (C28). Additionally, MP data show a lack of dissociation of GluII in the presence of C28 ([148]), which further supports our observation and implies the significance of sulfur in C61-GluIIα binding. The investigated interaction in question could be maintained in the presence of denaturing conditions, however, it could not after introducing TCEP. The data proves that disulfide bonding of the C61-ligand to a cysteine residue of GluIIα causes

GluII dimerisation interface disruption and, as a consequence, GluII dissociation and loss of activity, which could potentially influence the efficiency of viral biosynthesis.

5.4.2 GluII-thiopyridone interactions led by dock, lock and latch molecular mechanism

Supported by MP data, monitoring dissociation of different GluII mutants in the presence of C61 (acquired by Dr Fergus Bremner), and by MS-based studies carried out by Evotec ([148]) delivered enough evidence to put cysteine in position 822 as the main suspect, responsible for the formation of binding described above. To verify this hypothesis, we examined the C822A mutant in equivalent experiments.

If substituting that cysteine would cause an observable effect in covalent bond formation, it would mean that it is relevant for that interaction. However, the result shows the opposite. The mutant deprived of the cysteine, which is the closest to the dimerisation interface, is still able to form a covalent bond with the ligand. It means that either cysteine 822 does not take part in C61 binding or that when it is removed, the small molecule can bind to the closely neighbouring free cysteine (in position 423) with the same consequence.

This hypothesis is in agreement with the MP data, which reveals the activity of C61 against C822A mutant, but with lower potency and with MS data obtained by Evotec, which confirmed the presence of C822-including peptide bound to C61 ([148]). Furthermore, the activity assays conducted by Dr Fergus Bremner resulted in incoherent outcomes, which suggest inhibition of GluII by C61 *in vitro* but also imply the activity of C61 towards GluI. This is yet another evidence of C61 binding not in a specific manner, testifying for the above-proposed mechanism of ligand binding mode, but also raising a concern of the possibility of binding to other enzymes. These outcomes mean that, based on the presented data, we were not

able to clearly verify which cysteine is responsible for interacting with a potential antiviral drug (C61). In order to establish which exact residue is responsible, a double mutant with both cysteines 822 and 423 substituted should be examined.

The presented data are in alignment with a dock, lock, and latch molecular mechanism, whereby the C61 ligand gets to the interfacial pocket and covalently binds available cysteine through a disulfide bond, which induces structural rearrangement in the form of interface disruption, followed by dissociation and consequently lose of the GluII activity.

Moreover, the activity of thiopyridone C61 was tested in live cells. The nature of cytosolic glycans released from the ER of Huh7 cell lines was analysed. It determines the inhibition of GluI and GluII activity by the detection of triglucosylated and di- or monoglucosylated glycans, respectively. No activity of C61 was detected in the described assay([148]). The lack of GluII inhibition in live cells, possibly due to the inability to cross the cell or ER membrane, suggests the need to consider different delivery techniques, which would allow the ligand to pursue the above proposed binding mechanism.

5.5 Acknowledgments

This project is run in collaboration with the Zitzmann group (Department of Biochemistry, University of Oxford). Fergus Bremner introduced me to this work and provided all the samples for the experiments described here. His help has been invaluable, especially at the beginning when I joined the project, in familiarising me with this system and with the project in all its details. I also want to thank Prof. Weston Struwe for guidance and advice on glycoprotein work and Sean Burnap, a postdoctoral researcher in his group, who provided the glycomics data. Last but not

least, I am very grateful to Konstantin Zouboulis, who conducted all the LC-MS experiments and contributed significantly to this project.

5.6 Methods

5.6.1 nESI-MS of untreated and treated GluII

GluII samples (purified by Fergus Bremner from Zitzmann lab) were buffer-exchanged into 200-500 mM ammonium acetate (pH 6.8) using Bio-Spin®(Bio-Rad) centrifuge columns by performing two spinning cycles. The protein concentration was determined by measuring UV-absorbance (NanoDrop spectrophotometer), followed by dilution to around 10 µM concentration suitable for native MS. For treated samples, an appropriate C61 or C28 concentration was added and equilibrated at room temperature (equilibration times varied and are specified in the main text). For samples of treated GluII, with removed excess of the potential drug upon incubation, an additional 2-4 Bio-Spin cycles were performed. All native MS measurements were acquired using a Q Exactive UHMR mass spectrometer (ThermoFisher), with most parameters set as previously described (Chapter 2.6.4). The capillary temperature needed to be increased significantly (250°C) to achieve satisfactory transmission of the heterodimer, but for most experiments involving only the α-subunit, it was set to 50°C, which is closer to native conditions. The instrument was operated at a resolution of 12,500 (at 200 m/z), with a scan range set to 1,000-12,000 m/z. Some experimental parameters, such as capillary voltage and desolvation voltage were optimised on a sample-by-sample basis and ranged from 1.1 to 1.6 kV, and from -10 to -70 V, respectively. MS data were analysed using UniDec software [134]. For activation experiments, which aimed to induce dissociation of any non-covalent interactions, up to 180 eV was applied in the HCD cell. In tandem MS measurements first, a quadrupole isolation of a single peak of interest (as described in the main text) was conducted, whereupon a step-wise increase of energy was applied.

5.6.2 Characterisation of the ligand binding by LC-MS

An Agilent 1100 HPLC system (Agilent Technologies) was used for reverse-phase chromatography performed prior to MS measurements to limit the amount of salt adducts. Protein samples were diluted to around 1 mg/mL in 0.1 % formic acid in LC-MS grade water. Each time, a total sample volume of 50 µL was injected onto a 2.1 mm x 12.5 mm Zorbax 5um 300SB-C3 guard column. Solvent A (0.1 % formic acid in ultra-high purity water), and solvent B (0.1% formic acid in LC-MS grade methanol) were used in the liquid chromatography (LC) step and were run at 10% and 90%, respectively, at 1 mL/min flow rate, for 15 seconds, after which two-step linear gradient was applied. First, from 10% to 80% of solvent B reached within 45 seconds, and then further to 95% of solvent B over 3 seconds. It followed by an elution of 95% solvent B for 1 minute 12 seconds, and lastly 45-second equilibration at 90% B and 10%A. 1969 MSD-ToF electrospray ionisation orthogonal time-of-flight mass spectrometer (Agilent Technologies) was used for MS measurements. To obtain intact mass for tested samples the instrument was operated with a standard ESI source in positive ion mode and capillary voltage set to 4000 V. Ion optics voltages were optimised to maximise the transmission of intact denatured protein ions without prior fragmentation and were set to fragmentor 250 V, skimmer 60 V, and octopole RF 250 V. Other parameters were set as follows: nebulizer pressure 60 psig, drying gas 350°C, and drying gas flow rate 12 L/min.

6
Concluding remarks

The work presented in this thesis describes the *in vitro* interactions between various components of the protein homeostasis network. Focusing on the chaperone and ERQC machinery, we have gained insights into their structures, assemblies, and the evolution of their interfaces; additionally, we explored their pharmacological potential.

In Chapter 2, we characterised the biophysical properties of organelle sHsps from *A. thaliana*. The presented results demonstrate that they display a canonical for sHsps holdase activity. Despite reaching similar levels of chaperone activity toward a model substrate as CI sHsps„ the structure between cytosolic and organelle sHsps varies. Besides having uniquely long NTR, they also lack β6 strand, which is significant in dimerization, and show changes in the CTR regions. Those differences resulted in the quaternary structure diversification of the organelle sHsps. The most surprising finding was the gaining of polydispersity by both MTI members, as all previously studied plant sHsps are reported to exhibit monodispersity. Those alterations in the structure may be connected with adapting to novel functions in distinct sub-cellular locations. Regardless of changes in the secondary structure,

the self-assembly of MTII and CP sHsps is highly similar to CI and CII members, from which they evolved. However, we also detected other structural dissimilarities between mitochondrial and cytosolic sHsps. Based on the data previously reported for the latter, it was thought that the NTR is folded within the inner cavity of oligomeric structures. Here, with MS/MS experiments, we provide evidence for a dual architecture of NTR within the higher-ordered assemblies, where half of the monomers are oriented with their NTRs inside the oligomer, and the remaining monomers position their NTR on the exterior of the structure.

Next, we examined the ability to co-assemble between those proteins. We demonstrated that sHsps from different evolutionary classes, from the same and different membrane-bound cellular organelles, can co-assemble. Maintaining interactions between proteins physically separated for hundreds of millions of years indicates a high conservation of oligomerization interfaces of those sHsps, resulting in their compatibility. By utilizing a developed statistical-mechanical model, we determined that analysed pairs exhibit a range from a very slight to strong preference toward homo-interaction, revealing some degree of decrease in interface compatibility but not to the level that it would prevent their co-assembly. Moreover, native MS results detected that co-assembly between MTI&MTII induces structural changes of the latter, which might impact their activity *in vivo*.

To explore interactions of organelle-localised sHsps with even more distant relatives, we tested co-assembly between organelle and cytosolic sHsps. Interestingly, strongly favoured stoichiometries with even numbers of each sHsp subunit were observed. This proves a preference for a homo-dimer-based subunit exchange due to the incompatibility of dimerization interfaces between CI and MTI, MTII and CP members. It is caused by differences in the ACD domain, in particular, the lack of a $\beta 6$ strand in organelle sHsps, which is the reason for the dimerization

interface incompatibility and hence the formation of hetero-oligomer with mainly even stoichiometries. This points towards the possibility of novel roles that those organelle sHsps may have evolved in distant cellular compartments.

Results presented in these chapters illustrate a broad diversity of sHsps structures, self- and co-assembly preferences, and the exchange of subunits across various kinetic ranges. These preferences vary in each case and seem to depend on the evolutionary distance and compatibility of specific sequence regions. For example, HSP21, after incubation at a higher temperature, interacts in an almost unbiased manner with HSP26.5 but exhibits a strong preference for self-assembly in the presence of polydisperse HSP23.6 even after equilibration for 24 hours at 37°C. When mixed with CI members, it exchanges only dimeric subunits, even at room temperature. On the other hand, the mentioned HSP26.5 willingly exchanges subunits with polydisperse HSP23.6, forming hetero-oligomeric sizes that it does not create individually and interacts with CI members in the same way as HSP21, but reaches equilibrium much faster. This indicates that a direct correlation between kinetics, assembly preference, and dispersity was not observed and hints that the compatibility of specific sequence regions is more crucial for kinetics and assembly.

The last results chapter explores a key enzyme of ERQC, GluII, as a potential target of HTA. We aimed to determine the mode of interaction between GluII and a thiopyridone ligand C61, found in virtual screening and shown to provoke GluII dissociation upon treatment. The dimerization interface disruption was confirmed in literature and by our collaborators to inhibit GluII activity significantly. Utilizing a combination of MS-based approaches, we first determined the high affinity of the ligand to a GluIIα, which further binds covalently with this subunit. The interaction in question was further analysed and identified as disulfide bonding between the sulphur of C61 and one of the cysteines close to the dimerization interface. Next,

we intended to determine the residue interacting with the thiopyridone, but the obtained data are bilateral. They suggest a lack of specificity of the potential HTA, which shows the highest potency toward the closest cysteine to the dimerization interface (C822); however, in its absence, it interacts with a neighbouring C423. We propose a dock, lock, and latch molecular mechanism for the GluII-C61 interaction, in which the ligand, guided by the high affinity to the target protein, reaches the binding pocket, where it binds to the available cysteine by a disulfide bond, inducing structural rearrangements that lead to GluII dissociation.

This thesis furthered the understanding of significant components of the proteostasis network. The use of mass spectrometry techniques enabled us to examine heterogeneous mixtures while gaining detailed insights on a sub-stoichiometric level.

Native MS as a technique for structural investigation stands out due to its capacity for analyzing highly heterogeneous samples, providing in-depth knowledge through ultra-high mass resolution. It offers insights at the species-to-subspecies level rather than yielding average values for the entire system, unlike most biophysical and structural techniques (such as CD).

Native MS preserves the non-covalent interactions within complexes, allowing for the analysis of intact assemblies without the need for crystallization or extensive sample preparation. Moreover, native MS's unique feature, which allows the study of highly heterogeneous samples, provides insights into transient and low-abundance complexes. These complexes can be isolated in the gas phase and further manipulated, offering additional information about species that are difficult to capture with other methods. This technique enables the identification of various structural classes within the sample, elucidation of stoichiometries, detection of coexisting species, and real-time monitoring of reaction kinetics and dynamic systems. However, in comparison to X-ray crystallography and cryo-electron microscopy,

it does not provide information at the residue level, nor does it allow for the determination of the precise structures of the proteins under study.

In this instance, the approach facilitated a comprehensive examination of the composition of heterogeneous complexes composed of subunits with very similar mass.

References

[1] Brandon Ho, Anastasia Baryshnikova, and Grant W. Brown. "Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome". In: *Cell Systems* 6.2 (2018), pp. 192–205.

[2] Xiao Qiong Chen et al. "Protein homeostasis in aging and cancer". In: *Frontiers in Cell and Developmental Biology* 11.1143532 (2023), pp. 1–8.

[3] F. Ulrich Hartl and Manajit Hayer-Hartl. "Converging concepts of protein folding in vitro and in vivo". In: *Nature Structural and Molecular Biology* 16.6 (2009), pp. 574–581.

[4] Thomas R. Jahn and Sheena E. Radford. "The Yin and Yang of protein folding". In: *The FEBS journal* 272.23 (2005), pp. 5962–5970.

[5] Thorsten Hoppe and Ehud Cohen. "Organismal protein homeostasis mechanisms". In: *Genetics* 215.4 (2020), pp. 889–901.

[6] Anna Ainslie et al. "Genome instability and loss of protein homeostasis: Converging paths to neurodegeneration?" In: *Open Biology* 11.4 (2021).

[7] Claudio Hetz and Glimcher Laurie. "Protein homeostasis networks in physiology and disease". In: *Current opinion in cell biology* 23.1 (2012), pp. 123–125.

[8] F. Ulrich Hartl, Andreas Bracher, and Manajit Hayer-Hartl. "Molecular chaperones in protein folding and proteostasis". In: *Nature* 475.7356 (2011), pp. 324–332.

[9] Klaus Richter, Martin Haslbeck, and Johannes Buchner. "The Heat Shock Response: Life on the Verge of Death". In: *Molecular Cell* 40.2 (2010), pp. 253–266.

[10] Mohamed H. Al-Whaibi. "Plant heat-shock proteins: A mini review". In: *Journal of King Saud University - Science* 23.2 (2011), pp. 139–150.

[11] Kevin Reinle, Axel Mogk, and Bernd Bukau. "The Diverse Functions of Small Heat Shock Proteins in the Proteostasis Network: Functions and mechanisms of sHsps". In: *Journal of Molecular Biology* 434.1 (2022), p. 167157.

[12] Justin L.P. Benesch et al. "Small heat shock protein activity is regulated by variable oligomeric substructure". In: *Journal of Biological Chemistry* 283.42 (2008), pp. 28513–28517.

[13] Andrew J. Baldwin et al. "αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics". In: *Journal of Molecular Biology* 413.2 (2011), pp. 297–309.

[14] Georg K.A. Hochberg and Justin L.P. Benesch. "Dynamical structure of αB-crystallin". In: *Progress in Biophysics and Molecular Biology* 115.1 (2014), pp. 11–20.

[15] Garrett J. Lee et al. "A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state". In: *EMBO Journal* 16.3 (1997), pp. 659–671.

[16] Kenneth L. Friedrich et al. "Interactions between Small Heat Shock Protein Subunits and Substrate in Small Heat Shock Protein-Substrate Complexes". In: *Journal of Biological Chemistry* 279.2 (2004), pp. 1080–1089.

[17] Robert M Tanguay and Lawrence E Hightower Editors. *The Big Book on Small Heat Shock Proteins*. Vol. 8. 2015.

[18] Vierling E. 2012 Basha E, O'Neill H. "Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions." In: *Trends in Biochemical Sciences* 37.3 (2012), pp. 106–117.

[19] Eman Basha et al. "Mechanistic Differences between Two Conserved Classes of Small Heat Shock Proteins Found in the Plant Cytosol". In: *The Journal of biological chemistry* 285.15 (2010), pp. 11489–11497.

[20] Hernán Gabriel Bondino, Estela Marta Valle, and Arjen ten Have. "Evolution and functional diversification of the small heat shock protein/α-crystallin family in higher plants". In: *Planta* 235.6 (2012), pp. 1299–1313.

[21] Georg K.A. Hochberg et al. "The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.16 (2014).

[22] Eduardo Hilario et al. "Crystallization and preliminary X-ray diffraction analysis of XAC1151, a small heat-shock protein from Xanthomonas axonopodis pv. citri belonging to the α-crystallin family". In: *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 62.5 (2006), pp. 446–448.

[23] Rob L M Van Montfort et al. "Crystal structure and assembly of a eukaryotic small heat shock protein". In: *Nature Structural Biology* 8.12 (2001), pp. 1025–1030.

[24] Nathalie Braun et al. "Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.51 (2011), pp. 20491–20496.

[25] Alice R. Clark et al. "Terminal Regions Confer Plasticity to the Tetrameric Assembly of Human HspB2 and HspB3". In: *Journal of Molecular Biology* 430.18 (2018), pp. 3297–3310.

[26] Georg K A Hochberg et al. "Protein Paralogs To Evolve Distinct Functions". In: *Science* 359.6378 (2019), pp. 930–935.

[27] Arthur Laganowsky et al. “Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function.” In: *Protein science : a publication of the Protein Society* 19.5 (2010), pp. 1031–1043.

[28] Sanjay Mishra et al. “Engineering of a Polydisperse Small Heat-Shock Protein Reveals Conserved Motifs of Oligomer Plasticity”. In: *Structure* 26.8 (2018), 1116–1126.e4.

[29] Stefan Jehle et al. “N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.16 (2010), pp. 6409–6414.

[30] Ezelle T. McDonald et al. “Sequence, structure, and dynamic determinants of Hsp27 (HspB1) equilibrium dissociation are encoded by the N-terminal domain”. In: *Biochemistry* 51.6 (2012), pp. 1257–1268.

[31] Amanda F. Clouser et al. “Interplay of disordered and ordered regions of a human small heat shock protein yields an ensemble of ‘quasi-ordered’ states”. In: *eLife* 8 (2019), pp. 1–31.

[32] Annika Strauch et al. “The permanently chaperone-active small heat shock protein Hsp17 from Caenorhabditis elegans exhibits topological separation of its N-terminal regions”. In: *Journal of Biological Chemistry* 299.1 (2023), p. 102753.

[33] Gudrun Rutsdottir et al. “Chaperone-client interactions between Hsp21 and client proteins monitored in solution by small angle X-ray scattering and captured by crosslinking mass spectrometry”. In: *Proteins: Structure, Function and Bioinformatics* 86.1 (2018), pp. 110–123.

[34] Eman Basha, Kenneth L. Friedrich, and Elizabeth Vierling. “The N-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity”. In: *Journal of Biological Chemistry* 281.52 (2006), pp. 39943–39952.

[35] Sophia Ungelenk et al. “Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding”. In: *Nature Communications* 7 (2016), pp. 1–14.

[36] Di Wu et al. “The N terminus of the small heat shock protein HSPB7 drives its polyQ aggregation-suppressing activity”. In: *Journal of Biological Chemistry* 294.25 (2019), pp. 9985–9994.

[37] Yuya Hanazono et al. “Nonequivalence observed for the 16-meric structure of a small heat shock protein, SpHsp16.0, from Schizosaccharomyces pombe”. In: *Structure* 21.2 (2013), pp. 220–228.

[38] Roy A. Quinlan et al. "Changes in the quaternary structure and function of MjHSP16.5 attributable to deletion of the IXI motif and introduction of the substitution, R107g, in the α-crystallin domain". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1617 (2013), pp. 1–11.

[39] J. Andrew Aquilina et al. "Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.19 (2003), pp. 10611–10616.

[40] Christoph J.O. Kaiser et al. "The structure and oxidation of the eye lens chaperone αA-crystallin". In: *Nature Structural and Molecular Biology* 26.12 (2019), pp. 1141–1150.

[41] T. Reid Alderson, Justin L.P. Benesch, and Andrew J. Baldwin. "Proline isomerization in the C-terminal region of HSP27". In: *Cell Stress and Chaperones* 22.4 (2017), pp. 639–651.

[42] Eman Basha et al. "An unusual dimeric small heat shock protein provides insight into the mechanism of this class of chaperones". In: *Journal of Molecular Biology* 425.10 (2013), pp. 1683–1696.

[43] Frank Sobott et al. "Subunit exchange of multimeric protein complexes: Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry". In: *Journal of Biological Chemistry* 277.41 (2002), pp. 38921–38929.

[44] John den Engelsman et al. "The Small Heat-Shock Proteins HSPB2 and HSPB3 Form Well-defined Heterooligomers in a Unique 3 to 1 Subunit Ratio". In: *Journal of Molecular Biology* 393.5 (2009), pp. 1022–1032.

[45] Dominik Saman. "Quantitative description of co-assembly and evolution of small heat-shock proteins". PhD thesis. University of Oxford, 2021.

[46] Evgeny V. Mymrikov et al. "Regulation of small heat-shock proteins by hetero-oligomer formation". In: *Journal of Biological Chemistry* 295.1 (2020), pp. 158–169.

[47] Florian Stengel et al. "Quaternary dynamics and plasticity underlie small heat shock protein chaperone function". In: *Proceedings of the National Academy of Sciences* 107.5 (2010), pp. 2007–2012.

[48] Martin Haslbeck et al. "Some like it hot: The structure and function of small heat-shock proteins". In: *Nature Structural and Molecular Biology* 12.10 (2005), pp. 842–846.

[49] G. J. Lee and E. Vierling. "A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein". In: *Plant Physiology* 122.1 (2000), pp. 189–197.

[50] Masood Siddique et al. "The plant sHSP superfamily: Five new members in Arabidopsis thaliana with unexpected properties". In: *Cell Stress and Chaperones* 13.2 (2008), pp. 183–197.

[51] Mariela R. Escobar, Ivo Feussner, and Estela M. Valle. "Mitochondrial Small Heat Shock Proteins Are Essential for Normal Growth of Arabidopsis thaliana". In: *Frontiers in Plant Science* 12.600426 (2021).

[52] J. Horwitz et al. "Lens a-crystallin: function and structure". In: *Eye* 13 (1999), pp. 403–408.

[53] André Patrick Arrigo. "Human small heat shock proteins: Protein interactomes of homo- and hetero-oligomeric complexes: An update". In: *FEBS Letters* 587.13 (2013), pp. 1959–1969.

[54] Chen Hu et al. "Heat shock proteins: Biological functions, pathological roles, and therapeutic opportunities". In: *MedComm* 3.3 (2022), pp. 1–39.

[55] Joseph Horwitz. "Alpha-crystallin". In: *Experimental Eye Research* 76.2 (2003), pp. 145–153.

[56] Alice R. Clark, Nicolette H. Lubsen, and Christine Slingsby. "sHSP in the eye lens: Crystallin mutations, cataract and proteostasis". In: *International Journal of Biochemistry and Cell Biology* 44.10 (2012), pp. 1687–1697.

[57] Evgeny V. Mymrikov, Alim S. Seit-Nebi, and Nikolai B. Gusev. "Large potentials of small heat shock proteins". In: *Physiological Reviews* 91.4 (2011), pp. 1123–1159.

[58] Nicole Mounier and André-Patrick Arrigo. "Actin cytoskeleton and small heat shock proteins: How do they interact?" In: *Cell Stress and Chaperones* 7.2 (2002), pp. 167–176.

[59] Vaishali Kakkar et al. "Barcoding heat shock proteins to human diseases: Looking beyond the heat shock response". In: *DMM Disease Models and Mechanisms* 7.4 (2014), pp. 421–434.

[60] Christopher Woods et al. "Disordered region encodes α-crystallin chaperone activity toward lens client γD-crystallin". In: *Proceedings of the National Academy of Sciences* 120 (2023), p. 2017.

[61] Dae Heon Kim et al. "Small heat shock protein Hsp17.8 functions as an AKR2A cofactor in the targeting of chloroplast outer membrane proteins in arabidopsis". In: *Plant Physiology* 157.1 (2011), pp. 132–146.

[62] Nadja Wehmeyer et al. "Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation". In: *Plant Physiology* 112.2 (1996), pp. 747–757.

[63] Nadja Wehmeyer and Elizabeth Vierling. "The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance". In: *Plant Physiology* 122.4 (2000), pp. 1099–1108.

[64] Liesbeth Lemmens, Rani Baes, and Eveline Peeters. "Heat shock response in archaea". In: *Emerging Topics in Life Sciences* 2.4 (2018), pp. 581–593.

[65] Elizabeth R. Waters. "The evolution, function, structure, and expression of the plant sHSPs". In: *Journal of Experimental Botany* 64.2 (2013), pp. 391–403.

[66] S. Studer and F. Narberhaus. "Chaperone activity and homo- and hetero-oligomer formation of bacterial small heat shock proteins". In: *Journal of Biological Chemistry* 275.47 (2000), pp. 37212–37218.

[67] Stefanie S. Krajewski et al. "A tricistronic heat shock operon is important for stress tolerance of Pseudomonas putida and conserved in many environmental bacteria". In: *Environmental microbiology* 16.6 (2014), pp. 1835–1853.

[68] Elzbieta Ratajczak, Szymon Zietkiewicz, and Krzysztof Liberek. "Distinct Activities of Escherichia coli Small Heat Shock Proteins IbpA and IbpB Promote Efficient Protein Disaggregation". In: *Journal of Molecular Biology* 386.1 (2009), pp. 178–189.

[69] Martin Haslbeck and Elizabeth Vierling. "A first line of stress defense: Small heat shock proteins and their function in protein homeostasis". In: *Journal of Molecular Biology* 427.7 (2015), pp. 1537–1548.

[70] Kazutaka Araki and Kazuhiro Nagata. "Protein folding and quality control in the ER." In: *Cold Spring Harbor perspectives in biology* 4.8 (2012), pp. 1–25.

[71] Brant M. Webster, Holly K. Gildea, and Andrew Dillin. "Protein homeostasis from the outside in". In: *Nature Cell Biology* 22.8 (2020), pp. 911–912.

[72] Ari Helenius. "How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum". In: *Molecular Biology of the Cell* 5.3 (1994), pp. 253–265.

[73] Toru Shibatani et al. "Proteomic analysis of mammalian oligosaccharyltransferase reveals multiple subcomplexes that contain Sec61, TRAP, and two potential new subunits". In: *Biochemistry* 44.16 (2005), pp. 5982–5992.

[74] Catalina Ruiz-Canada, Daniel J. Kelleher, and Reid Gilmore. "Cotranslational and Posttranslational N-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms". In: *Cell* 136.2 (2009), pp. 272–283.

[75] Tadashi Satoh et al. "Structural basis for two-step glucose trimming by glucosidase II involved in ER glycoprotein quality control". In: *Scientific Reports* 6.20575 (2016), pp. 4–10.

[76] David B. Williams. "Beyond lectins: The calnexin/calreticulin chaperone system of the endoplasmic reticulum". In: *Journal of Cell Science* 119.4 (2006), pp. 615–623.

[77] Lori A. Rutkevich et al. "Functional relationship between protein disulfide isomerase family members during the oxidative folding of human secretory proteins". In: *Molecular Biology of the Cell* 21.18 (2010), pp. 3093–3105.

[78] Gillian R. Hilton and Justin L.P. Benesch. "Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry". In: *Journal of the Royal Society Interface* 9.70 (2012), pp. 801–816.

[79] Shibdas Banerjee and Shyamalava Mazumdar. "Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte". In: *International Journal of Analytical Chemistry* 2012 (2012), pp. 1–40.

[80] Michiel Van De Waterbeemd et al. "Dissecting ribosomal particles throughout the kingdoms of life using advanced hybrid mass spectrometry methods". In: *Nature Communications* 9.1 (2018).

[81] Tobias P. Wörner et al. "Adeno-associated virus capsid assembly is divergent and stochastic". In: *Nature Communications* 12.1 (2021), pp. 1–9.

[82] Tobias P. Wörner et al. "Mass Spectrometry-Based Structural Virology". In: *Analytical Chemistry* 93.1 (2021), pp. 620–640.

[83] Carol V Robinson. "From molecular chaperones to membrane motors: through the lens of a mass spectrometrist". In: *Biochemical Society Transactions* 45 (2017), pp. 251–260.

[84] Renato Zenobi et al. "Advancing Cyclic Ion Mobility Mass Spectrometry Methods for Studying Biomolecules: Toward the Conformational Dynamics of Mega Dalton Protein Aggregates". In: *Analytical Chemistry* 94.36 (2022), pp. 12435–12443.

[85] Justin L.P. Benesch et al. "The quaternary organization and dynamics of the molecular chaperone HSP26 are thermally regulated". In: *Chemistry and Biology* 17.9 (2010), pp. 1008–1017.

[86] Justin L.P. Benesch et al. "Protein complexes in the gas phase: Technology for structural genomics and proteomics". In: *Chemical Reviews* 107.8 (2007), pp. 3544–3567.

[87] Alexander J. Painter et al. "Real-Time Monitoring of Protein Complexes Reveals their Quaternary Organization and Dynamics". In: *Chemistry and Biology* 15.3 (2008), pp. 246–253.

[88] Helena Hernández and Carol V. Robinson. "Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry". In: *Nature Protocols* 2.3 (2007), pp. 715–726.

[89] ThermoScientific. *Exactive Series Operating Manual*. Tech. rep. 2017.

[90] Paul Kebarle and Udo H. Verkerk. "Electrospray: from ions in solution to ions in the gas phase, what we know now". In: *Mass Spectrom Rev* 28.6 (2009), pp. 898–917.

[91] Lars Konermann et al. "Unraveling the Mechanism of Electrospray Ionization". In: *Analytical chemistry* 85.1 (2013), pp. 2–9.

[92] Shahid Mehmood, Timothy M. Allison, and Carol V. Robinson. "Mass Spectrometry of Protein Complexes: From Origins to Applications". In: *Annual Review of Physical Chemistry* 66.1 (2015), pp. 453–474.

[93] J. Fernandez De La Mora. "Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism". In: *Analytica Chimica Acta* 406.1 (2000), pp. 93–104.

[94] Steve Nguyen and John B. Fenn. "Gas-phase ions of solute species from charged droplets of solutions". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.4 (2007), pp. 1111–1117.

[95] Lars Konermann, Antony D. Rodriguez, and Jiangjiang Liu. "On the formation of highly charged gaseous ions from unfolded proteins by electrospray ionization". In: *Analytical Chemistry* 84.15 (2012), pp. 6798–6804.

[96] Igor A. Kaltashov and Anirban Mohimen. "Estimates of protein surface areas in solution by electrospray ionization mass spectrometry". In: *Analytical Chemistry* 77.16 (2005), pp. 5370–5379.

[97] Jinyy Li et al. "Conformational effects in protein electrospray-ionization mass spectrometry". In: *Mass Spectrom Rev* 35.1 (2016), pp. 111–112.

[98] Igor V. Chernushevich, Alexander V. Loboda, and Bruce A. Thomson. "An introduction to quadrupole-time-of-flight mass spectrometry". In: *Journal of Mass Spectrometry* 36.8 (2001), pp. 849–865.

[99] Tim K. Esser et al. "Cryo-EM samples of gas-phase purified protein assemblies using native electrospray ion-beam deposition". In: *Faraday Discussions* 240 (2022), pp. 67–80.

[100] Justin L.P. Benesch. "Collisional Activation of Protein Complexes: Picking Up the Pieces". In: *Journal of the American Society for Mass Spectrometry* 20.3 (2009), pp. 341–348.

[101] Justin L.P. Benesch et al. "Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies". In: *Chemistry and Biology* 13.6 (2006), pp. 597–605.

[102] Norma J. Greenfield. "Using circular dichroism spectra to estimate protein secondary structure". In: *Nature Protocols* 1.6 (2007), pp. 2876–2890.

[103] Lee Whitmore and B. A. Wallace. "Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases". In: *Biopolymers* 89.5 (2007), pp. 392–400.

[104] Sharon M. Kelly, Thomas J. Jess, and Nicholas C. Price. "How to study proteins by circular dichroism". In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1751.2 (2005), pp. 119–139.

[105] Elizabeth R. Waters, Brian D. Aevermann, and Zipporah Sanders-Reed. "Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns". In: *Cell Stress and Chaperones* 13.2 (2008), pp. 127–142.

[106] Klaus Dieter Scharf, Masood Siddique, and Elizabeth Vierling. "The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing α-crystallin domains (Acd proteins)". In: *Cell Stress and Chaperones* 6.3 (2001), pp. 225–237.

[107] Antonio De Maio and Daniel Vazquez. "Extracellular heat shock proteins: a new location, a new function". In: *Physiology & behavior* 176.3 (2017), pp. 139–148.

[108] N. De Miguel, P. C. Echeverria, and S. O. Angel. "Differential subcellular localization of members of the Toxoplasma gondii small heat shock protein family". In: *Eukaryotic Cell* 4.12 (2005), pp. 1990–1997.

[109] Elizabeth R. Waters and Elizabeth Vierling. "Chloroplast small heat shock proteins: Evidence for atypical evolution of an organelle-localized protein". In: *Proceedings of the National Academy of Sciences of the United States of America* 96.25 (1999), pp. 14394–14399.

[110] Elizabeth R. Waters et al. "The recent evolution of a pseudogene: Diversity and divergence of a mitochondria-localized small heat shock protein in Arabidopsis thaliana". In: *Genome* 51.3 (2008), pp. 177–186.

[111] Elizabeth R. Waters and Ignatius Rioflorido. "Evolutionary analysis of the small heat shock proteins in five complete algal genomes". In: *Journal of Molecular Evolution* 65.2 (2007), pp. 162–174.

[112] Linlin Zhong et al. "Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in arabidopsis under heat stress". In: *Plant Cell* 25.8 (2013), pp. 2925–2943.

[113] Elizabeth R. Waters and Elizabeth Vierling. "The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses". In: *Molecular Biology and Evolution* 16.1 (1999), pp. 127–139.

[114] Stefan Rensing et al. "The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants". In: *Science* 319.5859 (2008), pp. 64–69.

[115] Elizabeth R. Waters and Elizabeth Vierling. "Plant small heat shock proteins – evolutionary and functional diversity". In: *New Phytologist* 227.1 (2020), pp. 24–37.

[116] Geneviève Morrow et al. "The small heat shock protein Hsp22 of Drosophila melanogaster is a mitochondrial protein displaying oligomeric organization". In: *Journal of Biological Chemistry* 275.40 (2000), pp. 31204–31210.

[117] Nathalie Banzet et al. "Accumulation of small heat shock proteins, including mitochondrial HSP22, induced by oxidative stress and adaptive response in tomato cells". In: *Plant Journal* 13.4 (1998), pp. 519–527.

[118] Martín D. Ré et al. "Small heat shock proteins and the postharvest chilling tolerance of tomato fruit". In: *Physiologia Plantarum* 159.2 (2017), pp. 148–160.

[119] Inbal Neta-Sharir et al. "Dual role for tomato heat shock protein 21: Protecting photosystem II from oxidative stress and promoting color changes during fruit maturation". In: *Plant Cell* 17.6 (2005), pp. 1829–1838.

[120] Katja Bernfur, Gudrun Rutsdottir, and Cecilia Emanuelsson. "The chloroplast-localized small heat shock protein Hsp21 associates with the thylakoid membranes in heat-stressed plants". In: *Protein Science* 26.9 (2017), pp. 1773–1784.

[121] Mastoureh Sedaghatmehr, Bernd Mueller-Roeber, and Salma Balazadeh. "The plastid metalloprotease FtsH6 and small heat shock protein HSP21 jointly regulate thermomemory in Arabidopsis". In: *Nature Communications* 7 (2016).

[122] Fionn McLoughlin et al. "Class I and II small heat shock proteins together with HSP101 protect protein translation factors during heat stress". In: *Plant Physiology* 172.2 (2016), pp. 1221–1236.

[123] Erik G. Marklund et al. "Correction to: Structural and functional aspects of the interaction partners of the small heat-shock protein in Synechocystis". In: *Cell Stress and Chaperones* 23.4 (2018), pp. 723–732.

[124] Afrooz Dabbaghizadeh et al. "Identification of proteins interacting with the mitochondrial small heat shock protein Hsp22 of Drosophila melanogaster: Implication in mitochondrial homeostasis". In: *PLoS ONE* 13.3 (2018), pp. 1–26.

[125] Mark Rutgers et al. "Substrates of the chloroplast small heat shock proteins 22E/F point to thermolability as a regulative switch for heat acclimation in Chlamydomonas reinhardtii". In: *Plant Molecular Biology* 95.6 (2017), pp. 579–591.

[126] Wei Ma et al. "Mitochondrial small heat shock protein mediates seed germination via thermal sensing". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.10 (2019), pp. 4716–4721.

[127] Gudrun Rutsdottir et al. "Structural model of dodecameric heat-shock protein Hsp21: Flexible N-terminal arms interact with client proteins while C-terminal tails maintain the dodecamer and chaperone activity". In: *Journal of Biological Chemistry* 292.19 (2017), pp. 8103–8121.

[128] G. J. Lee, N. Pokala, and E. Vierling. "Structure and in vitro molecular chaperone activity of cytosolic small heat shock proteins from pea". In: *Journal of Biological Chemistry* 270.18 (1995), pp. 10432–10438.

[129] András Micsonai et al. "Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.24 (2015), E3095–E3103.

[130] Lawrence A Kelley et al. "The Phyre2 web portal for protein modeling, prediction and analysis". In: *Nature Protocols* 10.6 (2016), pp. 845–858.

[131] Chuanyang Yu et al. "Structural basis of substrate recognition and thermal protection by a small heat shock protein". In: *Nature Communications* 12.1 (2021), pp. 1–11.

[132] Indu Santhanagopalan et al. “It takes a dimer to tango: Oligomeric small heat shock proteins dissociate to capture substrate”. In: *Journal of Biological Chemistry* 293.51 (2018), pp. 19511–19521.

[133] Mary Disa M. Raulfs et al. “Investigations of the Mechanism of the “Proline Effect” in Tandem Mass Spectrometry Experiments: The “Pipecolic Acid Effect””. In: *Journal of the American Society for Mass Spectrometry* 25.10 (2014), pp. 1705–1715.

[134] Michael T. Marty et al. “Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles”. In: *Analytical Chemistry* 87.8 (2015), pp. 4370–4376.

[135] Arvind S. Pillai et al. “Origin of complexity in haemoglobin evolution”. In: *Nature* 581.7809 (2020), pp. 480–485.

[136] Fériel Skouri-Panet et al. “Structural and functional specificity of small heat shock protein HspB1 and HspB4, two cellular partners of HspB5: Role of the in vitro hetero-complex formation in chaperone activity”. In: *Biochimie* 94.4 (2012), pp. 975–984.

[137] Donald T. Haynie. *Biological Thermodynamics, Chapter 2 The first law of thermodynamics*. 2001, p. 37.

[138] Elizabeth R. Waters and Elizabeth Vierling. “Plant small heat shock proteins – evolutionary and functional diversity”. In: *New Phytologist* 227.1 (2020), pp. 24–37.

[139] Anders Larsson. “AliView: A fast and lightweight alignment viewer and editor for large datasets”. In: *Bioinformatics* 30.22 (2014), pp. 3276–3278.

[140] Michel Pieren et al. “The use of calnexin and calreticulin by cellular and viral glycoproteins”. In: *Journal of Biological Chemistry* 280.31 (2005), pp. 28265–28271.

[141] Karin Wisskirchen et al. “New pharmacological strategies to fight enveloped viruses”. In: *Trends in Pharmacological Sciences* 35.9 (2014), pp. 470–478.

[142] Andrew C. Sayce, Joanna L. Miller, and Nicole Zitzmann. “Targeting a host process as an antiviral approach against dengue virus”. In: *Trends in Microbiology* 18.7 (2010), pp. 323–330.

[143] J. P. Martinez et al. “Antiviral drug discovery: Broad-spectrum drugs from nature”. In: *Natural Product Reports* 32.1 (2015), pp. 29–48.

[144] Marc F. Pelletier et al. “The heterodimeric structure of glucosidase II is required for its activity, solubility, and localization in vivo”. In: *Glycobiology* 10.8 (2000), pp. 815–827.

[145] Alessandro T. Caputo et al. “Structures of mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals”. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.32 (2016), E4630–E4638.

[146] Ivan Stigliano et al. "Glucosidase II beta Subunit Modulates N-Glycan Trimming in Fission Yeasts and Mammals". In: *Molecular Biology of the Cell* 20 (2009), pp. 3974–3984.

[147] Linda J. Olson et al. "Structure of the lectin mannose 6-phosphate receptor homology (MRH) domain of glucosidase II, an enzyme that regulates glycoprotein folding quality control in the endoplasmic reticulum". In: *Journal of Biological Chemistry* 288.23 (2013), pp. 16460–16475.

[148] Fergus Ross Bremner. "Investigating the Inhibition of Endoplasmic Reticulum α -Glucosidase II for the Development of Broad-Spectrum Antivirals". PhD thesis. University of Oxford, 2023.