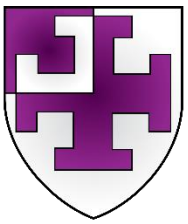


Cardiac small heat shock proteins and their roles in maintaining the integrity of the cytoskeletal system

Zihao Wang

Saint Cross College



University of Oxford

A thesis presented for the degree of

Doctor of Philosophy in Physical and Theoretical Chemistry

Trinity 2021

Abstract

Small heat shock proteins (sHsps) are a family of ubiquitous proteins that defend the first line for proteostasis. They have specific regulatory roles to fulfill their chaperone functions or maintain the physiological structures of tissues through interactions with their native or dysfunctional partners. This thesis will focus on the sHsps and their substrates in the human heart and skeletal muscles, using various biophysical and structural techniques to probe their structures and interactions. Among the sHsps, cardiovascular heat shock protein (cvHsp, HspB7) and HspB8 (Hsp22) perform their primary functions in cardiac tissues and skeletal muscles. Firstly, I find that, unlike the well-studied human sHsps like Hsp27 or αB-crystallin forming large oligomers based on dimeric building blocks, cvHsp is dominantly monomeric at dynamic conditions. I solve the crystal structure of cvHsp to explain that oligomerization and dimerization are much less favored for cvHsp. Then I study the actin-linking protein filamin C (FLNC) and confirm its dimerized domain 24 (d24) is the binding substrates of cvHsp. The monomeric cvHsp binds to the dimerized FLNCd24 and competes to its dimerization. This process can be regulated under stress conditions. Phosphorylation at FLNCd24 T2677 or Y2683 residues at different stress conditions has opposite effects in mediating the FLNC dimerization and interaction with cvHsp, which might be a potential target for mediating cancer development. To study this interaction at the atomic level, I reveal the binding surfaces for the cvHsp-FLNC interaction and obtain the crystal structure of the complex. cvHsp uses its β4 strand to bind FLNCd24 strand C, which is involved in the dimer interface. Some hydrophobic contacts within the interface make this interaction strong and specific. Both T2677 and Y2683 residues show up clearly in the structural analysis. We further research the FLNC clearance system, the chaperone-assisted selective autophagy (CASA) complex, which involves HspB8, BAG3, Hsp70, FLNC, and other components. We have obtained the stable construct of HspB8, and the purification of non-structured BAG3 is still going on. Overall, our studies provide new and deep insight into cardiac sHsps and their roles in maintaining proteostasis.

Statement of Originality

This thesis covers most of my work in the Department of Physical and Theoretical Chemistry, University of Oxford, between October 2017 and July 2021. All the texts, figures and tables in this thesis are generated based on my data and efforts unless being stated. In the acknowledgement section of each chapter, all the assistances of the works are clarified. The work and figures in this thesis have never been submitted for any degree at any institution.

Zihao Wang
August 2021

Acknowledgments

It has been an enjoyable period in my life because of the wonderful people I have met here.

There isn't a better supervisor than Justin Benesch that I can think of, which must be the luckiest thing during my years in Oxford. I am inspired and affected by him as a supervisor and a great person. I have experienced precious academic freedom and great support all at the appropriate moments from him. A conversation with him can always calm me down when uncertainty or pressure show up in my research. His generosity, passion for science and wisdom in life are also influencing me, which will benefit me in the future.

Miranda Collier was like an elder sister, I got tremendous and patient guidance in the lab from her when I first came to the lab, and her suggestions are always valuable. Her leaving after graduation at the end of my first year made me feel a bit stressed for a period, but that also pushed me to a new stage.

Thanks to the old instrument, I spent a lot of time with Xingyu Qiu in the PTCL common room, having crisps and coffee while waiting for our HDX experiments. We talked about a lot of funny things inside and outside the lab. Dominik Saman, I would describe him as my "comrade". We started our PhD on the same day, and we will carry on working together as close colleagues with Justin. Di Wu's suggestions have accompanied me all the time here, in my research and life.

Things become much easier with his experiences. The short periods with Reid Alderson in Oxford helped me a lot, and I could feel his passion for academia. Haiping Tang is not just my colleague but also my housemate for three years. Her working momentum always inspires me. Jingwen Li helped me a lot before I formally joined the lab, which I always appreciate. I'm thankful to Jani Bola, Timothy Allison, Shane Chandler, Michael McDonough and Jurgen Brem. They helped me a lot in structural biology or other areas of my projects. I enjoyed sharing the ideas with Guodong Cao, and I'm excited to see his molecular dynamic results of our structures. Shuang Liu and Xiao Liu shared their thoughts and helped me when things in our lab were unavailable. I am grateful to our collaborators include Katja Gehmlich in Birmingham and Bettina Warscheid in Freiburg. I can't get my data into this exciting story without works from their collaboration. I feel lucky to help supervise Wiktoria Sadowska, Elena Holden and Chloe Ridsdill-Smith in this lab. I hope my experiences can be helpful for them. I also miss the activities with Anna Howes, Fabian Soltermann, Francesco Fiorentino, Daniel Quetschlich, Denis Shutin and Siyun Chen. I hope these people are still able to meet and get a drink together in the future. I wouldn't have had any experiences in a nightclub without Yixuan Hu and Ziwen Xu. It was a funny and warm year with them. I am thankful to Todd Mize, Victoria Cullen and Andrew Dolan for arranging the excellent lab. I appreciate all those above and the other members from Benesch, Robinson, Baldwin, Rauschenbach and Kukura groups. I also feel lucky to be a member of St Cross College. I will miss the lunch and the

football team there.

In the end, I need to thank my family, my parents. I know they are always the ones that I can rely on and trust. I'm far away in the distance from them these years, but I never feel distant from them. I owe my parents too much. For me, they are the best parents in the world. They are my energy resources.

When I was in my childhood, I went to my grandma's place every summer and winter holidays. I realized something could never go back when she passed away in 2019 due to diabetes caused by dilated cardiomyopathy. Within a week, we phased cardiovascular heat shock protein structure after almost two years' efforts. I miss my grandma. I know my research is tiny and basic, and far away from helping cure those complicated diseases, but I still hope all these small steps can be accumulated for a breakthrough. I will appreciate that.

List of abbreviations

αBC	αB-crystallin
ABD	Actin binding domain
ACD	α-crystallin domain
ADP	Adenosine diphosphate
AP	Antiparallel
ATP	Adenosine triphosphate
BAG	BCL-2–associated athanogene
BME	β-Mercaptoethanol
BRIL	Bone-restricted IFITM-like
CASA	Chaperone-assisted selective autophagy
CCS	Collision cross-section
CHIP	C-terminus Hsp70 interacting protein
CID	Collision-induced dissociation
CMT	Charcot-Marie-Tooth
CRM	Charged residue model
CS	Citrate synthase
CTD	C-terminal domain
cryo-EM	Cryogenic electron microscopy
cvHsp	Cardiovascular heat shock protein
Da	Dalton
DC	Direct current
DCM	Dilated cardiomyopathy
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FILIP	Filamin A-binding protein
FL	Full-length
FLN(A/B/C)	Filamin A/B/C
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase

GPCR	G protein-coupled receptor
HDX	Hydrogen-deuterium exchange
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	Heart failure
HPLC	High Performance Liquid Chromatography
HSF	Heat shock factor
Ig	Immunoglobulin
IM	Ion mobility
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISO/PE	Isoprenaline/epinephrine
KO	Knock-out
LB	Lysogeny broth
LC	Liquid chromatography
MALDI	Matrix assisted laser desorption/ionization
MCP	Microchannel plate
MD	Molecular dynamics
MLP	Muscle LIM protein
MR	Molecular replacement
MS	Mass spectrometry
NEF	Nucleotide exchange factor
nESI	Nano-electrospray ionization
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PARK2	Parkinson disease 2
PBS	Phosphate-buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
PKC	Protein kinase C
PTM	Post-translational modification
RF	Radio frequency
RMSD	Root mean square deviation
RPM	Revolutions per minute

RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
sHsp	Small heat-shock protein
SNP	Single nucleotide polymorphism
TAC	Transverse aortic constriction
TB	Terrific broth
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
ToF	Time-of-flight
UPS	Ubiquitin proteasome system
WT	Wild type

Contents

1 **SMALL HEAT SHOCK PROTEINS AND BIOPHYSICAL TECHNIQUES USED IN CHARACTERIZATION 1**

1.1 ABSTRACT 1
1.2 BIOLOGICAL BACKGROUND 2
1.2.1 Cardiac small heat shock proteins in proteostasis 2
1.2.2 The structure of sHsps 4
1.3 TECHNICAL BACKGROUND 7
1.3.1 Native mass spectrometry 8
1.3.1.1 Nano-electrospray ionization 10
1.3.1.2 Ion transmission and analysis 12
1.3.1.3 Collision-induced dissociation and Ion mobility 13
1.3.2 Macromolecular X-ray crystallography 16
1.3.3 Hydrogen-deuterium exchange (HDX) MS 20

2 **CVHSP IS DOMINANTLY MONOMERIC AND UNIQUE IN THE FAMILY 38**

2.1 ABSTRACT 38
2.2 INTRODUCTION 39
2.3 RESULTS 42
2.3.1 cvHsp is dominantly monomeric 42
2.3.2 The crystal structure of cvHsp reveals that the dimeric form is unfavored 46
2.3.3 The CTD of cvHsp hinders its potential for self-oligomerization 50
2.3.4 The NTD is indispensable for cvHsp's potential chaperone activity 54
2.4 DISCUSSION 56
2.5 MATERIALS AND METHODS 62

3 **CVHSP BINDS FLNC TO COMPETE WITH ITS DIMER INTERFACE, AND THE INTERACTION IS REGULATED BY FLNC PHOSPHORYLATION 75**

3.1 ABSTRACT 75
3.2 INTRODUCTION 76
3.3 RESULTS 78
3.3.1 cvHsp binds FLNCd24 strongly and competes to its dimer interface 79
3.3.2 Structural analysis of cvHsp-FLNCd24 interaction 84
3.3.3 Phosphorylation at FLNCd24 T2677 and Y2683 have opposite effects in regulating FLNCd24 dimerization and its interaction with cvHsp 90
3.4 DISCUSSION 101
3.5 MATERIALS AND METHODS 106

4 **STRUCTURAL ANALYSIS OF HSPB8 AND ITS ROLE IN THE CHAPERONE ASSISTED SELECTIVE AUTOPHAGY (CASA) PROCESS** 120

4.1 ABSTRACT 120
4.2 INTRODUCTION 121
4.3 RESULTS 125
4.3.1 HspB8 is unstable and not interacting with cvHsp through the ACD-ACD interfaces *125*
4.3.2 Purification attempts of BAG3 ..... *129*
4.4 DISCUSSION 132
4.5 MATERIALS AND METHODS 135

5 **CONCLUDING REMARKS** 145

6 **APPENDIX** 148

1. GELS AND PROTEOLYSIS 148
2. CHROMATOGRAPHY 153
3. NATIVE MS 154
4. EQUATIONS FOR K_D CALCULATION 155

Chapter 1

Small heat shock proteins and biophysical techniques used in characterization

1.1 Abstract

Cardiac muscle function and morphogenesis are modulated systematically by a variety of chaperones. A specific family of chaperones is the small heat shock proteins (sHsps). The thesis focuses on small heat shock proteins targeted in the human heart and skeletal muscles and their interactions with native binding clients. This chapter describes an overview of the human cardiac sHsps about their cellular functions and general dynamic structures. To understand their specific behaviors on the molecular level, we applied different biophysical techniques to visualize the systems from different angles. I also introduce the main biophysical methods we used for this thesis. Native mass spectrometry (MS) and macromolecular X-ray crystallography provided the primary and valuable insights into the cardiac sHsps in the studies. Hydrogen-deuterium exchange (HDX) MS is also described in this chapter.

1.2 Biological background

1.2.1 Cardiac small heat shock proteins in proteostasis

The small heat shock proteins, also known as Hsp family B, are a group of chaperones that protect protein from aggregation at the early stage in cells to maintain proteostasis[1]. The absence, mutagenesis and over-expression of sHsps can be relevant to various disorders such as cancer, cardiomyopathy and neurodegenerative diseases[2–4]. First studies about some sHsps found that they are up-regulated upon stress, such as cancer, heat stress, and oxidative stress, as well as many other Hsps like Hsp70 and Hsp40, though later studies found that not all the sHsps are heat inducible[5,6]. Unlike the ATP-dependent heat shock proteins like Hsp70 or Hsp90, the sHsps perform their chaperone activities without the binding and hydrolysis of ATP[5]. They inhibit the formation of aggregation or hold the aggregates and pass them to the ATP-dependent chaperone machinery. The misfolded substrates can then be refolded through the consumption of ATP, or they can be ubiquitinated and degraded through another pathway[7].

The human genome encodes ten sHsps[8] (Table **1.1**). The most studied sHsps are Hsp27 (HspB1) and αB-crystallin (HspB5), which are ubiquitous and highly

expressed among the ten family members. HspB2, HspB3, HspB6, cvHsp (HspB7) and HspB8 are also highly expressed in the heart. cvHsp and HspB8 are mostly targeted in cardiac and skeletal tissues, and HspB8 can also be traced in brain[9–12].

Name	Synonymous names	Tissue	Promiscuous chaperones
HspB1	Hsp27	Ubiquitous	Yes
HspB2	MKBP	Muscle	No
HspB3	HspL27	Muscle	No
HspB4	αA-crystallin	Eye lens	Yes
HspB5	αB-crystallin	Ubiquitous	Yes
HspB6	Hsp20	Ubiquitous	No
HspB7	**cvHsp**	**Muscle**	**No**
HspB8	**Hsp22**	**Muscle and brain**	**No**
HspB9	CT51	Testis	No
HspB10	ODF1	Testis	No

Table 1.1. Basic properties of the ten human sHsps[5,6].

The cardiac sHsps defend the aggregates as chaperones and bind many native clients to maintain the function of cytoskeletal muscles or help the formation of the contractile units[13–15]. HspB8 is an indispensable component in the chaperone-assisted selective autophagy (CASA) pathway[16,17], which is a process to eliminate the over-stretched or aggregated filamins through the interaction between HspB8 and a co-chaperone BCL-2–associated athanogene 3 (BAG3)[18,19]. The function and

stability of HspB8 are also highly dependent on BAG3[20,21]. BAG3 links Hsp70 on the other side and thus, bridges the ATP-independent sHsp and Hsp70. The substrate carried by HspB8 can be transferred to the chaperone machinery[22–25]. The downstream autophagy can then be activated. HspB2, HspB6 and αB-crystallin also have a particular binding capacity to BAG3, but they can be replaced when HspB8 is overexpressed[9,21,26]. The function and expression rate of cvHsp is utterly independent of BAG3. The absence of cvHsp is fatal, but its prominent cellular role is improbable to be an ordinary chaperone[21,27]. The two crucial heart-specific proteins as cvHsp and HspB8 will be discussed more in detail in the following chapters.

1.2.2 The structure of sHsps

The members of sHsps are categorized based on a conserved α-crystallin domain (ACD), which locates in the middle region of the protein[8,12]. The ACD is comprised of several β-strands, as the immunoglobulin-like β-sandwich. Most of the previously studied sHsps tend to form dimeric units between the ACDs. In non-metazoan sHsps, a reciprocal dimer is formed by β2 and β6 strands[28,29]. However, in metazoan sHsps, the β6 is merged with β7 to create a longer β6+7 strand[30,31]. Anti-parallel dimer interfaces can form between the long β6+7 strands. Different registers (AP_I, AP_{II}, AP_{III}) which represent different overlapping areas on the anti-parallel interfaces have been observed, especially for αB-crystallin[30,32] (Figure **1.1**

e).

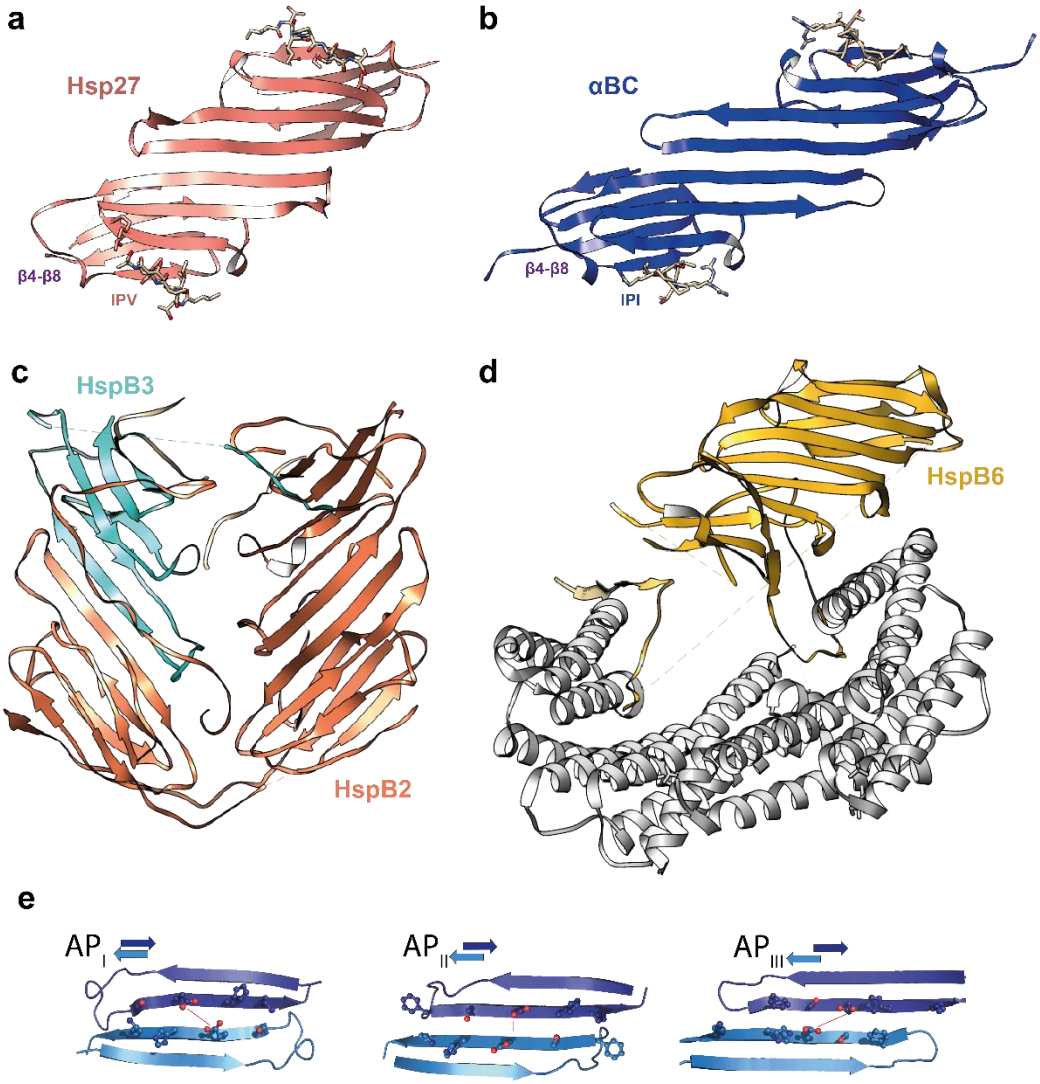


Figure 1.1. Structures of cardiac sHsps. a. Hsp27ACD (PDB: 4MJH) binds the CTD peptides containing the IPV motif[32]. **b.** αBCACD (PDB:4M5S) binds the CTD peptides containing the IPI motif[32]. **c.** HspB2 and HspB3 form the 3:1 complex (PDB:6F2R)[33]. **d.** Full-length HspB6 binds the signaling regulator 14-3-3 protein (PDB:5LTW)[34]. **e.** Three registers that exist in αBCACD. Figure adapted and edited from Hochberg et al[32].

The flexible N- and C-terminus flank by the ACD vary in length, which gives the sHsps different masses and characteristics. Some plant and mammalian sHsps

can form oligomers based on the dimer units. An indispensable I-X-I/V motif on the short C-terminal domain (CTD) can bind to a neighboring ACD dimer unit at the hydrophobic β4-β8 groove[32,35,36]. The non-structured N-terminal domain (NTD) also helps to stabilize the oligomer formed in this way[37,38]. Human heat-inducible sHsps such as Hsp27 and αB-crystallin can form homo- or hetero-oligomers of up to 50 subunits, which is termed as "polydispersity"[5,39]. Plants sHsps exist with much fewer oligomeric states[40]. Different sHsps may interact with each other to form hetero-oligomers, like the heterogeneity system between the polydisperse Hsp27 and αB-crystallin[5,37]. Some studies have shown that the homo- or hetero-oligomers will dissociate upon stress corresponding to heat or post-translational modifications (PTMs)[41] such as phosphorylation, which suggests the smaller species, rather than the oligomers, might be the ones to act as the genuine chaperones at stress conditions[41,42]. Another solid fact is many natural mutants causing the weakness of the formation of the sHsp oligomers can be disease-related, like the mutations on Hsp27 in Charcot-Marie-Tooth disease[2], which is an inherited neuropathy that affects the peripheral nervous system[43]. The dissociated sHsps also seem to have an enhanced chaperone activity[44,45]. The heart-specific sHsps, however, seem not to form larger molecular weight species[46,47]. HspB2 and HspB3 have been studied as a pair, and they can form a 3:1 complex[33]. HspB6 is dimeric dominantly[31,48] (Figure **1.1 a-d**). The behaviors of cvHsp and HspB8 will be mentioned in the following chapters.

Since the NTD is quite non-structured, and the oligomer is large and hard to be isolated, macromolecular X-ray crystallography and NMR techniques are hard to be adapted to the full-length sHsps. Hence the investigation towards the large polydisperse sHsps, about their functions and interactions, has enough research potential and challenges. Native mass spectrometry has been shown to be specifically compatible with resolving the polydisperse sHsps.

1.3 Technical background

To investigate the behavior of cardiac sHsps and their interactors, we approached a variety of biophysical techniques. The story of this thesis is mainly presented through MS and X-ray crystallography data. Native MS is a powerful tool to investigate the oligomeric states and interacting behaviors of dynamic protein complexes at native conditions. The sizes of the structured domains of sHsps and their binding partners are too small for cryo-EM to characterize, and that makes X-ray crystallography the proper technique to conduct detailed studies into them. In that way, I will introduce the background theories of native MS and X-ray crystallography that we primarily used in this thesis and briefly talk about hydrogen-deuterium exchange (HDX) MS.

1.3.1 Native mass spectrometry

Proteins or peptides need to get ionized first for MS analyzation. Ionization methods like matrix-assisted laser desorption/ionization (MALDI) denature the sample with organic acids in the matrix[49,50]. Native MS technique enables studying intact native proteins or even protein complexes and oligomers in the mass spectrometer[51]. Comparing to denaturing MS, native MS preserves the native protein structure and any non-covalent interactions within the complexes and oligomers in the gas phase[52].

The definitions of resolution in different techniques are different. The resolution in crystallography and cryo-EM are both measured in Å and can be expressed on the electron density maps. In crystallography, the resolution is defined as the minimum distance between crystal lattice planes that can produce measurable X-ray diffractions[53]. In cryo-EM, the resolution is measured in a similar way as the highest spatial frequency in Fourier space that can be interpreted[54,55]. In nuclear magnetic resonance (NMR) and MS, the definitions of resolution are similar to some extent. In MS, resolution represents the ability to separate ions with close mass to charge ratio (m/z). It can be calculated as $m/\Delta m$, where m is the nominal mass of the ion from the given peak, and Δm represents the width of the peak at half of the peak height[56]. In NMR, these concepts are replaced as chemical shifts in the unit of hertz. The high resolution and identification ability make native MS

a powerful tool to provide a new vision for structural biology and other techniques like X-ray crystallography, NMR and cryo-electron microscopy (cryo-EM) with enhanced resolution[35,36,57]. The applications of native MS now involve high molecular mass protein complexes and oligomers, glycoproteins, membrane proteins in detergents or native membranes, and virosome[58–62].

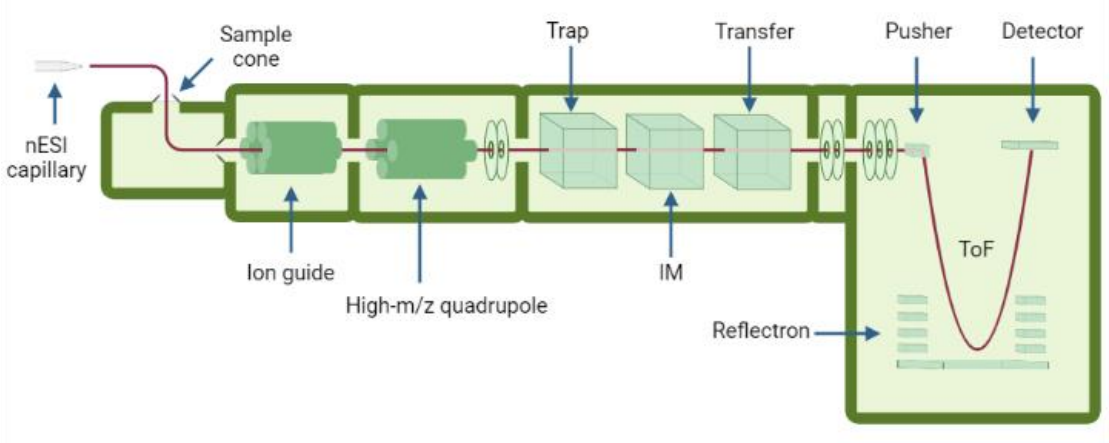


Figure 1.2. Schematic of the Waters Synapt G1 IM-MS instrument. The protein ions ionized by nESI will be focused and selected through the high pressure and the quadrupole. Then gas-phase manipulation can occur in the next stage. The protein ions can be activated in the trap or transfer cells or separated in the IM cell by size. Then the signal of the ions can be collected without the interference of kinetic energy difference by the ToF analyzer.

All the MS spectra we use in this thesis are acquired from the Waters Synapt G1 instrument. The workflow of this instrument is shown in Figure **1.2**. The sample is ionized and enters the gas phase from the nano-electrospray source, transmitted and selected in the chamber, and finally analyzed and recorded. The different parts in the native MS and the journey of ions inside the instrument are described below.

1.3.1.1 Nano-electrospray ionization

The very first step for any MS experiments is the ionization of the sample. Electrospray ionization (ESI) is a gentler way to ionize the protein compared to ionization methods like MALDI that denature the sample[51,63]. For ESI, a potential difference is added between the MS instrument with a conductive capillary containing the sample. The positive electrolytes move towards the capillary tip in the electric field with additional pressure to form a cone called the Taylor cone. Small droplets can form when the increasing charge makes the Columbic repulsion equal and exceeds the surface tension. The following desolvation and droplet fission take place before the analytes entering the sample cone[51,64]. Different mechanisms have been proposed for ion formation. The dominant one that explains the charge distribution of folded proteins is the charged residue model (CRM). This model postulate that gradual desolvation and droplet fission can reach the final droplet containing a single ion residue. The complete desolvation leaves the analyte ion with the remained charges[65,66]. The sample cone is typically perpendicular to the sample needle to decrease the number of neutral molecules entering the instrument (Figure **1.3**).

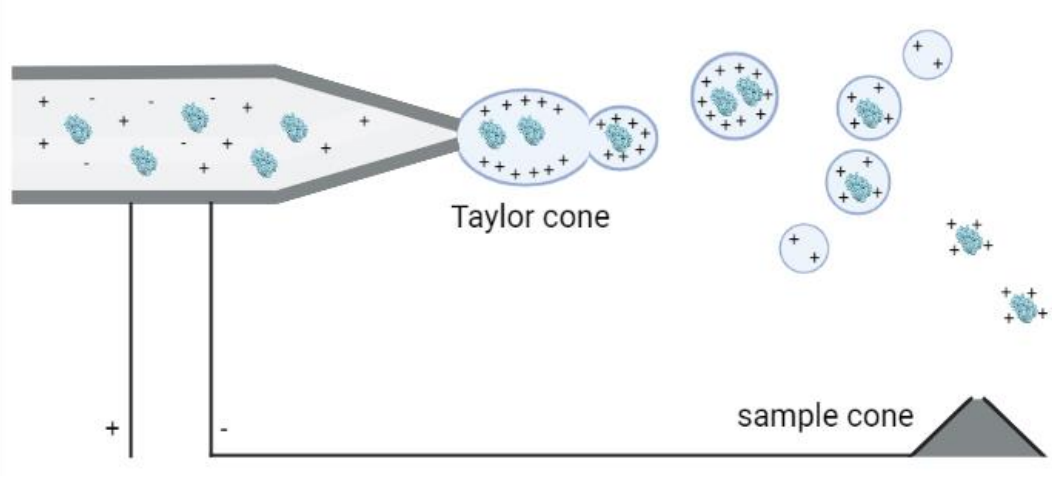


Figure 1.3. Schematic of nano-electrospray ionization of protein sample. A high voltage of around 1.5 kV is added to the gold-coated capillary to form a Taylor cone on the tip. Then fission and evaporation events can happen until one analyte ion remains in one droplet, which is described as the charged residue model (CRM). Then complete desolvation gives the charged analytes that will enter the orthogonal sample cone.

The nano-electrospray ionization (nESI) is revolutionary for the application of native MS to structural biology. It uses a coated capillary with a much smaller diameter and consumes just several μL of the sample for the measurement[67,68]. Compared to the conventional ESI, nESI is more tolerant to salt contamination, which is very important for resolving the peaks representing each charge state of the ions. This is because of a reduction of the initial droplet size formed from nESI. The non-volatile salt in the droplet doesn't evaporate during the fission as the volatile salt used in nESI, like ammonium acetate. Hence, the salt gets concentrated in the final droplet and makes the spectrum hard to be resolved. The smaller initial droplet in nESI improved the salt tolerance, and nESI is gentler and more consistent than ESI[68–70].

1.3.1.2 Ion transmission and analysis

The folded protein ions face a high pressure after they enter the ion guide to enhance the efficiency of macromolecular ions transmission. In the low radio frequency (RF) ion guide, the large protein ions experience more collision events with the neutral gas molecules under high pressure. The kinetic energy, as well as the velocity, are decreased for the ions, and the ions can be focused within their trajectory more easily[51,71,72].

The quadrupole is mainly used in native MS as the ion filter. Direct current (DC) is applied to the four paired parallel rods with an RF potential to select the ions within a specific m/z value region. The neighboring rods have reversed polarity. Ions with a larger or smaller m/z value can be destabilized and exit the trajectory to be neutralized. Only those ions in a specific m/z value region can be stable in the trajectory and pass the quadrupole chamber[73]. For the denaturing MS for the analysis of unfolded proteins with small m/z values or peptides, the quadrupole is enough to work as the analyzer. In native MS, transmission and analysis of ions at higher m/z region are required[72,74].

In the Synapt instruments, the Time-of-Flight (ToF) analyzer is used to analyze the ions for better resolution. It is designed perpendicular to the direction of the previous trajectory so that to reduce any initial kinetic energy differences of the ions[75,76]. Its working mechanism bases on the relationship,

$$zeV = \frac{1}{2}mv^2$$

Where z is the charge taken by an ion, e is the elementary charge, V is the electrical potential, m is the mass of the ion and v represents the velocity. When v is substituted by length d and time t, the relationship can also be presented as,

$$\frac{m}{z} = \frac{2eV}{d^2}t^2$$

Therefore, the theoretical m/z that ToF can analyze is unlimited[77]. But ions of the same m/z value still experience different collision events and gain different kinetic energy in the ToF chamber, giving a widened peak on the spectrum. The reflectron, also known as the "ion mirror", is used to reduce this effect. Ions with larger kinetic energy can penetrate the reflectron deeper to adjust the travelling time of ions of the same m/z value closer[78]. The ions can then be detected by hitting the microchannel plate (MCP), which works as the electron multiplier. The relevant current can be stimulated and recorded by the computer[51,79].

1.3.1.3 Collision-induced dissociation and Ion mobility

Collision-induced dissociation (CID) and ion mobility (IM) are two useful functions in the Synapt instruments specifically. They are not involved much in this thesis,

but they still have important applications in many significant native MS-relevant works. The dissociation of non-covalent interactions and the following analysis is an advantage of native MS compared to other structural biological techniques. CID is a sequential dissociation method[51,80]. The collision can happen either in the Trap or Transfer cells on both sides of the IM chamber. When the energy added reaches a level, one unit in the protein complex starts to be stripped off. In the spectrum, the stripped monomer shows as an unfolded series with more charge states, and the remained complex can be more separate for peak assignment (Figure **1.4**). The process is sequential with more energy added. CID can be a helpful tool for resolving the sHsp polydisperse oligomers and identifying the composition of complexes[80,81].

IM coupled to MS (IM-MS) is an informative technique working on the size estimation of protein molecules. In the gas phase, a collision-cross-section (CCS) value can be obtained as an estimation of the molecular size. The CCS value is linked to the drifting time of the ion, ions with a smaller CCS value experiences fewer collision events with gas molecules in the trajectory of the IM travel-wave drift tube and takes shorter time to leave the IM cell. Some standards need to be recorded at the same condition for travelling-wave IM for the estimation[82]. CCS values can be good references to study conformational changes[83]. When combined CID with the Trap cell, IM can also obtain the collision-induced unfolding information[81]. Different proteins or complexes with different conformation

stability can have unique unfolding plots with increasing collision voltage[13].

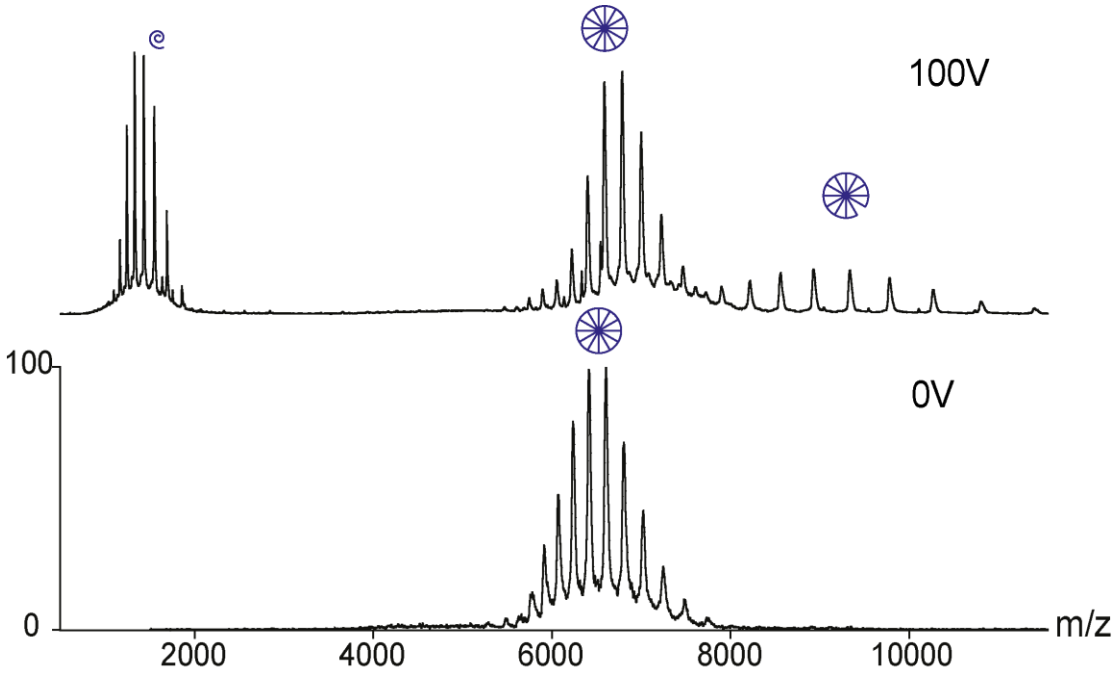


Figure 1.4. CID of a homo-oligomeric protein. Hsp23.5 from *Arabidopsis thaliana* exists as a dodecamer on native MS (bottom spectrum). When applying the energy in the collision cell to 100V, one unfolded subunit was ejected as the left peak series. The remained 11mer complex is right-shifted on the spectrum (upper spectrum).

The development of native MS makes the analysis of larger complexes achievable. The QExactive mass spectrometer with the orbitrap analyzer can provide more significant desolvation energy and the Eclipse orbitrap instrument makes the tandem ion selection much easier[57,84], which broadens the perspective of the systems that can be adapted for native MS.

1.3.2 Macromolecular X-ray crystallography

Despite the resolution revolution of cryo-EM[85,86], macromolecular X-ray crystallography is still an essential and indispensable technique in protein structural investigation. Unlike NMR or cryo-EM, crystallography is not limited by the mass of the target protein theoretically and can reach very high resolution[87,88]. Once above specific resolutions, it can provide folding, secondary structure, side chain twisting and even atomic location information[53]. The recent breakthrough of AlphaFold, which is a model for the structural prediction that has calculated the structures of almost the whole human proteome (98.5% of human proteins), is also a challenge to traditional protein crystallography[89,90]. But crystallography is still a dominant technique in studying the protein docking and the information about the interfaces of proteins of given mass ranges[91].

Successful crystallization depends on the homogeneity and rigidity of the protein constructs[92]. Pure protein sample (above 95%) is the basis of crystallography. The purification of the proteins to be crystallized needs to have a higher standard[93]. It can be more challenging to crystallize protein complexes or co-expressed products. However, just pure protein cannot guarantee a piece of diffractable crystal. Flexible regions in the protein hinder the periodic packing during the crystal formation, like the N-terminus in most sHsps. Though these flexible parts can be functional in the protein to be studied, they need to be removed to enhance the

crystallization. Site-specific limited proteolysis is a standard method to discover a rigid construct of the target protein[94]. Enzymes are incubated with the protein for specific time lengths to cut the flexible tails, then denaturing MS was used to confirm the sequence of the remained rigid construct. For some circumstances, some residues with flexible side chains or cysteines on the surface might also need to be mutated for crystallization. Some proteins that are not very stable by themselves might be co-crystallized with other proteins or ligands[95].

Thousands of matrices of commercial precipitation reagents and conditions are tested and optimized for one target for crystallization. The storage buffer containing the protein is mixed with the crystallization solution in a closed space, with a reservoir containing the same crystallization solution. Because the precipitant concentration in the droplet with the protein is lower than that in the reservoir, then net water movement from the droplet to the reservoir can form. The concentration of protein and precipitant both become increased gradually, and nucleation or precipitation can occur[96]. A single crystal that can be successfully diffracted may take a lot of effort and attempts to obtain.

Once the perfect single crystals form, they can be collected at the cryogenic conditions to reduce radiation damage when diffracted in the synchrotron. The X-ray source in the synchrotron is more stable and brighter than the in-house X-ray source, and it takes a much shorter time to do the complete diffraction of the crystal with higher resolution[97].

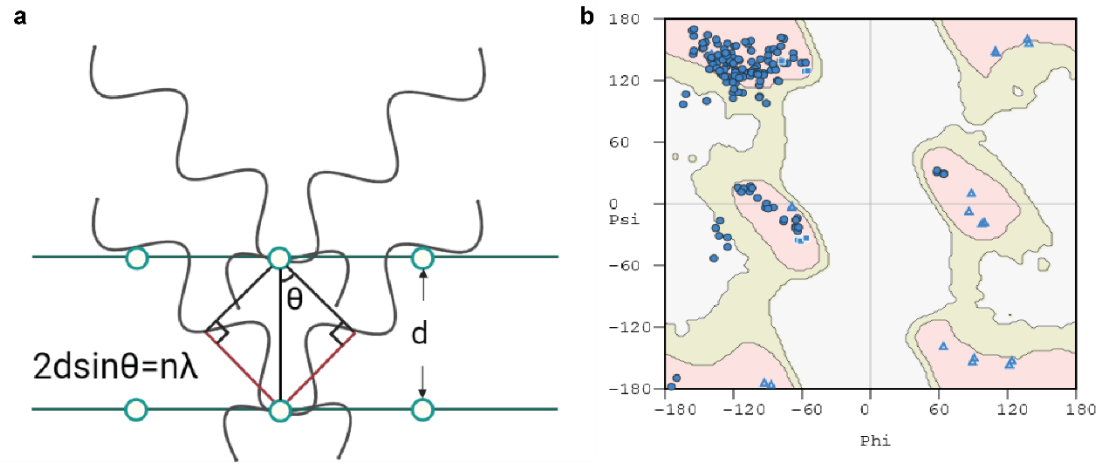


Figure 1.5. Diffraction and refinement in macromolecular X-ray crystallography. a. Bragg's Law describes the conditions that need to be fulfilled in the crystal for diffraction: only the extra travel length (2dsinθ) of the incident wave between the crystal planes are the integer number times of the wavelength (nλ), the diffraction can be amplified. **b.** The Ramachandran Plot of the well-polished FLNCd24 T2677D crystal structure. No outliers remain in the plot. 94% of the residues locate in the preferred regions, and the rest find in the allowed areas. The top left is the preferred region for β-strands.

The molecules in the crystal are packing periodically in the repeating unit cells. The molecules can be concentrated as lattice points in the asymmetric unit. Different planes that intersect the lattice points can be defined. The same set of planes have the same distance *d*. When the parallel X-ray is penetrating the plane at a given angle that satisfies Bragg's Law, the diffraction can be amplified between the lattice points of the same set of planes[98] (Figure **1.5a**). The individual diffraction amplitude and plane information can be recorded by the detector. However, this is not sufficient for the reverse Fourier Transformation to obtain the structure in the unit cell. The phase information of each diffraction is not recorded[53].

The most straightforward method to cope with the phase problem is through molecular replacement (MR), which is based on the Patterson Function's relationship. If any structure of the exact analogue of the target protein is available, then MR is worth the attempts. All the structures in this thesis were phased by MR. If there are no available structures for MR, then experimental phasing which requires anomalous scattering can be pursued. Heavy atoms like selenomethionine need to be introduced to the protein during expression to create the isomorphic crystal or soak some clusters like tantalum into the crystal[99,100].

The phased structure needs to be refined. Several criteria can be good references for the refinement. The angles and torsion angles of the twisted bonds and sheets are still limited to satisfy the secondary structural elements like α-helixes, β-sheets and loops. Ramachandran plot provides the relationship between φ and ψ of each amino acid in the phased structure, while ω is restricted[101]. The torsion angles of most residues locate in the allowed regions in the Ramachandran plot (Figure **1.5b**). The important factor for the validation of structure is the R-factor:

$$R = \frac{\sum\left||\boldsymbol{F}_{obs}|-|\boldsymbol{F}_{cal}|\right|}{\sum|\boldsymbol{F}_{obs}|}$$

Where $\mathbf{F}$ is the sum of the reflection amplitude of individual atoms. $\mathbf{F}_{obs}$ comes from the current model, while $\mathbf{F}_{cal}$ comes from the theoretical calculation. It gives information on how close the current model compares to the original dataset. A

smaller R value represents a better-refined model. The model can be overfitted to make the R value very small if too many parameters are used in the calculation. In this way, the other R_{free} value, which randomly picks some of the reflections can be supplementary to R_{work} value[53]. 5% of the reflections are chosen to calculate R_{free} in Phenix software that we used in this thesis[102].

1.3.3 Hydrogen-deuterium exchange (HDX) MS

We also used hydrogen-deuterium exchange (HDX) MS to supplement our crystal structures. HDX-MS is a structural technique that can be operated with functional buffering systems to study problems like protein folding, protein conformational changes and protein-protein or protein-ligand interactions[103,104]. In this way, it provides great supports to traditional structural techniques and unique insights into structural studies.

The working theory of HDX-MS is to incubate the protein or protein complex with deuterium buffer. The hydrogen on the backbone amide can undertake exchange with the deuterium in the buffer. The rate and speed of the exchange can be different for the residues in different parts of the protein, and some parts can be more buried while others can be more exposed. The deuterium exchange can be quenched at low pH, where the exchange rate is the slowest. The deuterium uptake information can be preserved. The proteins will undergo enzymatic

proteolysis by pepsin, which works efficiently at the low pH range. The peptides can be subjected to the liquid chromatography (LC) -MS for the lateral analyzation[51,105] (Figure **1.6**).

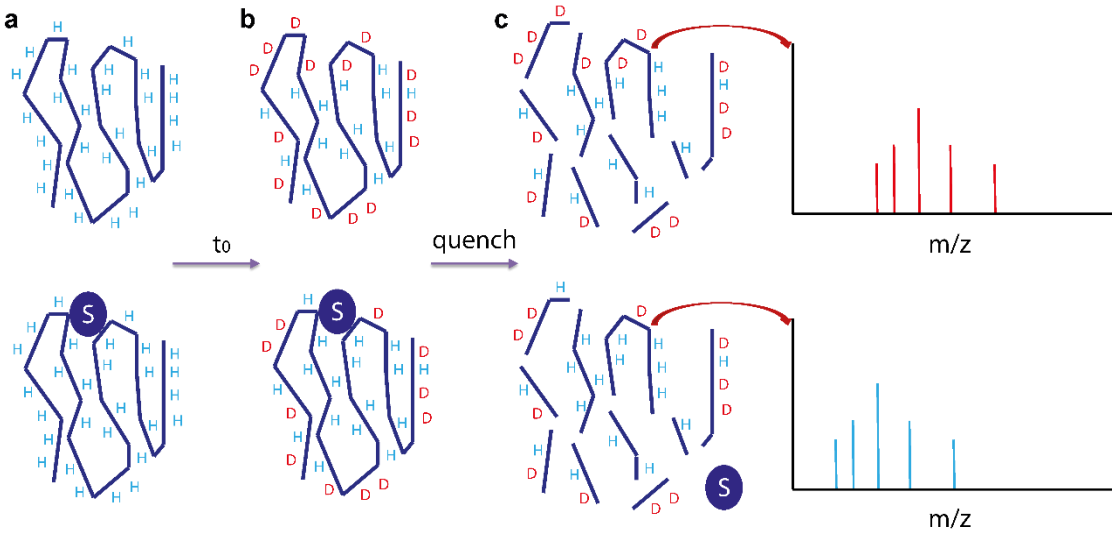


Figure 1.6. Schematic of using HDX-MS to confirm the binding site. a. The protein (apo-state) and the protein bounded with the substrate ligand have the hydrogens on backbone amides for exchange. **b.** The protein and the protein complex undergo deuterium exchange in the deuterium buffer for the same time length t_0. The binding site of the protein in the complex has a different rate of deuterium uptake. **c.** The exchanges are quenched at low pH, and the proteins are digested by enzymes before entering the LC-MS system for analysis. The raw MS data of the peptides that cover the binding site shows the difference in deuterium uptake.

The comparison of deuterium uptake is the basis of the structural information that HDX can provide. To study protein-ligand interactions, the protein itself, commonly named the apo-state, needs to experience the deuterium uptake at the same time lengths and conditions to work as the baseline. The data of the complex with ligand can be recorded in the same way. The same peptides but with different deuterium uptake can give information about the interaction region. The Woods

plot that can provide the information of protection or deprotection of each peptide by presenting deuterium uptake or relative fractional uptake for a specific time point is a good representation to visualize the HDX data for the whole protein. The heat map can provide the information of relative fractional uptake for all the time points or a specific time point that can be visualized on the structure[106–108].

Different techniques characterize the system from a variety of perspectives. X-ray crystallography, cryo-EM and structure prediction models provide the most detailed atomic information of the structure and interaction. Still, they cannot tell any of the dynamical parameters yet. So other techniques need to be combined for looking into nature from different angles. We can never know the extreme truth of the system that we are studying, but these different combinations of techniques can make us close to it.

Bibliography

1. Boelens, W. C. Structural aspects of the human small heat shock proteins related to their functional activities. *Cell Stress and Chaperones* **25**, 581–591 (2020).
2. Reid Alderson, T. *et al.* A weakened interface in the P182L variant of HSP27 associated with severe Charcot - Marie - Tooth neuropathy causes aberrant binding to interacting proteins. *The EMBO Journal* **40**, 1–23 (2021).
3. Cristofani, R. *et al.* The regulation of the small heat shock protein B8 in misfolding protein diseases causing motoneuronal and muscle cell death. *Frontiers in Neuroscience* **13**, 1–13 (2019).
4. Al-Tahan, S. *et al.* New family with HSPB8-associated autosomal dominant rimmed vacuolar myopathy. *Neurology: Genetics* **5**, 1–11 (2019).
5. Collier, M. P. & Benesch, J. L. P. Small heat-shock proteins and their role in mechanical stress. (2020).
6. Mymrikov, E. v., Daake, M., Richter, B., Haslbeck, M. & Buchner, J. The chaperone activity and substrate spectrum of human small heat shock proteins. *Journal of Biological Chemistry* **292**, 672–684 (2017).
7. Żwirowski, S. *et al.* Hsp70 displaces small heat shock proteins from aggregates to initiate protein refolding. *The EMBO Journal* **36**, 783–796 (2017).
8. Kappé, G. *et al.* The human genome encodes 10 α-crystallin-related small

heat shock proteins: HspB1-10. *Cell Stress and Chaperones* **8**, 53–61 (2003).

9. Morelli, F. F. *et al.* An interaction study in mammalian cells demonstrates weak binding of HSPB2 to BAG3, which is regulated by HSPB3 and abrogated by HSPB8. *Cell Stress and Chaperones* **22**, 531–540 (2017).

10. Li, F. *et al.* HSPB8 over-expression prevents disruption of blood–brain barrier by promoting autophagic flux after cerebral ischemia/reperfusion injury. *Journal of Neurochemistry* **148**, 97–113 (2019).

11. Doran, P., Gannon, J., O'Connell, K. & Ohlendieck, K. Aging skeletal muscle shows a drastic increase in the small heat shock proteins αB-crystallin/HspB5 and cvHsp/HspB7. *European Journal of Cell Biology* **86**, 629–640 (2007).

12. Lorenzo, P. *et al.* Identification and Characterization of cvHsp. *Journal of Biological Chemistry* **276**, 12201–12211 (1999).

13. Collier, M. P. *et al.* HspB1 phosphorylation regulates its intramolecular dynamics and mechanosensitive molecular chaperone interaction with filamin C. *Science Advances* **5**, (2019).

14. Juo, L. Y. *et al.* HSPB7 interacts with dimerized FLNC and its absence results in progressive m. *Journal of Cell Science* **129**, 1661–1670 (2016).

15. Liao, W. C., Juo, L. Y., Shih, Y. L., Chen, Y. H. & Yan, Y. T. HSPB7 prevents cardiac conduction system defect through maintaining intercalated disc integrity. *PLoS genetics* **13**, e1006984 (2017).

16. Arndt, V. *et al.* Chaperone-Assisted Selective Autophagy Is Essential for

Muscle Maintenance. *Current Biology* **20**, 143–148 (2010).

17. Ulbricht, A. *et al.* Induction and adaptation of chaperone-assisted selective autophagy CASA in response to resistance exercise in human skeletal muscle. *Autophagy* **11**, 538–546 (2015).

18. Ulbricht, A. & Höhfeld, J. Tension-induced autophagy: May the chaperone be with you. *Autophagy* **9**, 920–922 (2013).

19. Ulbricht, A. *et al.* Cellular mechanotransduction relies on tension-induced and chaperone-assisted autophagy. *Current Biology* **23**, 430–435 (2013).

20. Nivon, M. *et al.* NF-κB regulates protein quality control after heat stress through modulation of the BAG3-HspB8 complex. *Journal of Cell Science* **125**, 1141–1151 (2012).

21. Fang, X., Bogomolovas, J., Trexler, C. & Chen, J. The BAG3-dependent and -independent roles of cardiac small heat shock proteins. *JCI insight* **4**, 1–12 (2019).

22. Rauch, J. N. *et al.* BAG3 Is a Modular, Scaffolding Protein that physically Links Heat Shock Protein 70 (Hsp70) to the Small Heat Shock Proteins. *Journal of Molecular Biology* **429**, 128–141 (2017).

23. Ganassi, M. *et al.* A Surveillance Function of the HSPB8-BAG3-HSP70 Chaperone Complex Ensures Stress Granule Integrity and Dynamism. *Molecular Cell* **63**, 796–810 (2016).

24. Rauch, J. N. & Gestwicki, J. E. Binding of human nucleotide exchange factors to heat shock protein 70 (Hsp70) generates functionally distinct

complexes in vitro. *Journal of Biological Chemistry* **289**, 1402–1414 (2014).

25. Li, X. *et al.* Validation of the Hsp70-Bag3 protein-protein interaction as a potential therapeutic target in cancer. *Molecular Cancer Therapeutics* **14**, 642–648 (2015).

26. Fuchs, M. *et al.* Identification of the key structural motifs involved in HspB8/HspB6-Bag3 interaction. *Biochemical Journal* **425**, 245–255 (2010).

27. Vos, M. J. *et al.* HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. *Human Molecular Genetics* **19**, 4677–4693 (2010).

28. Hilario, E., Martin, F. J. M., Bertolini, M. C. & Fan, L. Crystal structures of xanthomonas small heat shock protein provide a structural basis for an active molecular chaperone oligomer. *Journal of Molecular Biology* **408**, 74–86 (2011).

29. van Montfort, R. L. M., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nature Structural Biology* **8**, 1025–1030 (2001).

30. Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein science : a publication of the Protein Society* **19**, 1031–1043 (2010).

31. Baranova, E. v. *et al.* Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *Journal of*

Molecular Biology **411**, 110–122 (2011).

32. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proceedings of the National Academy of Sciences of the United States of America* **111**, (2014).

33. Clark, A. R. *et al.* Terminal Regions Confer Plasticity to the Tetrameric Assembly of Human HspB2 and HspB3. *Journal of Molecular Biology* **430**, 3297–3310 (2018).

34. Sluchanko, N. N. *et al.* Structural Basis for the Interaction of a Human Small Heat Shock Protein with the 14-3-3 Universal Signaling Regulator. *Structure* **25**, 305–316 (2017).

35. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 1–13 (2013).

36. Baldwin, A. J. *et al.* Probing dynamic conformations of the high-molecular-weight αb-crystallin heat shock protein ensemble by NMR spectroscopy. *Journal of the American Chemical Society* **134**, 15343–15350 (2012).

37. Haslbeck, M., Weinkauf, S. & Buchner, J. Small heat shock proteins: Simplicity meets complexity. *Journal of Biological Chemistry* **294**, 2121–2132 (2019).

38. Jehle, S. *et al.* N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proceedings of the National Academy of Sciences of the United States of*

America **108**, 6409–6414 (2011).

39. Aquilina, J. A., Benesch, J. L. P., Bateman, O. A., Slingsby, C. & Robinson, C. v. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10611–10616 (2003).

40. Waters, E. R. & Vierling, E. Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist* **227**, 24–37 (2020).

41. Haslbeck, M., Peschek, J., Buchner, J. & Weinkauf, S. Structure and function of α-crystallins: Traversing from in vitro to in vivo. *Biochimica et Biophysica Acta - General Subjects* **1860**, 149–166 (2016).

42. Muranova, L. K., Ryzhavskaya, A. S., Sudnitsyna, M. v., Shatov, V. M. & Gusev, N. B. Small Heat Shock Proteins and Human Neurodegenerative Diseases. *Biochemistry (Moscow)* **84**, 1256–1267 (2019).

43. Almeida-Souza, L. *et al.* Small heat-shock protein HSPB1 mutants stabilize microtubules in Charcot-Marie-Tooth neuropathy. *Journal of Neuroscience* **31**, 15320–15328 (2011).

44. Jovcevski, B. *et al.* Phosphomimics destabilize Hsp27 oligomeric assemblies and enhance chaperone activity. *Chemistry and Biology* **22**, 186–195 (2015).

45. Alderson, T. R. *et al.* Local unfolding of the HSP27 monomer regulates chaperone activity. *Nature Communications* **10**, 1–16 (2019).

46. Vos, M. J., Kanon, B. & Kampinga, H. H. HSPB7 is a SC35 speckle resident

small heat shock protein. *Biochimica et Biophysica Acta - Molecular Cell Research* **1793**, 1343–1353 (2009).

47. Kim, M. v., Seit-Nebi, A. S., Marston, S. B. & Gusev, N. B. Some properties of human small heat shock protein Hsp22 (H11 or HspB8). *Biochemical and Biophysical Research Communications* **315**, 796–801 (2004).

48. Weeks, S. D. *et al.* Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6. *Journal of Structural Biology* **185**, 342–354 (2014).

49. Singhal, N., Kumar, M., Kanaujia, P. K. & Virdi, J. S. MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology* **6**, 1–16 (2015).

50. Nadler, W. M. *et al.* MALDI versus ESI: The Impact of the Ion Source on Peptide Identification. *Journal of Proteome Research* **16**, 1207–1215 (2017).

51. Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. & Robinsons, C. v. Protein complexes in the gas phase: Technology for structural genomics and proteomics. *Chemical Reviews* **107**, 3544–3567 (2007).

52. Robinson, C. v. Mass spectrometry: From plasma proteins to mitochondrial membranes. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 2814–2820 (2019).

53. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal* **275**, 1–21 (2008).

54. Beckers, M., Mann, D. & Sachse, C. Structural interpretation of cryo-EM image reconstructions. *Progress in Biophysics and Molecular Biology* **160**, 26–36 (2021).

55. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, (2020).

56. Tamara, S., den Boer, M. A. & Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chemical Reviews* (2021) doi:10.1021/acs.chemrev.1c00212.

57. Gault, J. *et al.* Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nature Methods* **17**, 505–508 (2020).

58. Yen, H. Y. *et al.* PtdIns(4,5)P2 stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* **559**, 423–427 (2018).

59. Wu, D., Struwe, W. B., Harvey, D. J., Ferguson, M. A. J. & Robinson, C. v. N-glycan microheterogeneity regulates interactions of plasma proteins. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 8763–8768 (2018).

60. Urner, L. H. *et al.* Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors. *Nature Communications* **11**, (2020).

61. Hochberg, G. K. A. *et al.* Structural principles that enable oligomeric small heat-shock protein paralogs to evolve distinct functions. *Science* **359**, 930–935 (2018).

62. Snijder, J., Rose, R. J., Veesler, D., Johnson, J. E. & Heck, A. J. R. Studying 18 MDa virus assemblies with native mass spectrometry. *Angewandte Chemie - International Edition* **52**, 4020–4023 (2013).

63. Cohen, L. R. H., Strupat, K. & Hillenkamp, F. Analysis of quaternary protein ensembles by matrix assisted laser desorption/ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **8**, 1046–1052 (1997).

64. Kebarle, P. & Tang, L. From Ions in Solution To Ions in the Gas Phase. *Analytical Chemistry* **65**, 972A-986A (1993).

65. Dole, M. & Mack, L. L. Molecular beams of macroions. *The Journal of Chemical Physics* **49**, 2240 (1968).

66. Kebarle, P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry* **35**, 804–817 (2000).

67. Wilm, M. S. & Mann, M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes* **136**, 167–180 (1994).

68. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry* **68**, 1–8 (1996).

69. Agasid, M. T., Sørensen, L., Urner, L. H., Yan, J. & Robinson, C. v. The Effects of Sodium Ions on Ligand Binding and Conformational States of G Protein-Coupled Receptors-Insights from Mass Spectrometry. *Journal of the*

American Chemical Society **143**, 4085–4089 (2021).

70. Juraschek, R., Dülcks, T. & Karas, M. Nanoelectrospray - More than just a minimized-flow electrospray ionization source. *Journal of the American Society for Mass Spectrometry* **10**, 300–308 (1999).

71. Douglas, D. J. & French, J. B. Collisional focusing effects in radio frequency quadrupoles. *Journal of the American Society for Mass Spectrometry* **3**, 398–408 (1992).

72. Hilton, G. R. & Benesch, J. L. P. Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *Journal of the Royal Society Interface* **9**, 801–816 (2012).

73. Labastie, P. & Doy, M. A high mass range quadrupole spectrometer for cluster studies. *International Journal of Mass Spectrometry and Ion Processes* **91**, 105–112 (1989).

74. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. v. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Analytical Chemistry* **74**, 1402–1407 (2002).

75. Dawson, J. H. J. & Guilhaus, M. Orthogonal-acceleration Time-of-flight. *Rapid Communications in Mass Spectrometry* **3**, 155–159 (1989).

76. Boyle, J. G. & Whitehouse, C. M. Time-of-Flight Spectrometry Electrospray. *Analytical Chemistry* **64**, 2084–2089 (1992).

77. Lössl, P., Snijder, J. & Heck, A. J. R. Boundaries of mass resolution in

native mass spectrometry. *Journal of the American Society for Mass Spectrometry* **25**, 906–917 (2014).

78. Sharon, M. & Robinson, C. v. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annual Review of Biochemistry* **76**, 167–193 (2007).

79. Ladislas Wiza, J. Microchannel plate detectors. *Nuclear Instruments and Methods* **162**, 587–601 (1979).

80. Benesch, J. L. P. Collisional Activation of Protein Complexes: Picking Up the Pieces. *Journal of the American Society for Mass Spectrometry* **20**, 341–348 (2009).

81. Allison, T. M. *et al.* Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nature Communications* **6**, (2015).

82. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S. J. & Robinson, C. v. Ion mobility-mass spectrometry analysis of large protein complexes. *Nature Protocols* **3**, 1139–1152 (2008).

83. Chandler, S. A. & Benesch, J. L. Mass spectrometry beyond the native state. *Current Opinion in Chemical Biology* **42**, 130–137 (2018).

84. Gault, J. *et al.* High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nature Methods* **13**, 333–336 (2016).

85. Scheres, S. H. w. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).

86. Cheng, Y. Single-particle cryo-EM-How did it get here and where will it go.

Science **361**, 876–880 (2018).

87. Wang, H. W. & Wang, J. W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Science* **26**, 32–39 (2017).

88. Frueh, D. P., Goodrich, A. C., Mishra, S. H. & Nichols, S. R. NMR methods for structural studies of large monomeric and multimeric proteins. *Current Opinion in Structural Biology* **23**, 734–739 (2013).

89. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* (2021) doi:10.1038/s41586-021-03828-1.

90. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

91. Elez, K., Bonvin, A. M. J. J. & Vangone, A. Biological vs. Crystallographic protein interfaces: An overview of computational approaches for their classification. *Crystals* **10**, 1–15 (2020).

92. Wang, H. W. & Wang, J. W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Science* **26**, 32–39 (2017).

93. Kim, Y. *et al. High-Throughput Protein Purification for X-Ray Crystallography and NMR. Advances in Protein Chemistry and Structural Biology* vol. 75 (Elsevier Masson SAS, 2008).

94. Wernimont, A. & Edwards, A. In Situ proteolysis to generate crystals for structure determination: An update. *PLoS ONE* **4**, (2009).

95. Müller, I. Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallographica Section D: Structural Biology* **73**, 79–92

(2017).

96. Dessau, M. A. & Modis, Y. Protein crystallization for X-ray crystallography. *Journal of Visualized Experiments* **9**, 1–6 (2010).

97. Hendrickson, W. A. Synchrotron crystallography. *Trends in Biochemical Sciences* **25**, 637–643 (2000).

98. Bragg, W. H. & Bragg, W. L. The Reflection of X-rays by Crystals. *Proceedings of the Royal Society of London . Series A* **88**, 428–438 (1913).

99. McCoy, A. J. & Read, R. J. Experimental phasing: Best practice and pitfalls. *Acta Crystallographica Section D: Biological Crystallography* **66**, 458–469 (2010).

100. Evans, P. & McCoy, A. An introduction to molecular replacement. *Acta Crystallographica Section D: Biological Crystallography* **64**, 1–10 (2007).

101. Hollingsworth, S. A. & Karplus, P. A. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts* **1**, 271–283 (2010).

102. Shabalin, I. G., Porebski, P. J. & Minor, W. Refining the macromolecular model – achieving the best agreement with the data from X-ray diffraction experiment agreement with the data from X-ray diffraction experiment. *Crystallography Reviews* **3508**, 236–262 (2018).

103. Jia, R. *et al.* Hydrogen-deuterium exchange mass spectrometry captures distinct dynamics upon substrate and inhibitor binding to a transporter. *Nature Communications* **11**, (2020).

104. Fiorentino, F. *et al.* Dynamics of an LPS translocon induced by substrate and an antimicrobial peptide. *Nature Chemical Biology* **17**, 187–195 (2021).

105. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nature Methods* **16**, 595–602 (2019).

106. Woods, V. L. & Hamuro, Y. High resolution, high-throughput amide Deuterium Exchange-Mass Spectrometry (DXMS) determination of protein binding site structure and dynamics: Utility in pharmaceutical design. *Journal of Cellular Biochemistry* **84**, 89–98 (2001).

107. Lau, A. M. C., Ahdash, Z., Martens, C. & Politis, A. Deuteros: Software for rapid analysis and visualization of data from differential hydrogen deuterium exchange-mass spectrometry. *Bioinformatics* **35**, 3171–3173 (2019).

108. Lau, A. M., Claesen, J., Hansen, K. & Politis, A. Deuteros 2.0: Peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics* **37**, 270–272 (2021).

Chapter 2

cvHsp is dominantly monomeric and unique in the family

2.1 Abstract

cvHsp is mainly targeted in heart and skeletal muscles, and it belongs to the human small heat shock protein family. The absence of cvHsp is fatal and can cause the collapse of the cytoskeletal structure. Our studies using native MS find that unlike the canonical human sHsps forming oligomers based on dimeric units, cvHsp is dominantly monomeric. We report the first crystal structure of cvHsp and observe that cvHsp is not packing as dimer units in the crystal due to a distinct β5-β6+7 region, which is the canonical dimer interface of human sHsps. cvHsp does not form large oligomers because the indispensable interaction between its ACD and CTD is very weak. It does not have the proper IPI/V motif on its CTD to bind the β4-β8 groove on its ACD, though the conformation and hydrophobic residues are highly conserved at its β4-β8 groove compared to the polydispersed Hsp27 in the family. We also confirm that the disordered NTD is indispensable for the chaperone activity of cvHsp for inhibiting the aggregation of heat-induced standard substrate citrate synthase. These observations based on biophysical studies show the feature of a crucial cardiac-specific protein and explain its unique characteristics at the atomic level.

2.2 Introduction

Cardiovascular heat shock protein (cvHsp, HspB7) was first identified in 1999 and classified as a member of the human small heat shock protein (sHsp) family for having a conserved α-crystallin domain (ACD)[1,2]. cvHsp is mostly targeted in heart and skeletal muscles and associates with the nuclear speckles, which are the small subnuclear membraneless organelles or structures for RNA splicing at non-stress conditions[1,3,4]. cvHsp locates adjacent to the Z-line, which separates and links the sarcomeres at the intercalated discs that connect heart muscle cells in the striated muscle[5,6]. The expression of cvHsp increases with age, and the expression level of cvHsp at the myocardium and diaphragm can be biomarkers of heart failure and Duchenne muscular dystrophy[7–9]. The deficiency of cvHsp genes can be disease-related. Single nucleotide polymorphisms (SNP) on cvHsp genes are highly associated with different cardiomyopathies, Cheyne-Stokes respiration with sleep apnea and heart failure (HF)[10–16].

cvHsp participates in a variety of biological modulations to maintain the normal functions of the cytoskeletal system. It regulates the length of thin filaments at the heart developing stage and hinders monomeric actin (G-actin) from aggregation[17]. cvHsp is involved in the early stage of cardiac morphogenesis and heart migration in the zebrafish models[18,19]. Genetic regulation of cvHsp is closely related to muscle atrophy and autophagic processes[20]. cvHsp also plays a vital

role in the osteogenic differentiation of adipose stem cells[21]. The knockout of cvHsp is fatal in mice and can cause severe cardiomyopathy and defects in diaphragm muscle, including sarcomeres and intercalated discs[5,6]. The absence of cvHsp results in the upregulation and aggregation of the actin-linking protein filamin-C (FLNC)[5,6,22], but the clearance of aggregated FLNC is independent of cvHsp[23], which is discussed in the following chapters.

cvHsp is relatively unstudied on the molecular level compared to other ubiquitous human sHsps. It has the conserved ACD, flanked by the flexible N-terminal and C-terminal domains (NTD and CTD)[24] as the other sHsps. cvHsp has been observed as a small protein, not forming large polydispersed oligomers like other human sHsp family members like Hsp27 or αB-crystallin (αBC) [25,26]. Rat cvHsp was revealed to distribute between dimers and dodecamers, while human cvHsp has been reported to be small, likely to be monomeric and dimeric[25–27]. Unlike the other cardiac-specific sHsps such as αBC, HspB6 and HspB8, the function and stability of cvHsp is independent of the heat-inducible co-chaperone B-cell lymphoma 2 (Bcl-2)-associated athanogene (BAG3)[25,28]. BAG3 has two IPV motifs to bind the β4-β8 groove on sHsp ACD and competes with the IPI/V of sHsp CTD for the oligomer formation[28,29]. There are no interactions between cvHsp and BAG3, which suggests a unique role of cvHsp in heart and skeletal muscles compared to the other cardiac sHsps[25,28]. The specific interaction between BAG3 and HspB8, which is essential in the chaperone-assisted selective autophagy

(CASA), is discussed in the last chapter of this thesis. Interestingly, HspB8 is the only member within the human sHsp family that has been reported to be a binding client of cvHsp to form heteromers[24], and the size of HspB8 is also controversial.

cvHsp is not heat-inducible and has no effect inhibiting the aggregation of temperature-induced standard substrates like luciferase, rhodanese and MDH (but has a limited impact on citrate synthase and GAPDH), unlike the other canonical chaperones like Hsp27, αA-crystallin and αBC[26,30]. However, cvHsp was the most effective chaperone to suppress the aggregation of the neurodegenerative disease-related polyQ protein[25]. The disordered and less conserved NTD has been suggested to be an indispensable region for cvHsp's potential chaperone activity towards anti-polyQ aggregation[4,31]. cvHsp is also effective to suppress the aggregation of the Parkinson's disease-associated PARK2 C289G mutant[32]. However, the localization of cvHsp makes it unlikely to perform these functions as a conventional chaperone to inhibit the aggregation of these substrates[22]. One of the realistic roles for cvHsp is to inhibit the aggregation of cytoskeletal proteins FLNC and G-actin through a non-chaperone mechanism[5,6,33]. Notably, though cvHsp can suppress aggregations in specific circumstances, it has never been reported to be capable of clearing formed aggregates by itself like Hsp27. Autophagic processes are found necessary and accompanied to degrade the relevant aggregates[6,22].

In this chapter, we report our high-resolution biophysical and structural analysis

of cvHsp. Using mostly crystallography and native MS, we show that cvHsp favors to be monomeric. cvHsp dimer is barely observed due to a unique β5-β6+7 region. The distinct cvHspCTD hinders its binding to the ACD, which makes cvHsp unlikely to form oligomers. We also confirmed the NTD is indispensable for cvHsp to inhibit the aggregation of heat-induced substrate citrate synthase (CS). These observations suggest a unique feature of an important cytoskeletal specific chaperone.

2.3 Results

2.3.1 cvHsp is dominantly monomeric

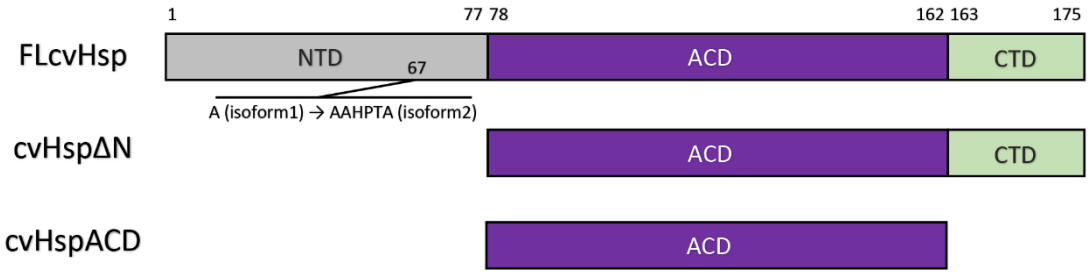


Figure 2.1. Schematics for cvHsp constructs. Isoform 2 is used as the full-length construct. cvHspΔN does not have the NTD, and cvHspACD contains only the α-crystallin domain. The difference of isoforms 1 and 2 comes from the splicing at A67.

There are three different isoforms of cvHsp generated by alternative splicing[1,2]. The longest isoform (isoform 2, 175 residues, all the followings in this thesis are based on isoform 2) was used to study in this thesis. Isoform 1 was not

successfully purified from several species by previous group member. Two truncated constructs cvHspACD (residues 78-162 of isoform 2) and cvHspΔN (residues 78-175 of isoform 2) were designed (Figure **2.1**), which are both conserved in isoform 1 and 2, while isoform 3 only has the NTD (Appendix figure **1.1-1.2**). cvHsp is small and is generally considered to be dimeric[25,31], since the dimeric form is the stable state for most of the well-studied sHsps, which has been supported by thermal stability, circular dichroism and molecular dynamics[34]. Native MS can give an accurate and consistent feature of the oligomeric states of given proteins at functional ranges of concentrations. By performing native MS experiments on cvHsp constructs, we surprisingly found that cvHsp is dominantly monomeric. Even dimeric form is hardly observed in neither the full-length nor the truncated constructs (Figure **2.2**). The spectrum of FLcvHsp shows strong signals at 2107, 2370 and 2711 m/z, representing a protein with a mass of 19 kDa, which is the mass of the FLcvHsp monomer. The other two truncated constructs also showed to be monomeric by native MS. A cysteine (C131) is presumably located on the edge of the β6+7 strand[35], which is the conventional dimer interface for human sHsps. Still, it cannot help in forming covalent homodimers at non-reducing conditions as the Hsp27 C137 locates in the center of the Hsp27ACD β6+7 strand[36].

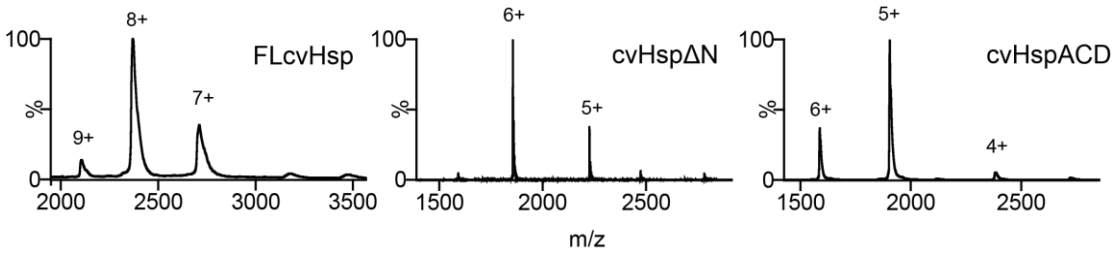


Figure 2.2. cvHsp is dominantly monomeric on native MS. Native MS of cvHsp constructs at 5 μM. The peaks series suggest that FLcvHsp exists as a monomer of 18953 Da. The main charge states are from 9+ to 7+. Peaks of cvHspΔN represent a monomeric protein of 11135 Da, and the cvHspACD also shows as a monomeric protein, with a mass of 9525 Da. Dimer peaks are negligible in all three constructs.

The challenge to handle with cvHsp and perform experiments on native MS was that all the three constructs have the same and very small extinction coefficient values (around 1490), which made the measurement of concentration based on UV-absorption (280 nm) very inaccurate when the protein concentration was not high enough. Some non-specific dimers or trimers could exist in the cvHspACD and cvHspΔN constructs if the concentrations were not measured correctly. A more consistent way to measure the concentration of cvHsp was to start from a high concentration and do some sequential dilutions. In this way, the concentration of cvHsp could be measured more precisely, and the native MS data was convincing. Another method that worked well to help quantify cvHsp was using its binding partner FLNC dimerized domain 24 (this is discussed in the following chapters), which has a large extinction coefficient value and can form a very tight 1:1 complex with cvHsp.

The full-length cvHsp (FLcvHsp) is not a stable construct due to the non-

structured NTD, and it was partially unfolded observed by native MS (Appendix figure **3.1**). The storage of this protein was also a pain since it tended to get self-truncated at -80 °C. The problems might come from a poly-serine region on the NTD. The role of this region hasn't been clarified clearly yet. The process of obtaining a stable FLcvHsp is still going on in the group. The cvHspACD and cvHspΔN constructs are stable at most conditions.

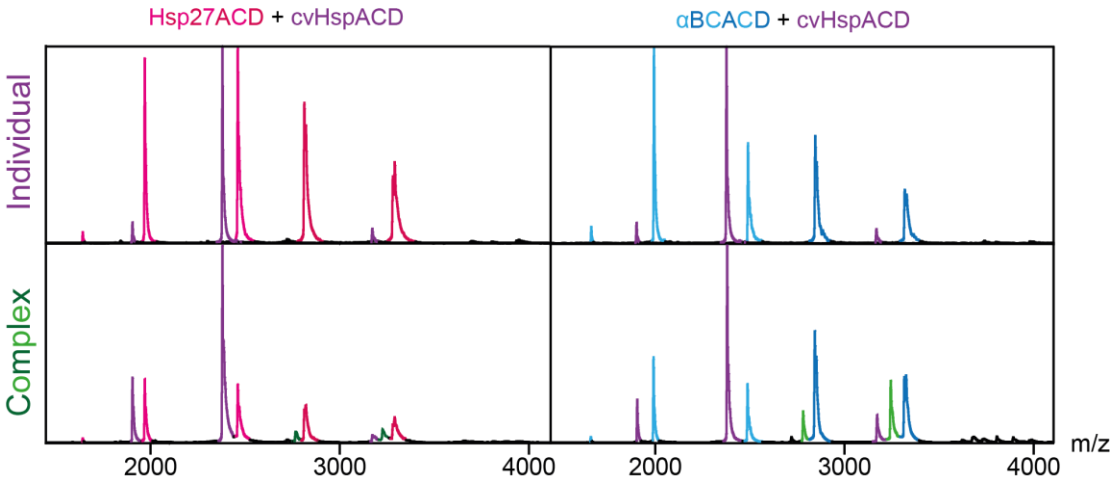


Figure 2.3. cvHspACD can form heterodimers with Hsp27ACD and αBCACD. Native MS of ACD constructs and the interactions between cvHspACD and the other two human sHsp ACD constructs at reducing environment with the addition of 200 mM DTT. The upper row is combined spectra of individual proteins, Hsp27ACD and αBCACD distribute as an equilibrium of monomers and dimers. Peaks in darker colors represent the dimer species. The bottom row shows the interactions between proteins mixed at the monomer ratio of 1:1. The green peaks represent the newly forming heterodimer complex.

The interactions within the sHsps can provide an insight into the behaviors of sHsp at native and functional conditions. Though cvHsp hasn't been reported to interact with the ubiquitous Hsp27 and αBC[4], we observed the heterodimer interfaces between cvHspACD and the ACD constructs of Hsp27 and αBC (Figure

2.3). We mixed the proteins at the monomer ratio of 1:1, and a new peak series came out for both coupling experiments, suggesting the formation of the new complex between the two proteins. The experiments were conducted under reducing conditions as Hsp27ACD can form covalent dimers. More heterodimers were observed between cvHspACD and αBCACD, and the formation of heterodimers didn't induce the homodimer formation of cvHspACD. The interaction between the full-length proteins still requires the participation of both CTD and NTD of both proteins, which needs to be further confirmed.

2.3.2 The crystal structure of cvHsp reveals that the dimeric form is unfavored

The unique characteristic of cvHsp makes it very valuable to pursue its structure at the atomic level. However, sHsps can be polydisperse and have dynamic terminals[37]. Crystallization is generally difficult for sHsps. Regardless of the species that the proteins come from, only a few human sHsps (HspB2, HspB3 and HspB6) have been crystallized with a nearly complete sequence[38,39]. The relatively conserved and rigid ACD is an ideal construct compared to the full-length protein for crystallization attempts. The ACD constructs of Hsp27 and αBC have been crystallized with the bound CTD peptides[36]. Inspired by this, we set up our crystal trials using both the cvHspACD and cvHspΔN constructs.

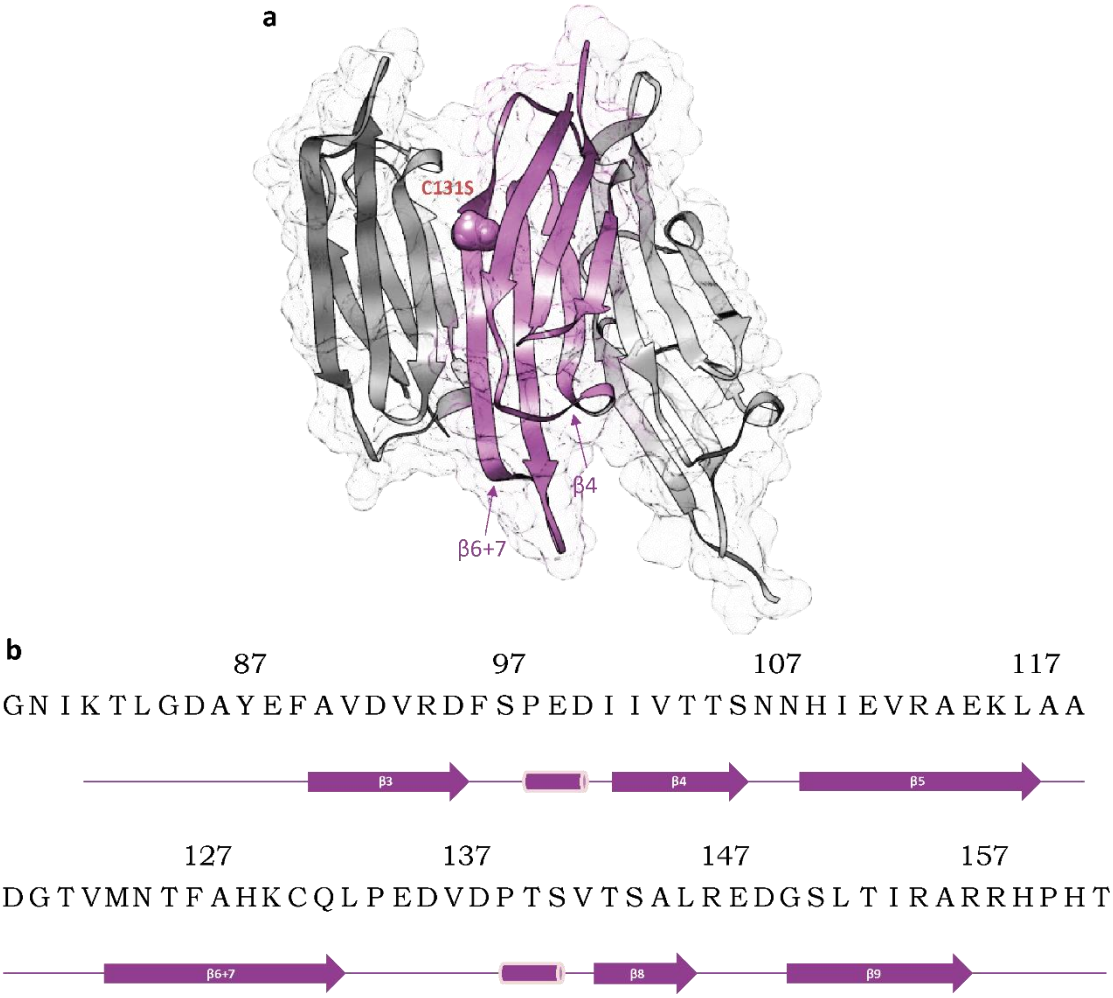


Figure 2.4. The framework of cvHspACD crystal structure. a. cvHspACD C131S crystal structure. The asymmetric unit contains three chains, and chain A is highlighted in the middle. The C131S residue is shown as the sphere. **b.** The sequence and corresponding secondary structural elements of cvHspACD C131S crystal structure. The C131S residue locates on the edge of the β6+7 strand in cvHspACD.

However, we never obtained any crystals with these two purified proteins, no matter how we manipulated the composition of storage buffer (from high-salt buffer to low-salt buffer), pH (from pH 6 to pH 8), and protein concentration (from around 5 to 30 μM). At first, we thought the terminals of the selected sequences might be flexible and inhibited the molecular packing during the crystal formation.

Site-specific enzymatic proteolysis was then applied to the cvHspΔN protein, using trypsin and chymotrypsin to incubate with the protein for different time lengths[40]. The sequence of the bulky protein after the removal of flexible tails on both ends by the enzyme was confirmed by HPLC-MS (Appendix figure **1.11**). Unfortunately, this new shorter ACD construct was failed to be expressed.

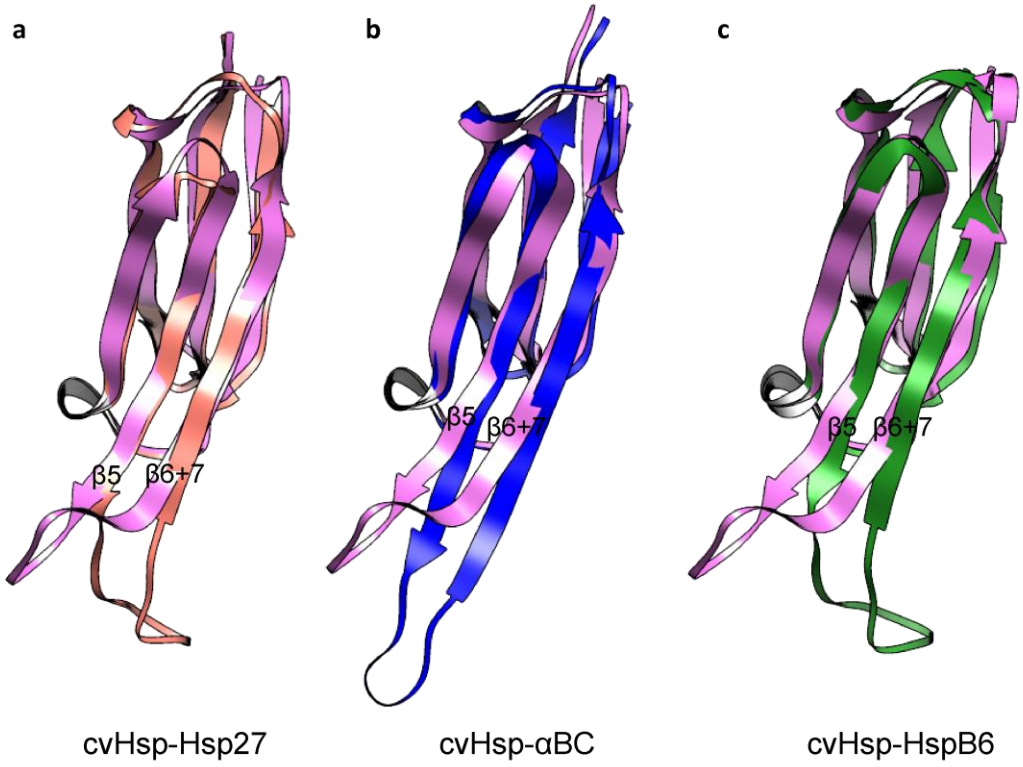


Figure 2.5. cvHsp is distinct at β5-β6+7 region. a-c. Structure alignments of cvHspACD C131S (orchid) with Hsp27ACD (4MJH, pink), αBCACD[36] (4M5S, blue) and HspB6ACD[41] (4JUS, green). The backbones of the structures are mostly aligned except for the flexible region between β5 and β6+7. **d.** Sequence alignments of the β5-β6+7 region of cvHsp, Hsp27, αBC and HspB6. The K/RHE motif is absent on cvHsp β5, which is the key factor to make cvHsp monomeric.

It then came to our mind that residue C131 presumed to be on the edge of the canonical dimer interface might cause some non-specific intermolecular hindrance in the crystal by forming covalent disulfide bonds at high concentrations. We then introduced a cysteine-to-serine (C131S) mutation at this residue for both constructs (Appendix figure **1.3-1.4**). Then magic happened once we set up the crystal trials after just one night. We obtained crystals in several conditions for the cvHspACD C131S protein. No crystals were observed for the cvHspΔN C131S construct, which suggested the CTD of cvHsp might be flexible and free from the ACD. The freshness of the protein was also a key to getting the crystals.

The single crystal being fished and successfully phased by molecular replacement after diffraction gave the structure in the $P4_12_12$ space group at 2.18Å. One asymmetric unit contains three chains, contacting through neighboring anti-parallel β4 and β6+7 strands, and this sheet continues to extend out, which suggests that forming antiparallel dimers through the β6+7 strands is not as stable as packing between the antiparallel β4 and β6+7 strands for cvHspACD molecules in the crystal (Figure **2.4**). It is also consistent with the native MS data showing that cvHspACD is dominantly monomeric. Residues LAADGTV (residues 112-118 of Isoform 2, the loop linking β5 and β6+7) of chain C is relatively disordered and lost most of the density on the 2Fo-Fc map, making the backbone hard to assign, which suggests that this loop can be a flexible region. Chains A

and B are generally well resolved (chain A is used for the following analysis).

Compared to the single chains in the dimeric HspB27ACD, αBCACD and HspB6ACD from PDB by structure alignments, our cvHspACD structure mostly overlaps with them. The RMSD values are 1.035, 1.336, 1.218 respectively, which suggests the structure of cvHspACD is closest to the Hsp27ACD structure (Figure **2.5a-c**). Compared to the other sHsp structures, the most significant difference of the cvHspACD structure comes from the β5 to β6+7 region. A conserved K/RHE motif on β5 in the other three proteins is absent in cvHsp, the shorter β5 gives a shorter β6+7 on the anti-parallel β sheet, while the other three human sHsps can form anti-parallel dimer pairing by their longer β6+7 strands (Figure **2.5d**). The loop between β5 and β6+7 bends outside the structure while this part of the other three structures is quite flat, though we know this region is flexible from chain C in the unit cell and from previous studies on other sHsps structures. In comparison to the β6+7 sequences of the other three proteins, the residues on this strand of cvHsp are varied and less hydrophobic for the side chain packing in dimer formation (Figure **2.5d**). The shorter and distinctive β5-β6+7 region can be the critical factor to explain that cvHsp is dominantly monomeric.

2.3.3 The CTD of cvHsp hinders its potential for self-oligomerization

The binding between the IPI/V motif of CTD and the β4-β8 hydrophobic groove of the ACD is essential for the oligomerization of sHsps[37]. The neighboring dimer units contact through the groove on ACD with flexible CTD tails to build the large oligomers[36,37]. The co-chaperone BAG3 also competes with the sHsp CTD to bind this groove and leads to the dissociation of sHsps oligomers[28,29,42].

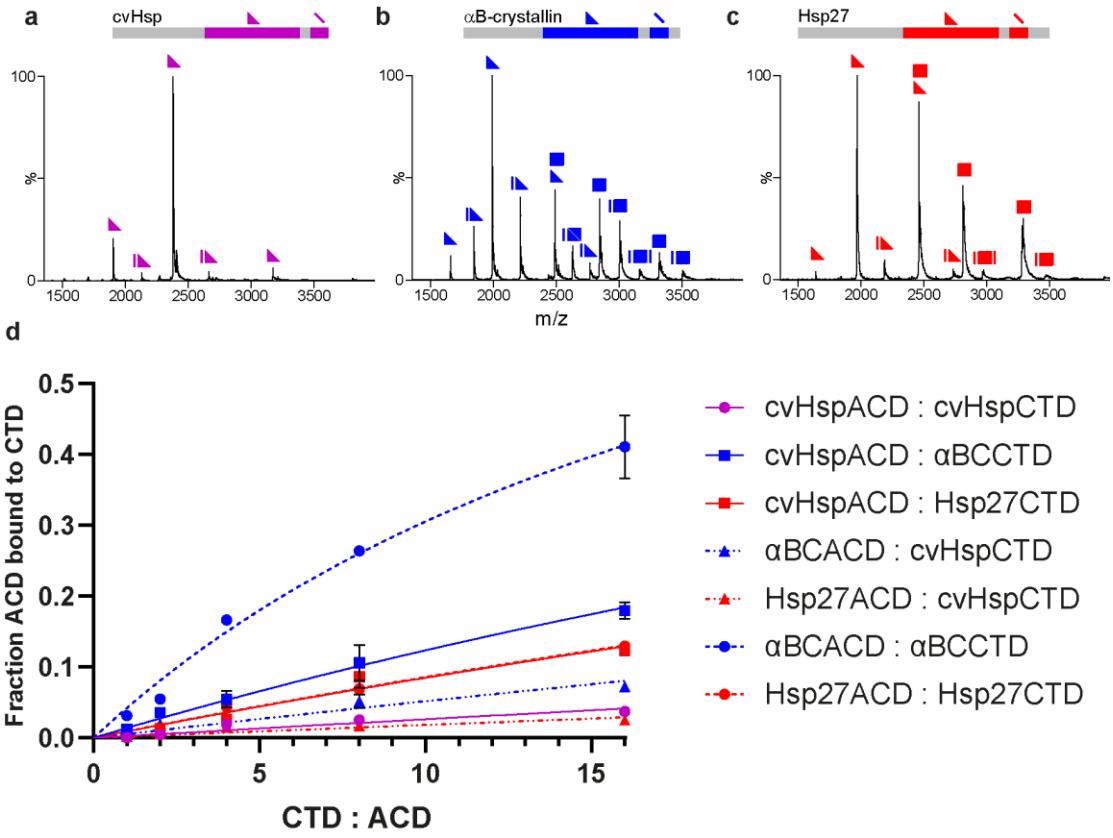


Figure 2.6. cvHsp is binding its CTD weakly. a-c. Native MS of 5 μM (monomer concentration) cvHspACD, αBCACD and Hsp27ACD interacting with 40 μM of their own CTD peptides. The small triangle represents the ACD, and the stick represents the CTD. **d.** Plots of the CTD and ACD titration experiments. ACD constructs remain at 5 μM. cvHspCTD performs weakly in binding all the sHsp ACDs.

To examine the ACD-CTD interface within cvHsp, we used a peptide to mimic the

eight most C-terminal residues (residue 168-175, TFRTE**IKI**) containing the sequence IKI, which can be a very similar candidate for the IPI/V motif. The IKI residues are far enough from the ACD to be connected via a flexible hinge. We conducted native MS titration experiments to test the binding affinities between the ACD constructs and the CTD peptides from Hsp27 (residue 179-185, IT**IPV**TF), αBC (residue 68-153, ERT**IPI**TRE) and cvHsp (Figure **2.6a-c**). In contrast to the ACD-CTD self-interaction of Hsp27 and αBC, the binding between cvHspCTD with cvHspACD is very weak. The peaks representing the binding can be neglected even the CTD peptide are several times in excess (Figure **2.6a**). cvHspCTD peptide is also poor at binding the ACD constructs of the other two proteins. αBCCTD with the motif IPI shows the highest binding affinities to the ACDs. The IPV motif in Hsp27CTD is less favorable.

	K_D (μM)	**ΔG (kJ/mol)**
cvHspACD : cvHspCTD	378.7 ± 57.4	-19.55 ± 0.38
cvHspACD : αBCCTD	71.33 ± 6.07	-23.67 ± 0.21
cvHspACD : Hsp27CTD	109.0 ± 11.5	-22.62 ± 0.26
αBCACD : cvHspCTD	186.2 ± 25.4	-21.30 ± 0.34
Hsp27ACD : cvHspCTD	558.6 ± 111.2	-18.61 ± 0.50
αBCACD : αBCCTD	22.84 ± 1.89	-26.49 ± 0.21
Hsp27ACD : Hsp27CTD	106.9 ± 5.2	-22.66 ± 0.12

Table 2.1. Thermodynamic parameters of the ACD-CTD interfaces.

The bindings of cvHspACD and Hsp27ACD with the CTD peptides are highly

similar with almost no difference in ΔG (Figure **2.6d**, Table **2.1**), so we look to the structure alignments again and focus on the β4-β8 hydrophobic grooves. The K_D and ΔG calculations are shown in appendix **4**, which can also be adapted to the rest of the calculations in the thesis.

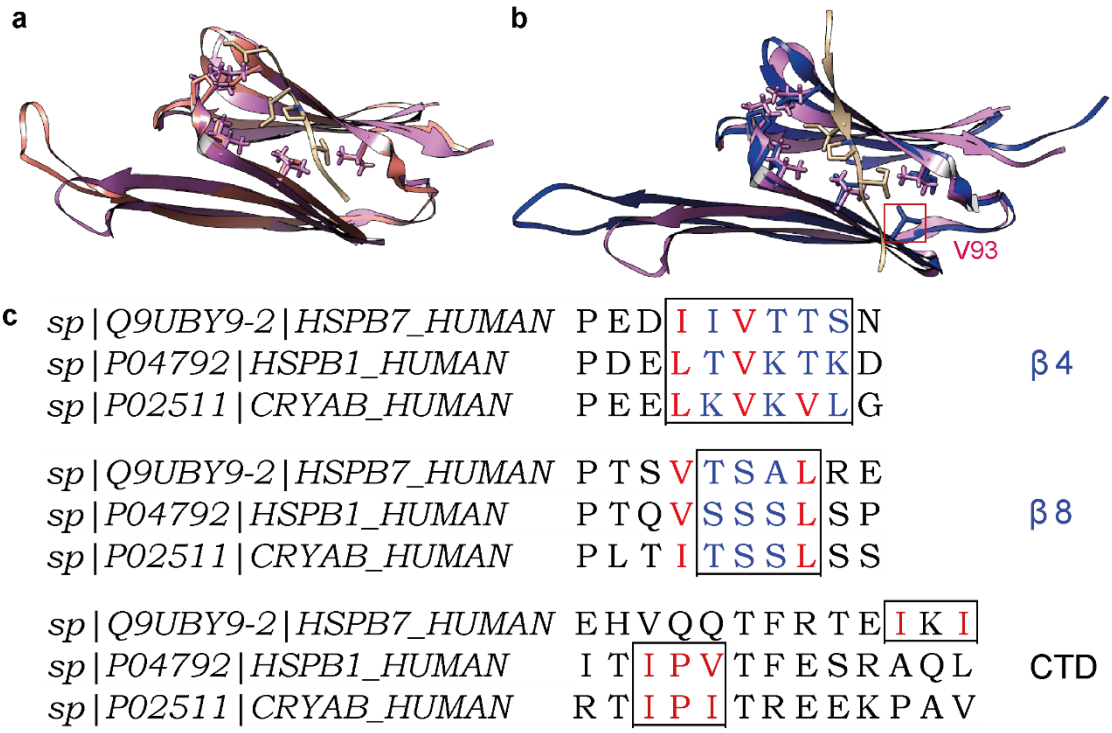


Figure 2.7. The hydrophobic residues are highly conserved on the β4-β8 grooves between cvHsp and Hsp27, and cvHspCTD does not have the proper IPI/V motif. a-b. Structure alignments of the β4-β8 grooves of cvHspACD C131S (orchid) with Hsp27ACD (4MJH, pink) and αBCACD[36] (4M5S, blue). The conformation and hydrophobic residues are mostly reserved between cvHsp, Hsp27 and αBC. αBC has one more hydrophobic residue on β4 (V93). The CTD peptides of Hsp27 and αBC in the 4MJH and 4M5S structures and hydrophobic residues are shown as sticks. **c.** Sequence alignments of the β4, β8 and CTD regions of cvHsp, Hsp27 and αBC. Hydrophobic residues within the strand are labeled to be red, and the residues within the strand are framed.

By comparing the structures, the conformations of the β4-β8 grooves are mostly identical between cvHspACD, HspB27ACD and αBCACD. And the hydrophobic

residues to bind the CTD peptides are also mostly conserved to be hydrophobic between the three structures, while αBCACD has one more hydrophobic residue at V93 (Figure **2.7a, b**). These explain the similar binding capabilities between cvHspACD and HspB27ACD, and αBCACD has a more hydrophobic β4-β8 groove. The tighter binding between cvHspACD and αBCCTD proves that the distinct cvHspCTD is the key factor to its weak interaction with cvHspACD. Through sequence alignments, the IKI residues are also far from the typical locations of IPI/V motifs (Figure **2.7c**).

2.3.4 The NTD is indispensable for cvHsp's potential chaperone activity

As a member of human sHsp, the chaperone activity of cvHsp can be a good reference to probe its real-life function. We tested the chaperone activity of cvHsp using the three constructs (FLcvHsp, cvHspACD and cvHspΔN). Citrate synthase (CS) from pig heart is of the few substrates that can be suppressed from aggregation by cvHsp *in vitro* from the previous studies[26]. We picked this heat-induced standard substrate and tested its aggregation without chaperones in PBS and HEPES buffer. We found the performance of CS was more consistent in HEPES as a standard buffer. The absorbance was recorded for the whole night, and the aggregation took about 2 hours to reach the end, which is the flat region of the curve. CS was induced at 45 °C with the addition of reducing agent DTT

(Figure **2.8**).

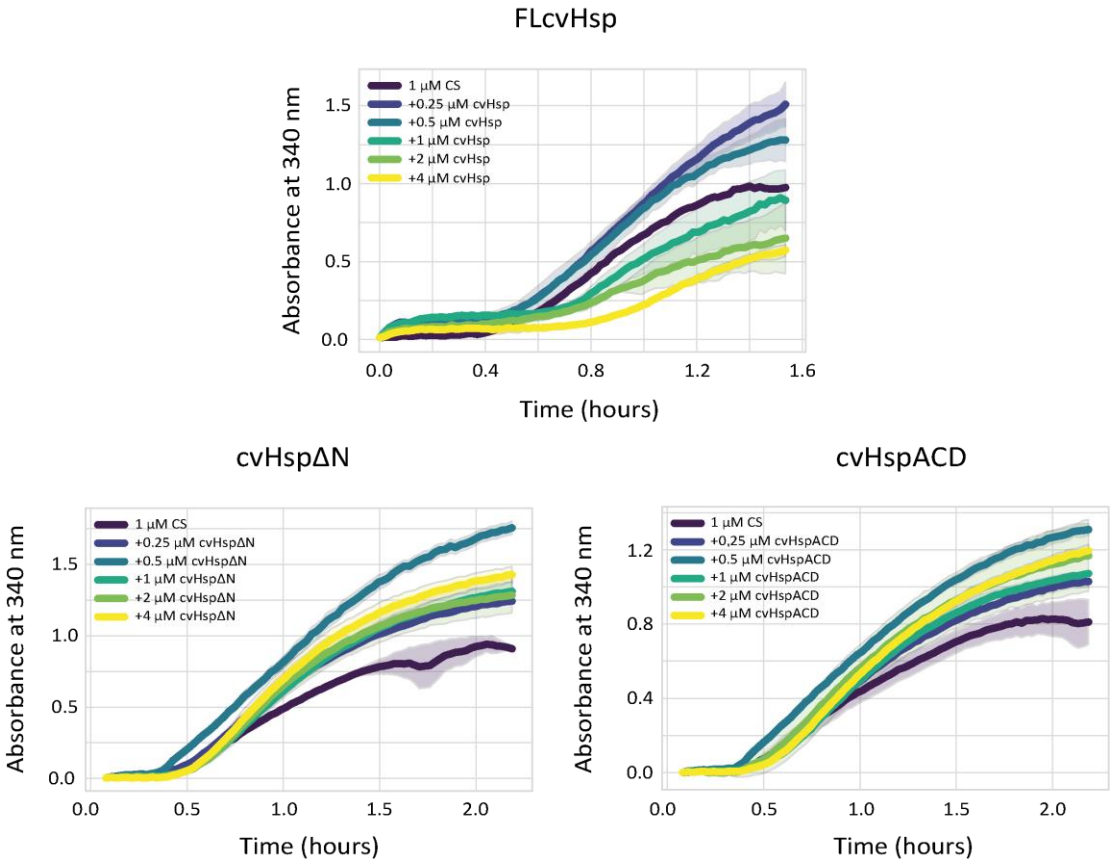


Figure 2.8. The full-length cvHsp is chaperone-active towards CS with the indispensable NTD. The chaperone assays of FLcvHsp, cvHspΔN and cvHspACD using the heat-induced CS as the substrate. The purple curve represents the absorbance of the aggregation of CS without the chaperone. FLcvHsp starts to act as a chaperone towards CS once its concentration is above 0.5 μM, while cvHspΔN and cvHspACD without the NTD lose the chaperone function.

The same condition was adapted when the chaperones were added. When the concentration of FLcvHsp was lower than the CS concentration, it did not work as the chaperone and showed no suppression to the aggregates, and it just increased the complexity of the system. The absorbance curves went higher compared to the CS aggregation without the chaperone on the plot. But once the

ratio of FLcvHsp and CS was equal to and above 1:2, cvHsp started to work as a chaperone and gave a better suppression effect with more protein added. On the plot, these curves are lower than the aggregation of CS itself. These results suggested that the FlcvHsp can perform chaperone activity towards the aggregation of CS. We also found that cvHspACD and cvHspΔN have no effect in suppressing the aggregation of CS. The addition of cvHspACD and cvHspΔN only increased the heterogeneity of the solution. Thus, we know that the NTD of cvHsp is the critical region to perform the chaperone activity and suppress the aggregation of CS.

2.4 Discussion

The oligomeric states of cvHsp were dubious before our study and were generally considered to be an equilibrium between monomers and dimers based on experiences from other sHsps[4,31]. However, we have shown that cvHsp is dominantly monomeric by native MS and explained that dimerization is less favorable because of its unique canonical β5-β6+7 region on the ACD from the crystal structure. This situation should apply to all three isoforms of cvHsp, as the variations all come from the NTD. Hsp27 has also been crystallized in the non-dimeric (PDB 3Q9P) unit packing form[43]. The parallel ACD chains contact through very few residues between the neighboring β4 and β6+7 strands to form a

hexameric ring. Notably, there are two Glu to Ala mutations in the 3Q9P structure to facilitate this hexamer formation, and the two Glu residues are absent in cvHsp, as we discussed before. In the cvHsp structure, the neighboring chains are anti-parallel, but β6+7-β4 contacting is more favorable than the β6+7-β6+7 packing in the crystal. This is also solid evidence showing that cvHsp is less favorable to be dimeric.

The prevalent AlphaFold system uses the potential of mean force to describe the initial shape of protein to further predict protein folding structure rapidly and accurately[44]. The structures of almost the whole proteome (98.5% of human proteins) have been predicted in the AlphaFold protein structure database[45]. The database also gave the structural prediction of full-length human cvHsp isoform 1. Our crystal structure of cvHspACD can align well with the Alphafold structure, and the RMSD value is 1.200. We noticed that the β6+7 strand of cvHsp is predicted to be attached with a peptide (FPARPGGAG, residues 65-73 of isoform 1) from NTD, which is the part before ACD β2 (Figure **2.9**). F65 and P66 are close to V123 and M124 to form hydrophobic contacts, which can be the most significant interactions within this surface. In the Hsp27 and αBC AlphaFold predicted structures, the β6+7 strands are exposed and free from any part of the protein, suggesting that dimerization through the β6+7 is available to stabilize the protein. The structural prediction result is another evidence showing that cvHsp can be stabilized by itself to keep monomeric through this part of the peptide from

the NTD. It is also interesting that the splicing starts from residue A67. Insertion of 5 amino acids from this residue gives cvHsp isoform 2. What effect does this insertion do on cvHsp β6+7 strand deserves further exploration, and there can be further experiments on cvHspACD with this part of peptides.

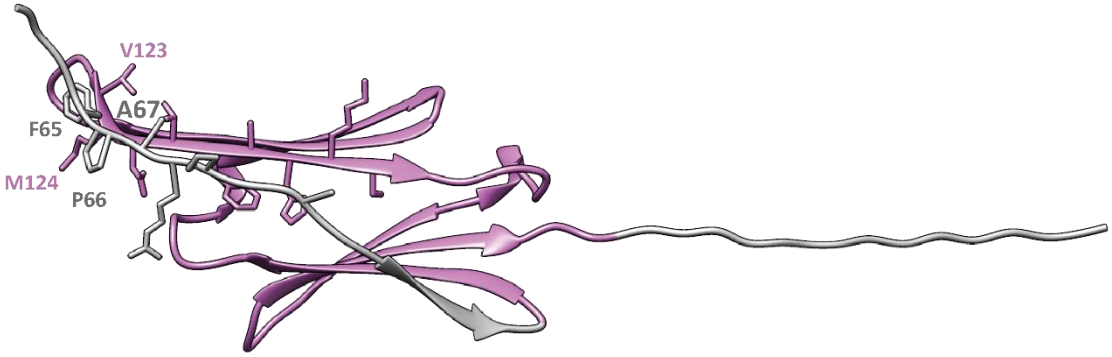


Figure 2.9. The AlphaFold-predicted structure of cvHsp isoform 1. The ACD of the structure is colored to be orchid. Residues 1-60 have been removed, and the rest of the protein is shown in grey. The residues on β6+7 and on NTD that participate in the contact are shown as sticks[45].

We also found that cvHspACD is very similar compared to the Hsp27ACD structure, especially at the β4-β8 groove. The two structures have highly conserved hydrophobic residues in the pocket, and they have similar binding capacities to the CTD peptides. The IKI residues on cvHspCTD look like a similar candidate to the IPI/V motif but still different, and the CTD hinders the probable oligomerization of cvHsp. In the cvHsp AlphaFold predicted structure, the CTD is also free from the ACD, which can also support our data (Figure **2.9**). This suggests that the proline within the motif is also important. In Hsp27, this proline to leucine variant can weaken the interaction between the ACD and CTD to loosen

the oligomer formation, and this mutant is related to Charcot-Marie-Tooth disease[46]. This conserved ACD can also be a reference to explain why both cvHsp and Hsp27 are weak binding partners for the co-chaperone BAG3[28]. BAG3 uses two IPV motifs to compete with the sHsp CTD to bind the β4-β8 groove[29]. The proline-to-leucine mutation on one of the IPV motifs in BAG3 is also myopathy related[47]. HspB8, HspB6 and αBC have much higher binding affinities compared to cvHsp and Hsp27[28]. The distinct ACDs can be the key factor to influence this interaction. Some clues show that the heart-specific cvHsp, HspB8 and HspB6 don't form large oligomers, which might provide an interesting insight into the human sHsp genome evolution. The key modification to getting the crystal of cvHspACD is the C131S mutation, and this residue is located on the edge of the β6+7 strand, which is now confirmed by the crystal structure. Though C131 cannot influence the dimerization of cvHsp, its location on the outer surface still suggests that this residue might have some independent roles, likely in electrophile signaling, which might be worth further investigations[35].

The NTD is the key factor for cvHsp to perform its chaperone activity to suppress the aggregation of CS, which is consistent with the studies about cvHsp in the inhibition of polyQ proteins[25,31]. It is notable that cvHsp has the isoform with only the NTD, suggesting that cvHsp is performing different roles in different compartments or stages. cvHsp has been observed to target the aggregates at a late stage and form a ring around to prevent further aggregation. The NTD might

be indispensable for recognition and targeting of the aggregates and formation of the ring for further autophagic clearance[25].

Our study of cvHsp on the atomic level provides a deeper insight into this specific member of the human sHsp family. We reveal that cvHsp is dominantly monomeric because of its unique ACD, but its CTD does not have the exact motif at the proper position to bind its preserved β4-β8 groove. The chaperone function towards CS does exist and mostly depends on its NTD, but the cellular role of this cardiac-specific protein has more to be explored.

Acknowledgements

I want to thank Miranda Collier for leading me into this exciting and worth-exploring project when I started my DPhil. She developed the current purification protocol for the FLcvHsp. I thank Jani Bola, Michael McDonough and especially Timothy Allison for their great help on the structural determination of cvHsp. I'm also grateful to Reid Alderson for the Python code for the aggregation assay and help me with the assay experiments during his short stay in Oxford. Guodong Cao has been working on the molecular dynamics (MD) analysis on the ACD structures to support our structural analysis mentioned in this chapter, which is not included in this thesis.

2.5 Materials and methods

Expression and purification of cvHsp truncated constructs

Human cvHspACD (residues 78-162 of isoform 2, WT and C131S), cvHspΔN (residues 78-175 of isoform 2, WT and C131S) encoded in the pET28a(+) vector with N-terminal 6×His-tag were transformed in *Escherichia coli* BL21(DE3) cells (Agilent). A Gly-Ser overhang was left between the tobacco etch virus (TEV) protease recognition sites and the NTD of the constructs after the TEV proteolysis. Cells were used to inoculate a 10 mL lysogeny broth (LB) media with kanamycin (30 μg/mL) at 37°C overnight and transferred to a 1 L large LB culture with the same antibiotics. The culture was shaking at 180 RPM at 37°C until the OD600 reached 0.6-0.8 and followed by the induction with IPTG to a final concentration of 0.5 mM. After another overnight shaking at 22°C, the cells were harvested, pelleted and lysed using French Press Microfluidizer with the addition of EDTA-free protease inhibitor cocktail and buffer containing 300 mM NaCl, 50 mM Tris and 20 mM imidazole (pH 8.0). The lysate was spun at 20,000 g for 20 min at 4°C and filtered through 0.22 μm filter to clarify before loading to the HisTrap HP column. The protein was eluted with the buffer with 300 mM NaCl, 50 mM Tris and 500 mM imidazole (pH 8.0) and dialyzed to the loading buffer with TEV protease at room temperature overnight. The sample was again loaded to the HisTrap HP column to remove the His-tag and exchanged to buffer containing 50

mM Tris (pH 8.0) for cvHsp constructs using 3.5 kDa Amicon centrifugal concentrator. cvHsp constructs were loaded to the HiTrap Q anion exchange column and eluted with a gradient of 150-200 mM NaCl in the elution buffer. The pure protein was then buffer exchanged to PBS buffer with 5 mM TCEP for storage.

Expression and purification of FLcvHsp Isoform 2

Full-length human cvHsp (Isoform 2, A67 to AAHPTA, 175 residues) encoded in pET23a was transformed into BL21(DE3) pLysS cells and expressed in LB media (with 0.5% glucose, 50 mg/mL ampicillin, and 37 mg/mL chloramphenicol) for 16 hours at 22°C following cell culture growth to OD600 between 0.6-0.8 and induction with 0.5 mM IPTG. Cells were lysed using French Press Microfluidizer in 20 mM Tris, 1 mM EDTA, 5 mM β-mercaptoethanol (BME) at pH 7.4 with the EDTA-free protease inhibitor cocktail. The lysate was centrifuged at 20,000 g for 20 min at 4°C and filtered through 0.22 μm filter, which pelleted the desired protein in inclusion bodies. The pellet was washed with the lysis buffer supplemented with 1% triton X-100 2 times and then 1 M urea to remove membrane phase contaminants. The protein was then solubilized in 6 M urea and 10 mM sodium phosphate at 4°C and refolded by removing urea gradually in dialysis buffers containing 3 M, 2 M, and 1 M urea with 100 mM NaCl, 20 mM Tris, 5 mM BME (pH 7.5) buffering the process.

Crystallization and structure determination of cvHspACD C131S

Fresh, unfrozen cvhspACD C131S protein was buffer-exchanged to 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 8.0) and concentrated to 10 mg/mL. 96-well sitting-drop vapor diffusion plates (Hampton Research) were used for crystallography. The crystal for structural determination of cvHspACD C131S came out after one day, and it was collected from the condition with 0.2 M magnesium formate dihydrate and 20% w/v polyethylene glycol (PEG) 3350. The crystal was cryo-protected by adding 10% glycerol and then fished and flash-frozen in liquid nitrogen.

Property	**cvHspACD C131S**
Space group	$P4_12_12$
Cell constants	57.86Å 57.86Å 167.34Å
a, b, c, α, β, γ	90.00° 90.00° 90.00°
Resolution (Å)	54.68-2.18
Data completeness (%)	99.1
(in resolution range)	(54.68-2.18)
Refinement program	PHENIX 1.18.2_3874
R, R_{free}	0.265, 0.305
Average B, all atoms ($Å^2$)	70.0
Total number of atoms	3508

Table 2.2. Diffraction data and refinement statistics of the cvHspACD C131S structure.

The cryo-protected crystal was diffracted at beamline I04 at Diamond Light Source, UK. The diffraction data were integrated, merged and scaled automatically by

Xtriage. The phase was determined using the comparative model based on HspB6 (4JUS) structure at Robetta comparative modeling server[48,49]. Phenix Molecular Replacement (MR) and Autobuild programs were used to phase the structure. The side chain of chain C Arg147 was removed for a significant B factor. The refinement was conducted using the Phenix Refinement tool and the PDB validation server as a reference, and the structure was finalized with the R and R_{free} values at 0.265 and 0.305.

Native MS

The full-length cvHsp was buffer-exchanged into 1 M ammonium acetate to get stabilized, while the other proteins were bio-spun into 200 mM ammonium acetate (pH 6.9). The peptides were dissolved in water and added to the proteins in ammonium acetate straightforward. Gold-coated capillaries were pulled and coated in house. The needles were pulled from the P-1000 Flaming / Brown Micropipette Puller (Sutter Instrument) and coated from the Agar Auto Sputter Coater (Sutter Instrument). All the spectra were recorded from the Synapt G1 mass spectrometer modified for the transmission of intact non-covalent protein complexes (Waters Corp.). A reservoir of acetonitrile was introduced at the electro-spray source to reduce the charge states and stabilize the protein-peptide interaction[50]. The parameters were set as: capillary, 1.5 kV; sampling cone, 20-35 V; extraction cone, 3.0-3.3 V; backing pressure, 3.8-4.1 mbar; trap gas (argon) flow, 1.5-3.0 mL/min; trap energy, 6-8 V; transfer energy, 6-12 V. Three technical

repeats were collected for each ratio in the titration experiments. The peak intensities were extracted manually in UniDec[51]. The dissociation constants were analyzed using GraphPad Prism 8.0.1 through the specific binding with the Hill slope model.

CS aggregation assay

The reaction buffer was 40 mM HEPES with 1mM DTT, pH 7.5. CS from the porcine heart (Sigma-Aldrich Co Ltd.) and cvHsp constructs were buffer-exchanged into the HEPES buffer. The final concentration of CS was maintained at 1 μM while the concentrations of cvHsp constructs were 0.25-4 μM. The samples were transferred to the 96-microcell plates. The aggregation of CS was induced at 45 °C and recorded for at least two hours in the Fluostar Optima plate reader (BMG Lab Technologies) at 340 nM wavelength for absorbance change. Three technical repeats were conducted for each ratio.

Bibliography

1. Lorenzo, P. *et al.* Identification and Characterization of cvHsp. *Journal of Biological Chemistry* **276**, 12201–12211 (1999).
2. Kappé, G. *et al.* The human genome encodes 10 α-crystallin-related small heat shock proteins: HspB1-10. *Cell Stress and Chaperones* **8**, 53–61 (2003).
3. Golenhofen, N., Perng, M. der, Quinlan, R. A. & Drenckhahn, D. Comparison of the small heat shock proteins αb-crystallin, MKBP, HSP25, HSP20, and cvHSP in heart and skeletal muscle. *Histochemistry and Cell Biology* **122**, 415–425 (2004).
4. Vos, M. J., Kanon, B. & Kampinga, H. H. HSPB7 is a SC35 speckle resident small heat shock protein. *Biochimica et Biophysica Acta - Molecular Cell Research* **1793**, 1343–1353 (2009).
5. Juo, L. Y. *et al.* HSPB7 interacts with dimerized FLNC and its absence results in progressive m. *Journal of Cell Science* **129**, 1661–1670 (2016).
6. Liao, W. C., Juo, L. Y., Shih, Y. L., Chen, Y. H. & Yan, Y. T. HSPB7 prevents cardiac conduction system defect through maintaining intercalated disc integrity. *PLoS genetics* **13**, e1006984 (2017).
7. Doran, P., Martin, G., Dowling, P., Jockusch, H. & Ohlendieck, K. Proteome analysis of the dystrophin-deficient MDX diaphragm reveals a drastic increase in the heat shock protein cvHSP. *Proteomics* **6**, 4610–4621 (2006).

8. Doran, P., Gannon, J., O'Connell, K. & Ohlendieck, K. Aging skeletal muscle shows a drastic increase in the small heat shock proteins αB-crystallin/HspB5 and cvHsp/HspB7. *European Journal of Cell Biology* **86**, 629–640 (2007).

9. Rüdebusch, J. *et al.* Dynamic adaptation of myocardial proteome during heart failure development. *PLoS ONE* **12**, 1–13 (2017).

10. Stark, K. *et al.* Genetic association study identifies HSPB7 as a risk gene for idiopathic dilated cardiomyopathy. *PLoS Genetics* **6**, 1–9 (2010).

11. Matkovich, S. J. *et al.* Cardiac signaling genes exhibit unexpected sequence diversity in sporadic cardiomyopathy, revealing HSPB7 polymorphisms associated with disease. *Journal of Clinical Investigation* **120**, 280–289 (2010).

12. Cappola, T. P. *et al.* Common variants in HSPB7 and FRMD4B associated with advanced heart failure. *Circulation: Cardiovascular Genetics* **3**, 147–154 (2010).

13. Fang, X. *et al.* Loss-of-function mutations in co-chaperone BAG3 destabilize small HSPs and cause cardiomyopathy. *Journal of Clinical Investigation* **127**, 3189–3200 (2017).

14. Garnier, S. *et al.* Involvement of BAG3 and HSPB7 loci in various etiologies of systolic heart failure: Results of a European collaboration assembling more than 2000 patients. *International Journal of Cardiology* **189**, 105–107 (2015).

15. Wang, M. *et al.* Association between polymorphisms of the HSPB7 gene and Cheyne–Stokes respiration with central sleep apnea in patients with dilated cardiomyopathy and congestive heart failure. *International Journal of Cardiology* **221**, 926–931 (2016).

16. Esslinger, U. *et al.* Exome-wide association study reveals novel susceptibility genes to sporadic dilated cardiomyopathy. *PLoS ONE* **12**, (2017).

17. Wu, T. *et al.* HSPB7 is indispensable for heart development by modulating actin filament assembly. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 11956–11961 (2017).

18. Rosenfeld, G. E., Mercer, E. J., Mason, C. E. & Evans, T. Small heat shock proteins Hspb7 and Hspb12 regulate early steps of cardiac morphogenesis. *Developmental Biology* **381**, 389–400 (2013).

19. Lahvic, J. L. *et al.* Small heat shock proteins are necessary for heart migration and laterality determination in zebrafish. *Developmental Biology* **384**, 166–180 (2013).

20. Tobin, S. W. *et al.* Regulation of Hspb7 by MEF2 and AP-1: Implications for Hspb7 in muscle atrophy. *Journal of Cell Science* **129**, 4076–4090 (2016).

21. Jin, C., Shuai, T. & Tang, Z. HSPB7 regulates osteogenic differentiation of human adipose derived stem cells via ERK signaling pathway. *Stem Cell Research and Therapy* **11**, 1–13 (2020).

22. Mercer, E. J., Lin, Y. F., Cohen-Gould, L. & Evans, T. Hspb7 is a

cardioprotective chaperone facilitating sarcomeric proteostasis. *Developmental Biology* **435**, 41–55 (2018).

23. Ulbricht, A. *et al.* Cellular mechanotransduction relies on tension-induced and chaperone-assisted autophagy. *Current Biology* **23**, 430–435 (2013).

24. Sun, X. *et al.* Interaction of Human HSP22 (HSPB8) with Other Small Heat Shock Proteins. *Journal of Biological Chemistry* (2004) doi:10.1074/jbc.M311324200.

25. Vos, M. J. *et al.* HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. *Human Molecular Genetics* **19**, 4677–4693 (2010).

26. Mymrikov, E. v., Daake, M., Richter, B., Haslbeck, M. & Buchner, J. The chaperone activity and substrate spectrum of human small heat shock proteins. *Journal of Biological Chemistry* **292**, 672–684 (2017).

27. Yang, Z., Wang, Y., Lu, Y. & Zhao, X. Molecular characterization of rat cvHsp/HspB7 in vitro and its dynamic molecular architecture. *Molecular Medicine Reports* **4**, 105–111 (2011).

28. Fang, X., Bogomolovas, J., Trexler, C. & Chen, J. The BAG3-dependent and -independent roles of cardiac small heat shock proteins. *JCI insight* **4**, 1–12 (2019).

29. Fuchs, M. *et al.* Identification of the key structural motifs involved in HspB8/HspB6-Bag3 interaction. *Biochemical Journal* **425**, 245–255 (2010).

30. Mymrikov, E. v., Bukach, O. v., Seit-Nebi, A. S. & Gusev, N. B. The pivotal

role of the β7 strand in the intersubunit contacts of different human small heat shock proteins. *Cell Stress and Chaperones* **15**, 365–377 (2010).

31. Wu, D. *et al.* The N terminus of the small heat shock protein HSPB7 drives its polyQ aggregation-suppressing activity. *Journal of Biological Chemistry* **294**, 9985–9994 (2019).

32. Minoia, M., Grit, C. & Kampinga, H. H. HSPA1A-Independent Suppression of PARK2 C289G Protein Aggregation by Human Small Heat Shock Proteins. *Molecular and Cellular Biology* **34**, 3570–3578 (2014).

33. Ke, L. *et al.* HSPB1, HSPB6, HSPB7 and HSPB8 Protect against RhoA GTPase-Induced Remodeling in Tachypaced Atrial Myocytes. *PLoS ONE* **6**, (2011).

34. Santhanagopalan, I. *et al.* It takes a dimer to tango: Oligomeric small heat shock proteins dissociate to capture substrate. *Journal of Biological Chemistry* **293**, 19511–19521 (2018).

35. Surya, S. L. *et al.* Cardiovascular Small Heat Shock Protein HSPB7 Is a Kinetically Privileged Reactive Electrophilic Species (RES) Sensor. *ACS Chemical Biology* **13**, 1824–1831 (2018).

36. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proceedings of the National Academy of Sciences of the United States of America* **111**, (2014).

37. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philosophical Transactions of the Royal Society B:*

Biological Sciences **368**, 1–13 (2013).

38. Clark, A. R. *et al.* Terminal Regions Confer Plasticity to the Tetrameric Assembly of Human HspB2 and HspB3. *Journal of Molecular Biology* **430**, 3297–3310 (2018).

39. Sluchanko, N. N. *et al.* Structural Basis for the Interaction of a Human Small Heat Shock Protein with the 14-3-3 Universal Signaling Regulator. *Structure* **25**, 305–316 (2017).

40. Wernimont, A. & Edwards, A. In Situ proteolysis to generate crystals for structure determination: An update. *PLoS ONE* **4**, (2009).

41. Weeks, S. D. *et al.* Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6. *Journal of Structural Biology* **185**, 342–354 (2014).

42. Muranova, L. K., Ryzhavskaya, A. S., Sudnitsyna, M. v., Shatov, V. M. & Gusev, N. B. Small Heat Shock Proteins and Human Neurodegenerative Diseases. *Biochemistry (Moscow)* **84**, 1256–1267 (2019).

43. Baranova, E. v. *et al.* Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *Journal of Molecular Biology* **411**, 110–122 (2011).

44. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

45. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* (2021) doi:10.1038/s41586-021-03828-1.

46. Reid Alderson, T. *et al.* A weakened interface in the P182L variant of HSP27 associated with severe Charcot - Marie - Tooth neuropathy causes aberrant binding to interacting proteins. *The EMBO Journal* **40**, 1–23 (2021).

47. Myers, V. D. *et al.* The Multifunctional Protein BAG3: A Novel Therapeutic Target in Cardiovascular Disease. *JACC: Basic to Translational Science* vol. 3 122–131 (2018).

48. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).

49. Raman, S. *et al.* Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins: Structure, Function and Bioinformatics* **77**, 89–99 (2009).

50. Hopper, J. T. S., Sokratous, K. & Oldham, N. J. Charge state and adduct reduction in electrospray ionization-mass spectrometry using solvent vapor exposure. *Analytical Biochemistry* **421**, 788–790 (2012).

51. Marty, M. T. *et al.* Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Analytical Chemistry* **87**, 4370–4376 (2015).

Chapter 3

cvHsp binds FLNC to compete with its dimer interface, and the interaction is regulated by FLNC phosphorylation

3.1 Abstract

Filamin-C (FLNC) is a dimerized actin-linking protein that is mainly expressed in striated muscles. The knock-out of cvHsp can cause the aggregation of FLNC and the collapse of the cytoskeletal system. We observed upregulation and interaction of cvHsp and FLNC from three heart stress models in mice. We confirmed that the monomeric cvHsp is binding the FLNC dimerized domain 24 (d24) specifically and competing with FLNC dimerization. Through crystallography and HDX, we demonstrated that cvHsp uses its β4 to bind strand C, which is one of the dimer interfaces of FLNCd24. An extreme hydrophobic interface between the two proteins makes this interaction intense and specific. The binding surface also involves cvHspACD β5 and FLNCd24 strand D. Several residues on FLNC d24 can be phosphorylated at different stress modes. We found the two most significant phosphorylation sites at FLNC T2677 and Y2683 in response to electric stimulation and cancer respectively. Phosphorylation at FLNC T2677 strengthens the dimerization of FLNCd24 and weakens its interaction with cvHsp, while the phosphorylated Y2683 gives completely opposite effects. Both residues are shown to be important on the FLNC dimer interface and complex interface in the crystal structures. Combined with studies on other domains of FLNC, we conclude that FLNC dimer tends to be more rigid with less motility to deal with acute stress, and the dimer is interrupted to make FLNC more mobile in tumor tissues. This study shows a specific chaperone-substrate interaction and how the process is regulated by post-translational modifications (PTMs) at stress conditions.

3.2 Introduction

Filamins are a family of actin-linking proteins. They cross-link the 3-dimensional F-actin cytoskeleton and maintain the rigidity and flexibility of the plasma membrane of cells[1]. Among the family, filamin C (FLNC) is mainly targeted in Z-discs and sarcomeres of striated muscle and heart[2,3], and its two other homologous isoforms, filamin A and filamin B (FLNA and FLNB), are ubiquitous but not dominantly expressed in striated muscles[2,4,5]. FLNC has important roles in signaling pathways, cell migration and differentiation and helping the assembly of Z-disc[6–10]. Natural mutations on FLNC are closely related to dilated cardiomyopathy (DCM) and muscle-skeletal myopathies[11–14].

The actin-binding domain of FLNC locates on the N-terminal, followed by 24 immunoglobulin-like (Ig) domains (d1-d24). There are two hinges between d15 and d16 (hinge 1), and d23 and d24 (hinge 2), while hinge 1 can be absent during myogenesis caused by splicing. D1-d15 forms a rigid rod, and d16-d24 perform most of the functions of FLNC through interactions with a variety of binding partners (Figure **3.1a**). D20 contains a large insertion of 82 amino acids to localize the protein in Z-discs. It has significant contributions together with the hinges to make FLNC a mechano-sensitive scaffold protein towards tensions[1,6,15]. The last domain d24 mediates the homo-dimerization of FLNC[13,16]. The study based on purified full-length FLNC is challenging, but the available structural

investigations towards the individual filamin Ig domains reveal that the Ig repeats share highly similar frames. Each Ig domain is made of seven strands (A-G), forming two β-sheets. D24 forms the anti-parallel dimer through strands C and D[16]. Interestingly, the binding sites for interacting partners at other domains have also been found to be limited within these two strands[6,17–22]. Domain pairs exist within the single chain of filamin, d16-d17, d18-d19 and d20-d21 pack in pairs[23–26]. In the FLNA d20-d21 pair, the strands C-D surface of d21 is blocked by strand A of d20. The application of mechano-force can shift strand A to expose the surface for the binding partners[23].

The behavior of FLNC is mediated by post-translational modification (PTM). Phosphorylations at Ser, Thr, Tyr or other residues are canonical PTMs at stress conditions[27,28]. Protease calpain-1 cleaves hinge 2 to release FLNC from Z-disc to other compartments with reserved functions from other domains. This process can be inhibited effectively by protein kinase C α (PKCα) -mediated phosphorylation at S2623/S2624, while the calpain-1 cleavage site is Y2625[29–31]. The insertion region at d20 is the binding site for filamin A-binding protein 1 (FILIP1). FILIP1 binding can trigger chaperone assisted selective autophagy (CASA) in the filamin degradation process. FILIP1 and FLNC are both upregulated during skeletal muscle cell differentiation, while FLNA is down-regulated. Akt and PKCα phosphorylate two sites individually at the FLNC insertion region to inhibit its interaction with FILIP1, thus protecting FLNC from being degraded[32–34].

Phosphorylation also exists on the binding partners to regulate FLNC. Phosphorylated Hsp27 (S78 and S82) obtains a large exposure area to bind the FLNC insertion region, making the insertion more compact and rigid[35] (Figure **3.12**). cvHsp is the binding partner of FLNC dimerized d24. The knockout of cvHsp causes FLNC aggregation in mouse models, but cvHsp doesn't seem to take part in the clearance of FLNC aggregates[36,37], which suggests that cvHsp is protecting FLNC from aggregation through a non-chaperone mechanism.

Our experimental approaches confirm the interaction between cvHspACD and FLNCd24, and we observe that the FLNCd24 dimer dissociated during the binding. We investigated the binding interfaces between the two proteins and revealed indispensable hydrophobic contacts within the interface. There are two representative phosphorylation sites on FLNC T2677 and Y2683 in response to different stress modes that can give reversed effects to FLNC dimerization and interaction with cvHsp. Our studies point out a proteostasis-associated interaction regulated by PTM in the cardiac system.

3.3 Results

3.3.1 cvHsp binds FLNCd24 strongly and competes to its dimer interface

We investigated the interaction between cvHsp and FLNC in the heart. To mimic the natural environments of the interaction, three cardiac biomechanical stress models in mice were established to test the location and expression level of the target proteins at specific stress conditions. The first model is the muscle LIM protein (MLP) knockout mice. MLP is also targeted specifically in striated muscles as FLNC, and the absence of this protein results in the morphological loss in muscle[35,38]. The second mouse model is through the transverse aortic constriction (TAC) surgery, which can overload the pressure in cardiac tissue to induce heart failure (HF). The third one is the mice with isoprenaline/epinephrine (ISO/PE) treatment, which gives a drug-induced blood pressure increase effect.

We observed that cvHsp and FLNC were both upregulated and interacting in the MLP KO model from immunoprecipitation (IP). Both proteins were upregulated compared to their controls respectively from Western Blot, and the MLP knockout model had the most significant response (Figure **3.1b, c**). The frozen cardiac ventricular tissues from the three mice models were also stained for the immunofluorescence experiments. cvHsp was found more intense at sarcomeres and intercalated discs. The co-localization and upregulation of both proteins were confirmed (Figure **3.1d**).

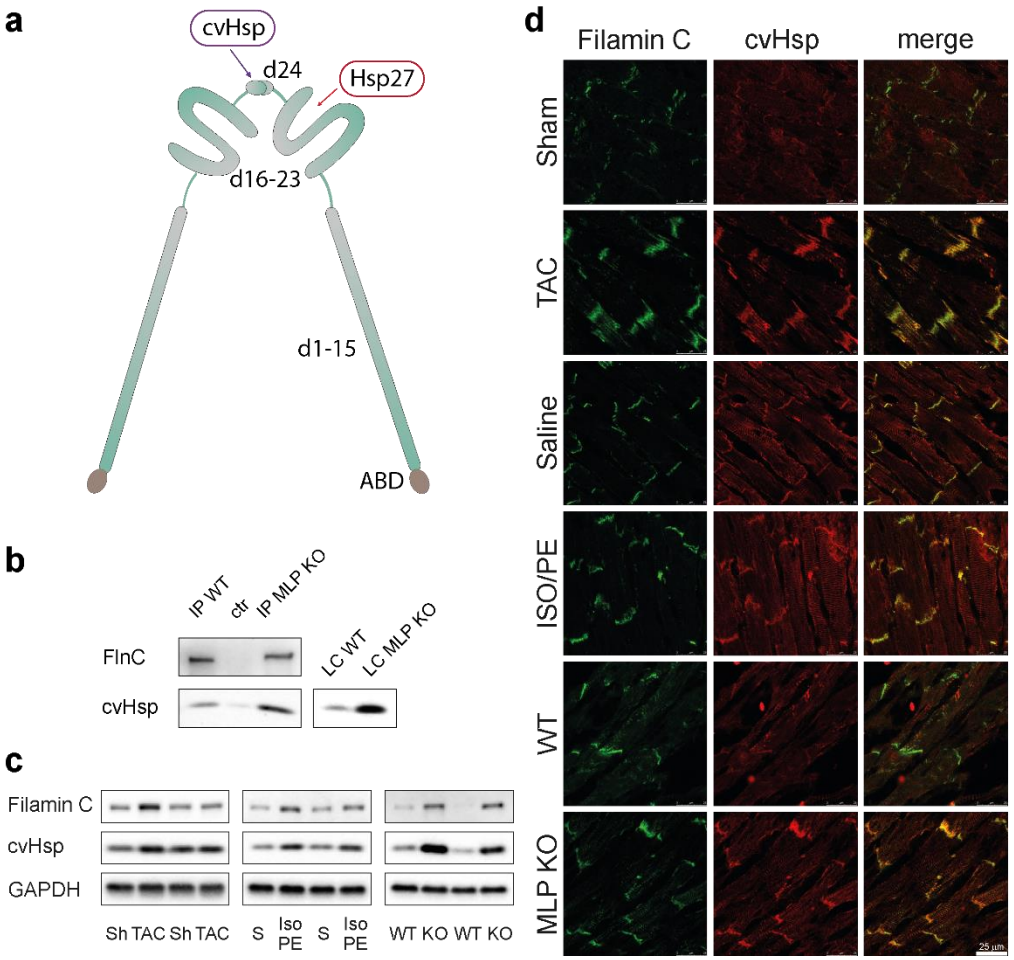


Figure 3.1. cvHsp and FLNC are upregulated and interacting in the heart stress models in mice. a. Schematic of dimeric FLNC protein and domains. Figure modified from the original of Miranda Collier. cvHsp has been supposed to be the binding partner of FLNCd24, and Hsp27 interacts with the insertion region[35,37]. **b.** Immunoprecipitation of cvHsp and FLNC from MLP KO mice hearts ventricular tissue, isotype control and lysate control (LC). The co-precipitation of two proteins suggests their interaction. **c.** Western Blots for FLNC and cvHsp in TAC, ISO/PE and MLP KO mice hearts. GAPDH works as the loading control (Sh: Sham, S: Saline. Specific controls for TAC and ISO/PE). **d.** Immuno-fluorescence of cvHsp and FLNC from frozen sections of ventricular tissues from TAC, ISO/PE and MLP KO mice models. Sections were stained for FLNC and cvHsp and counter-stained for both. Both proteins are intense at sarcomeres and intercalated discs upon stress, and TAC gives the most significant effect. Data was generated by Katja Gehmlich and Charlotte Hooper.

FLNCd24 has been reported to be the binding domain for cvHsp[37]. To test the interaction, we expressed and purified the FLNCd24 protein (Appendix figure **1.5**).

On the spectrum, the peaks at 1794 and 2153 m/z represent charge states 6+ and 5+ for a protein with a mass of 10759 Da, which can be assigned to the FLNCd24 monomer. The other peaks at 2392, 2691 and 3075 m/z represent charge states 9+, 8+ and 7+ of a protein with a mass of 21519 Da, which is the FLNCd24 dimer. FLNCd24 existed to be dominantly dimeric by native MS, and a dissociated monomer was also observable (Figure **3.7a**). We mixed FLNCd24 with the cvHsp constructs (cvHspACD, cvHspΔN and FLcvHsp). A new peak series showed up after mixing FLNCd24 with cvHsp constructs, which suggested the interactions between FLNCd24 and cvHsp. As for the interaction between FLNCd24 and cvHspACD, peaks at 1585 and 1904 m/z represents cvHspACD monomer with a mass of 9509 Da. Signals on the spectrum at 2151, 2390 and 2693 represent the FLNCd24 dimer with a mass of 21512 Da. After mixing these two proteins, a new peak series came out with strong signals at 2252 and 2535 m/z, representing a protein with a mass of 20265 Da, which can be assigned to the protein complex formed by cvHsp monomer and FLNCd24 monomer. No peaks represent the formation of a complex that contains cvHsp monomer and FLNCd24 dimer. The bindings were intense, and almost no individual species of FLNCd24 or cvHsp constructs could be observed at the ratio of 1:1 (Figure **3.2**).

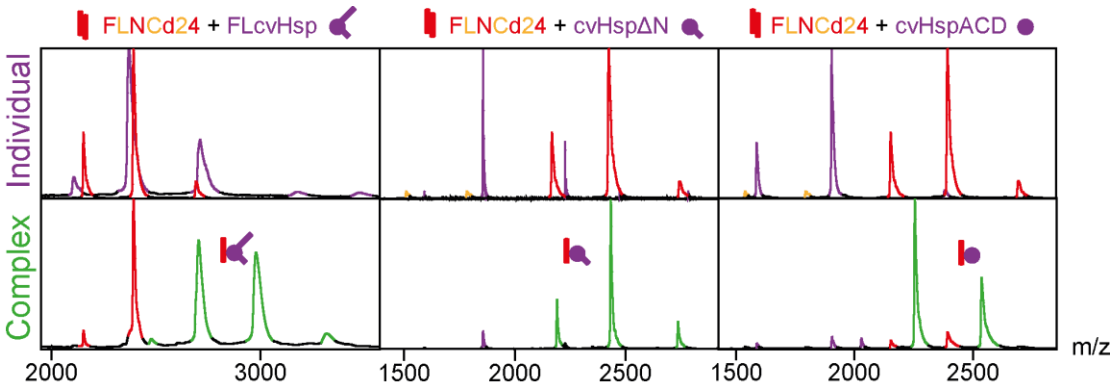


Figure 3.2. cvHsp is binding FLNCd24 and competes to FLNC dimerization using its ACD. Native MS experiments show that FLNCd24 (dimer: red, monomer: orange) can form tight complexes with FLcvHsp, cvHspΔN and cvHspACD (orchid). FLNCd24 is dimeric dominantly with a mass of 21570 Da. Green peak series represent the complex formed by FLNCd24 and cvHsp constructs. The complex contains FLNCd24 monomer and monomeric cvHsp constructs. FLNCd24 dimer is not observed in the formed complex. The upper row is combined spectra of individual proteins, the bottom row shows the interactions between proteins mixed at the monomer ratio of 1:1. FLcvHsp is partially unfolded, so FLNCd24 is in excess in the coupling on the spectrum.

Because of the instability of cvHsp NTD, FLcvHsp is partially unfolded, as we mentioned in Chapter 2, so FLNCd24 was in excess in the spectrum of the interaction with FLcvHsp. The strengths of the bindings were at similar levels, which suggests the binding site of cvHsp for the FLNCd24 interaction locates within the ACD.

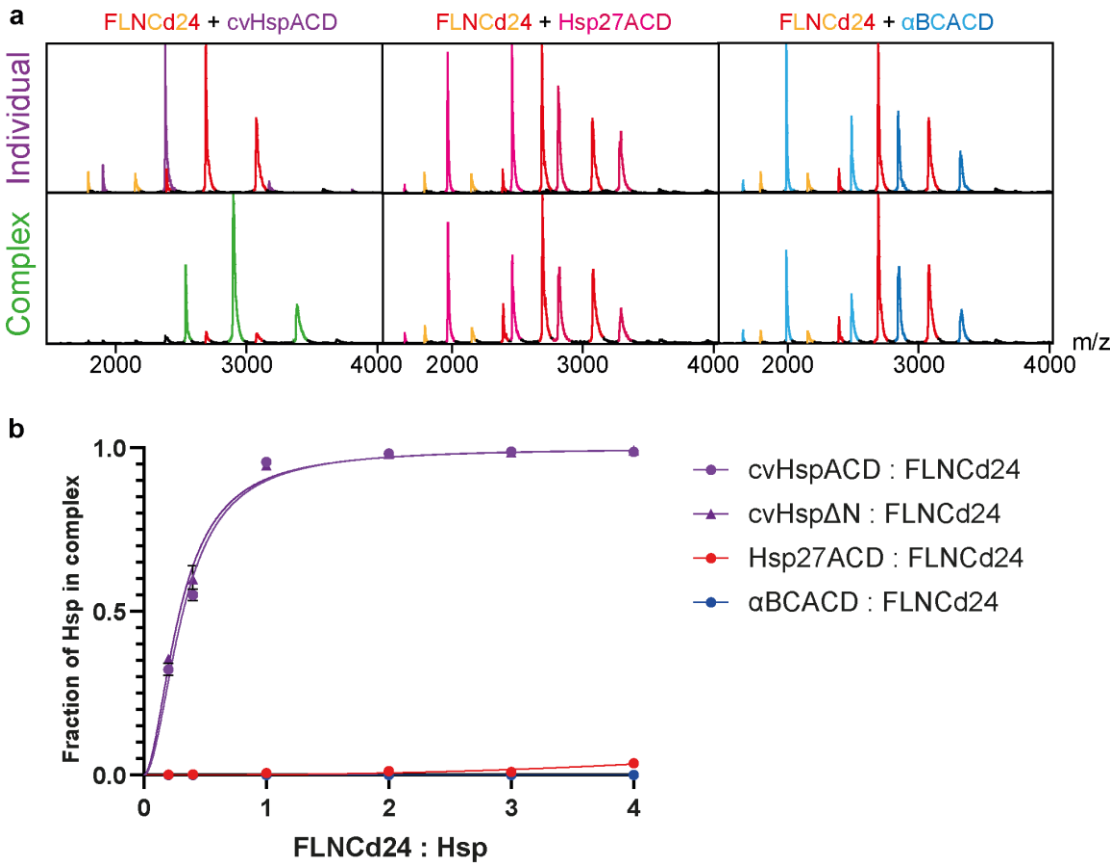


Figure 3.3. FLNCd24 is targeting cvHspACD specifically. a. Native MS experiments show that FLNCd24 binds cvHspACD strongly but does not interact with Hsp27ACD and αBCACD at reducing conditions. Hsp27ACD and αBCACD distribute as an equilibrium of dimers and monomers. Dimer peaks are labelled in darker colors. The upper row is combined spectra of individual proteins, the bottom row shows the interactions between proteins mixed at the monomer ratio of 1:1. **b.** Plots of native MS titration experiments of the interactions between FLNCd24 and sHsp constructs. cvHspACD has a similar binding capacity towards FLNCd24 compared to cvHspΔN. The interactions of FLNCd24 with αBCACD and Hsp27ACD are negligible. The concentrations of Hsp monomer are maintained at 5 μM, and the concentrations of FLNCd24 are varied.

Surprisingly, no FLNCd24 dimer was observed in the complex binding with cvHsp. All the peak series representing the complex suggest the mass of cvHsp plus monomeric FLNCd24. The result means cvHsp is binding FLNCd24 and out competes its dimerization. To quantify this interaction, we performed native MS titration experiments on FLNCd24 with cvHspACD and cvHspΔN but without the

FLcvHsp (Figure **3.5b**). The similar ΔG values of both interactions are around -37 kJ/mol, suggesting that the binding affinities with FLNCd24 are close between cvHspACD and cvHspΔN (Table **3.1**). The CTD and NTD are unlikely to have essential roles in this interaction.

Hsp27ACD and αBCACD were also tested as control experiments with FLNCd24 by native MS. Almost no new peak series represented the complex could be observed (Figure **3.3**), which means there are no interactions between FLNCd24 and the ACD of these two sHsps. These results presented preliminary information that cvHspACD is binding FLNCd24 close to its dimer interface, and this interaction is very strong and specific. We further attempted to study this interaction in more detail.

	K_D (nM)	**ΔG (kJ/mol)**
cvHspACD : FLNCd24	321.7 ± 24.8	-37.05 ± 0.19
cvHspΔN : FLNCd24	294.4 ± 19.0	-37.26 ± 0.16
Hsp27ACD : FLNCd24	41418 ± 33610	-26.33 ± 2.80
αBCACD : FLNCd24	/	/

Table 3.1. Thermodynamic parameters of the sHsp-FLNCd24 interactions.

3.3.2 Structural analysis of cvHsp-FLNCd24 interaction

The behavior of the cvHsp-FLNCd24 interaction makes it very valuable and essential to conduct structural studies. It took almost two years to obtain the

crystal structure of the cvHspACD-FLNCd24 complex. The cvHspACD C131S mutation was still introduced to facilitate crystallization, as we mentioned in Chapter 2. More cvHspACD C131S were mixed with FLNCd24 protein to break dimeric FLNCd24 molecules into monomeric with cvHspACD in the complex because we assumed that the cvHspACD-FLNCd24 complex has a similar mass with FLNCd24 dimer. The concentrated protein mixture was separated by SEC, and the proteins isolated from the peak representing the complex from the chromatography was used for crystallography (Appendix figure **2.1**).

The structure was phased and refined in space group $P3_221$ at 2.85Å. cvHspACD β4 and FLNCd24 strand C form parallel β-sheet (Figure **3.4a**). Three groups of hydrophobic contacts on the interface make this interaction strong and specific. The MMM (residue 2667-2669) motif on FLNCd24 strand C aligns exactly with IIV (residue 101-103) on cvHspACD β4. FLNC M2667 and M2669 also contact L146 and V142 on cvHspACD β8 respectively, and cvHsp I102 is close to V2684 on the FLNCd24 strand D (Figure **3.4b**). Through the sequence alignments with the other human sHsps (Hsp27, αBC and HspB6), most of the hydrophobic residues on cvHsp that are essential in the interaction with FLNCd24 remain to be hydrophobic in other sHsps. cvHsp I102 is the only important residue in binding FLNCd24 that is not conserved to be hydrophobic in other human sHsps (Figure **3.4b, c**). This side chain of this residue is pointing inside the β4-β8 groove, so it is not involved in the CTD binding as we discussed in the previous chapter.

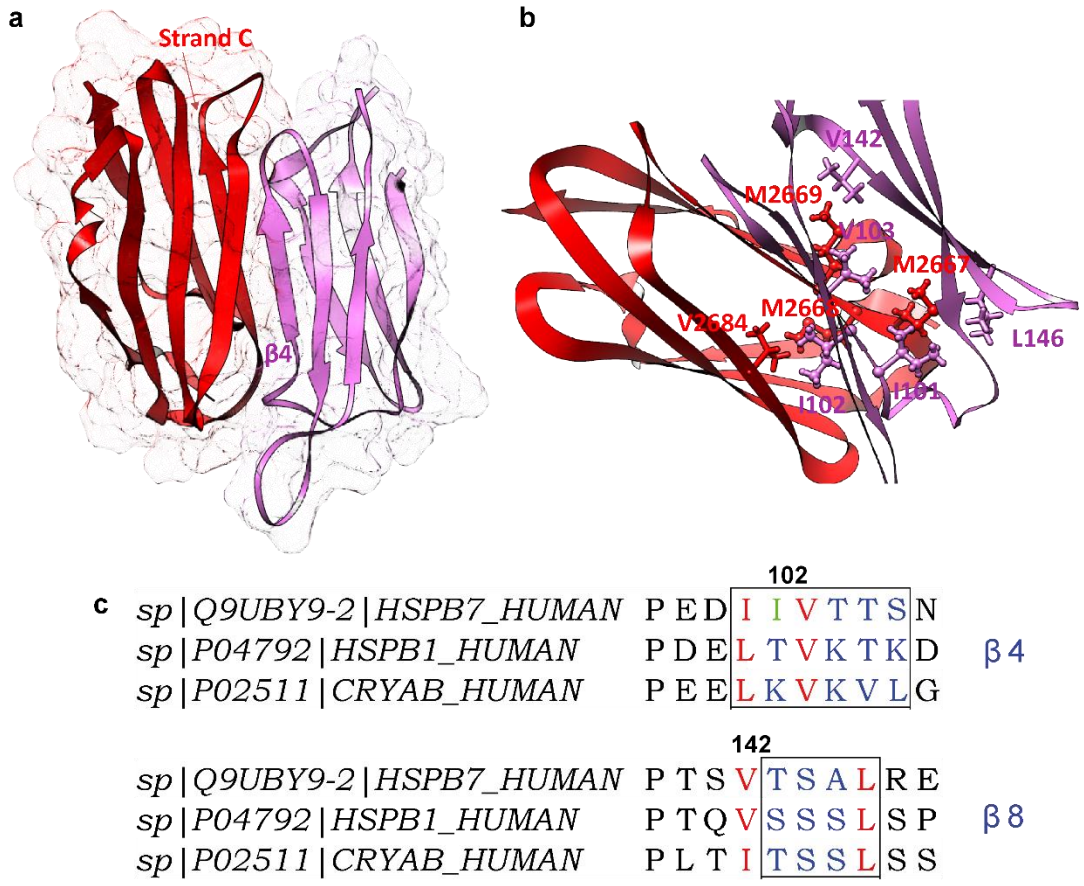


Figure 3.4. The complex structure shows that the cvHspACD-FLNCd24 interaction is specific. a. The cvHspACD (orchid)-FLNCd24 (red) complex crystal structure with the surface. The contact interface is between cvHspACD β4 and FLNCd24 strand C. **b.** Hydrophobic contacts between the complex interfaces. FLNC M2667 is packing with cvHsp V103 and V142. FLNC M2669 is closest to cvHsp I101 and L146. The specific hydrophobic residue cvHsp I102 is contacting with FLNC M2668 and V2684. **c.** Sequence alignments of the β4 and β8 regions of cvHsp, Hsp27 and αBC. Compared to cvHsp, most of the hydrophobic residues (orange) essential for the interaction with FLNC are conserved in Hsp27 and αBC except for cvHsp I102 (light green).

Considering the highly similar structural conformations of the β4-β8 regions in cvHsp, Hsp27 and αBC, it is obvious that cvHsp I102 is the critical residue that evolved to bind FLNCd24 during genome evolution. Another residue pair between FLNC R2690 and cvHsp E99 form salt-bridge contact, which also contributes to

the interaction. The binding interface also involves the packing of cvHspACD β5 and FLNCd24 strand D.

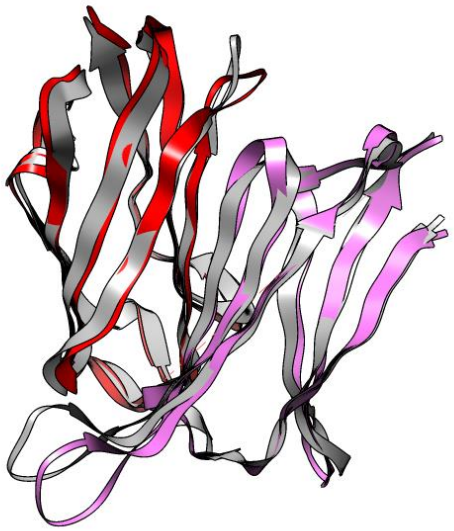


Figure 3.5. The backbones of cvHspACD and FLNC are mostly conserved in the complex structure compared to their apo states. Structure alignment of the complex (cvHspACD chain: orchid; FLNCd24 chain: red) with cvHspACD chain A (dark grey) and FLNCd24 (1V05, dark grey).

Compared to the apo-cvHsp (chain A) and FLNCd24 structures (PDB 1V05), the backbones of the two proteins in the complex are conserved mostly. The RMSD values are 1.189 for cvHspACD and 0.753 for FLNCd24. For cvHsp, most differences come from the loop between β5 and β6+7. This region is flexible, as we discussed in Chapter 2. The β4 strand elongates in the complex for hydrogen bondings with FLNCd24 strand C. The loop between FLNCd24 strand C and D twists in the complex (Figure **3.5**), this region is found important in regulating the cvHsp-FLNC interaction in the next paragraph. The side chains in FLNCd24 fluctuate more than those in cvHsp. Molecular dynamics (MD) are being processed

in the meantime to quantify these changes.

We also did hydrogen-deuterium exchange (HDX) MS experiments to characterize the binding interface and support the crystal structure. HDX can be conducted in the functional buffer, which can be regarded as a more “native” condition. The sequence coverages of both proteins were well achieved on the HDX-MS platform, which guaranteed the basis for the later experiments. Only the first two amino acids on N-terminal for FLNCd24 and F89, S144 and A145 for cvHspACD were missed. The sequence coverage values for cvHspACD and FLNCd24 were 96.6% and 98.0% respectively. The binding site of cvHpsACD was studied when mixed at the ratio of 1:1 with FLNCd24 (25 μM of monomer concentration before quenching). Data were collected at five exposure time points from 30s to 5000s. Compared to the apo cvHspACD, the most protected peptide in all the time points in cvHspACD covered residues 101-115 (IIVTTSNNHIEVRAE), which is almost the whole region of β4-β5. The most protected region from deuterium uptake in cvHspACD matches with the binding site of cvHspACD in the complex crystal structure. The second most shielded region covers residues 153-162, which is the last part of β9. Residues 91-100, which involves the flexible region between β3 and β4, was also protected at this level. This part is close to the main binding interface of cvHspACD, so it is understandable that it can be protected slightly (Figure **3.6a**).

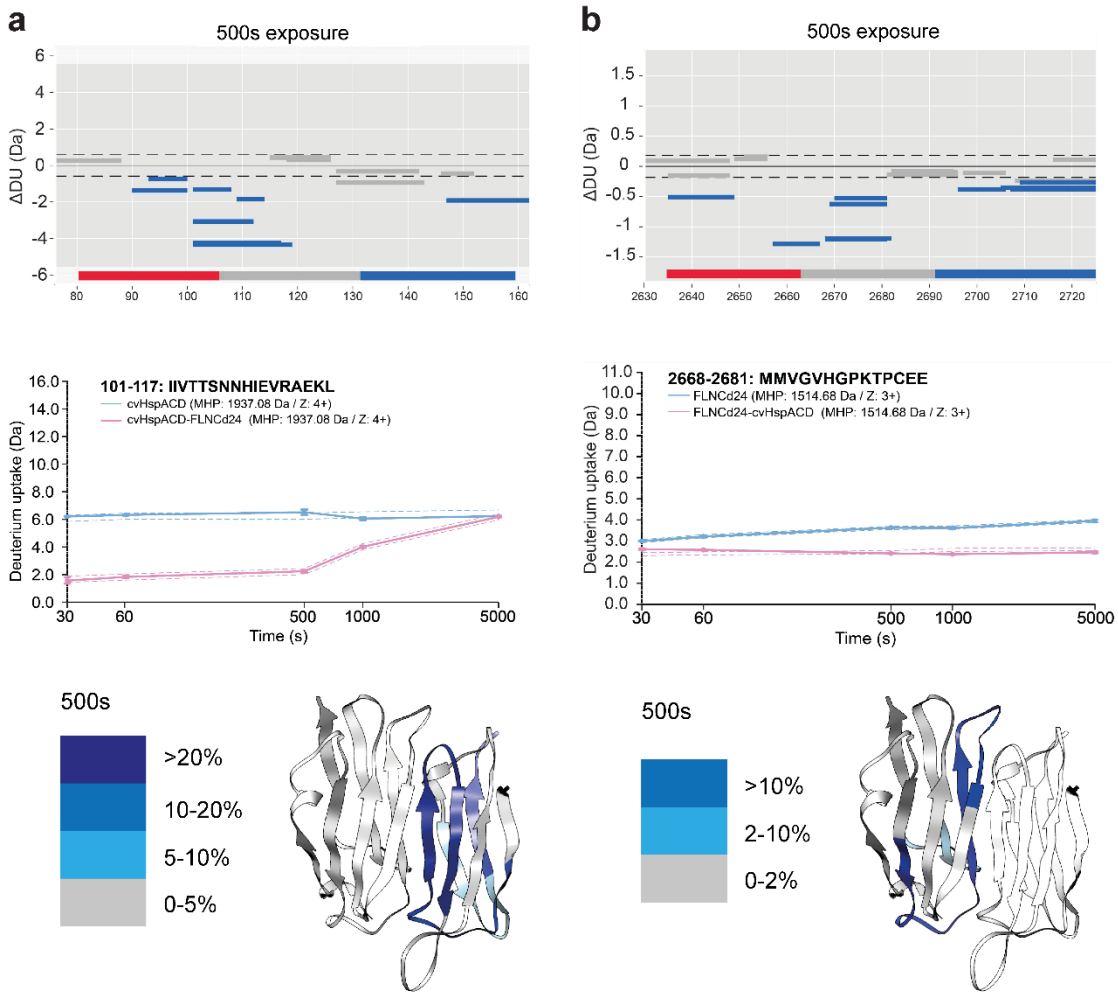


Figure 3.6. HDX studies of cvHspACD and FLNCd24 in the complex compared to their apo-states. a. Deuterium uptake studies of cvHspACD in the complex. The top figure is the Woods plot showing the comparison of deuterium uptake of cvHspACD peptides from the complex and apo-protein (as the basis) with the exposure time of 500s with deuterium. The middle figure is the plot of deuterium uptake versus exposure time of the most shielded peptide of cvHspACD in the complex (covers β4 and β5) compared to this peptide of apo cvHspACD. The bottom figure shows the heat map of relative fractional uptake of cvHspACD peptides from the complex and the apo-protein at 500s with deuterium on the complex crystal structure. **b.** The deuterium uptake studies of FLNCd24 in the complex. Woods plot and heat map on the crystal structure are plotted using the 500s data. The plot of deuterium uptake versus exposure time is for the peptide covering strand C in FLNCd24 in the complex compared to apo FLNCd24. Data generated by Xingyu Qiu.

To study the binding interface of FLNCd24 in this interaction, FLNCd24 was

added to four equivalents of cvHsp (FLNCd24 12.5 μM, cvHspACD 50 μM before the addition of labeling buffer) before deuterium labelling to avoid any FLNCd24 in dimeric form. Compared to the apo FLNCd24, the most protected part in all the time slots covers residues from 2658-2682 (VDCSKAGTNMMMVGVHGPKTPCEEV), which starts from the middle of strand B till the first part of strand D (Figure **3.6b**). This part covers the binding site of FLNCd24 in the complex crystal structure. Because the cutting sites for pepsin are limited, the peptides cannot be further cut into shorter ones that exactly covers the binding sites, so it is reasonable that the protection region from HDX is a bit more redundant than the information provided by the crystal structure. To eliminate the interference from dimer dissociation during the dilution, we also did titration experiments on FLNCd24 at different concentrations, and almost no effect can be observed on the dimer interface (data not shown here).

3.3.3 Phosphorylation at FLNCd24 T2677 and Y2683 have opposite effects in regulating FLNCd24 dimerization and its interaction with cvHsp

Several new phosphorylation sites (S2632, S2655, T2677, S2718) on FLNCd24 were discovered by phospho-proteomics when the skeletal tissues were treated with electrical stimulation in the Warscheid Lab. These phosphorylations on FLNCd24 have the effect of weakening the interaction with cvHsp.

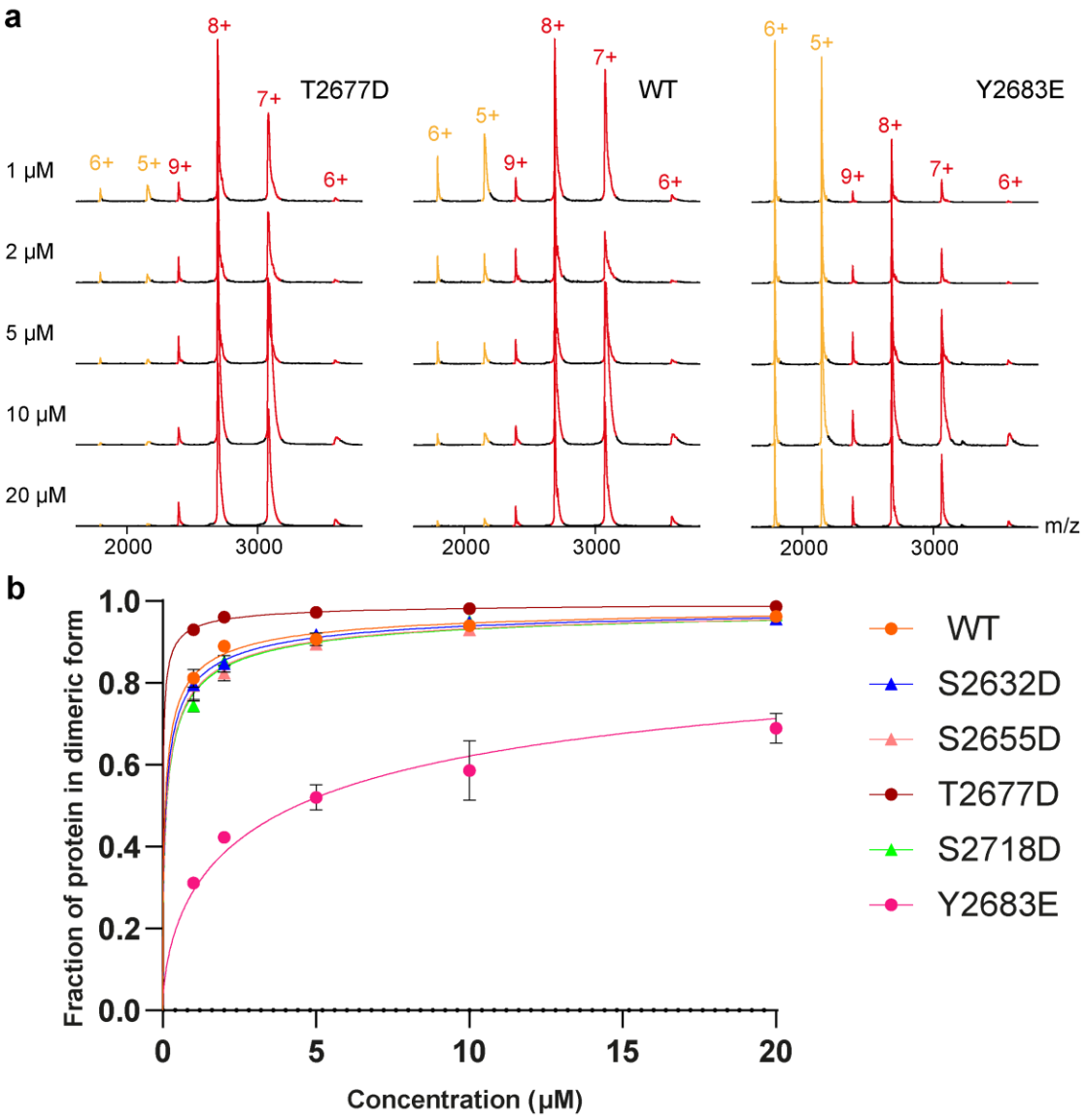


Figure 3.7. Phosphorylation regulates the strength of FLNCd24 dimerization. a. Native MS titration experiments of FLNCd24 WT and phosphorylation mimics T2677D and Y2683E at different concentrations. Monomer peaks are colored yellow, and dimer peaks are red. In the T2677D mutant, the monomer peaks are generally much lower than those in WT, suggesting that T2677D strengthens the dimerization of FLNCd24. In the Y2683E mutant, the monomer peaks are much higher, representing its weakening effect on the dimer interface. **b.** Plots of FLNCd24 WT and phosphomimics titrations. The concentrations are based on FLNCd24 monomer concentrations.

To conduct the *in vitro* study, we used an aspartic acid mutation to mimic the

phosphorylation for the individual site (Appendix figure **1.6**). The phosphomimic mutants were expressed and characterized by native MS. Surprisingly, we observed that one of the phosphomimics, T2677D, has a strengthened dimer through the titration experiments of a variety of concentrations, while the other three behaved similarly as FLNCd24 WT. The peaks representing FLNCd24 T2677D monomers are much lower than those representing FLNCd24 WT monomers at the same concentrations (Figure **3.7**). From the titration experiments, we obtained the thermodynamical parameters of the FLNCd24 dimer interfaces, the ΔΔG of the dimer interfaces between T2677D mutant with WT and other constructs is around 5 kJ/mol (Table **3.2**).

	K_D (nM)	**ΔG (kJ/mol)**
WT	83.17 ± 13.44	-40.42 ± 0.40
S2632D	103.6 ± 11.1	-39.86 ± 0.27
S2655D	124.7 ± 21.0	-39.42 ± 0.42
T2677D	12.21 ± 1.63	-45.17 ± 0.33
S2718D	130.3 ± 21.8	-39.31 ± 0.42
Y2683E	4444 ± 784	-30.57 ± 0.44

Table 3.2. Thermodynamic parameters of the FLNCd24 WT and phosphomimics dimer interfaces.

We then tried the mixing of the four mutants with cvHspACD. The interaction still competes with the dimer interfaces of the FLNCd24 constructs. Compared to the FLNCd24 WT, S2655D almost has no difference in binding cvHspACD, and

S2632D and S2718D weaken the interaction slightly. The most significant difference comparing to WT FLNCd24 still comes from the T2677D mutant (Figure **3.9**), making it meaningful to get a deeper insight into its structural analysis.

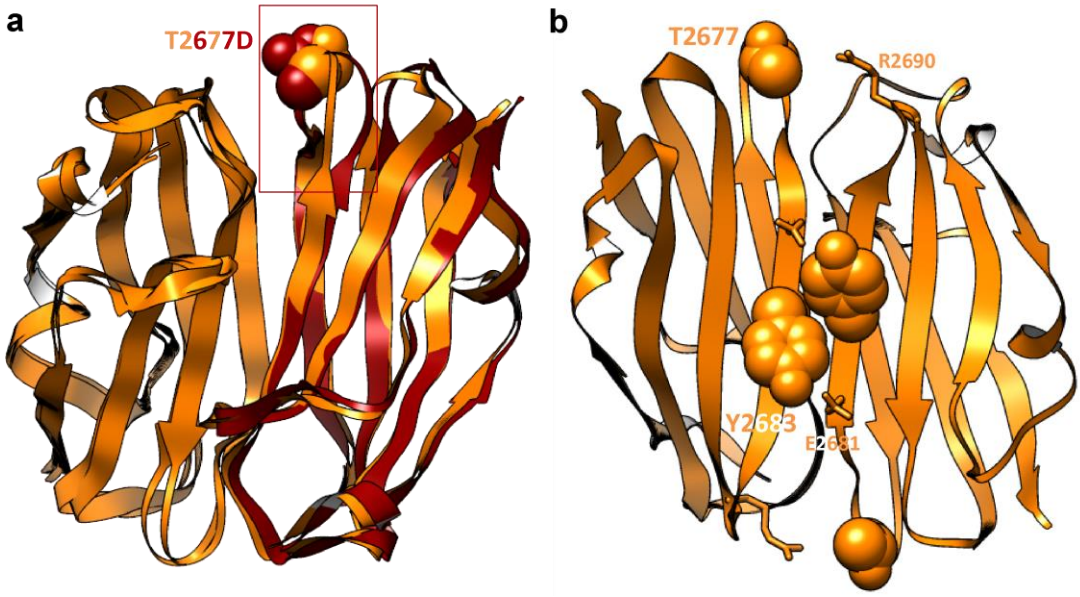


Figure 3.8. Structural visualizations of the phosphorylation sites on the FLNCd24 dimer crystal structures. a. Structure alignment of FLNCd24 WT (1V05, orange) with FLNCd24 T2677D (dark red). The phosphorylation-mimic on T2677 twists the loop between strand C and D, which is the most significant conformational change compared to the WT structure. **b.** The locations of T2677 and Y2683 on the FLNCd24 WT dimer structure, where phosphorylation can happen and manipulate the dimerization of FLNCd24. R2690 and E2681 are the closest residues to T2677 and Y2683 respectively in space on the other chain.

We attempted to observe the differences caused by the mutation on the crystal structure of the FLNC mutant dimer. The crystal structure of FLNCd24 T2677D was obtained without many obstacles. The structure was phased at the resolution of 1.59Å in the $P6_1$ space group. The RMSD value is 0.897 compared to the dimer WT FLNCd24 structures (1V05). Because of the T2677D point mutation, the most

significant and relevant fluctuation comes from the loop between strand C and D (Figure **3.8a**). The backbones and side chains of residues H2673-P2678 all shift to cause the twist of the loop in the structure.

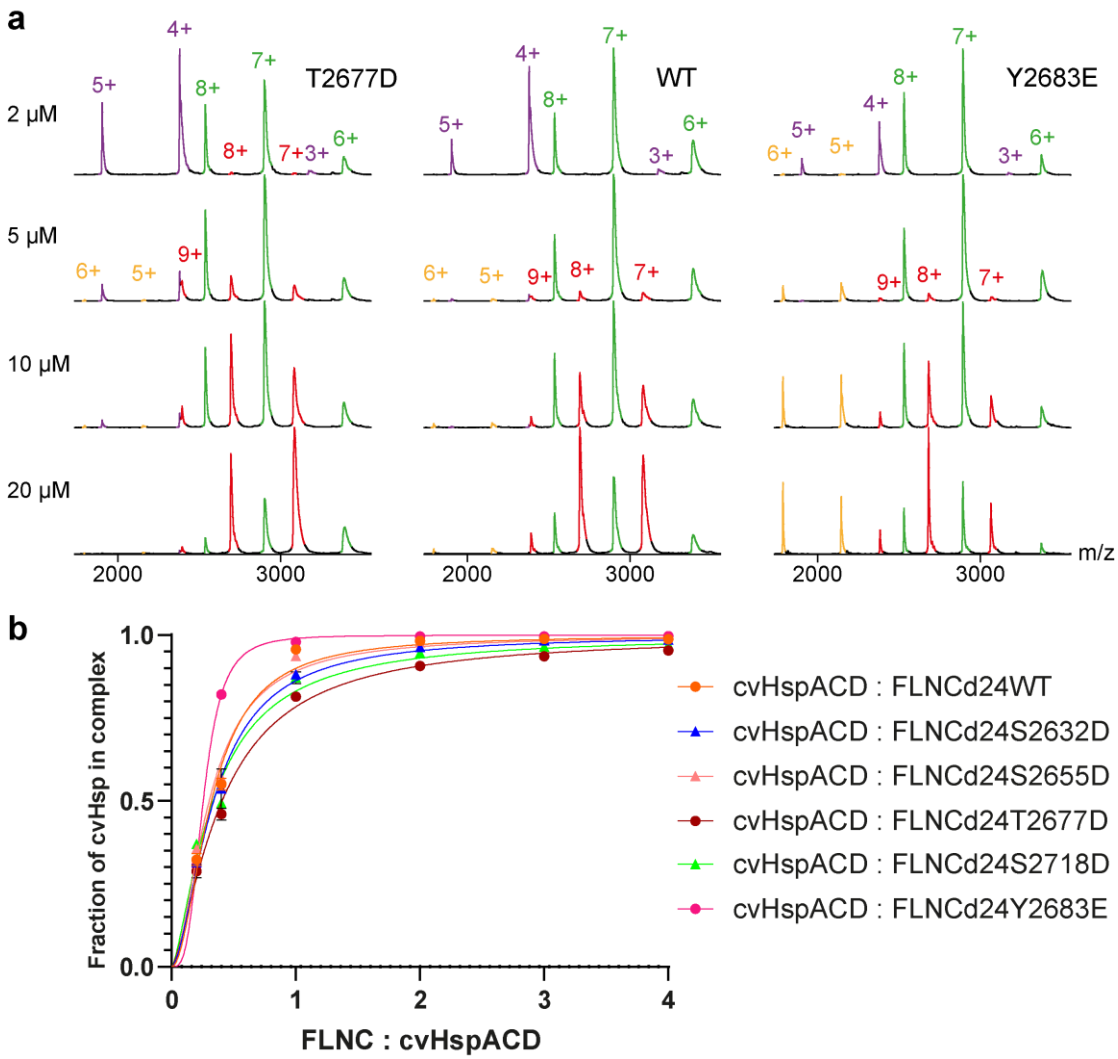


Figure 3.9. Phosphorylation-mimic on specific residues of FLNCd24 manipulates its interaction with cvHspACD. a. Native MS titration experiments of cvHspACD with FLNCd24 WT and phosphomimics. Complex peaks are colored green. Peaks representing cvHspACD is colored orchid. Red peaks represent FLNCd24 dimer, and monomer is colored yellow. T2677D mutant leaves more dissociated cvHspACD on the spectrum, while the Y2683E mutant can bind more cvHspACD at the same condition. **b.** Plots of the cvHspACD-FLNCd24 phosphomimics titration experiments. cvHspACD is maintained at 5 μM, and the concentrations of FLNCd24 constructs are varied.

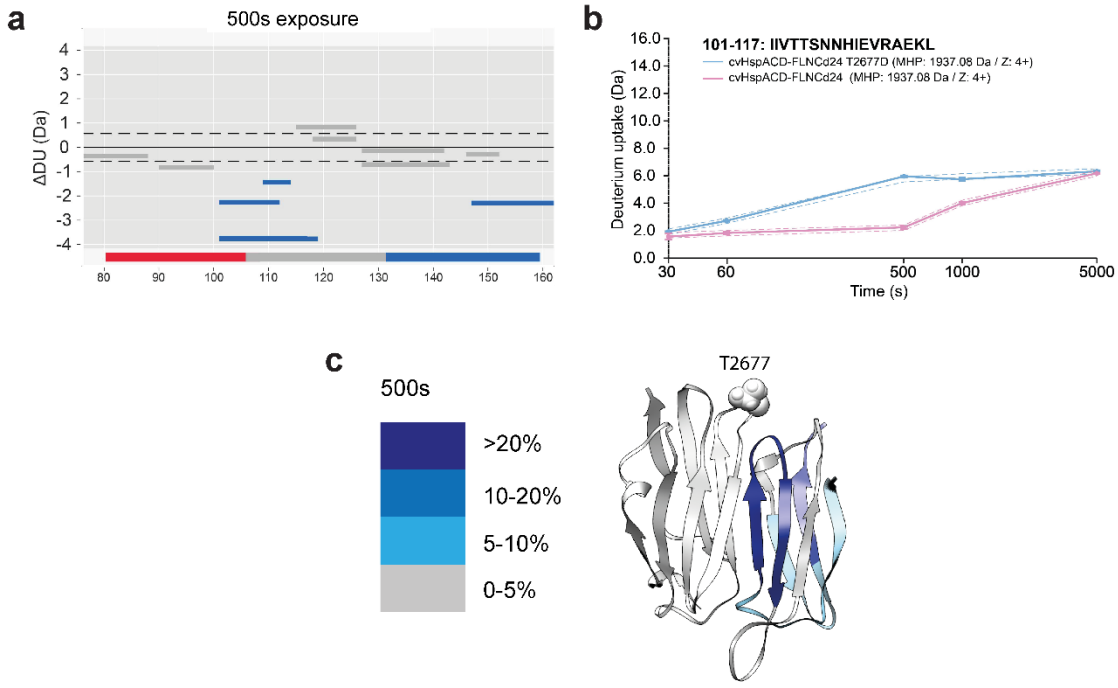


Figure 3.10. HDX experiments reveal the interference from the FLNCd24 T2677D mutation by comparing the deuterium uptake of cvHspACD in the complex. a. Woods plot gives the comparison of deuterium uptake of cvHspACD peptides from the FLNCd24 WT complex and the FLNCd24 T2677D complex (as the basis) after the exposure time 500s with deuterium. The peptide that covers β4 and β5, which is the main binding site from the crystal structure, shows the most significant difference when being mutated. **b.** The plots of deuterium uptake versus exposure time of the most shielded peptide of cvHspACD in the FLNCd24 WT complex compared to this peptide from the FLNCd24 T2677D mutant complex. The peptide is more protected in the WT complex compared to it in the FLNCd24 T2677D mutant complex. **c.** The heat map of relative fractional uptake at 500s of cvHspACD peptides from the WT complex and the FLNCd24 T2677D complex on the complex crystal structure. Data generated by Xingyu Qiu.

From native MS titrations, we know the dimerization is strengthened after the T2677D mutation. The closest residue to T2677 on the other chain is R2690 (Figure **3.8b**), so it is likely that the strengthening of the dimerization is through the salt-bridge interaction between phosphorylated T2677 with R2690. But the side chains of the T2677D residue and R2690 residue are not close to each other

on the well-resolved 2Fo-Fc map.

It is understandable that crystal packing is not the most native environment for proteins. The electron density of the FLNC T2677 region does not resolve perfectly. And the aspartic acid mutation is a good mimic but still not natural phosphorylation. MD calculations can be a good supplement and reference by calculating how these side chains move under a mimic native condition to solve the problems came from crystallography. From the preliminary MD results, the two side chains are moving closer as predicted in the mimic native environment.

	K_D (nM)	**ΔG (kJ/mol)**
cvHspACD : FLNCd24WT	321.7 ± 24.8	-37.05 ± 0.19
cvHspACD : FLNCd24S2632D	338.9 ± 21.5	-36.92 ± 0.16
cvHspACD : FLNCd24S2655D	303.1 ± 21.1	-37.19 ± 0.17
cvHspACD : FLNCd24T2677D	406.8 ± 23.6	-36.46 ± 0.14
cvHspACD : FLNCd24S2718D	334.6 ± 34.7	-36.95 ± 0.26
cvHspACD : FLNCd24Y2683E	255.1 ± 3.3	-37.61 ± 0.03

Table 3.3. Thermodynamic parameters of the cvHspACD-FLNCd24 interactions.

In the complexed structure, the T2677 also shows to be a vital residue near the interface. Interestingly, the region around T2677 is almost the only area that is hard to be well resolved on the density map. The closest residue to T2677 on cvHspACD is N107 (Figure **3.11a**), and the density of N107 is also missed on the map. The phosphorylated threonine or aspartic acid can be much more repulsive

with asparagine, which can cause a weakened interaction with cvHsp. The more conclusive explanation will be followed by the MD analysis.

We also did HDX to compare the difference of deuterium uptake of the interfaces between the WT and mutant complexes. Because of the change of the mass of the peptide caused by the mutation, it was straightforward to compare the difference of deuterium uptake of cvHspACD peptides in the WT complex and the FLNCd24 T2677D complex. We did experiments at five exposure time points from 30s to 5000s. The general feature of the comparison based on cvHspACD peptides is similar to the previous HDX study on the binding interfaces (Figure **3.10a**). The most significant peptide still covers cvHspACD β4-β5. For this peptide, the difference caused by the FLNCd24 T2677D mutation was not apparent at the early stage of the exposure to deuterium and later became significant at around 500s. At around 5000s, the deuterium uptake had reached equilibrium for both complexes (Figure **3.10b**). On the heat map, it is notable that the part on cvHspACD that is far away from the T2677D mutation on FLNCd24 is not significant in comparison with that in the WT complex (Figure **3.10c**). In general, the T2677D mutation on FLNCd24 gave stronger deuterium uptake in cvHspACD in the complex than the WT, which matches the idea that this mutation on FLNCd24 weakens its interaction with cvHsp.

Another phosphorylation site on FLNCd24 discovered earlier in sarcoma and tumor tissues with adequate evidence is Y2683[39]. This residue locates on the

center of strand D, which is in the dimer interface (Figure **3.8b**). The hydrogen bonding between Y2683 and E2681 on the other chain contributes to the anti-parallel β-sheet between strands D of FLNCd24 dimer. When Y2683 is phosphorylated by tyrosine kinase during tumor growth, those acidic residues can have strong repulsion on the dimer interface (together with E2680), and the FLNC dimer can be much weakened. We introduced the tyrosine to glutamic acid mutant on this residue to mimic this phosphorylation. Native MS titration experiments revealed that the FLNCd24 dimer is much weakened. More dissociated monomer peaks can be observed in the spectrum for the FLNCd24 Y2683E mutant compared to the WT (Figure **3.7**). The ΔΔG of the dimer interfaces is around 10kJ/mol between the Y2683E mutant and the WT, which is much higher than the difference caused by the T2677D mutation on the other side (Table **3.2**). The crystallography of the Y2683E mutant was not as straightforward as the strengthened T2677D dimer. The interference of the anti-parallel β-sheet between strands D on the dimer interface hindered the crystal packing. We did not obtain any Y2683E crystals suitable for diffraction.

On the other hand, in the complexed structure, the Y2683 residue is on the FLNCd24 strand D. The cvHspACD β5 is also contacting with the FLNCd24 strand D. R113 on the cvHspACD β5 strand is the closest residue to Y2683 (Figure **3.11a**). FLNC with phosphorylated Y2683 can interact with cvHsp much stronger because of this new salt-bridge contact. The titration experiment between the FLNCd24

Y2683E mutant with cvHspACD confirms a tighter binding between the two proteins.

	ΔG (kJ/mol)
cvHspACD : FLNCd24WT	-16.84 ± 0.39
cvHspACD : FLNCd24S2632D	-16.99 ± 0.29
cvHspACD : FLNCd24S2655D	-17.48 ± 0.38
cvHspACD : FLNCd24T2677D	-13.88 ± 0.31
cvHspACD : FLNCd24S2718D	-17.30 ± 0.47
cvHspACD : FLNCd24Y2683E	-22.33 ± 0.25

Table 3.4. Free energies of the interactions between cvHspACD with dissociated FLNCd24 monomers.

To elucidate the thermodynamic relationships of the interactions between cvHspACD and the FLNCd24 constructs from the native MS titration experiments, the dissociation of FLNCd24 dimer during the interaction with cvHspACD must be considered.

$$\Delta G_{cvHspACD+FLNCd24\ mono} = \Delta G_{cvHspACD+FLNCd24} - \Delta G_{FLNCd24\ dimer}$$

After eliminating the factor of FLNCd24 dimer dissociation, the free energy values for cvHspACD interacting with dissociated monomeric FLNCd24 WT and S2632D, S2655D, S2718D mutants are around -17.0 kJ/mol. The T2677D mutant weakens the interaction, with the free energy around -13.9 kJ/mol.

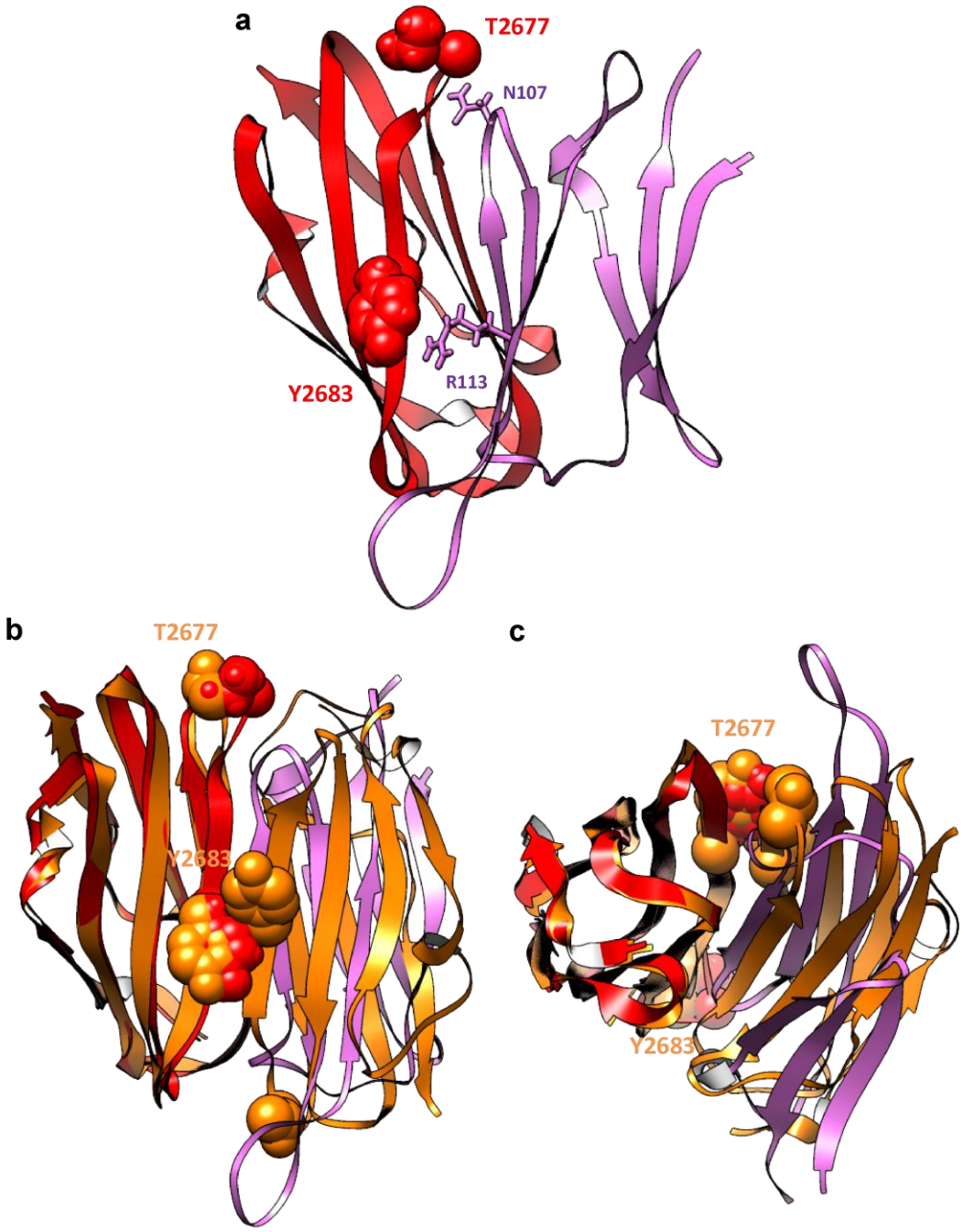


Figure 3.11. The cvHspACD-FLNCd24 structure reveals the binding and regulating mechanisms. a. FLNC T2677 and Y2683 residues are shown as the sphere on the cvHspACD-FLNCd24 structure. Their closest residues N107 and R113 on cvHsp is shown as sticks. S2632, S2655 and S2718 are not highlighted on the structure because they are not close to either FLNC dimer interface or complex interface with cvHsp. **b-c.** Structure alignment of the complex with FLNCd24 WT (1V05). T2677 and Y2683 residues are highlighted.

The Y2683E mutant regulates the interaction to the other direction and lifts the free energy to 22.3 kJ/mol through the salt bridge between the glutamate acid and arginine (Table **3.4**). Figure **3.11b-c** shows the structural alignment of cvHspACD binding FLNCd24 and competing to its dimer interface.

3.4 Discussion

We investigated and confirmed the specific interaction between cvHsp and FLNCd24 and how this interaction is regulated by FLNC phosphorylation. Our structural analysis reveals that cvHspACD β4 strand binds FLNCd24 strand C to form a new parallel β-sheet between the two proteins. The interaction interferes with the homo-dimerization of FLNCd24, which involves FLNCd24 strands C and D. The genome evolution of cvHsp in the human sHsp family not just prevents this protein from forming large oligomers as we discussed in Chapter 2, but also makes it a specific binding partner for FLNCd24 to participate in modulating this large cytoskeletal protein. The key residue compared to other human sHsps in this interaction is the I102 residue, which enables the hydrophobic contacts with FLNC M2668 and V2684. The other important hydrophobic residues of cvHsp on the interface are conserved to be hydrophobic in other sHsps. The β4-β8 groove is the most functional region for sHsp ACD. When binding the peptide containing IPI/V motif, the peptide locates in the pocket between β4 and β8[40], with a much

weaker binding affinity compared to cvHsp-FLNCd24 interaction, though the hydrophobic residues on the cvHspACD β8 strand also stabilize the two methionine residues from FLNC. Strand C and D are also the binding sites for partners in all the other FLNC domains with a structure from previous reports[6,17–22].

Phosphorylation at T2677 strengthens the dimerization of FLNC itself and weakens its interaction with cvHsp, while the phosphorylated Y2683 mutant has the opposite effect. Notably, the phosphorylations on serine and threonine mediated by kinases like PKC have identical effects in general to the whole FLNC protein[31,32,35], making the long dimeric protein to become more rigid and stable, with limited mobility. Serine and threonine phosphorylation on FLNC or binding partners hinder the cleavage between d23 and d24; prevent the insertion region from over-extension through interaction with phosphorylated Hsp27; inhibit the binding with FILIP1 to the further degradation of FLNC; strengthen d24 dimerization and weaken the interaction with cvHsp (Figure **3.12**). Most of the serine and threonine phosphorylation identifications were carried out after electrical stimulation, which is a representative acute stress mode. It is notable that we studied the phosphorylation sites *in vitro* individually in this thesis, but in tissue, phosphorylation can happen on all the potential sites simultaneously. Though charge repulsion can avoid phosphorylation on the sites that are too close on one molecule at one time, the phosphorylation effect still needs to be

considered from a holistic perspective. The combining effect of serine and threonine phosphorylations on FLNCd24 towards acute stress is to weaken its interaction with cvHsp.

However, phosphorylation on tyrosine is more common during cancer in tumor tissues through activation of tyrosine kinases[39,41–44]. The weakened dimer enhances the dynamic of FLNC to participate in cancer cell locomotion and differentiation (Figure **3.12**), which might be a common behavior during chronic stress. The mediation of FLNC dimerization and motility can be a potential target to control certain types of cancer development.

The clearance of FLNC depends on the BAG3-mediated CASA process. The CASA complex also contains HspB8, Hsp70 and chaperone-associated ubiquitin ligase CHIP protein[33,34]. D19-21 have been reported to be involved in the CASA complex interaction. The most frequent myopathy-related mutation on FLNCd24 is W2710X. This mutant loses the last 16 residues on the C terminus at d24 and fails in dimerization. It was no longer interacting with cvHsp and tending to aggregate, though the non-aggregated protein was still functional. BAG3 was still recruited into the aggregates, but the autophagic clearance of mutant FLNC is blocked[13,45]. The function and stability of cvHsp are independent of BAG3, but HspB8 and cvHsp have been reported to be binding partners, which might bridge cvHsp, FLNC, HspB8 and BAG3 together[46,47].

In summary, our study shows that cvHsp is binding to compete to the dimerization of FLNCd24. This interaction and FLNCd24 dimerization are regulated by serine and threonine or tyrosine phosphorylations at different stress modes. Our study provides a new chaperone-client proteostasis mechanism and a potential cancer treatment target.

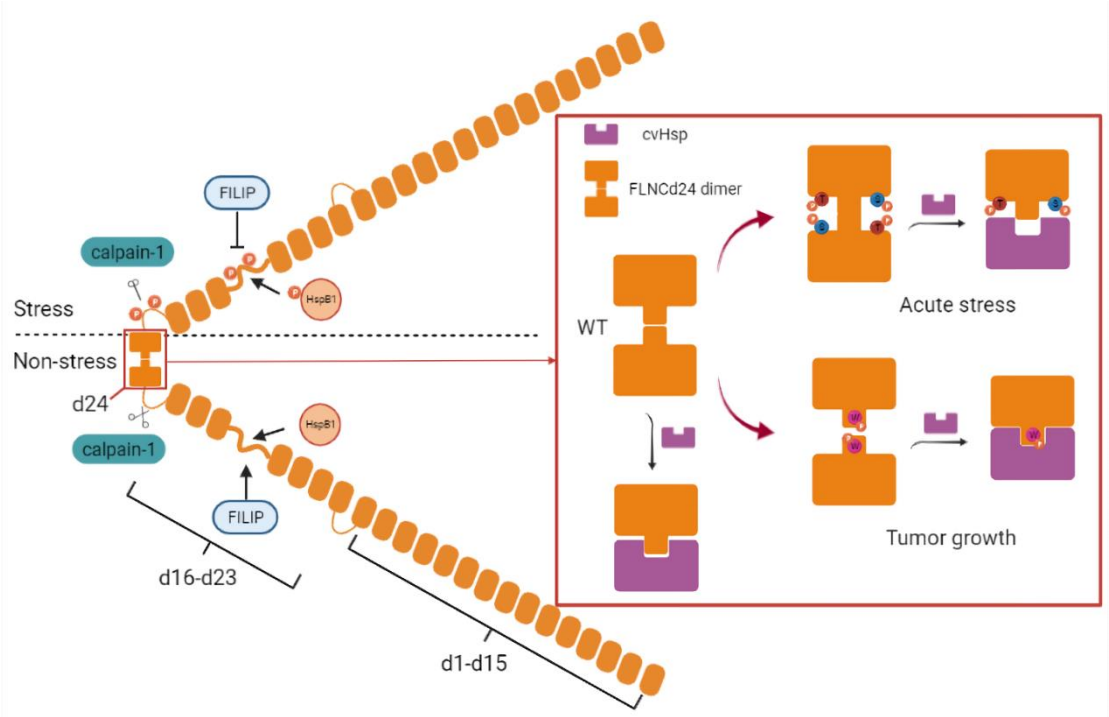


Figure 3.12. Schematic representation of the FLNCd24-cvHsp interaction and full-length FLNC regulations at different conditions. Phosphorylation upon electric stimulation on given serine and threonine residues on FLNCd24 strengthens its dimerization but weakens its interaction with cvHsp, and the phosphorylation on Y2683 as a response to tumor growth has reversed effects in FLNCd24 dimerization and its interaction with cvHsp. The full-length FLNC can also be regulated by phosphorylation on other domains in response under different conditions.

Acknowledgments

I appreciate our collaborators Katja Gehmlich, Charlotte Hooper and Bettina Warscheid for the mouse model, immunoprecipitation works, and FLNC phosphorylation sites information. I'm thankful to Xingyu Qiu, who conducted the HDX-MS experiments within a long period and helped to generate the HDX figures, which was indeed a great effort. I thank Michael McDonough, Jurgen Brem and Patrick Rabe for helping me acquire the diffraction data for the structures. I am grateful to Timothy Allison for helping me with the structural determination and refinements. Guodong Cao has been making great efforts in molecular dynamics (MD) to support the structures, which is great support.

3.5 Materials and methods

Expression and purification of FLNCd24 constructs

Human FLNCd24 (residues 2630-2724, WT and phosphomimic mutants S2632D, S2655D, T2677D, S2718D and Y2683E) encoded in the pET28a(+) vector with N-terminal 6×His-tag were transformed in *Escherichia coli* BL21(DE3) cells (Agilent). A Gly-Ser overhang was left between the TEV protease recognition sites and the N-terminal of the constructs after the TEV proteolysis. Cells were used to inoculate a 10 mL LB media with kanamycin (30 μg/mL) at 37°C overnight and transferred to a 1 L large LB culture with the same antibiotics. The culture was shaking at 180 rpm at 37°C until the OD600 reached 0.6-0.8 and followed by the induction with IPTG to a final concentration of 0.5 mM. After another overnight shaking at 22°C, the cells were harvested, pelleted and lysed using French Press Microfluidizer with the addition of EDTA-free protease inhibitor cocktail and buffer containing 300 mM NaCl, 50 mM Tris and 20 mM imidazole (pH 8.0). The lysate was then spun at 20,000 g for 20 min at 4°C and filtered through 0.22 μm filter to get clarified before loading to the HisTrap HP column. The protein was eluted with the buffer with 300 mM NaCl, 50 mM Tris and 500 mM imidazole (pH 8.0) and dialyzed to the loading buffer with TEV protease at room temperature overnight. The sample was again loaded to the HisTrap HP column to remove the His-tag and exchanged to PBS buffer with 5 mM TCEP.

Crystallization and structure determination of FLNCd24 T2677D and cvHspACD C131S – FLNCd24 WT complex

Fresh, unfrozen FLNCd24 T2677D protein was buffer-exchanged to 20 mM HEPES, pH 8.0 and concentrated to 10 mg/mL. 96-well sitting-drop vapor diffusion plates (Hampton Research) were used for crystallography. The crystal for structural determination of FLNCd24 T2677D showed up after one day, and it was collected from the condition with 0.2 M sodium chloride, 0.1 M sodium cacodylate, pH 6.5 and 2 M ammonium sulfate. Crystals were cryo-protected by adding 10% glycerol and then fished and flash-frozen in liquid nitrogen.

Property	**FLNCd24 T2677D**
Space group	$P6_1$
Cell constants	87.39Å 87.39Å 64.75Å
a, b, c, α, β, γ	90.00° 90.00° 90.00°
Resolution (Å)	49.20-1.59
Data completeness (%)	100
(in resolution range)	(49.20-1.59)
Refinement program	PHENIX 1.18.2_3874
R, R_{free}	0.183, 0.195
Average B, all atoms ($Å^2$)	45.0
Total number of atoms	3069

Table 3.5. Diffraction data and refinement statistics of the FLNCd24 T2677D structure.

The cryo-protected crystal was diffracted at beamline I04 at Diamond Light Source,

UK. The diffraction data were merged and scaled by Xtriage automatically. The structure was phased by the Phenix Molecular Replacement (MR) tool based on the FLNCd24 WT (1V05) structure. Phenix Refinement tool was used to finalize the structure with R and R_{free} values at 0.183 and 0.195 respectively. PDB validation server was used as a reference for the refinement.

Property	**cvHspACD C131S-FLNCd24**
Space group	$P3_221$
Cell constants	94.13Å 94.13Å 51.60Å
a, b, c, α, β, γ	90.00° 90.00° 120.00°
Resolution (Å)	47.06-2.85
Data completeness (%)	99.5
(in resolution range)	(47.06-2.85)
Refinement program	PHENIX 1.18.2_3874
R, R_{free}	0.224, 0.291
Average B, all atoms ($Å^2$)	66.0
Total number of atoms	2670

Table 3.6. Diffraction data and refinement statistics of the cvHspACD C131S-FLNCd24 structure.

The crystallization of the cvHspACD C131S-FLNCd24 WT complex requires both proteins to be purified on the same day. Then cvHspACD C131S and FLNCd24 WT were mixed at a ratio roughly at 4:1, which was to reduce the dimeric FLNCd24 that has a similar mass to the complex. The mixed proteins were then loaded to the size exclusion chromatography (SEC) column (Superdex 200, 16/600). The isolated protein complex was concentrated to 10 mg/mL in the 20

mM HEPES storage buffer (pH 8.0). In the 96-well sitting drop crystallography plates, the crystal took a week to grow. The crystal for the diffraction was acquired from the condition with 0.1 mM HEPES and 20% w/v PEG 8000, pH 7.0. No cryo-protectant was added for the crystal fishing.

The crystal was diffracted at beamline I03 at Diamond Light Source, UK, and then merged and scaled by Xtriage automatically. Phenix MR tool was used to phase the FLNC chain in the asymmetric unit first. Then all the unresolved models were deleted. The remained FLNC model was treated as the fixed existing component, and cvHspACD C131S chain A was searched as the second model using the CCP4 MR tool. Phenix Refinement tool and PDB validation server were used to finalize the structure with R and R_{free} values at 0.224 and 0.291.

Native MS

Proteins were bio-spun into 200 mM ammonium acetate (pH 6.9). All the spectra were recorded from the Synapt G1 mass spectrometer modified for the transmission of intact non-covalent protein complexes (Waters Corp.). The parameters were set as capillary, 1.5 kV; sampling cone, 20-35 V; extraction cone, 3.0-3.3 V; backing pressure, 3.8-4.1 mbar; trap gas (argon) flow, 1.5-3.0 mL/min; trap energy, 6-8 V; transfer energy, 6-12 V. Three technical repeats were collected for each ratio in the titration experiments. The peak intensities were extracted manually in UniDec. The dissociation constants were analyzed using GraphPad

Prism 8.0.1 through the specific binding with Hill slope model[48].

Hydrogen–deuterium exchange mass spectrometry (HDX-MS)

cvHspACD, FLNCd24 and cvHspACD-FLNCd24 complex were all prepared in equilibration buffer, which contains 40 mM HEPES at pH 7.5. To form the complex, cvhspACD and FLNCd24 mixed at 1:1 (for the measurement of cvHspACD) or 4:1 (for the measurement of FLNCd24) molar ratio prior to HDX-MS. To carry out the HDX-MS experiments, 5 µL of proteins were first exposed in 55 µL of D_2O buffer (40mM HEPES, pD 7.5) for a time course of 30, 60, 500, 1000 or 5000 seconds, followed by mixing with one equivalent of pre-chilled quenching solution (60 µL, 40 mM HEPES, 5 mM TCEP, pD 2.0). Triplicates of datasets were collected for each time point. 80 µL of quenched samples were then injected into the nanoACQUITY UPLC System (Waters corp.) and online digested by Enzymate™ BEH Pepsin Column (2.1 x 30 mm, Waters corp.) at room temperature. The digested peptides were trapped by a BEH C18 trap column (1.7 µm, 2.1 x 5 mm, Waters corp.) for 3 minutes to remove the salt. BEH C18 analytical column (1.7 µm, 1 x 100 mm, Waters corp.) was then used to separate the peptides with a linear gradient of buffer B (acetonitrile with 0.1 % formic acid) from 3% to 35% at a flow rate of 40 µL/min. To avoid back exchange, the temperature was kept below 4 ℃ after digestion. Wash buffer (1.5 M guanidinium chloride pH 2.8, 4 % acetonitrile, 0.8% formic acid) is used to clean the pepsin column between each run. MS acquisition of peptides was achieved by using

Synapt G2-Si HDMS mass spectrometer (Waters corp.) in positive mode. MS/MS spectra were obtained in MSE mode. Peptides from un-deuterated samples were analyzed and identified by ProteinLynx Global Server 2.5.1 (Waters Corp.). DynamX 3.0 (Waters Corp.) was used to analyze the HDX-MS data and to generate the heat maps. The Wood plots and plots of deuterium uptake of individual peptides were generated using Deuteros 2.0[49,50]. The peptide-level significance test was used as the model, and the alpha value was set at 0.01.

Bibliography

1. Stossel, T. P. *et al.* Filamins as integrators of cell mechanics and signalling. *Nature Reviews Molecular Cell Biology* vol. 2 138–145 (2001).
2. van der Ven, P. F. M. *et al.* Indications for a Novel Muscular Dystrophy Pathway. *Journal of Cell Biology* **151**, 235–248 (2000).
3. van der Ven, P. F. M. *et al.* Characterization of muscle filamin isoforms suggests a possible role of γ-Filamin/ABP-L in sarcomeric Z-disc formation. *Cell Motility and the Cytoskeleton* **45**, 149–162 (2000).
4. Feng, Y. *et al.* Filamin a (FLNA) is required for cell-cell contact in vascular development and cardiac morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 19836–19841 (2006).
5. Zhou, X. *et al.* Filamin B deficiency in mice results in skeletal malformations and impaired microvascular development. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3919–3924 (2007).
6. Razinia, Z., Mäkelä, T., Ylänne, J. & Calderwood, D. A. Filamins in Mechanosensing and Signaling. *Annual Review of Biophysics* **41**, 227–246 (2012).
7. Zhou, A. X., Hartwig, J. H. & Akyürek, L. M. Filamins in cell signaling,

transcription and organ development. *Trends in Cell Biology* **20**, 113–123 (2010).

8. Dalkilic, I., Schienda, J., Thompson, T. G. & Kunkel, L. M. Loss of FilaminC (FLNc) Results in Severe Defects in Myogenesis and Myotube Structure. *Molecular and Cellular Biology* **26**, 6522–6534 (2006).

9. Goetsch, S. C., Martin, C. M., Embree, L. J. & Garry, D. J. in Developing and Regenerating Skeletal Muscle. *Stem Cells and Development* **187**, 181–187 (2005).

10. Prill, K. & Dawson, J. F. Assembly and maintenance of sarcomere thin filaments and associated diseases. *International Journal of Molecular Sciences* vol. 21 1–20 (2020).

11. Vorgerd, M. *et al.* A mutation in the dimerization domain of filamin c causes a novel type of autosomal dominant myofibrillar myopathy. *American Journal of Human Genetics* **77**, 297–304 (2005).

12. Shatunov, A. *et al.* In-frame deletion in the seventh immunoglobulin-like repeat of filamin C in a family with myofibrillar myopathy. *European Journal of Human Genetics* **17**, 656–663 (2009).

13. Schuld, J. *et al.* Homozygous expression of the myofibrillar myopathy-associated p.W2710X filamin C variant reveals major pathomechanisms of sarcomeric lesion formation. *Acta Neuropathologica Communications* **8**, 1–19 (2020).

14. Duff, R. M. *et al.* Mutations in the N-terminal actin-binding domain of

filamin C cause a distal myopathy. *American Journal of Human Genetics* **88**, 729–740 (2011).

15. Mao, Z. & Nakamura, F. Structure and function of filamin c in the muscle z-disc. *International Journal of Molecular Sciences* vol. 21 17–19 (2020).

16. Pudas, R., Kiema, T. R., Butler, P. J. G., Stewart, M. & Ylänne, J. Structural basis for vertebrate filamin dimerization. *Structure* **13**, 111–119 (2005).

17. Nakamura, F., Stossel, T. P. & Hartwig, J. H. The filamins: Organizers of cell structure and function. *Cell Adhesion and Migration* **5**, 160–169 (2011).

18. Kiema, T. *et al.* The molecular basis of filamin binding to integrins and competition with talin. *Molecular Cell* **21**, 337–347 (2006).

19. Ithychanda, S. S. *et al.* Migfilin, a molecular switch in regulation of integrin activation. *Journal of Biological Chemistry* **284**, 4713–4722 (2009).

20. Lad, Y. *et al.* Structural basis of the migfilin-filamin interaction and competition with integrin β tails. *Journal of Biological Chemistry* **283**, 35154–35163 (2008).

21. Smith, L. *et al.* Biochemical basis of the interaction between cystic fibrosis transmembrane conductance regulator and immunoglobulin-like repeats of filamin. *Journal of Biological Chemistry* **285**, 17166–17176 (2010).

22. Takala, H. *et al.* β2 integrin phosphorylation on Thr758 acts as a molecular switch to regulate 14-3-3 and filamin binding. *Blood* **112**, 1853–1862 (2008).

23. Lad, Y. *et al.* Structure of three tandem filamin domains reveals auto-

inhibition of ligand binding. *EMBO Journal* **26**, 3993–4004 (2007).

24. Sethi, R. *et al.* A novel structural unit in the N-terminal region of filamins. *Journal of Biological Chemistry* **289**, 8588–8598 (2014).

25. Nakamura, F., Osborn, T. M., Hartemink, C. A., Hartwig, J. H. & Stossel, T. P. Structural basis of filamin A functions. *Journal of Cell Biology* **179**, 1011–1025 (2007).

26. Tossavainen, H. *et al.* Model of a six immunoglobulin-like domain fragment of filamin a (16-21) built using residual dipolar couplings. *Journal of the American Chemical Society* **134**, 6660–6672 (2012).

27. Pérez-Mejías, G. *et al.* Exploring protein phosphorylation by combining computational approaches and biochemical methods. *Computational and Structural Biotechnology Journal* **18**, 1852–1863 (2020).

28. Hornbeck, P. v., Chabra, I., Kornhauser, J. M., Skrzypek, E. & Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561 (2004).

29. Fabrice, R., Carole, J. N., Anne, M., Dieter, F. & Yves, B. Calpain 1-γ filamin interaction in muscle cells: A possible in situ regulation by PKC-α. *International Journal of Biochemistry and Cell Biology* **38**, 404–413 (2006).

30. Gu, X. & Bishop, S. P. Increased protein kinase C and isozyme redistribution in pressure-overload cardiac hypertrophy in the rat. *Circulation Research* **75**, 926–931 (1994).

31. Reimann, L. *et al.* Myofibrillar Z-discs are a protein phosphorylation hot

spot with protein kinase C (PKCα) modulating protein dynamics. *Molecular and Cellular Proteomics* **16**, 346–367 (2017).

32. Reimann, L. *et al.* Phosphoproteomics identifies dual-site phosphorylation in an extended basophilic motif regulating FILIP1-mediated degradation of filamin-C. *Communications Biology* **3**, 1–19 (2020).

33. Arndt, V. *et al.* Chaperone-Assisted Selective Autophagy Is Essential for Muscle Maintenance. *Current Biology* **20**, 143–148 (2010).

34. Ulbricht, A. *et al.* Cellular mechanotransduction relies on tension-induced and chaperone-assisted autophagy. *Current Biology* **23**, 430–435 (2013).

35. Collier, M. P. *et al.* HspB1 phosphorylation regulates its intramolecular dynamics and mechanosensitive molecular chaperone interaction with filamin C. *Science Advances* **5**, (2019).

36. Liao, W. C., Juo, L. Y., Shih, Y. L., Chen, Y. H. & Yan, Y. T. HSPB7 prevents cardiac conduction system defect through maintaining intercalated disc integrity. *PLoS genetics* **13**, e1006984 (2017).

37. Juo, L. Y. *et al.* HSPB7 interacts with dimerized FLNC and its absence results in progressive m. *Journal of Cell Science* **129**, 1661–1670 (2016).

38. Arber, S. *et al.* MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. *Cell* **88**, 393–403 (1997).

39. Bai, Y. *et al.* Phosphoproteomics identifies driver tyrosine kinases in sarcoma cell lines and tumors. *Cancer Research* **72**, 2501–2511 (2012).

40. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proceedings of the National Academy of Sciences of the United States of America* **111**, (2014).

41. Linardic, C. M. & Crose, L. E. S. Receptor tyrosine kinases as therapeutic targets in rhabdomyosarcoma. *Sarcoma* **2011**, (2011).

42. Shor, A. C., Agresta, S. v., D'Amato, G. Z. & Sondak, V. K. Therapeutic potential of directed tyrosine kinase inhibitor therapy in sarcomas. *Cancer Control* **15**, 47–54 (2008).

43. Du, Z. & Lovly, C. M. Mechanisms of receptor tyrosine kinase activation in cancer. *Molecular Cancer* **17**, 1–13 (2018).

44. Stommel, J. M. *et al.* Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318**, 287–290 (2007).

45. Ruparelia, A. A., Oorschot, V., Ramm, G. & Bryson-Richardson, R. J. FLNC myofibrillar myopathy results from impaired autophagy and protein insufficiency. *Human Molecular Genetics* **25**, 2131–2142 (2016).

46. Fang, X., Bogomolovas, J., Trexler, C. & Chen, J. The BAG3-dependent and -independent roles of cardiac small heat shock proteins. *JCI insight* **4**, 1–12 (2019).

47. Sun, X. *et al.* Interaction of Human HSP22 (HSPB8) with Other Small Heat Shock Proteins. *Journal of Biological Chemistry* (2004) doi:10.1074/jbc.M311324200.

48. Marty, M. T. *et al.* Bayesian deconvolution of mass and ion mobility spectra:

From binary interactions to polydisperse ensembles. *Analytical Chemistry* **87**, 4370–4376 (2015).

49. Lau, A. M. C., Ahdash, Z., Martens, C. & Politis, A. Deuteros: Software for rapid analysis and visualization of data from differential hydrogen deuterium exchange-mass spectrometry. *Bioinformatics* **35**, 3171–3173 (2019).

50. Lau, A. M., Claesen, J., Hansen, K. & Politis, A. Deuteros 2.0: Peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics* **37**, 270–272 (2021).

Chapter 4

Structural analysis of HspB8 and its role in the chaperone assisted selective autophagy (CASA) process

4.1 Abstract

One of the main mechanisms to remove the over-stretched and aggregated FLNC is chaperone-assisted selected autophagy (CASA). The CASA complex involves HspB8, BAG3, Hsp70, and other components. The co-chaperone BAG3 links the ATP-independent sHsps and the ATP-dependent Hsp70. The stability and functions of HspB8 are highly dependent on BAG3. The BAG3-HspB8 and BAG3-Hsp70 interactions can be regulated by disease-related mutations or inhibitors under different circumstances. Native MS and other biophysical techniques can be adapted to the studies of CASA-relevant interactions. Our preliminary data reveal that HspB8 can interact with cvHsp through the ACD-CTD interface rather than the canonical ACD-ACD interface. The non-structured BAG3 is not straightforward to be purified at MS quality. Reversed-phased chromatography has been attempted for BAG3 purification. There are more to be explored for this project based on our current progress.

4.2 Introduction

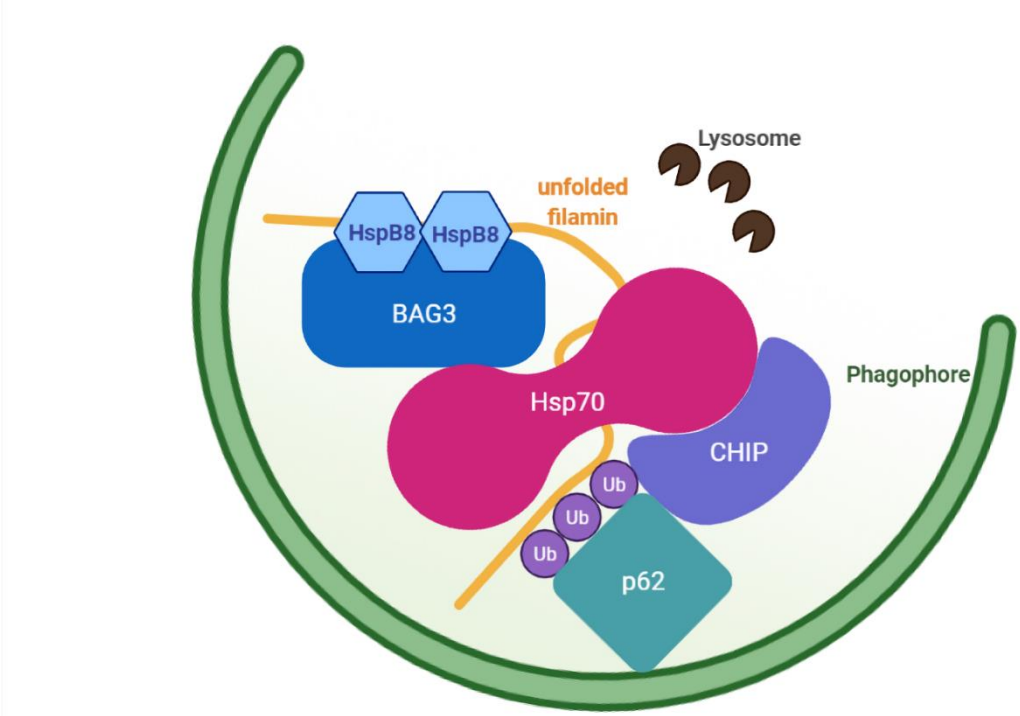


Figure 4.1. Schematic presentation of the CASA complex. Ub, ubiquitin.

Filamins with structural defects due to over-extension during mechano-activities need to be removed to maintain muscular proteostasis[1–3]. Chaperone-assisted selective autophagy (CASA) is a ubiquitin-dependent tension-induced autophagic process to clear non-functional clients including filamins, which is distinct from the ubiquitin-independent chaperone-mediated autophagy[1,4]. Not any other sHsps, but HspB8 is the essential component of the CASA complex. The interaction between HspB8 and the heat shock factor-1 (Hsf1) mediated co-chaperone B-cell lymphoma 2 (Bcl-2)-associated athanogene (BAG3) initiates the CASA process[2,5]. BAG3 works as the scaffold factor to bridge the sHsp with the ATP-dependent Hsc70 (or stress-inducible Hsp70)[2,3,6]. The aggregated clients held

by HspB8 are handed over to Hsp70 and then be ubiquitinated by the C-terminus Hsp70 interacting protein (CHIP) ligase. Hsp70 is induced to target the autophagy adaptor p62/SQSTM1, which will recruit the structure of the autophagosome formation and the fusion of the targets[1,6,7]. The ubiquitinated aggregates and the CASA complex can be degraded together in the autophagosome by the lysosomes[2,3,6] (Figure **4.1**).

The BAG proteins (BAG1-6) with a conserved BAG domain are a family of nucleotide exchange factors (NEFs) for Hsp70[8–10]. BAG1 and BAG3 bind Hsp70 in a competitive mechanism to initiate ubiquitin-proteasome system (UPS) or autophagic pathways for the clearance of aggregates. BAG3 uses its BAG domain to bind the ATPase domain on the N-terminus of Hsp70, which can induce the Hsp70 C-terminal substrate-binding domain to release the bound substrate, including nucleotides and unfolded substrates from being degraded by UPS[9–12]. Since both BAG3 and Hsp70 are up-regulated in cancer cells, and their interaction can protect the clearance of aggregates to inhibit apoptosis. The blockage of this interaction can be a potential target for drug treatment. A small-molecule JG-98 has been reported to be an effective Hsp70 inhibitor to target the Hsp70 ATPase domain in a similar manner as the Hsp70-ADP interaction[13,14], thus, to block the binding between Hsp70 and BAG3. The JG-98 treatment has been found to stop the growth of cancer cells[15]. On the other hand, the E455K natural mutation on BAG3 has been observed in dilated cardiomyopathy (DCM)

patients[16,17]. The mutation is in the BAG region and inhibits the binding between Hsp70 and BAG3. Though the mutation is not taking place in the sHsps binding region on BAG3, the expression level and stability of sHsps including HspB8 decrease significantly[17].

Though the structure of HspB8 has not been determined yet, it is believed to have the most hydrophobic β4-β8 groove among the sHsp family[10,18,19]. HspB8 does not have the IPI/V motif on its CTD, so it is unlikely to form large oligomers, and it is observed to form small-mass species[20]. BAG3 has two IPV motifs between its CASA-relevant WW and proline-rich PXXP regions to bind two HspB8 molecules at the same time[21]. The stability and function of HspB8 are highly dependent on BAG3. Other sHsps like HspB6 and αBC also have certain binding affinities with BAG3, while Hsp27 and cvHsp do not bind BAG3. But when HspB6 is incubated with BAG3 together with HspB8, no HspB6 can be observed in the complex. The expression level and interaction of BAG3 and HspB8 can also be regulated by the transcription factor NF-κB during the heat shock recovery[22]. Together with p62/SQSTM1, BAG3, and HspB8, but not Hsp70, are also involved in the actin spindle placement during mitosis[23]. The K141E and K141N mutations on HspB8 are relevant to the Charcot-Marie-Tooth (CMT) disease. Though K141 residue is presumed to be located on the β6+7 strand, which is far away from the binding interface with BAG3, these mutations at this residue still disrupt the interaction of HspB8 with BAG3[24–28]. The myopathy-related P209L mutation on one of the

IPV motifs of BAG3 blocks the autophagy clearance of aggregated FLNC and hinders the interaction between BAG3 and p62. The mechanism of this mutant still needs to be further explored[8,29]. The BAG3-HspB8 interaction can also be mediated by the serine/threonine-protein kinase 38 (STK38), which is an important factor of the Hippo pathway that regulates tissue growth[30–32]. The binding with STK38 induces the conformational change of BAG3 and inhibits its interaction with HspB8, but the cytoskeletal system assembly can be stimulated at the mean time[33].

Our study reveals that HspB8 is not a stable protein. We tried four HspB8ACD constructs with the same length but in different vectors, and we managed to obtain one stable protein that works well by native MS. We proved that the β4-β8 groove of this construct is functional. No interaction was observed between HspB8ACD and cvHspACD, but HspB8ACD can bind cvHsp through the CTD. Traditional purification methods cannot get rid of the adducts from BAG3, and we are seeking some non-canonical methods such as reversed-phase chromatography to purify BAG3. The CASA complex as well as the BAG3-HspB8 and BAG3-Hsp70 interactions and their inhibitors can be good systems to study on native MS quantitatively and on other accessible biophysical techniques.

4.3 Results

4.3.1 HspB8 is unstable and not interacting with cvHsp through the ACD-ACD interfaces

We first intended to express and purify the full-length and ACD constructs of HspB8 but found that truncations always existed from the cell lysis step. The ACD of sHsps is a conserved and stable domain, but the ACD of HspB8 is not as stable as ACDs from other sHsps. The main truncation was likely taken place between E126 and E127 on the ACD β5 strand. The intact constructs of HspB8 seemed challenging to obtain. We attempted to add a BRIL peptide on the N-terminal of the HspB8ACD to stabilize the target protein from self-truncation. The apocytochrome b_{562} RIL (BRIL, RIL indicates the three mutations from protein engineering) protein is an engineered peptide that is commonly used to stabilize and enhance the solubility of the target proteins such as G-protein coupling receptors (GPCRs) in cryo-EM and crystallography. The BRIL is 12 kDa and forms as a 4-helix bundle, which facilitates the orientation of protein molecular packing during crystal formation[34–36] (Figure **4.2d**).

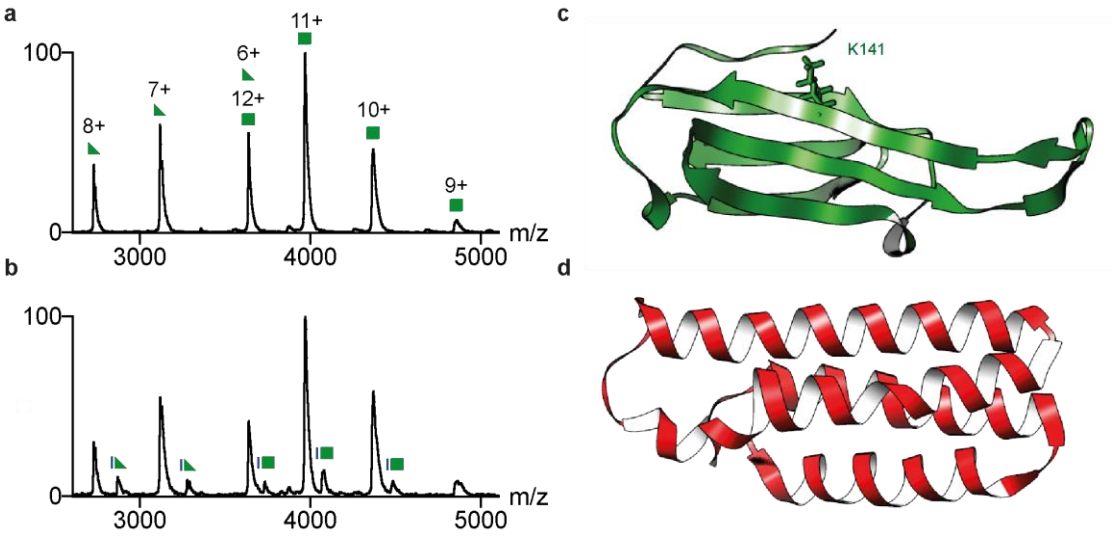


Figure 4.2. BRIL-HspB8ACD construct is stable and functional. a. BRIL-HspB8ACD exists as an equilibrium between monomers and dimers on native MS. The green triangle represents monomeric BRIL-HspB8ACD. **b.** Native MS experiment shows that the β4-β8 groove of BRIL-HspB8ACD can bind αBCCTD peptide, which means the construct is functional. The blue stick represents αBCCTD. **c.** HspB8ACD structure from comparative modeling based on Hsp27ACD structure (4MJH)[37,38]. The disease-related mutagenesis site K141 is shown as the stick. **d.** The crystal structure of *Escherichia coli* cytochrome b_{562} RIL protein (PDB: 1M6T), which forms as the four-helices bundle[36].

We added the BRIL peptide to the full-length and ACD constructs of HspB8. The BRIL-FLHspB8 still could not get expressed properly, but we acquired the BRIL-HspB8ACD construct with no truncations (Appendix figure **1.7**). The purified BRIL-HspB8ACD was partially unfolded as observed by native MS (Appendix figure **3.2**), but the protein could mostly be refolded in PBS overnight. The BRIL-HspB8ACD exists as an equilibrium between monomers and dimers on native MS (Figure **4.2a**), which is identical to previous reports. The peaks on the spectrum at 2729, 3118 and 3638 m/z correspond to a protein with a mass of 21821 Da, which can be assigned to the BRIL-HspB8ACD monomer. And the peaks at 3638, 3969, 4369 and 4853 m/z represent charge states 12+, 11+, 10+ and 9+ of a

protein with a mass of 43645 Da, which can be assigned to the BRIL-HspB8ACD dimer. Since the IPI/V motif is absent on the CTD, HspB8 is unlikely to form homo-oligomers, and the full-length protein should have similar behavior as the ACD construct.

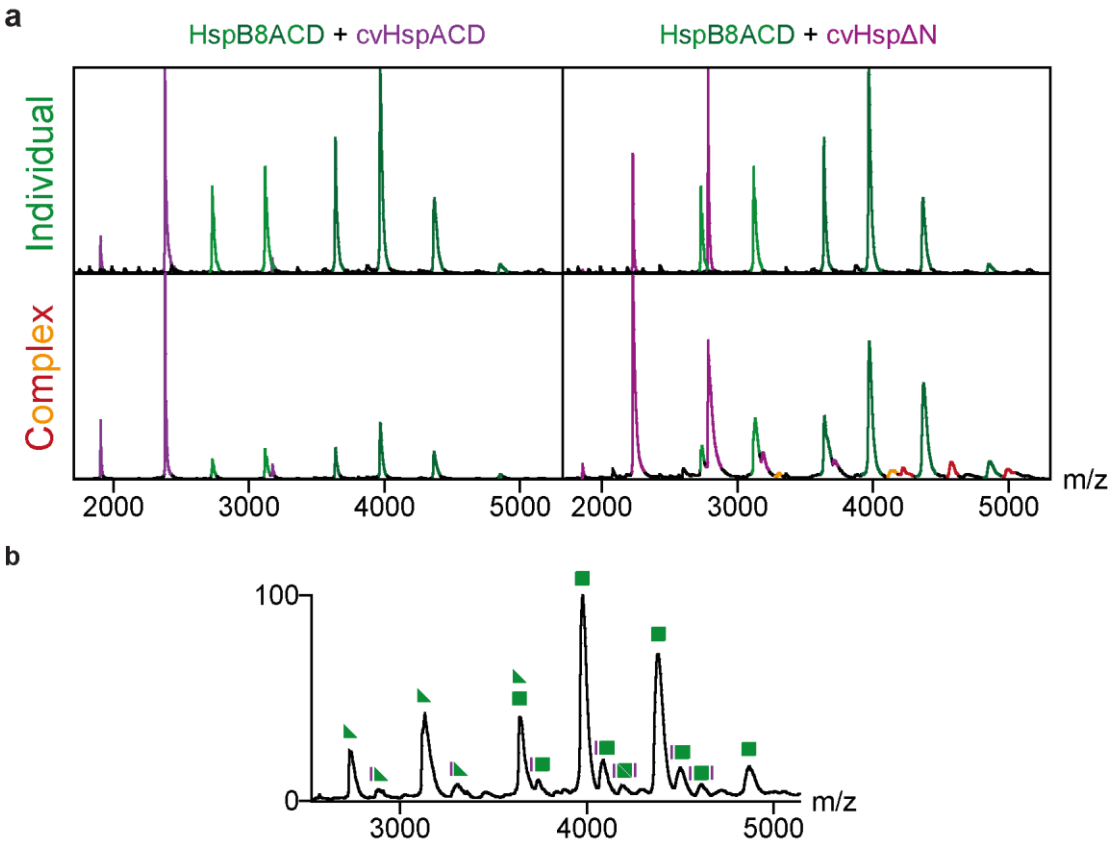


Figure 4.3. HspB8ACD is interacting with cvHsp through an ACD-CTD interface. a. Native MS experiments show that HspB8ACD cannot bind cvHspACD but can interact with cvHspΔN. Red peak series represent complex of HspB8ACD dimer binding with cvHspΔN monomer, while orange peaks represent complex made of HspB8ACD monomer and cvHspΔN monomer. The upper row is combined spectra of individual proteins, and the bottom row shows the interactions between proteins mixed. **b.** Native MS experiment shows that HspB8ACD can bind cvHspCTD. The orchid stick represents the cvHspCTD peptide.

We then tested whether the β4-β8 hydrophobic groove of this HspB8ACD

construct is functional for the lateral experiments with BAG3. We used the αBCCTD peptide with the IPI motif that we used in Chapter 2, which has been proved to have the highest affinity to bind sHsp ACD β4-β8 groove, to incubate with the BRIL-HspB8ACD protein. By native MS, we observed the interaction between αBCCTD peptide with BRIL-HspB8ACD (Figure **4.2b**). Aside from the main peak series, the shoulder peak series represent the complex made of αBCCTD peptide and BRIL-HspB8ACD. The interaction between the protein and αBC peptide confirms that the β4-β8 groove of the HspB8 construct is functional.

cvHsp has been reported to be the interacting partner with HspB8 without the involvements of both NTDs[39], though in some other reports, HspB8 could not bind any other sHsps[40]. We tested the bindings between the BRIL-HspB8ACD with cvHspACD and cvHspΔN. On native MS, we could not observe any new species formed by cvHspACD and BRIL-HspB8ACD. Interestingly, HspB8ACD has a limited binding affinity to cvHspΔN, and we could observe peaks represent the complex of cvHspΔN monomer with BRIL-HspB8ACD monomer and BRIL-HspB8ACD dimer (Figure **4.3a**). This suggests that there are no ACD-ACD interfaces between cvHspACD and BRIL-HspB8ACD, but cvHspCTD may have some roles in linking these two proteins. We then used cvHspCTD peptide to mix with BRIL-HspB8ACD. We observed new peak species formed by cvHspCTD and BRIL-HspB8ACD. (Figure **4.3b**). Our data suggest that there are no interactions between cvHsp and HspB8 through the ACD-ACD interfaces, but the β4-β8 groove

of HspB8ACD is hydrophobic enough to bind cvHspCTD. We have been also working on the crystallography of HspB8ACD to probe the unique behaviors of this protein, but more efforts on such as mutagenesis or storage conditions are still needed to obtain the crystals.

4.3.2 Purification attempts of BAG3

We chose HspA1A, which is a stress-inducible member of the Hsp70 family, for our study. The expression and purification of HspA1A was straightforward (Appendix figure **1.8**), and the purified full-length HspA1A was distributed as monomer and dimer peak series by native MS (Figure **4.4a**). The peaks on the spectrum at 4457, 4760, 5100, 5492 and 5949 m/z correspond to charge states 16+, 15+, 14+, 13+ and 12+ of a protein with a mass of 71362 Da, which can be assigned to the HspA1A monomer. And the peaks at 6493, 6804, 7144, 7522 and 7927 m/z represent charge states 22+, 21+, 20+,19+ and 18+ of a protein with a mass of 142822 Da, which can be assigned to the HspA1A dimer. We also planned to purify the N-terminal construct of HspA1A, which will be used for the interaction with JG-98 and crystallography.

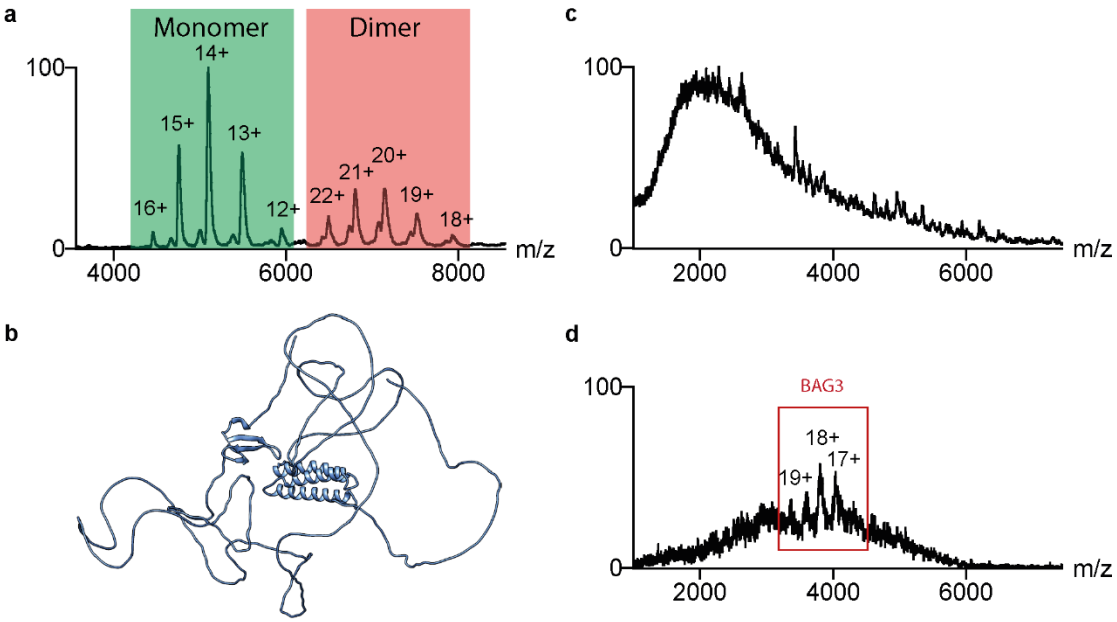


Figure 4.4. BAG3 protein is non-structured and can be resolved on native MS after reversed-phase chromatography. a. HspA1A protein distributes as monomer and dimer on native MS. **b.** AlphaFold predicted the structure of full-length human BAG3 protein[41,42]. The protein is predicted to be mostly non-structured except for the BAG domain (helices bundle) and the WW domain (β-sheet). **c.** BAG3 protein was shown as a hump and could not get resolved on native MS due to interactions with impurities. **d.** Peaks represent BAG3 protein becomes resolvable after reversed-phase chromatography.

The purification of BAG3 was challenging as expected since the protein is non-structured and can bind a variety of proteins from *E. coli*, such as Dnak (Figure **4.4b**). After the nickel-column, anion exchange, and SEC (Appendix figure **1.9**), the protein still existed as a big smooth hump on the MS. The hump did not show up in the big or small m/z value regions, which suggests that the BAG3 protein does not oligomerize or aggregate in the gas phase. The bounded adducts should be removed to improve the quality of purification. We tried eluting the protein with ATP to at least get rid of Dnak, which is the *E. coli* chaperone protein, but no effect was achieved. Since the impurities from the cell bounded to BAG3 tightly, the

additional purification step should be aiming to unfold and separate the protein and impurities like SDS-PAGE. The protein also needs to be recycled and refolded properly. A practical technique is reversed-phase chromatography HPLC. The column is made of non-polar aliphatic chain-bonded silica (C8 or C18) as the stationary phase. The protein is first denatured by organic acid like formic acid. The non-polar solvent that is miscible with water can be used for elution as the mobile phase. The pre-treated denatured protein can bind to the column and be eluted through a gradient by the solvent. The proteins with subtle differences on polarity can be separated by reversed-phase chromatography at optimized conditions. The protein can be recycled and refolded in functional buffer[43]. Since BAG3 is mostly non-structured, this method might work for its purification. Followed by this method (Appendix figure **1.10, 2.2**), we could see the peak series represent BAG3 protein on native MS (Figure **4.4c, d**). The peaks on the spectrum at 3611, 3804 and 4035 m/z correspond to charge states 19+, 18+ and 17+ of a protein with a mass of 68541 Da roughly, which can be assigned to the BRIL-BAG3 monomer. Though the theoretical mass of BRIL-BAG3 should be around 73 kDa, considering the weak signal intensity and the error caused by the broad peaks, this is still enough to identify the protein. The peaks were still not ideal, and the recovering rate of the protein from refolding is very low at this stage, but this has already been good progress. More efforts will be made to get the purified and functional BAG3 from the optimization of reversed-phase chromatography or some other methods like co-expression and co-purification.

4.4 Discussion

The function and stability of HspB8 are highly dependent on BAG3[18,44]. In the CASA process, HspB8 is also degraded together with the CASA complex and unfolded filamins in the autophagosome. The short lifetime of HspB8 *in vivo* suggests that this protein is unstable. Our experiences in dealing with the cardiac sHsp *in vitro* (cvHsp and HspB8) reveal that the NTD of cvHsp is unstable and tends to truncate by itself, while HspB8 is even unstable with just the ACD. We found that the HspB8ACD construct without BRIL truncated by itself at specific sites (data not shown). It is understandable that the up-regulated cardiac-specific sHsps at stress conditions need to be eliminated during stress recovery, but the detailed mechanisms and how the proteins get self-truncated are awaiting further studies.

Our data suggest that cvHsp is not interacting with HspB8 through the ACD-ACD contact. The BRIL-HspB8ACD distributes as monomers and dimers, which suggests the canonical β6+7 dimer interface is accessible and is not hindered by the BRIL. The β4-β8 groove of HspB8ACD guarantees the CTD-ACD interface between HspB8 and cvHsp. It is likely that BAG3 with more favorable motifs can easily compete with this interaction for the hydrophobic pocket on HspB8. The knock-out of cvHsp leads to the aggregation of FLNC, and HspB8 targets d19-21

of non-functioned FLNC in CASA. cvHsp has never been reported to exist in the CASA complex. What's the state of FLNCd24 in CASA can be a big factor to influence the role of cvHsp in the CASA process.

BAG3 is a non-structured co-chaperone and has certain affinities to bind a variety of proteins, which makes its purification difficult. It is the key scaffold of the CASA complex. On the other hand, the adaptive characteristic of BAG3 might make its function easy to be recovered from unfolded state. The BAG3 bound to Hsp70 and sHsps might also stabilize the protein itself and makes it more rigid. A combining study of BAG3 with Hsps might be achievable for structural studies.

Overall, the study will combine cvHsp and FLNC, which has already been discussed in the previous chapters, together with HspB8, BAG3, and the other CASA complex components as well as the inhibitors to get a deeper insight into this heart-specific process.

Acknowledgments

I thank Jani Bola for providing the vectors used in this chapter and help in design the sequences. I appreciate Di Wu for his help in establishing the Agilent instrument for reverse-phase chromatography on BAG3 purification.

4.5 Materials and methods

Protein expression and purification

All the constructs encoded in the pET15 vector with the C-terminal 6×His tag were transformed in *Escherichia coli* BL21(DE3) cells (Agilent). TEV protease cleavage site was designed at the C-terminal. Cells (BRIL-HspB8ACD and HspA1A, but not BRIL-BAG3) were used to inoculate a 10 mL LB media with ampicillin (100 μg/mL) at 37°C overnight and transferred to a 1 L large LB culture with same antibiotics. The culture was shaking at 180 rpm at 37°C until the OD600 reached 0.6-0.8 and followed by the induction with IPTG to a final concentration of 0.5 mM. After another overnight shaking at 22°C, the cells were harvested, pelleted and lysed using French Press Microfluidizer with the addition of EDTA-free protease inhibitor cocktail and buffer containing 300 mM NaCl, 50 mM Tris, and 20 mM imidazole (pH 8.0). The lysate was then spun at 20,000 g for 20 min at 4°C and filtered through the 0.22 μm filter to get clarified before loading to the HisTrap HP column. The protein was eluted with 300 mM NaCl, 50 mM Tris and 500 mM imidazole (pH 8.0) and dialyzed to the loading buffer with TEV protease at room temperature overnight. The sample was again loaded to the HisTrap HP column to remove the His-tag. The purified HspA1A was buffer-exchanged to PBS with 5 mM TCEP. BRIL-HspB8ACD was concentrated and loaded to the SEC column (Superdex 200, 16/600) in PBS. The partially unfolded BRIL-HspB8 protein was

then dialyzed in PBS at 4°C overnight and then concentrated and stored.

The BRIL-BAG3 cell was cultured in terrific broth (TB) media with the same antibiotics and protocol. The cell lysis and IMAC remained the same as mentioned above. The protein was then exchanged to the buffer with 20 mM NaCl and 50 mM Tris (pH 8.0) and loaded to the HiTrap Q anion exchange column, and eluted with 1 M NaCl, 50 mM Tris (pH 8.0). The eluted protein was then concentrated for SEC (Superdex 200, 16/600).

Reversed-phase chromatography

The BRIL-BAG3 protein purified after SEC was concentrated and incubated with 0.1% formic acid to get denatured. The C8 reverse phase column (Agilent Technologies, 5 µm) was connected to the Agilent HPLC instrument. The denatured sample was loaded to the column equilibrated with 0.1% formic acid. Then the column with bounded protein was eluted with buffer containing 0.1% formic acid and 80% acetonitrile. The BRIL-BAG3 protein was eluted when the elution buffer contains 0.1% formic acid and 51.2% acetonitrile. The eluted purified BRIL-BAG3 protein was transferred for dialysis in PBS overnight and recycled for use.

Native MS

BRIL-HspB8ACD can be bio-spun into 200 mM ammonium acetate (pH 6.9), while

HspA1A and BRIL-BAG3 were buffer-exchanged into ammonium acetate. All the spectra were recorded from the Synapt G1 mass spectrometer modified for the transmission of intact non-covalent protein complexes (Waters Corp.). The parameters were set as: capillary, 1.4-1.6 kV; sampling cone, 20-35 V; extraction cone, 3.0-3.3 V; backing pressure, 3.8-4.1 mbar; trap gas (argon) flow, 1.5-3.0 mL/min; trap energy, 6-8 V; transfer energy, 6-12 V. A reservoir of acetonitrile was put in the sample cone to facilitate the signal intensity when peptides were being sprayed.

Bibliography

1. Arndt, V. *et al.* Chaperone-Assisted Selective Autophagy Is Essential for Muscle Maintenance. *Current Biology* **20**, 143–148 (2010).
2. Ulbricht, A. *et al.* Cellular mechanotransduction relies on tension-induced and chaperone-assisted autophagy. *Current Biology* **23**, 430–435 (2013).
3. Ulbricht, A. *et al.* Induction and adaptation of chaperone-assisted selective autophagy CASA in response to resistance exercise in human skeletal muscle. *Autophagy* **11**, 538–546 (2015).
4. Cuervo, A. M. & Wong, E. Chaperone-mediated autophagy: Roles in disease and aging. *Cell Research* **24**, 92–104 (2014).
5. Franceschelli, S. *et al.* Bag3 Gene Expression Is Regulated By Heat Shock Factor 1. *Journal of Cellular Physiology* **215**, 575–577 (2008).
6. Ulbricht, A. & Höhfeld, J. Tension-induced autophagy: May the chaperone be with you. *Autophagy* **9**, 920–922 (2013).
7. Gamerdinger, M. *et al.* Protein quality control during aging involves recruitment of the macroautophagy pathway by BAG3. *EMBO Journal* **28**, 889–901 (2009).
8. Myers, V. D. *et al.* The Multifunctional Protein BAG3: A Novel Therapeutic Target in Cardiovascular Disease. *JACC: Basic to Translational Science* vol. 3 122–131 (2018).

9. Rauch, J. N. & Gestwicki, J. E. Binding of human nucleotide exchange factors to heat shock protein 70 (Hsp70) generates functionally distinct complexes in vitro. *Journal of Biological Chemistry* **289**, 1402–1414 (2014).

10. Rauch, J. N. *et al.* BAG3 Is a Modular, Scaffolding Protein that physically Links Heat Shock Protein 70 (Hsp70) to the Small Heat Shock Proteins. *Journal of Molecular Biology* **429**, 128–141 (2017).

11. Ganassi, M. *et al.* A Surveillance Function of the HSPB8-BAG3-HSP70 Chaperone Complex Ensures Stress Granule Integrity and Dynamism. *Molecular Cell* **63**, 796–810 (2016).

12. Cicardi, M. E. *et al.* Autophagic and proteasomal mediated removal of mutant androgen receptor in muscle models of spinal and bulbar muscular atrophy. *Frontiers in Endocrinology* **10**, 1–14 (2019).

13. Arakawa, A., Handa, N., Shirouzu, M. & Yokoyama, S. Biochemical and structural studies on the high affinity of Hsp70 for ADP. *Protein Science* **20**, 1367–1379 (2011).

14. Massey, A. J. *et al.* A novel, small molecule inhibitor of Hsc70/Hsp70 potentiates Hsp90 inhibitor induced apoptosis in HCT116 colon carcinoma cells. *Cancer Chemotherapy and Pharmacology* **66**, 535–545 (2010).

15. Li, X. *et al.* Validation of the Hsp70-Bag3 protein-protein interaction as a potential therapeutic target in cancer. *Molecular Cancer Therapeutics* **14**, 642–648 (2015).

16. Villard, E. *et al.* A genome-wide association study identifies two loci

associated with heart failure due to dilated cardiomyopathy. *European Heart Journal* **32**, 1065–1076 (2011).

17. Fang, X. *et al.* Loss-of-function mutations in co-chaperone BAG3 destabilize small HSPs and cause cardiomyopathy. *Journal of Clinical Investigation* **127**, 3189–3200 (2017).

18. Fang, X., Bogomolovas, J., Trexler, C. & Chen, J. The BAG3-dependent and -independent roles of cardiac small heat shock proteins. *JCI insight* **4**, 1–12 (2019).

19. Vos, M. J. *et al.* HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. *Human Molecular Genetics* **19**, 4677–4693 (2010).

20. Kim, M. v., Seit-Nebi, A. S., Marston, S. B. & Gusev, N. B. Some properties of human small heat shock protein Hsp22 (H11 or HspB8). *Biochemical and Biophysical Research Communications* **315**, 796–801 (2004).

21. Fuchs, M. *et al.* Identification of the key structural motifs involved in HspB8/HspB6-Bag3 interaction. *Biochemical Journal* **425**, 245–255 (2010).

22. Nivon, M. *et al.* NF-κB regulates protein quality control after heat stress through modulation of the BAG3-HspB8 complex. *Journal of Cell Science* **125**, 1141–1151 (2012).

23. Fuchs, M. *et al.* A Role for the Chaperone Complex BAG3-HSPB8 in Actin Dynamics, Spindle Orientation and Proper Chromosome Segregation during Mitosis. *PLoS Genetics* **11**, 1–32 (2015).

24. Mymrikov, E. v., Daake, M., Richter, B., Haslbeck, M. & Buchner, J. The chaperone activity and substrate spectrum of human small heat shock proteins. *Journal of Biological Chemistry* **292**, 672–684 (2017).

25. Kasakov, A. S., Bukach, O. v., Seit-Nebi, A. S., Marston, S. B. & Gusev, N. B. Effect of mutations in the β5-β7 loop on the structure and properties of human small heat shock protein HSP22 (HspB8, H11). *FEBS Journal* **274**, 5628–5642 (2007).

26. Shemetov, A. A. & Gusev, N. B. Biochemical characterization of small heat shock protein HspB8 (Hsp22)-Bag3 interaction. *Archives of Biochemistry and Biophysics* **513**, 1–9 (2011).

27. Al-Tahan, S. *et al.* New family with HSPB8-associated autosomal dominant rimmed vacuolar myopathy. *Neurology: Genetics* **5**, 1–11 (2019).

28. Muranova, L. K., Ryzhavskaya, A. S., Sudnitsyna, M. v., Shatov, V. M. & Gusev, N. B. Small Heat Shock Proteins and Human Neurodegenerative Diseases. *Biochemistry (Moscow)* **84**, 1256–1267 (2019).

29. Quintana, M. T. *et al.* Cardiomyocyte-Specific Human Bcl2-Associated Anthanogene 3 P209L Expression Induces Mitochondrial Fragmentation, Bcl2-Associated Anthanogene 3 Haploinsufficiency, and Activates p38 Signaling. *American Journal of Pathology* **186**, 1989–2007 (2016).

30. Yu, F. X. & Guan, K. L. The Hippo pathway: Regulators and regulations. *Genes and Development* **27**, 355–371 (2013).

31. Martin, A. P. *et al.* STK 38 kinase acts as XPO 1 gatekeeper regulating the

nuclear export of autophagy proteins and other cargoes . *EMBO reports* **20**, 1–20 (2019).

32. Martin, A. P. J. & Camonis, J. H. The hippo kinase STK38 ensures functionality of XPO1. *Cell Cycle* **19**, 2982–2995 (2020).

33. Klimek, C. *et al.* The Hippo network kinase STK38 contributes to protein homeostasis by inhibiting BAG3-mediated autophagy. *Biochimica et Biophysica Acta - Molecular Cell Research* **1866**, 1556–1566 (2019).

34. Mukherjee, S. *et al.* Synthetic antibodies against BRIL as universal fiducial marks for single−particle cryoEM structure determination of membrane proteins. *Nature Communications* **11**, 1–14 (2020).

35. Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20**, 967–976 (2012).

36. Chu, R. *et al.* Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography. *Journal of Molecular Biology* **323**, 253–262 (2002).

37. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).

38. Raman, S. *et al.* Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins: Structure, Function and Bioinformatics* **77**, 89–99 (2009).

39. Sun, X. *et al.* Interaction of Human HSP22 (HSPB8) with Other Small Heat Shock Proteins. *Journal of Biological Chemistry* **279**, 2394–2402 (2004).

40. Boelens, W. C. Structural aspects of the human small heat shock proteins related to their functional activities. *Cell Stress and Chaperones* **25**, 581–591 (2020).

41. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* (2021) doi:10.1038/s41586-021-03828-1.

42. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

43. Žuvela, P. *et al.* Column Characterization and Selection Systems in Reversed-Phase High-Performance Liquid Chromatography. *Chemical Reviews* **119**, 3674–3729 (2019).

44. Carra, S., Seguin, S. J., Lambert, H. & Landry, J. HspB8 chaperone activity toward poly(Q)-containing proteins depends on its association with Bag3, a stimulator of macroautophagy. *Journal of Biological Chemistry* **283**, 1437–1444 (2008).

Concluding remarks

I have focused on cardiac sHsps (especially cvHsp and HspB8) and their interacting partners in this thesis. Our detailed structural and biophysical analysis provide new understandings of the mechanisms of these essential cardiac and skeletal proteostasis relevant interactions.

In Chapter 2, I proved that cvHsp is dominantly monomeric, which is unique in the human sHsps family. By analyzing the crystal structure of cvHspACD, I found that its CTD hinders the oligomer formation of cvHsp, and the shortened and distinct β5-β6+7 region makes dimerization less favored for cvHsp. I confirmed the ACD-ACD interactions between cvHsp with Hsp27 and αBC, and I revealed that the chaperone function of cvHsp locates on its NTD.

In Chapter 3, I observed the strong and specific interaction between cvHsp and its binding partner FLNCd24. The main binding interfaces involve cvHsp β4 and FLNCd24 strand C, and the interaction competes for the dimerization of FLNCd24. Phosphorylation on specific serine and threonine residues of FLNCd24 weakens its interaction with cvHsp, which is an effect of acute stress like electric stimulation. The most significant phosphorylation on T2677 strengthens FLNCd24 dimerization as well. In sarcoma and tumor tissues, tyrosine kinases

are activated. Y2683 residue can be phosphorylated at such conditions, and the highly charged strand D makes FLNCd24 dimer no longer stable. But the binding with cvHsp is strengthened when a new salt-bridge interaction is formed between the phosphorylated Y2683 and cvHsp R113 on the β5 strand.

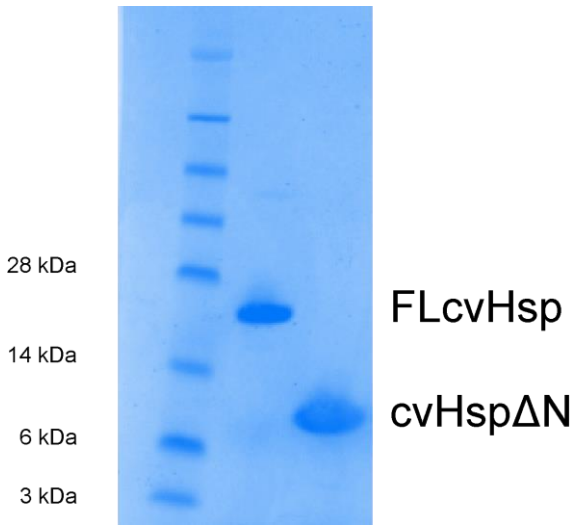
The last chapter is about the preliminary research on the cardiac-specific HspB8 and the CASA complex for filamin clearance. HspB8 was found to be very unstable, even for the ACD construct. I showed that HspB8ACD exists as an equilibrium between monomers and dimers. The ACD-ACD interaction between cvHspACD and HspB8ACD was not observed, but HspB8ACD can bind the CTD of cvHsp. The purification of BAG3, which is the key factor of the CASA complex, has been found difficult because this non-structured co-chaperone is binding a lot of impurities from the cell. The purification is now being optimized by unfolding and refolding this protein.

In general, both heart-specific cvHsp and HspB8 participate in maintaining the proteostasis of the cytoskeletal protein FLNC. cvHsp binds native FLNCd24 to keep FLNC functional and protects it from aggregation. Combined with other reports relevant to other domains of FLNC and its binding partners, the figure for FLNC can be: 1) In response to acute stress conditions, FLNC tends to be more rigid for the regions as a dimer. This is mostly through serine and threonine phosphorylations on FLNC itself and other relevant proteins in the Z-discs. 2) In tumor tissues, FLNC needs to be more mobile to take part in the cytoskeletal

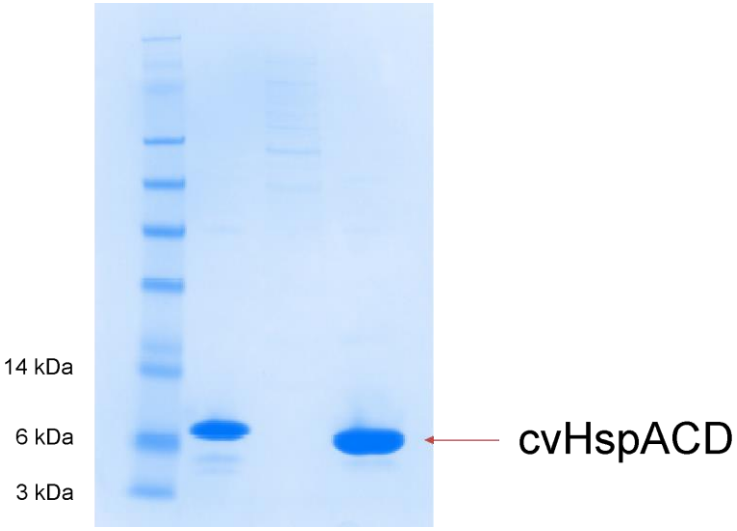
system formation through cell growth and differentiation. The activation of tyrosine kinases is one of the signs in sarcomas and tumors, and the dimeric FLNC is more likely to dissociate under these circumstances. More needs to be studied about FLNC in cancer. 3) At normal conditions, all these processes are maintained at a balanced level. FLNC is a largely consumable protein as a sensor towards mechano-force. The over-stretched FLNC can be degraded through CASA, which involves HspB8, BAG3, Hsp70, and other components. HspB8 binds aggregated FLNC straightforward at the insertion region. The work in this thesis provides some mechanistic insight into muscular proteostasis, and the research potential of this system is still worth developing.

Appendix

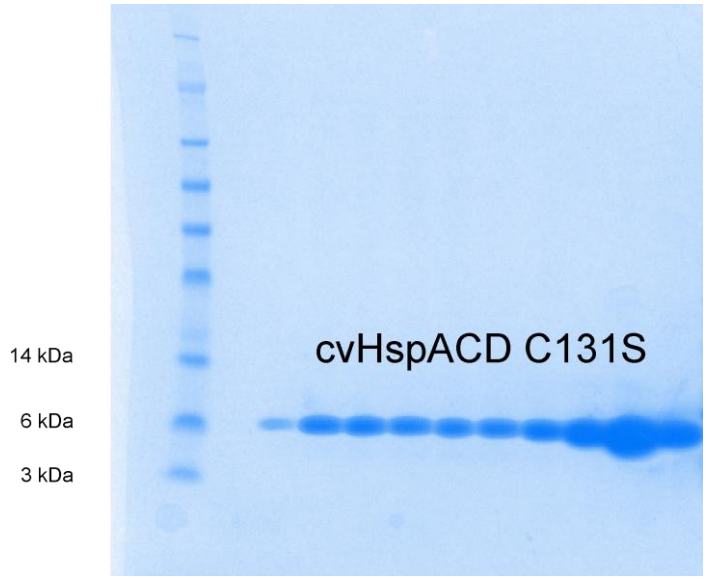
1. Gels and proteolysis



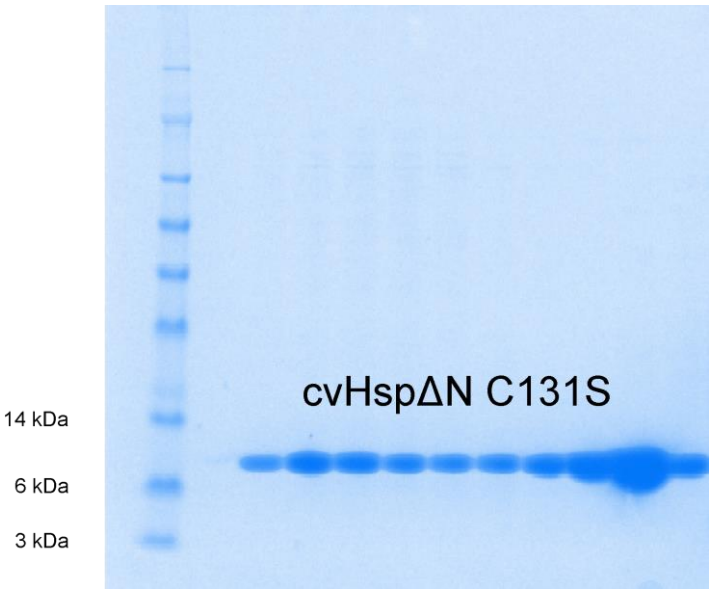
Appendix figure 1.1. SDS-PAGE of purified FLcvHsp and cvHspΔN.



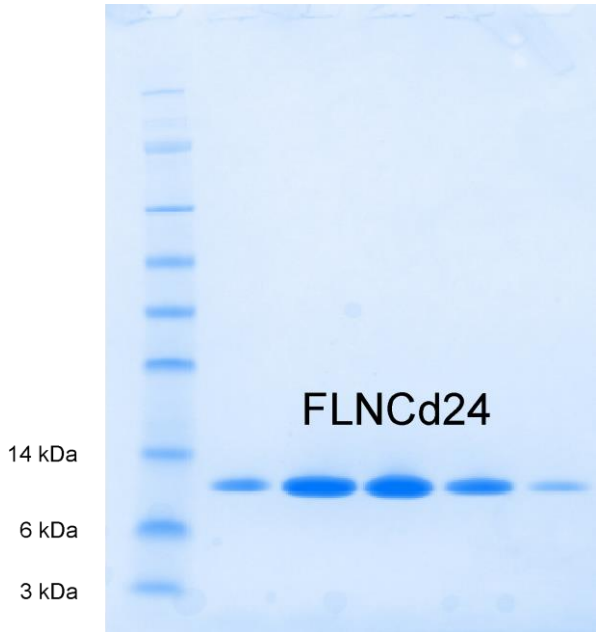
Appendix figure 1.2. SDS-PAGE of purified cvHspACD.



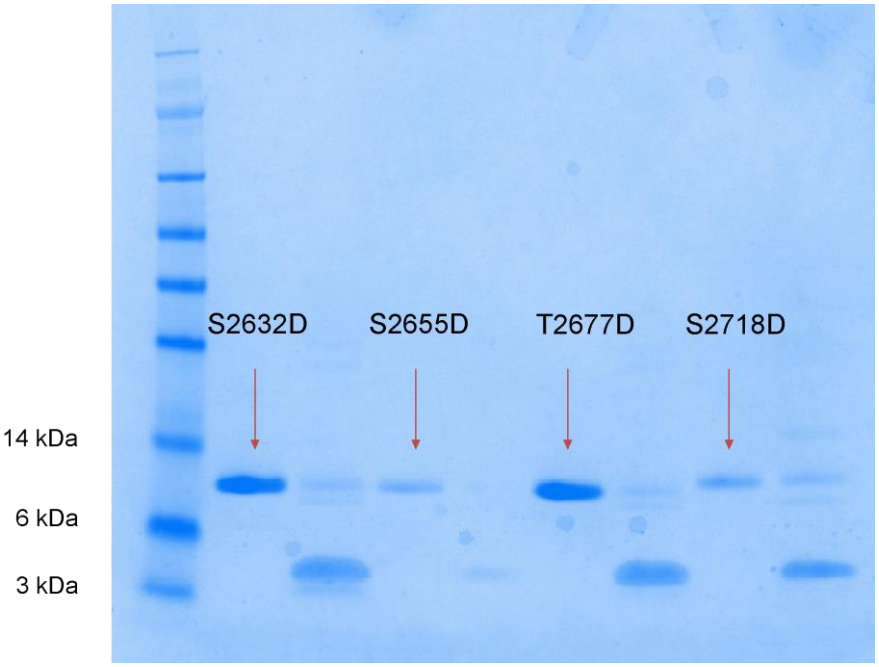
Appendix figure 1.3. SDS-PAGE of purified cvHspACD C131S.



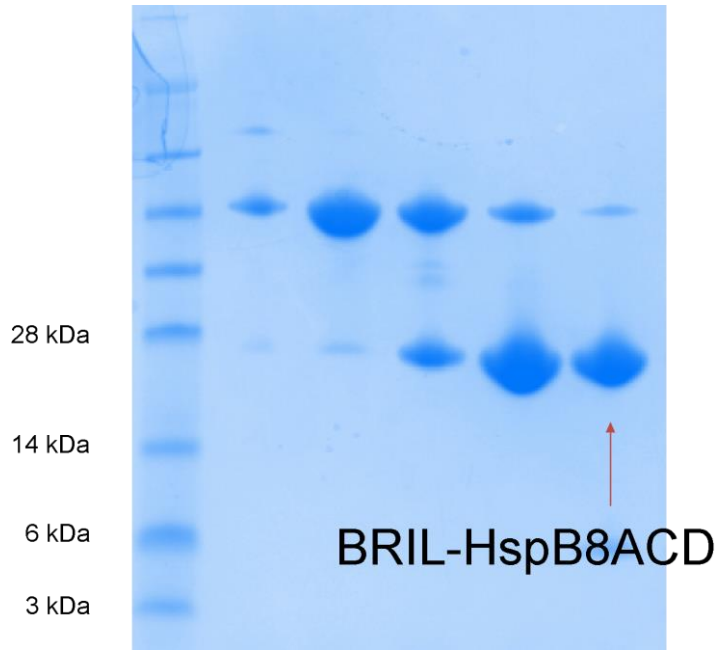
Appendix figure 1.4. SDS-PAGE of purified cvHspΔN C131S.



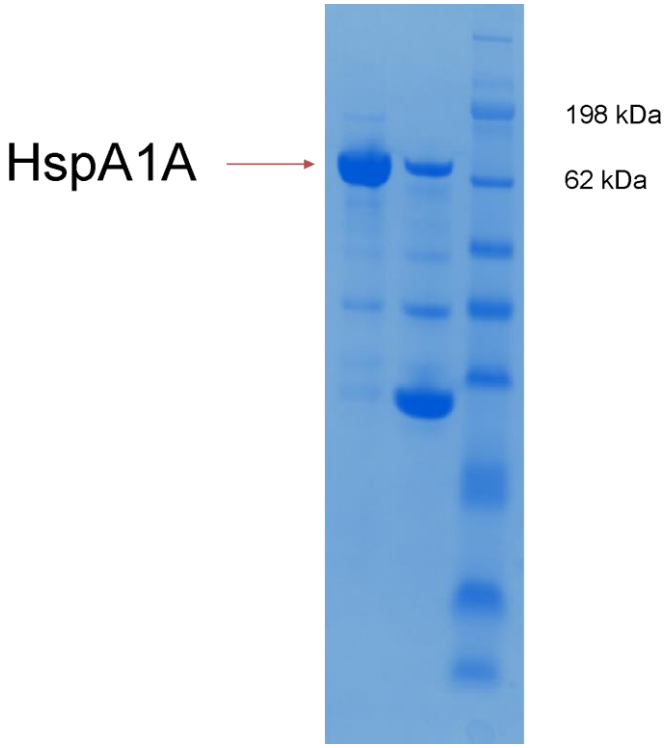
Appendix figure 1.5. SDS-PAGE of purified FLNCd24.



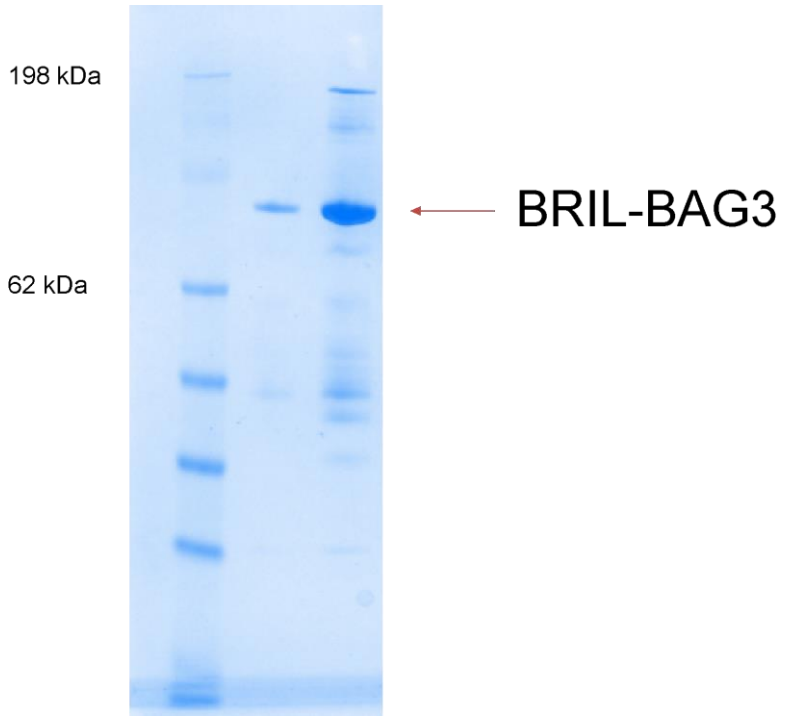
Appendix figure 1.6. SDS-PAGE of purified FLNCd24 phosphorylation mimics.



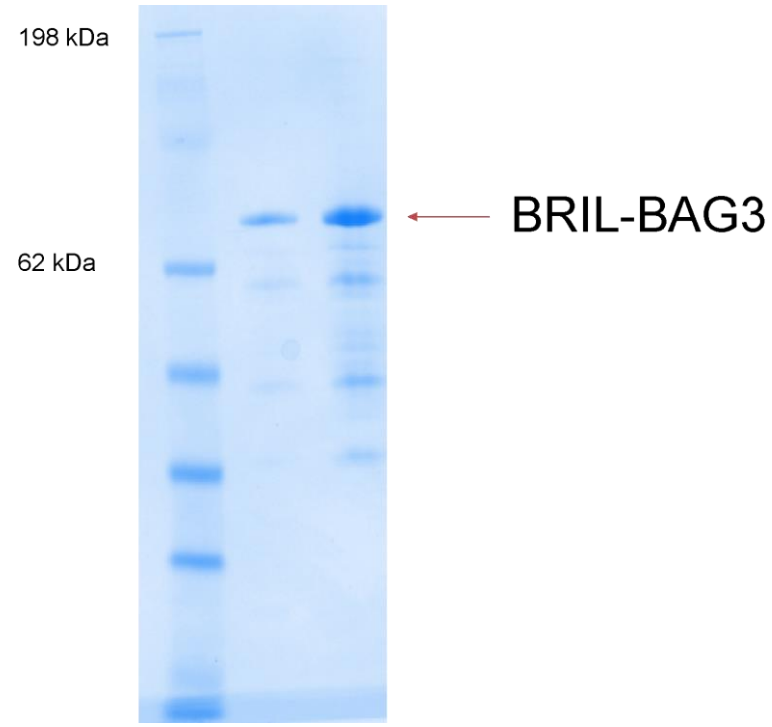
Appendix figure 1.7. SDS-PAGE of purified BRIL-HspB8ACD. Some dimers are visible on the gel due to the high loading concentration.



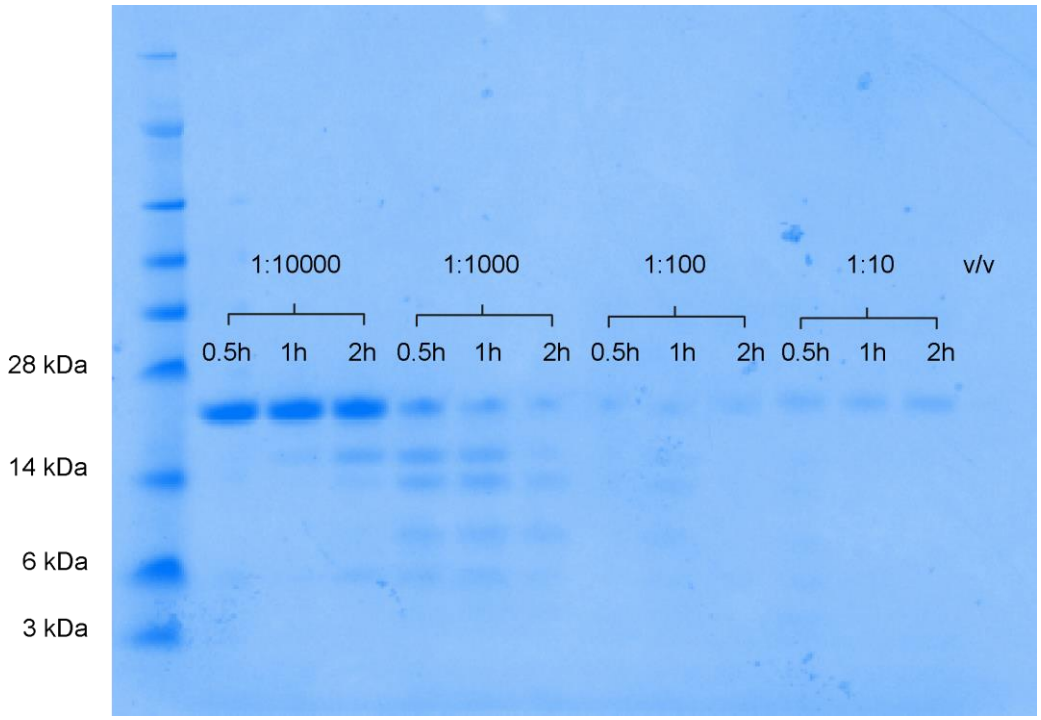
Appendix figure 1.8. SDS-PAGE of purified HspA1A.



Appendix figure 1.9. SDS-PAGE of purified BRIL-BAG3 (pre-reversed phase).

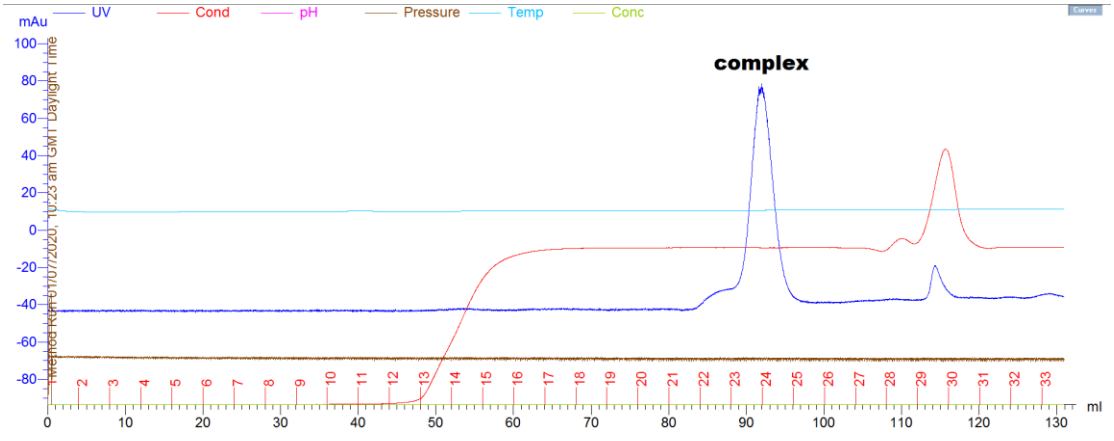


Appendix figure 1.10. SDS-PAGE of purified BRIL-BAG3 (post-reversed phase). Some high molecular weight impurities were eliminated after the reversed phase chromatography.



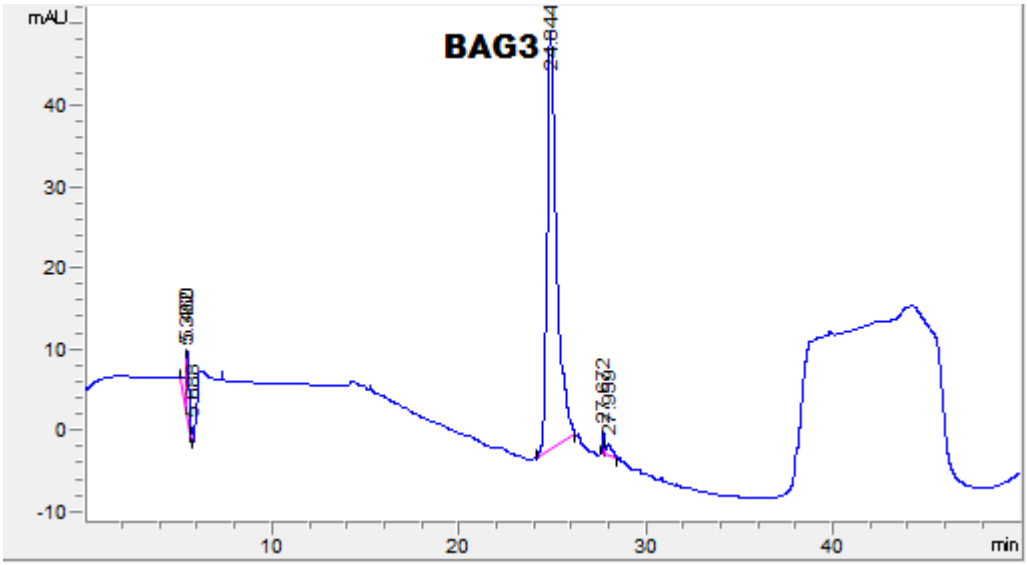
Appendix figure 1.11. Trypsin proteolysis of FLcvHsp. Trypsin was dissolved into 1 mM HCl and 2 mM $CaCl_2$. FLcvHsp was maintained at 10 µM. Coomasie and 70% of formic acid were used to stop the reactions.

2. Chromatography



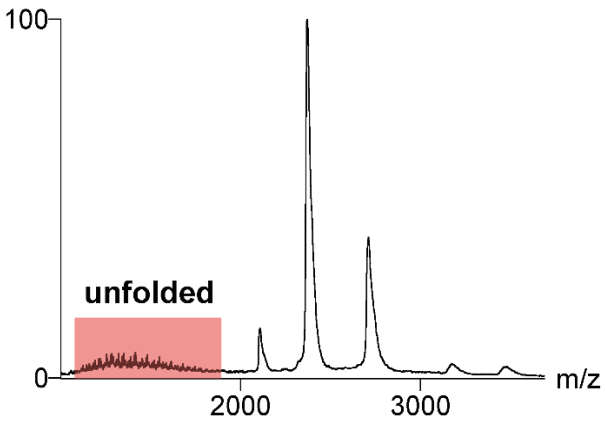
Appendix figure 2.1. Chromatography of SEC for the separation of cvHspACD

C131S-FLNCd24 complex.

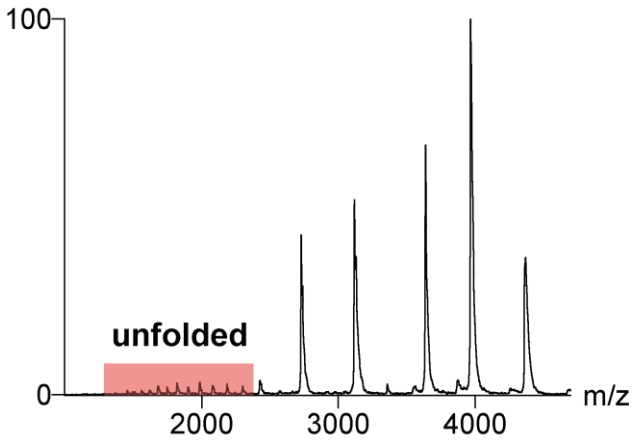


Appendix figure 2.2. Chromatography of reversed-phase chromatography of BRIL-BAG3 purification.

3. Native MS



Appendix figure 3.1. Native mass spectrum showing the unfolded species of FLcvHsp.



Appendix figure 3.2. Native mass spectrum showing the unfolded species of BRIL-HspB8ACD.

4. Equations for K_D calculation

All the K_D and ΔG values of the titrations in the thesis were calculated in the same way. Firstly,

$$K_D = \frac{[L][R]}{[com]}$$

where $[L]$ represents the concentration of ligand, $[R]$ represents the concentration of free receptors (the protein defined as receptor gives the specific binding on Y axis), and $[com]$ represents the concentration of complex. The value on the Y axis can be presented as

$$Y = \frac{[com]}{[com] + [R]}$$

The equation can now also be written as

$$Y = \frac{\frac{[L][R]}{[K_D]}}{\frac{[L][R]}{[K_D]} + [R]}$$

$$Y = \frac{[L]}{[L] + K_D}$$

So K_D can be calculated as

$$K_D = \frac{[L]}{Y} - [L]$$

And Gibbs free energy ΔG can be obtained through

$$\Delta G = -RTlnK_D$$

In which R is the ideal gas constant (8.314 $J \cdot K^{-1} \cdot mol^{-1}$), T is temperature (commonly used as 298 K).