Molecular dynamics computational analysis of protein interactions involving small heat shock proteins in human cardiac muscle tissue



Guodong Cao

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2024

Acknowledgements

During the course of this degree, I have received a lot of help from my family, colleagues and friends. I am extremely grateful to all new friends I have met and every person I have spent time with during this adventure.

I would not be able to fulfil any targets in science without the collaboration and help from my colleagues who generously share their knowledge and experience with me. I started my course together with Angela Gehrckens, Wiktoria Sadowska, Josh Bishop and Luke Smith, and we have shared a great time together in the group. Since the beginning of my DPhil course, Zihao Wang and Dominik Šaman have been sharing their knowledge in science with me, and also supportive words when I received negative results or error messages. Later, I had a fortune to cross paths with DPhil students including Ellie Holden, Laila Shah, Navoneel Sen, Olav Johannesen, Konstantin Zouboulis, Francesca Naughton-Allen, Alexander O'Shea and Sulaiman Sulaiman, and enjoyed activeness and laughter they brought to me and good luck to the rest of their DPhil courses. Thanks to Sabine Huth-Rauschenbach, Andy Dolan and everyone else in the management group of Department of Chemistry and Kavli Institute for help running the group and dealing with common issue.

Many collaborators have shared their understanding in computational and experimental biochemistry with me. Many many thanks to Matteo Degiacomi for his guidance on implementing MD calculation and patience while I bothered him with my poor scripts, and his enthusiasm in Chinese martial arts. Thanks to Joana Filipa Costeira Paulo for her knowledge in steered MD and her nice paintings. Thanks to Miranda Collier, Xingyu Qiu, Georg Hochberg, Katja Gemlich and to everyone who has spent their time and effort together with me in the projects.

I owe the most immense debt of gratitude to Justin Benesch for his support and instruction in science throughout the course, and his optimism in scientific research and everyday life. Thanks him for offering me the chance to become a scientific researcher and the opportunity to learn the skills and to explore in the field of science freely. Also, he has often reminded me, especially when I became perplexed about the progress or the significance of my work, that what I was doing was not merely a simulation or a series of data analysis, but a contribution to a great research, which I could be more proud of.

I am grateful to Linacre College, which has a great atmosphere and a group of kind and professional staff, and Oxford University Go Club, where I had a lot of fun during my leisure time.

I would like to pay gratitude to my friends, especially to Shuwen, Yuru, Xinrong, Shiyuan, Jiacheng, Jingyu and Zhongyi, whom I met in high school and kept in touch since then. I am always so excited every time I meet everyone you in person or receive your message while I wake up which you send by mid-day.

Many thanks to Liangti, now Dr. Dai, and I wish you enjoy everything in your life. Finally, thanks to my parents for being my parents and I am so glad to be your son.

Abstract

Small heat-shock proteins (sHSPs) are a class of molecular chaperones that are essential for the development and functionaility of human cardiac muscle tissue. Various types of sHSPs are expressed and localised within the sarcomere, the fundamental functional unit of striated muscle tissue, where they play a significant role in the regulation of sarcomeric proteins. To investigate the interactions between sHSPs and cardiac muscle proteins, we employ molecular dynamics (MD) simulations to model the behaviour of proteins and protein complexes.

Our studies initially examine the interaction between HspB1 and B5, which is essential for regulating the chaperoning functions of both sHSPs in cardiac muscle tissues. We confirm that HspB1 and B5 are capable of both self-assembly and co-assembly, resulting in the formation of hetero-oligomers. Notably, the affinities of the interactions between the α-Crystallin domain (ACD) and the C-terminal domain (CTD) of HspB1 and HspB5 are not unifrom. Due to their varying interaction affinities, HspB1 and B5 can associate in different assembly states, which can modulate the chaperoning capabilities of both proteins. Based on binding energies and inter-residue distances calculated from MD simulations, we find that the CTD of HspB5 demonstrates greater binding affinities with ACDs compared to the CTD of HspB1. This discrepancy is attributed to the presence of additional arginine residues in the sequence, a conclusion that is corroborated by simulations of complexes involving the last common ancestral (LCA) protein of HspB1 and B5.

Next, we examine the regulatory effect of HspB7 on filamin C (FLNC), a protein that is essential for the formation of the sarcomeric backbone. The interaction between HspB7 and the FLNC domain 24 (FLNCd24) competes with the dimerisation of FLNCd24, a process that is vital for the functionality of filamin C as a linker of actin filaments. Additionally, our findings indicate that two phosphorylation sites within FLNCd24 modulate the equilibrium between two types of interactions. Specifically, phosphorylation of FLNCd24 at T2677 enhances homo-dimerisation, thereby surpassing the interaction between FLNCd24 and HspB7, which is critical for the normal function of FLNC. Conversely, pathogenic phosphorylation at Y2683 produces completely opposite effects.

Our studies also investigate the interaction between HspB5 and titin, with a particular emphasis on how this interaction modulates the mechanical properties

of titin. Titin, a fibrous protein that serves as the structural backbone of the sarcomere, exhibits elastic properties and can be elongated in response to external forces. Our findings suggest that the binding of the immunoglobulin-like domain 26, an elastic region of titin, to the C-terminal segment of HspB5 enhances its resistance to mechanical stress. However, this regulatory effect is significantly reduced in the presence of the R157H mutation in HspB5.

Contents

List of Figures **xi**

List of Tables **xv**

List of Abbreviations **xvii**

1 Small Heat-Shock Proteins **1**

1.1 Small heat-shock proteins protect proteins from denaturation 2

1.1.1 Molecular chaperones ensure correct protein folding and preventing protein aggregation 2

1.1.2 Small heat-shock proteins are a class of molecular chaperones 3

1.1.3 Structural characteristics of human sHSPs 5

1.2 sHSPs modulate the mechanical properties of cardiovascular protein complexes 9

1.2.1 Sarcomeres are the fundamental structural units of cardiac muscle 9

1.2.2 Various types of sHSPs are expressed within the sarcomere to regulate the function of sarcomeric proteins 10

1.2.3 Modifications of sHSPs and sarcomeric proteins modulate cardiovascular functions 12

2 Molecular Dynamics Simulations of Proteins and Protein Complexes **15**

2.1 Implementation of molecular dynamics for the simulation of protein structures 16

2.1.1 Molecular simulations and molecular dynamics 16

2.1.2 Fundamental theory of MD simulations 21

2.2 MD simulations of proteins and protein-ligand complexes 25

2.2.1 Identifying the sequences and initial structures of proteins . 26

2.2.2 Preparation of input files and parameters 26

2.2.3 MD simulation process 27

2.3 Analysis of protein conformations and protein-protein interaction in a MD trajectory 29

2.3.1 Structural alignment and analysis of conformational stability 30
2.3.2 Principal component analysis . . . 31
2.3.3 Molecular Mechanics Poisson-Boltzmann (or Generalised Born) Surface Area . . . 33
2.4 Steered molecular dynamics: a method to investigate protein behaviour under stress . . . 35
2.5 Supporting Techniques . . . 39
2.5.1 X-ray Crystallography . . . 39
2.5.2 Native Mass Spectrometry . . . 41
2.5.3 AlphaFold . . . 44

3 HspB1 and HspB5 hetero-oligomerise via ACD-CTD interactions 47
3.1 Abstract . . . 48
3.2 Introduction . . . 49
3.2.1 Functional properties of HspB1 and B5 . . . 49
3.2.2 The co-assembly of HspB1 and B5 influences the bioactivity of sHSPs. . . . 50
3.2.3 Plan of experiments and analysis . . . 52
3.3 Methods . . . 53
3.3.1 Preparation of the initial structures . . . 53
3.3.2 MD simulations . . . 54
3.3.3 Trajectory preprocessing and data analysis . . . 54
3.4 Results and discussions . . . 56
3.4.1 HspB1-HspB5 co-assembly is biased as a result of ACD-CTD interactions. . . . 56
3.4.2 HspB1 CTD forms less intense ACD-CTD interactions than HspB5 CTD. . . . 59
3.4.3 The LCA CTD demonstrates similarity to the HspB5 CTD due to conserved arginines . . . 65
3.5 Conclusions . . . 72

4 Phosphorylations at single residues regulate the dimerisation of filamin C domain 24. 75
4.1 Abstract . . . 76
4.2 Introduction . . . 76
4.2.1 FLNCd24 plays a crucial role in cardiac muscle tissue . . . 76
4.2.2 HspB7 is a predominantly monomeric sHSP . . . 78
4.2.3 Residue phosphorylations modify the function of FLNCd24. 81
4.2.4 Plan of experiments and analysis . . . 83
4.3 Methods . . . 83

4.3.1 Preparation of initial structures 83
4.3.2 MD simulations 84
4.3.3 Preprocessing and data analysis 84
4.4 Results and discussions 86
4.4.1 HspB7 is dominantly monomeric, unlike most other sHSPs . 86
4.4.2 Phosphorylations of T2677 and Y2683 regulate the protonation state of H2673. 88
4.4.3 The competition between FLNCd24-HspB7 interaction and FLNCd24 dimerisation. 90
4.4.4 The phosphorylation of T2677 and Y2683 results in distinct regulatory effects on the homo- and hetero-interactions . . . 95
4.5 Conclusions 109

5 The R157H mutation of HspB5 CTD regulates the mechanosensitivity of the titin immunoglobulin-like domain 26 113
5.1 Abstract 114
5.2 Introduction 115
5.2.1 The role of titin in regulating the mechanical properties of cardiac muscle 115
5.2.2 HspB5 modulates the retraction of titin through its interaction with titin Ig26 117
5.2.3 The mutation point: R157 or R163 118
5.2.4 Plan of experiments and analysis 118
5.3 Methods 119
5.3.1 MD simulations 119
5.3.2 Steered MD simulations 120
5.4 Results and discussions 121
5.4.1 Native-MS experiments verifies the interactions between titin Ig26 domain and HspB5 CTD peptide, which is weakened by the R157H mutation of the CTD peptide. 121
5.4.2 Determination of pulling velocity and identification of R-H mutation point 123
5.4.3 HspB5 CTD peptide binds titin Ig26 domain, and the interaction is weakened by R157H mutation. 126
5.4.4 Interaction with the HspB5 CTD peptide protects titin Ig26 against unfolding under mechanical force, which is diminished by the R157H mutation. 131
5.5 Conclusions 140

6 Concluding remarks **145**
6.1 Conclusion of the work 145
6.2 Future potential applications of our technique 148
6.2.1 Co-assembly of HspB1 and B6 148
6.2.2 The structure of HspB8 and its function in the sarcomere . . 148

A Appendix **151**

References **159**

List of Figures

1.1 The conformation of the human sHSP ACD region 5
1.2 AP interfaces of sHSP dimer 7
1.3 The conformation of a HspB5 monomer in HspB5 24-mer 9
1.4 The structure of a sarcomere 10

2.1 Improper torsion of a covalent bond 19

3.1 Structural models of HspB1 and B5 large oligomers. 50
3.2 The LCA of HspB1 and B5 52
3.3 Reconstruction of LCA 53
3.4 Ensembles of HspB1-B5 hetero-oligomers. 57
3.5 Homo- and hetero-dimerisation of HspB1 and HspB5 ACD. 58
3.6 Native-MS results of mixtures of heterogeneous ACD and CTD. . . 59
3.7 RMSD and RMSF of HspB1/B5 ACD-CTD complexes. 61
3.8 P-values from t-tests among the binding energy values generated from the repeats. 62
3.9 Frequencies of close contacts between intermolecular residue pairs in HspB1/B5 ACD-CTD complexes. 63
3.10 Snapshots of the interfaces of HspB1/B5 ACD-CTD complexes . . . 65
3.11 RMSD and RMSF of LCA-related ACD-CTD complexes. 66
3.12 P-values from t-tests among the binding energy values generated from the repeats of LCA-associated ACD-CTD pairs. 68
3.13 Frequencies of close contacts between intermolecular residue pairs in ACD-CTD complexes containing LCA segment. 69
3.14 Snapshot of interfaces between LCA ACD and HspB5/LCA CTD . 71

4.1 FLNC in Z-disc 77
4.2 Structure of FLNCd24 dimer. 79
4.3 The sequence alignment of HspB1, B5 and B7 80
4.4 Aligned structures of HspB1, B5, and B7 ACD monomers 81
4.5 Results of the native-MS experiment of 1:1 mixed HspB7 ACD and CTD, alongside controlling of HspB1 and B5. 86
4.6 RMSD and RMSF of residues in HspB1, B5, and B7 monomers. . . 88

4.7 Conformations of HspB7 and HspB1 ACD dimer 89
4.8 Residue RMSD and RMSF of phosphorylated and protonated FLNCd24-HspB7ACD complexes. 90
4.9 Experiments demonstrate the interaction between FLNCd24 and HspB7 ACD. 91
4.10 Intermolecular residue-residue distances of the FLNCd24 dimer and the FLNCd24-HspB7ACD complex. 93
4.11 The distances between all residues in FLNCd24 and its interacting partners. 94
4.12 The probabilities of hydrogen bonds between intermolecular residue pairs in the FLNCd24 dimer and the FLNCd24-HspB7ACD complex. 96
4.13 Results of Native MS experiment of mixture containing WT and mutated FLNCd24 and HspB7 ACD. 97
4.14 The distances between residues in a FLNCd24 molecule and the interactors, analysed for both WT and T2677-modified structures. . 100
4.15 The differences in hydrogen bond probabilities between inter- and intra-molecular residue pairs in WT and pT2677 FLNCd24 homodimers. 101
4.16 The differences in hydrogen bond probabilities between intermolecular residue pairs in WT and T2677 modified FLNCd24-HspB7ACD complexes 103
4.17 The distances between residues in a FLNCd24 molecule and the interactors, analysed for both WT and Y2683-modified structures. . 105
4.18 The differences in hydrogen bond probabilities between inter- and intra-molecular residue pairs in WT and Y2683 modified FLNCd24-HspB7ACD complexes 106
4.19 The differences in hydrogen bond probabilities between intermolecular residue pairs in WT and Y2683 modified FLNCd24-HspB7ACD complexes 107
4.20 A ccatter map of hydrogen bonds based on their probabilities in phosphorylated and phosphomimicking mutated structures. 110

5.1 Structure of the sarcomere and titin, along with a schematic representation of titin isoforms in cardiac muscle tissue. 116
5.2 Composition of N2A and N2B regions of titin 117
5.3 Native MS results of the interaction of titin Ig26 and HspB5 CTD . 122
5.4 Time points of separation of all beta strands pairs within Ig domains unfolded at different velocities. 125
5.5 Differences of accumulated work of WT and R-H mutated complexes 126
5.6 The positions of HspB5 CTD peptides, either WT or R157H-mutated, within the complexes. 127

5.7 Residue RMSD and RMSF of WT and R157H complexes and free titin Ig26 128
5.8 The occupancy ratio of inter-residue contacts among inter-molecular residue pairs in the complexes. 130
5.9 PCA of backbone atom coordinates of titin Ig26 domains 131
5.10 The differences in intramolecular residue-residue distances within titin Ig26 molecules. 132
5.11 Instantaneous force and accumulated work of the two complexes and the free titin Ig26 domain. 133
5.12 The intramolecular inter-residue distances of the titin Ig26 domain at 6 and 8 ns during the simulations 135
5.13 The intermolecular inter-residue distances of the WT and R157H complexes at 6 and 8 ns within the SMD simulations 137
5.14 onformational snapshots of the folded regions within the titin Ig26 domain at 6 and 8 ns during the SMD simulation. 138
5.15 The breaking points of intermolecular inter-residue contacts in the complexes. 139
5.16 The breaking points of intramolecular inter-residue contacts in the titin Ig26 domains. 141

A.1 Distributions of hydrogen bond probabilities in WT and phosphorylated complexes. 155
A.2 Differences between distributions of hydrogen bond probabilities in WT and phosphorylated complexes. 156

List of Tables

1.1 Nomenclature and properties of sHSPs 4

2.1 Universal parameters of MD simulations presented in this thesis . . 28

3.1 Binding energies of ACD-CTD complexes consisting of HspB1 and B5 segments, calculated by mmPBSA. 61
3.2 Double-sided P-values from t-test between values of binding energies of ACD-CTD complexes 62
3.3 Frequencies of intermolecular hydrogen bonds in the HspB1/B5 ACD-CTD complexes 64
3.4 Binding energies of ACD-CTD complexes containing LCA segments, calculated by mmPBSA. 67
3.5 Double-sided P-values from t-test between values of binding energies of ACD-CTD complexes related and unrelated to LCA segments . . 67
3.6 Frequencies of hydrogen bonds between LCA ACD and all CTDs . 70
3.7 Frequencies of hydrogen bonds between the HspB1/B5 ACD and the LCA CTD 71

4.1 Binding energies of the sHSP ACD dimers 88
4.2 Binding energies of the WT FLNCd24 dimer and the WT FLNCd24-HspB7ACD complex, and p-values from t-tests for the variation of the repeating simulations 92
4.3 Binding energies of the pT2677 and T2677D FLNCd24 dimer and FLNCd24-HspB7ACD complex, p-values from t-tests between binding energies of WT and modified structures, and from t-tests for the variation of repeating simulations. 98
4.4 Binding energies of the pY2683 and Y2683E FLNCd24 dimer and FLNCd24-HspB7ACD complex, p-values from t-tests between binding energies of WT and modified structures, and from t-tests for the variation of repeating simulations. 104
4.5 T-test of binding energies of T2677 and Y2683 modified complexes . 108
4.6 Correlation of h-bond occupancies between phosphorylated and phosphomimicking protein complexes, revealed by p-values calculated by the Wilcoxon signed-rank test 109

5.1 The binding energies of the WT and R157H titin Ig26-HspB5 CTD complex, along with the p-values obtained from t-tests utilised to evaluate the variations across the repeating simulations. 129

A.1 Raw data of binding energies of ACD-CTD complexes consisting of HspB1 and B5 segments, calculated by mmPBSA. 152
A.2 Raw data of binding energies of ACD-CTD complexes consisting of LCA segments, calculated by mmPBSA. 153
A.3 Binding energies of WT and modified FLNCd24 homodimers and FLNCd24-HspB7ACD complexes 154
A.4 Raw data of binding energies between titin Ig26 and HspB5 CTD peptides, calculated by mmPBSA. 157

List of Abbreviations

ACD . . . . . α-Crystallin like Domain

AMBER . . . Assisted Model Building with Energy Refinement

ASR . . . . . Ancestral Sequence Reconstruction

ATP . . . . . Adenosine triphosphate

atm . . . . . standard atmosphere

BAG3 . . . . . Bcl-2-associated athanogene 3

Bcl-2 . . . . . B-cell lymphoma 2

CASA . . . . . chaperone-assisted selected autophagy

CHARMM . . Chemistry at Harvard Macromolecular Mechanics

CoM . . . . . Centre of mass

Cryo-EM . . . Cryogenic electron microscopy

CTD . . . . . C-terminal domain

FLNC . . . . . Filamin C

FLNCd24 . . . FLNC domain 24

HSP . . . . . heat-shock protein

Ig . . . . . (titin) Immunoglobulin-like domain

LCA . . . . . lowest common ancestor

MAPK . . . . mitogen-activated protein kinase

MD . . . . . molecular dynamics

MM . . . . . molecular mechanics

mmGBSA . . Molecular Mechanic Generalised Born Surface Area

mmPBSA . . . Molecular Mechanic Poisson-Boltzmann Surface Area

MP . . . . . mass photometry

MS . . . . . mass spectrometry

MSA . . . . . multiple sequence alignment

NAMD . . . . Nanoscale Molecular Dynamics

NMR . . . . . Nuclear magnetic resonance

NTD . . . . . N-terminal domain

PCA . . . . . Principal component analysis

PME . . . . . Particle Mesh Ewald

PTMs . . . . . Post-translational modifications

QM . . . . . Quantum mechanics

RMSD . . . . . Root Mean Square Deviation

RMSF . . . . . Root Mean Square Fluctuation

sHSPs . . . . . small heat shock proteins

SMD . . . . . Steered Molecular Dynamics

VMD . . . . . Visual Molecular Dynamics (software)

WT . . . . . wild type

1

Small Heat-Shock Proteins

Contents

1.1 Small heat-shock proteins protect proteins from denaturation 2

1.1.1 Molecular chaperones ensure correct protein folding and preventing protein aggregation 2

1.1.2 Small heat-shock proteins are a class of molecular chaperones 3

1.1.3 Structural characteristics of human sHSPs 5

1.2 sHSPs modulate the mechanical properties of cardiovascular protein complexes 9

1.2.1 Sarcomeres are the fundamental structural units of cardiac muscle 9

1.2.2 Various types of sHSPs are expressed within the sarcomere to regulate the function of sarcomeric proteins 10

1.2.3 Modifications of sHSPs and sarcomeric proteins modulate cardiovascular functions 12

1.1 Small heat-shock proteins protect proteins from denaturation

1.1.1 Molecular chaperones ensure correct protein folding and preventing protein aggregation

The conformation of a protein is significantly influenced by both its amino acid sequence, commonly referred to as the primary structure, as well as by the manner in which this sequence folds to form higher-order structures. The higher-order structures of proteins are classified into three categories: secondary structure, which pertains to the organisation of amino acids within a localised region of a polypeptide chain; tertiary structure, which describes the three-dimensional organisation of the entire polypeptide chain; and quaternary structure, which involves the assembly of multiple polypeptide chains. These higher-order structures are critical in determining the functionality of proteins. It is notable that most proteins exhibit a consistent pattern in their higher-order structures, which is essential for the attainment of their normal functions[1].

Proteins may fail to maintain their native folding state, which can result in dysfunction[2, 3], toxicity[4], and harmful aggregation[5], all of which can significantly contribute to the etiology of various diseases. The factors that contribute to protein misfolding can be categorised as internal, including genetic mutations[6] and post-translational modifications (PTMs)[7], or external, such as cellular stress[8]. Notably, protein folding is intricately associated with the assembly and aggregation of multi-subunit proteins[5]. Certain proteins function as oligomers, making it essential to maintain the correct assembly pattern to ensure their functionality. In contrast, misfolded proteins with exposed hydrophobic surfaces can promote the formation of harmful aggregates[2, 3, 4].

Molecular chaperones represent a specific class of proteins that facilitate the proper folding of other proteins through their binding interactions. These molecular chaperones not only assist in the correct folding of their interactors but also stabilise their correctly folded conformations. In instances of protein misfolding, molecular chaperones can aid in the refolding process or recognise irreversibly

misfolded proteins, subsequently directing them toward degradation[9, 10, 11]. Furthermore, molecular chaperones are frequently implicated in the processes of protein assembly and disassembly, as well as in the prevention or reversal of pathogenic protein aggregation[12, 13, 14].

Heat-shock proteins (HSPs) are a class of molecular chaperones that were initially identified due to their functional upregulation in response to rapid increases in temperature, a phenomenon commonly referred to as "heat shock"[15, 16]. Subsequent research has demonstrated that HSPs also play a critical role in the protection and repair of dysfunctional proteins that become misfolded as a result of various external factors[17, 18, 19, 20]. Through their interactions with a diverse species of proteins, HSPs are implicated in numerous biochemical processes, including the monitoring and degradation of problematic proteins, intracellular transportation, antigen presentation, and the regulation of cardiovascular functions[21, 22, 23, 24]. HSPs are classified into several families based on their masses, including the family of small heat-shock proteins[25, 26].

1.1.2 Small heat-shock proteins are a class of molecular chaperones

Small heat-shock proteins (sHSPs) are a class of HSPs characterised by masses ranging from approximately 15 to 30 kDa[27]. These sHSPs engage in ATP-independent interactions with their target proteins, thereby preventing their misfolding or unfolding. Although sHSPs do not possess the ability to refold their interactors through ATP-dependent mechanisms, they can recognise misfolded proteins and facilitate their refolding in collaboration with other molecular chaperones from the Hsp70 and Hsp100 families[28, 29].

Interactions with sHSPs are essential for the proper folding and normal functioning of specific proteins, including certain cardiovascular proteins[30, 31]. Among the various classes of molecular chaperones, sHSPs are particularly notable due to their ability to oligomerise; these oligomers may acquire distinct functions in response to their higher-order conformations[32]. The expression of sHSPs is ubiquitous across

nearly all organisms, ranging from archaea and bacteria to all forms of eukaryotes, including humans. The diversity of sHSP types varies among different organisms; for instance, the human genome encodes ten types of sHSPs (table1.1), which is more than most prokaryotes but fewer than certain animal and plant species.

Protein	Oligomerisation	Distribution	Stress-inducibility	Promiscuity
HspB1	oligomer	ubiquitous	yes	yes
HspB2	small oligomer	brain and muscle	no	no
HspB3	dimer and trimer	brain and muscle	no	no
HspB4	oligomer	crystalline lens	no	yes
HspB5	oligomer	ubiquitous	yes	yes
HspB6	dimer	ubiquitous	no	no
HspB7	monomer	striated muscle	no	no
HspB8	monomer	ubiquitous	cell-type dependent	no
HspB9	unknown	testis	no	no
HspB10	unknown	testis	no	no

Table 1.1: Nomenclature and properties of sHSPs

Among the ten types of human sHSPs, HspB1, B5, B6, and B8 are widely distributed throughout the human body and are associated with conferring resistance against external stress, including heat shock and other environmental pressures, which has an impact on numerous proteins[33]. In contrast, the remaining sHSPs are expressed exclusively in specific tissues and are associated with the normal functioning of those tissues[34, 35]. A range of brain and muscle disorders, including those related to cellular autophagy and apoptosis[36, 37], as well as neurodegenerative diseases such as Alzheimer's disease[38], Parkinson's disease[39] and multiple sclerosis[40], have been proven to be linked to the pathogenic aggregation of misfolded proteins. Consequently, sHSPs present in tissues involved in the detection and degradation of harmful protein aggregates are associated with the monitoring and treatment of these diseases. Furthermore, certain sHSPs have been reported to be highly expressed in cardiac and skeletal muscle tissues, where they play a critical role in the functional regulation of muscle proteins[41, 42].

1.1.3 Structural characteristics of human sHSPs

All human sHSPs exhibit a universal structural composition comprised of a coiled N-terminal domain (NTD), an α-Crystallin domain (ACD) composed of β strands, and a flexible C-terminal domain (CTD) that is shorter than the NTD[43], as displayed in fig1.1. In human sHSPs, the ACD is the only domain that retains an explicit secondary structure, consisting of seven β strands, commonly labelled from β2 to β9, including an extended strand referred to as β6+7. These strands are clearly divided into two β sheets: one sheet contains β4, β5, and the extended strand, while the other contains all the remaining strands. The integration of β6 and β7 is conserved among animal and higher eukaryotic sHSPs; however, in plant and prokaryotic sHSPs, β6 and β7 strands are separated[44]. Regardless of the integration of β6 and β7, these strands function as one of the primary binding sites for sHSP ACDs, particularly in relation to their homogeneous interactions. In prokaryotic and plant systems, two sHSP monomers can interact by exchanging their short β6 strand; conversely, in animal and higher eukaryotic sHSPs, two merged β6+7 strands bind to one another to form an interface.

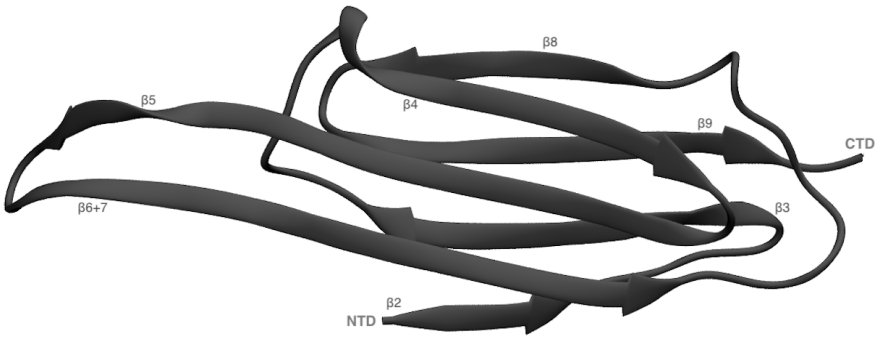


Figure 1.1: The conformation of the human sHSP ACD region. Human sHSP ACD region consists of two β sheets, one containing four strands, β2-β3 and β8-β9, and the other one containing β4, β5 and the extended β6+7.

Interactions between two β6+7 strands facilitate the anti-parallel docking of two ACD domains, resulting in the formation of an sHSP dimer[45]. In addition to acquiring distinct biological functions, certain sHSP dimers can aggregate into larger oligomers in specific tissues within the human body, including the brain, heart,

and skeletal muscle[46]. The conformation of the dimer interface is determined by interactions, primarily hydrogen bonds, between residues located on or adjacent to the β6+7 strands, and is classified based on the interacting surface area. Three registers of the anti-parallel interfaces (AP) in human sHSP oligomers have been identified and designated as AP_I, AP_{II} and AP_{III}(fig1.2), in accordance with the size of the dimer interface. The dimer interfaces corresponding to the AP_I register exhibit the greatest overlap between the two β6+7 strands, whereas those of the AP_{III} register demonstrate the least overlap. Within different oligomers, a single type of sHSP monomer can dimerise according to different registers. The AP_I dimer interface contains a significant number of positively charged residues and thus necessitates the stabilising effect of small anions. Consequently, the AP_I interface is commonly observed only in crystal structures of sHSP dimers, accompanied by inorganic anions located between the interfaces. For example, the AP_I dimer interface in an HspB5 homodimer (PDB 3l1g) is stabilised by hydrogen bonds formed between a sulphate ion and R116 or R120[47]. The other two registers, AP_{II} and AP_{III}, are formed throught salt bridges and hydrogen bonds that connect residues across the interface.

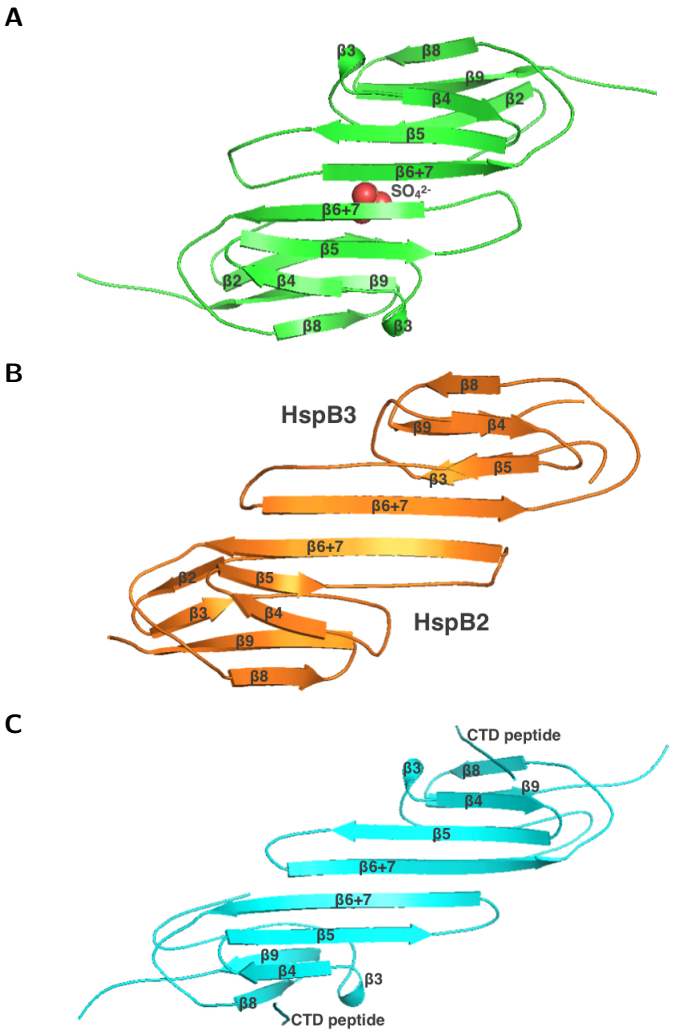


Figure 1.2: **A** An AP_I dimer interface of the HspB5 dimer, extracted from PDB 3L1G, which is stabilised by a sulfate ion. Within an AP_I dimer interface, the entire β6+7 strand, along with a part of the loop, contributes to the formation of the interface. **B** An AP_{II} dimer interface of HspB2-B3 dimer, extracted from PDB 6F2R. The AP_{II} dimer interface contains all, or the overwhelming majority, of the residues within the β6+7 strands. As showed in the structure, the dimer interface comprises the complete β6+7 strand of HspB3 monomer and all but one of the residues located on the β6+7 strand of the HspB2 monomer, since it has one more residue assigned to the β6+7 strand compared to the HspB3 monomer. **C** An AP_{III} dimer interface of the HspB5 dimer, extracted from PDB 4M5S. In an AP_{III} interface, not all residues of the β6+7 strands are occupied; specifically, nine of the eleven residues in each β6+7 strand overlap at the interface in the 4M5S structure.

In addition to the β6+7 strand, human sHSPs exhibit various ligand-binding sites, including a surface formed by the β4 and β8 strands, as well as exposed regions within the NTD and CTD, which are capable of binding to another sHSP or heterogeneous interactors[48]. These binding sites are predominantly hydrophobic and demonstrate less sequential and conformational consistency than the dimerisation site across different types of sHSPs. Subsequent to this diversity, these binding sites have a higher selectivity for target proteins than the dimerisation site, thereby ensuring the interaction specificity of sHSPs. The β4-β8 surface, together with adjacent loops, forms a highly hydrophobic pocket that serves as an optimal receptor for short hydrophobic peptides[49]. A common class of ligands that fit within this pocket is a highly hydrophobic motif, I-X-I or I-X-V (where X is ideally a hydrophobic amino acid), which located at the terminus, particularly within the CTDs, of sHSPs[50]. Interactions between the β4-β8 pockets and CTD peptides facilitate the bridging between sHSP dimers during the assembly of sHSP oligomers [51]. Short peptides containing the hydrophobic motif in the central region are frequently employed as substitutes for full-length CTDs in studies investigating ACD-CTD interactions. In contrast to CTDs, NTDs are generally more disordered, structurally unstable, and less conserved among different types of sHSPs[52]. As a result, few conformations of NTD-involved interfaces have been resolved, regardless that ACD-NTD and NTD-NTD interactions are critical for the formation of large sHSP oligomers[53].

HspB1 and B5 are the only human sHSPs identified that can form large homo-oligomers *in vivo*[54]. Additionally, HspB4 has been observed to co-assemble with HspB5 within the eye lenses[55], HspB6 can form large hetero-oligomers with HspB1 and/or B5 in striated muscle tissue, and HspB2 and B3 can hetero-oligomerise in a strict 3:1 ratio[56]. The secondary structures of sHSP monomers in dimers or small oligomers commonly exhibit a high resemblance to the regular secondary structure shown in fig1.1. However, in large oligomers, sHSP monomers frequently undergo conformational changes in that enhance the inter-dimer interactions, thereby stabilising the oligomer structure (fig1.3).

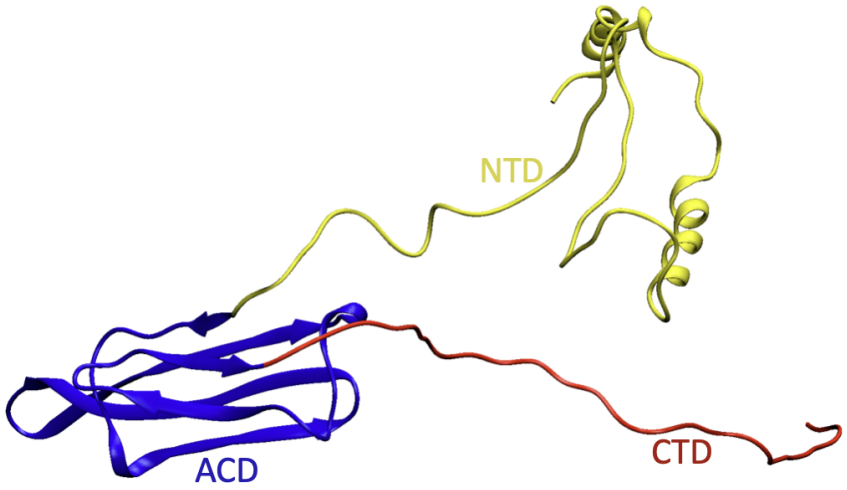


Figure 1.3: The conformation of a full-length HspB5 monomer within an HspB5 24-mer (PDB ID:3J07). Notably, the conformation of the ACD in this context, in contrast to the conformation observed in the homo-dimer as shown in fig1.1, comprises only six shorter β strands. Consequently, the number of residues contained within the β strands are much fewer, suggesting that the contribution of coiled flexible regions to the formation of large sHSPs oligomers is greater than that within dimer or smaller oligomers.

1.2 sHSPs modulate the mechanical properties of cardiovascular protein complexes

1.2.1 Sarcomeres are the fundamental structural units of cardiac muscle

Cardiac muscle, also called the myocardium, is a significant component of the human heart wall and is primarily composed of sarcomeres[57], which are long band-like units that contain various types of cardiovascular proteins. During the process of contraction-excitation coupling in cardiac muscle, which facilitates the capacity of heart to pump blood, signals are conducted among sarcomeric proteins in the form of electrical signals and mechanical forces[58, 59]. Consequently, various species of cardiovascular proteins have a high sensitivity to mechanical force, and their responses to mechanical stress are essential for the proper functioning of cardiac muscle tissue.

The sarcomere is multi-protein elastic complex that measures approximately 2 μm in length when at rest and can elongate to nearly double its length under

mechanical stress[60, 61]. As illustrated in fig1.4, this complex is composed of backbone proteins, including myosin, actin, and titin, as well as various linker proteins that connect these backbone proteins[62]. The sarcomere is the smallest functional unit capable of completing a contraction-relaxation cycle within the muscle tissue. Each branch extending from a thick filament is attached to an actin molecule on the thin filament, and within a contraction-excitation coupling, each branch moves along the thin filament from one binding site to another, to achieve the re-location of the myosin fibre[63, 64]. Concurrently, the elastic filament composed of titin, the largest protein in the human body, undergoes reversible shortening or elongation as a result of the compression and extension of its elastic regions[65, 66, 67].

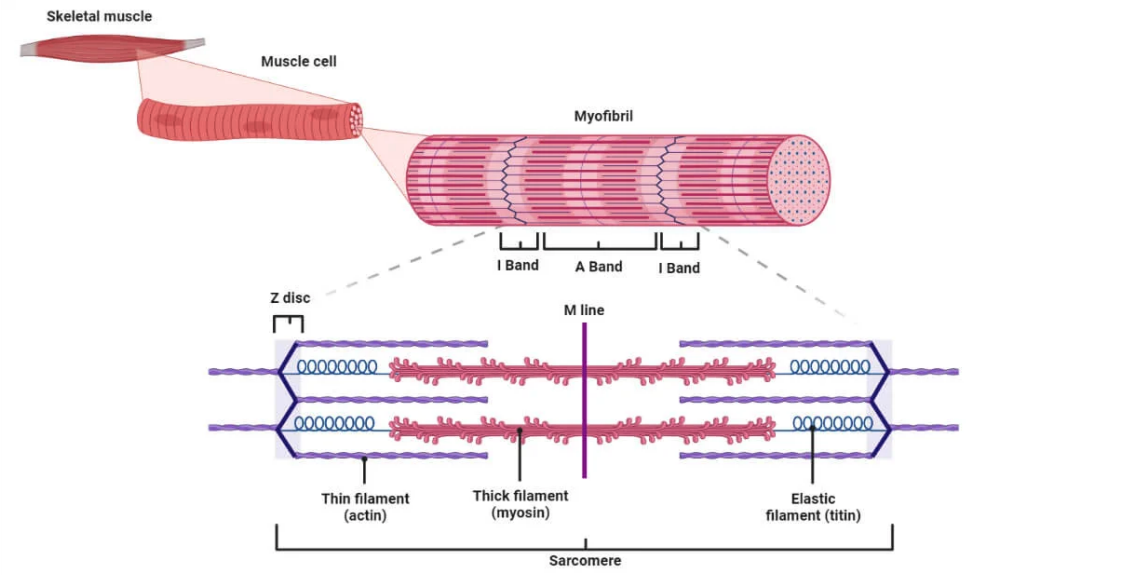


Figure 1.4: The structure of a sarcomere, image adapted from Richfield et al[68]. Sarcomere is the fundamental repeating unit of striated muscle and is composed of backbone proteins, including titin, myosin (thick filament), and actin (thin filament), along with various linker proteins connecting these backbone components.

1.2.2 Various types of sHSPs are expressed within the sarcomere to regulate the function of sarcomeric proteins

In muscle tissue, various types of proteins undergo simultaneous conformational changes, which is fundamental to the contraction and relaxation cycle of muscle fibres. These conformational changes may occur spontaneously to counteract

external mechanical force, as either a result of residue phosphorylation, which is often ATP-dependent, or ligand binding[69]. In cardiac muscle tissue, any disruption of these conformational changes, or the emergence of unexpected conformational alterations, can lead to cardiovascular disorders, which are frequently associated with cardiomyopathy and heart failure[70, 71, 72]. To maintain the proper functioning of cardiovascular protein complexes, sHSPs, along with other molecular chaperones such as Hsp70 and ubiquitin, are expressed in cardiac muscle tissue to bind to proteins and regulate their conformational transformation[73, 74].

Various sHSPs can be expressed within cardiovascular tissue as regulators of muscle proteins[41]. As detailed in table1.1, HspB7 is exclusively localised in striated muscle, while HspB1, B2, B3, B5, B6, and B8 also play significant roles in the biochemical processes occurring in muscle tissue. Although expressed throughout the body, HspB1, B5, and B6 are often expressed at a high level in muscle, serving as unique interactors for proteins in the muscle tissue[51]. The high concentrations of these three sHSPs also facilitate their assembly into co-oligomers, with HspB1 being a essential component for binding HspB5 and/or HspB6. This interaction regulates the interaction equilibrium between each of the three sHSPs and its interactors[75]. Furthermore, HspB2 and B3 can form a dual-source oligomer exclusively in muscle tissues[56], whereas the functions of HspB7 and B8 in muscle tissue are not directly relevant to other sHSPs.

In vivo studies have demonstrated the co-localisation of multiple combinations of sHSPs and muscle proteins, wherein sHSPs play a role in the detection and response to mechanical stress[27]. Various proteins, including actin[76], titin[77], and actin-binding filamins[78], can interact with sHSPs, and these interactions can be regulated by modification of sHSPs in response to mechanical forces exerted on the tissue. Both HspB1 and B5 are involved in actin polymerisation, which contributed to the formation of the thin filaments within the sarcomere. HspB1 has been reported to localise to the region bearing the highest traction forces within thin filaments[76, 79], and a similar recruitment of HspB5 to actin filaments has also been observed, particularly in response to heat stress[80]. In addition to actin,

titin is also a target protein of both sHSPs, and the interactions were initially considered as a response to muscle ischemia but were subsequently confirmed to occur following mechanical stress[77, 81, 82].

Despite its co-localisation with HspB1 and B5, HspB6 has not yet been observed to interact directly with any sarcomeric protein[83]. However, it has been reported that the expression of HspB6 in cardiomyocytes is significant for the contractility and stress resistance of cardiac muscle tissue[84]. Similarly, co-assembled HspB2-B3 hetero-oligomers exhibit little chaperoning ability and low thermal sensitivity, yet they play a role in myogenic development and muscle differentiation[85]. HspB7 and B8, the only two monomeric sHSPs expressed in muscle tissue, can both bind to filamins, the actin-linking proteins, which also interact with HspB1. Both HspB1 and B7 are relevant to the functions of filamin[86, 87], while HspB8 is involved in the detection and immobilisation of filamin in chaperone-assisted selective autophagy (CASA), a process that degrades misfolded filamin proteins[88].

1.2.3 Modifications of sHSPs and sarcomeric proteins modulate cardiovascular functions

Reversible residue phosphorylation is one of the most prevalent modifications influencing the interactions between sHSPs and cardiovascular proteins, particularly as a result of transient mechanical forces[27, 89, 49, 90]. This modification is also a requisite condition for certain interactions[49, 90]. Specifically, the recruitment of HspB1 to actin polymers can be activated by mechanical stress, which induces phosphorylation of HspB1 through the action of p38 mitogen-activated protein kinase (MAPK)[79]. Furthermore, the phosphorylation of HspB1 at Ser82 is essential for its regulatory role concerning filamin C (FLNC)[91]. FLNC itself is also subject to phosphorylation, which regulates its interaction with HspB7[86]. This regulatory mechanism is one of the key topics discussed in this thesis.

Unlike phosphorylations, mutations are permanent sequential alteration of proteins, often resulting in disruptive or pathogenic modifications in their interactions. For example, arginine mutations in HspB5, specifically R120G[92, 93]

and R157H[94], impede the interaction between HspB5 and the elastic regions of titin, which is associated with the development of cardiomyopathy. A series of mutations occurring at the dimerisation site and domain 7 of FLNC obstruct its interaction with HspB1, as evidenced by the presence of HspB1 aggregates in the sample from patients diagnosed skeletal and cardiac myofibrillar myopathy[95]. Furthermore, a truncating mutation in FLNC has also been linked to an increased risk of dilated and arrhythmogenic cardiomyopathy[96]. In addition to their indirect effects on chaperoning interactions, mutations of sHSPs can also directly influence their assembly. For example, the phase separation of HspB2, which is essential for its function, can be regulated by HspB3; however, the regulation is not observed with two HspB3 mutants identified in samples from myopathy patients[97]. Similarly, interactions between HspB1 and B5 or B6 can be affected by point mutations in one of the two interacting components[92, 98].

In this thesis, I present three projects that investigate protein-protein interactions involving sHSPs and/or sarcomeric proteins, all of which regulate the function of cardiac muscle tissues. The first project examines the interactions between HspB1 and B5, both of which are expressed in cardiovascular tissues. The findings indicate that the interactions between these two proteins modulate their chaperoning capabilities. The other two projects focus on the influence of protein modifications on the interactions between sHSPs and sarcomeric proteins through the analysis of protein-protein complexes. In the second project, which studies the complex formed by HspB7 and FLNC, I investigate how the interaction between FLNC domain 24 (FLNCd24) and HspB7 regulates the dimerisation of FLNC, as well as how this regulatory function is affected by phosphorylation at specific sites on FLNCd24. The final project explores the complex comprising of HspB5 and titin immunoglobulin-like domain 26, where I analyse how their interaction regulates the mechanical resistance of titin and how the regulatory effect is weakened by the R157H mutation in HspB5.

2

Molecular Dynamics Simulations of Proteins and Protein Complexes

Contents

2.1 Implementation of molecular dynamics for the simulation of protein structures 16
2.1.1 Molecular simulations and molecular dynamics 16
2.1.2 Fundamental theory of MD simulations 21
2.2 MD simulations of proteins and protein-ligand complexes 25
2.2.1 Identifying the sequences and initial structures of proteins 26
2.2.2 Preparation of input files and parameters 26
2.2.3 MD simulation process 27
2.3 Analysis of protein conformations and protein-protein interaction in a MD trajectory 29
2.3.1 Structural alignment and analysis of conformational stability 30
2.3.2 Principal component analysis 31
2.3.3 Molecular Mechanics Poisson-Boltzmann (or Generalised Born) Surface Area 33
2.4 Steered molecular dynamics: a method to investigate protein behaviour under stress 35
2.5 Supporting Techniques 39
2.5.1 X-ray Crystallography 39
2.5.2 Native Mass Spectrometry 41
2.5.3 AlphaFold 44

2.1 Implementation of molecular dynamics for the simulation of protein structures

2.1.1 Molecular simulations and molecular dynamics

Molecular dynamics (MD) is the primary technique implemented in this thesis to simulate the conformations and behaviours of protein molecules and protein-protein complexes. MD, as well as other molecular simulation methods such as Monte Carlo (MC) simulations, aims to model the behaviours of target molecules or macromolecule complexes at the atomic level[99]. In our projects, MD simulations are used to refine protein conformations, model the interfaces in protein-protein complexes, and elucidate the conformational and behavioural transformations induced by protein modifications.

In all-atom MD simulations, an atom is represented by a spherical rigid body interacting with other atoms in the system through covalent bond and non-bonded interactions. The interactions can be represented by a potential, and since the MD simulations are implemented based on classical mechanics without accounting for quantum effects within the system, the potentials acting to each atom in the system are independent and can be derived as functions to interatomic distances[100, 101]. Force fields, which include all parameters required for the calculation of the potentials[102], are developed through the integration of data generated from various techniques, including infrared and Raman spectrometry, NMR, X-Ray crystallography, thermodynamic measurements, quantum mechanical simulations, and semi-empirical corrections of all of the above. Commonly, the overall potential acting to an atom can be derived as:

$$U_{all} = \overbrace{U_{bond} + U_{angle} + U_{dihedral}(+U_{improper})}^{bonded} + \overbrace{U_{VdW} + U_{Coulomb}}^{unbonded} \tag{2.1}$$

In eq2.1, U_{bond} is the cumulative potential energy contributions resulting from the displacement between covalently bonded atoms. The displacement of the interatomic distance from the equilibrium position is typically minimal due to the significant energy barrier to break the covalent bond. In a case that the displacement is not

comparable to the bond length, a covalent bond can be modelled as a spring, and the potential energy can be described in the form of elastic potential energy as follows:

$$U_{bond} = \sum_{bond} k_b(r_b - r_0)^2 \tag{2.2}$$

in which k_b is the spring constant of the spring modelling the bond, r is the instantaneous distance between the covalently bonded atoms, and r_0 is the equilibrium distance.

U_{angle} refers to the sum of potential energy contributions as a result of the harmonic vibrational motion between two adjacent covalent bonds, which can also be approximately modelled as a spring model. The energy can be quantified by measuring the angular displacement from the equilibrium state as follows:

$$U_{angle} = \sum_{angle} k_\theta(\theta - \theta_0)^2 \tag{2.3}$$

in which k_θ is the spring constant of the spring model modelling the angle formed by two covalent bonds, θ is the instantaneous inter-bond angle, and θ_0 is that angle in the equilibrium state.

$U_{dihedral}$ denotes the potential energy distribution resulting from an angular displacement of the dihedral angle. In the molecular structure, a dihedral angle refers to the angle formed between two covalent bonds that are adjacent to a common covalent bond. Given that if four atoms, A, B, C, and D, are connected by covalent bonds in the sequence A-B-C-D, the dihedral angle is the angle between two planes, one determined by the A-B and B-C bonds and the other determined by the B-C and C-D bonds. As the central bond rotates, the dihedral angle formed by the two adjacent bonds deviates from its equilibrium state, leading to the potential energy that can be described by the following equation:

$$U_{dihedral} = \sum_{dihedral} k_d[1 + cos(n\psi_d - \sigma)] \tag{2.4}$$

in which k_d is the force constant of the model, n refers to the multiplicity, ψ_d is the instantaneous dihedral angle, and σ is the phase offset. The dihedral angle

is contingent upon the configuration, while all other parameters are predefined within the selected force field.

In certain MD simulation engines, an additional energy contribution associated with covalent bonds is calculated separately, defined as the improper torsion energy term, in order to maintain coplanarity. One of the most common improper torsion in protein molecule templates occurs at an sp^2 carbon within an amide or aromatic group, which should be coplanar with all the three atoms which the carbon atom is bonded to (fig2.1). The improper torsion energy term quantifies the potential energy related to the deviation from coplanarity within the simulation[103]. Various computational algorithms have been implemented across different engines for the calculation of the improper torsion term: some propose that its energy contribution should be treated as an independent term in eq2.1, modelled as an angular displacement string model, and calculated using an equation analogous to that for the calculation of the angle term[104], described as:

$$U_{improper} = \sum_{improper} k_\phi(\phi - \phi_0)^2 \tag{2.5}$$

which is applied in MD simulations introduced in this thesis following the application of the Amber force field; others suggest that it can be calculated in the form identical to the dihedral term[105], thereby described as:

$$U_{improper} = \sum_{improper} k_\phi[1 + cos(n_i\phi - \sigma_i)] \tag{2.6}$$

or even integrated with the dihedral term and not calculated independently in the potential energy calculation.

In MD simulations, noncovalent interactions are calculated as a combination of van der Waals forces and electrostatic interactions[100]. Theoretically, noncovalent interactions can occur between any two atoms within the system, which renders the calculation of these interactions is notably more computationally intensive than that of covalent interactions. In practical simulations, two primary approximations are often applied to reduce the computational cost. First, extremely short-range noncovalent interactions are excluded, as their effects are accounted for in the

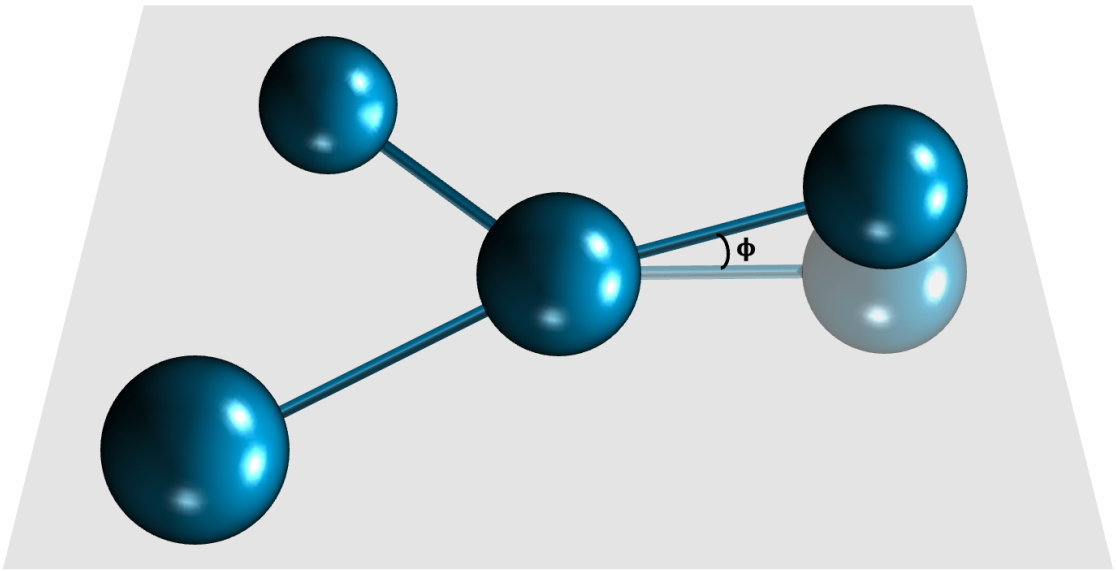


Figure 2.1: A sketch of the improper torsion of a covalent bond. The four atoms, for example, an sp^2 carbon atom and the three atoms to which it is covalently bonded. These atoms should be coplanar in accordance to chemical principles; however, in certain conformations, one atom may deviate from the plane, which is defined as "improper torsion", resulting in an increase of the potential energy.

simulation of covalent interactions. Since the strengths of these interactions are comparable to those of covalent interactions, their effects are already integrated into the calculations of covalent interactions utilising parameters derived from experimental measurements and quantum mechanics computation. Hence, short-range noncovalent interactions should be completely or partially disregarded. The second approximation is associated with the noncovalent interactions between long-distance atom pairs, which diminish rapidly with increasing interatomic distance. These interactions can either be completely omitted or approximated by a much less computationally demanding algorithm, which both require the selection of a cut-off distance for the methods.

The van der Waals force is commonly defined as the combination of dipole-dipole forces, dipole-inducing polarisation forces, and dispersion forces among particles, including atoms and molecules. The value of the van der Waals force is distance-relevant: two particles repel each other within an extremely short distance, and the interaction shifts to attraction at a longer distance, ultimately declining to zero at indefinite distance. The Lennard-Jones potential is a simplified mathematical

model of the potential energy of van der Waals force:

$$U_{VdW} = \sum 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] \tag{2.7}$$

in which ϵ is the depth of the energy well, σ is the distance at which the value of the function is zero, and r is the interatomic distances[106, 107, 108]. In practical applications, calculation of van der Waals interaction energy in MD simulations adheres to the principle that:

- calculation of the van der Waals energy between two atoms is declined when two atoms are:
  - covalent-bonded (1-2),
  - covalent-bonded to a common atom (1-3),
  - not located within a predetermined cut-off distance,
- and is either declined or scaled when two atoms are:
  - covalently bonded to two atoms, which are covalently bonded (1-4).

$U_{Coulomb}$ represents the electrostatic potential energy between charged atoms and is quantified by the Coulomb' Law as follows:

$$U_{Coulomb} = \sum \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{q_1 q_2}{r} \tag{2.8}$$

In practice, calculation of $U_{Coulomb}$ follows the same principle as U_{VdW} in terms of diminishing or scaling the short-range potential energies. However, long-range Coulomb interactions cannot be disregarded, as they are less sensitive to interatomic distance than the long-range van der Waals forces and therefore have a more significant effect to the particles. As an alternative approach, the Coulomb potential energy can be approximately calculated with the Particle Mesh Ewald (PME) algorithm[109, 110, 111, 112].

The PME algorithm takes advantage of periodic boundary conditions (PBCs), which are commonly implemented in solution-phase MD simulations. In PBCs, the unit cell that contains the entire system is replicated in each direction to minimise

side effects and the influence of collisions between atoms and the walls of the unit cell. When a particle exits the unit cell, it reappears in the adjacent cell. For MD simulations using PBCs, it is essential to manually select the dimensions of the periodic unit cell and to neutralise the charge of the system.

The PME method, implemented in the simulations under the premise of PBCs, subsitutes the total Coulomb potential energy with an electrostatic potential energy, and calculates the energy as the sum of two terms, the short-range energy, which is determined in real space (called the "particle" term), and the long-range electrostatic potential energy, which is calculated in reciprocal space (called the "mesh" term). The electrostatic energy can then be described as follows:

$$E_{PME} = \sum_{i,j} \varphi(\mathbf{r}_j - \mathbf{r}_i) = E_{particle} + E_{mesh} \tag{2.9}$$

$$E_{particle} = \sum_{i,j} \varphi_{sr}(\mathbf{r}_j - \mathbf{r}_i) \tag{2.10}$$

$$E_{mesh} = \sum_{\mathbf{k}} \tilde{\Phi}_{lr}(\mathbf{k}) |\tilde{\rho}(\mathbf{k})|^2 \tag{2.11}$$

in which φ represents the electrostatic potential, $\mathbf{r}_i$ and $\mathbf{r}_j$ are position vectors of the charges, in which i typically refers to a charge located in the central unit cell while j refers to a charge in the lattice, $\tilde{\Phi}_{lr}(\mathbf{k})$ is the Fourier transform of the electrostatic potential, and $\tilde{\rho}(\mathbf{k})$ represents the Fourier transform of the charge density. The Fourier term is calculated by the fast Fourier transform algorithm, and therefore requires the unit cell to be infinitely periodically replicated in all directions. Also, the unit cell should be charge neutral to prevent the occurrence of divergent sums in the Fourier term.

2.1.2 Fundamental theory of MD simulations

The aim of protein MD simulations is to produce a sequence of structural conformations, commonly called a trajectory, of the protein, to examine the evolution of atomic conformations of the protein molecule over a specific timescale. A standard MD simulation computes the trajectory by solving Newton's equations of motion

with a conservative force field. If the total potential energy of the ensemble at time t is $U(t)$, the force exerted on the ensemble at time t, $f(t)$, is:

$$f(t) = -\nabla U(t) \tag{2.12}$$

Given that the coordinate set of the ensemble at time t is $r(t)$, the velocity $v(t)$ is:

$$v(t) \simeq \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} \tag{2.13}$$

The evolution of the trajectory is the aggregation of atomic displacements, and to compute the displacements, MD simulations require the velocity of each atom in every configuration. The Verlet integration method is applied in most MD algorithms to derive the velocities [113, 114, 115]:

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 \tag{2.14}$$

$$r(t+2\Delta t) = r(t+\Delta t) + v(t+\Delta t)\Delta t + \frac{f(t+\Delta t)}{2m}\Delta t^2 \tag{2.15}$$

then:

$$r(t+2\Delta t) \simeq 2r(t+\Delta t) - r(t) + \frac{f(t+\Delta t)}{m}\Delta t^2 \tag{2.16}$$

$$v(t+\Delta t) \simeq v(t) + f(t+\Delta t) + \frac{f(t)}{2m}\Delta t \tag{2.17}$$

Therefore, as long as the atomic coordinates, r, and the velocities, v, at time t are given, the atomic coordinates and velocities at a subsequent time step $t + \Delta t$ can be calculated. Hence, a conformation corresponding to a local minimum on the potential energy surface of the system can be generated by iterating the potential energy function. Since each atom has distinct velocity vectors, an additional restraint algorithm is applied to restrain the lengths and angles of covalent bonds, based on the internal coordinates that describe bond lengths, bond angles, and dihedral angles for each atom. These constraints are commonly formulated in the form of Lagrange multiplier equations, and in MD simulations of large molecules including proteins, the SHAKE algorithm is implemented to solve these equations effectively, particularly those involving hydrogen atoms.

Standard MD simulation

A standard MD simulation is conducted in an NVE ensemble, which indicates that the state variables of the system, including the total number of particles (N), volume (V) and total energy (E), are constant, while the other variables, such as system temperature (T) and pressure (P), fluctuate. The total energy is divided into kinetic energy and potential energy, both of which are functions of atomic velocities and coordinates. Since the system in a standard MD simulation is an isolated system, according to the ergodic hypothesis, the system can experience all possible conformations if simulated over an infinite timescale. However, in MD simulations where atoms are physically constrained, it is difficult for the molecular model to overcome energy barriers and transition between local minima, particularly for macromolecules such as proteins or protein complexes. Hence, the selection of initial conformations determines the quality of the MD simulations.

The initial conformation of a protein can be obtained through either experimental techniques, including X-Ray crystallography, NMR[116], and cryo-electron microscopy (Cryo-EM), or computational modelling methods such as AlphaFold[117], Modeller[118], and RosettaCommons[119]. In addition, the temperature of the system must be determined to assign the initial atomic velocities based on a Maxwell-Boltzmann distribution[120]. The time step applied in protein MD simulations is often set to 2 fs, which is adequate to record the fastest atomic motions in a protein structure, the vibrations of X-H covalent bonds, which have a maximum frequency of approximately 3700 cm^{-1}[121], and a corresponding time scale of about 9 fs. For MD simulations of proteins in water solution, the protein molecule is solvated in a water box that is comparable in all dimensions to the periodic cell, and is charge-neutralised and stabilised by sodium chloride.

Non-standard MD simulations

Protein MD simulations are frequently implemented along with experimental techniques to study target proteins. Therefore, the production phase of an MD simulation is usually implemented using the NPT (or NVT) ensemble to replicate

the real *in vivo* and *in vitro* environments. Both NPT and NVT simulations require the specification of a constant temperature value; additionally, for the NPT ensemble, a constant pressure value for the system is also essential to be selected.

In simulations of an ensemble maintained at a constant temperature, the instantaneous temperature is ascertained by the kinetic energy:

$$T = \frac{2}{3Nk_B} \sum_{i=1}^{N} E_{k,i} \tag{2.18}$$

in which N is the number of particles in the system, k_B is the Boltzmann parameter, and $E_{k,i}$ is the kinetic energy of the i-th particle. In MD simulations, the instantaneous temperature is regulated by a specific algorithm to gradually attain the target temperature. Multiple methods can be applied to regulate the temperature, including introducing a heat bath coupled to the system to enable heat transfer, scaling the atomic velocities at a specified rate corresponding to the ratio between the instantaneous temperature and the desired temperature, as well as other techniques that are more precise, although incurring additional computational costs, such as Langevin Dynamics.

Langevin Dynamics is a method to regulate kinetic energy and, consequently, regulate the temperature of a system[122]. This method is often applied in simulations primarily focusing on a single molecule rather than the entire system. Langevin Dynamics describes the force exerted on the molecule by introducing a damping constant:

$$m\frac{d^2r(t)}{dt^2} = -\nabla U - \gamma m\frac{dr(t)}{dt} + \sqrt{2m\gamma k_B T}R(t) \tag{2.19}$$

in which γ is the damping constant and the final term that contains a stationary Gaussian process, $R(t)$, represents the random force applied to the molecule as an approximation of the consequence of the molecule-solvent interactions.

Additionally, for simulations requiring a constant temperature, such as the NPT ensemble, Langevin Dynamics can be integrated with the Nosé-Hoover constant pressure method to regulate the pressure of the system, which functions as a barostat[123, 124]. The method introduces a piston into the system, of which the

fluctuation is regulated by Langevin Dynamics, while pressure control is achieved by Nosé-Hoover algorithm. Certain parameters should be specified for this method, including the oscillation period, the damping time scale that does not exceed the oscillation period, and the target temperature corresponding to the target temperature specified for the Langevin constant temperature algorithm when implementing MD simulations with the NPT ensemble.

2.2 MD simulations of proteins and protein-ligand complexes

MD simulation has been a widely used computational technique in the fields of structure biology and proteomics, and is often implemented together with *in vivo*, *in vitro* and other *in silico* techniques to study the target proteins and protein-related complexes. In multi-technique research, structural biology techniques such as X-ray crystallography, NMR and computational modelling methods can produce one or multiple potential conformations of target proteins, which can be used as the initial structures for protein MD simulations; while conversely, MD simulations can generate extensive trajectories containing a great amount of conformation samples, which can be used to validate conformation hypotheses, modify existing conformations, and re-evaluate the most probable conformations of the proteins.

In addition to the determination and modification of protein structures, MD simulations are frequently applied to examine the intermolecular interactions between protein and ligand molecules, including small molecules, lipids, carbohydrates, nucleic acids, short peptides, and other protein molecules. As a computational method emphasising the simulation of noncovalent interactions, MD simulations can investigate the protein-ligand binding at the atomic level, which is one of the notable advantages of this technique.

In all projects introduced in this thesis, we introduce a standardised pipeline to implement MD simulations to simulate and analyse the structure and behaviour of the target proteins:

1. Identify the sequences and initial structures of the proteins.

2. Prepare input files, algorithms, and parameters.

3. Implement MD simulations.

4. Analyse the trajectories and outputs of the simulations.

2.2.1 Identifying the sequences and initial structures of proteins

The structures of most wild type (WT) human proteins are accessible through online databases, primarily the Protein Data Bank (PDB)[125]. Certain structures of proteins and protein-ligand complexes can also be modelled by experimental techniques including X-ray crystallography, NMR and cryo-EM. If the structures remain undetermined after conducting experiments and queries of existing databases, we can use computational structure prediction software and engines, including AlphaFold, RosettaCommons[119], and Modeller[118], to generate a prediction of the conformation model. For proteins without determined sequences, we utilise sequence reconstruction techniques to predict their sequences, and then generate corresponding structure models using the aforementioned prediction software.

2.2.2 Preparation of input files and parameters

In the projects presented in this thesis, Amber force fields are used to transform protein structural conformations to mathematical models, and MD simulations are implemented by the Nanoscale Molecular Dynamics (NAMD) engine[126]. NAMD requires two input files providing atomic information, a topology file and a coordinate file, to run an MD simulation, which can be derived from a single PDB file prepared in the previous step. A cleanup is compulsory for most PDB files prior to their input into a modelling engine, such as adding hydrogen atoms to the crystal structures, excluding the multiple coordinate sets except the appropriate one for the multi-conformational regions in the NMR structures, and selecting one single conformation according to the confidence and rationality from all conformations produced by

computational modelling methods. In our projects, the cleanup is performed by the Visual Molecular Dynamics (VMD) engine with either its plugin or homemade TCL scripts. VMD is also used to complete residue mutation and substitution, which helps to generate modified protein structures for investigations into protein modification. Upon completion of the structural cleanup, VMD exports a PDB file containing the essential information regarding the protein structure.

The clean PDB file is then imported into AmberTools[127], employing AMBER force fields, specifically either ff14SB or ff19SB[128], to model the molecule(s). The model is then solvated within a water box containing a 0.15M NaCl water solution and, in case that the system is not charge-neutral, the model is neutralised by supplementary sodium or chloride ions. Both water molecules and ions are modelled by the TIP3P water model[129]. The solvated protein models are then documented in a topology file and a coordinate file, formatted as .prmtop and .inpcrd, for later integration into the MD simulation engine.

Accuracy of the product of MD simulations are highly related to the algorithms and parameters that are manually selected and integrated into the input file submitted to the MD simulation engine. Certain parameters are universal across all simulations presented in this thesis, as they represent common requirements for protein MD simulations and are essential for maintaining uniformity among the simulations. These parameters are listed in table2.1.

2.2.3 MD simulation process

In all projects presented in this thesis, classical MD simulations adhere to a standardised procedure to generate the trajectories, starting with a potential energy minimisation step which lasts $10^4 \sim 10^5$ time steps. The minimisation process accommodates atomic positions and covalent bonds, particularly for structures complemented or modified by VMD. These modifications often introduce atoms or covalent bonds that contradict chemistry principles, resulting in higher difficulties and therefore longer minimisation timescale for the conformations to be reallocated.

Keyword	Parameter	Keyword	Parameter
set temperature	300(K)	**PME parameters**	
Force field parameters		PMEGridSpacing	1(Å)
exclude[a]	scaled1-4[b]	**Langevin dynamics**	
1-4scaling	5/6	LangevinDamping	1(ps^{-1})
cutoff[c]	12(Å)	**Data saving and output**	
switch[d]	on	restartfreq[i]	500(steps)
switchdist	10(Å)	dcdfreq	500(steps)
pairlistdist[e]	13.5(Å)	xstFreq	500(steps)
zeromomentum	on	outputEnergie	50(steps)
Integrator parameters		outputPressure	50(steps)
timestep	2.0 fs	**Nosé-Hoover Langevin method**	
rigidbonds	all	reassignFreq	500(steps)
nonbondedFreq[f]	1(step)	reassignTemp[j]	1(K)
fullElectFrequency[g]	2(steps)	reassignIncr[k]	1(K)
stepspercycle	20(steps)	langevinPiston	on
Periodic boundary conditions		langevinPistonTarget	1.01325($*10^5$pa)
wrapWater[h]	on	langevinPistonPeriod	200(fs)
wrapAll	on	langevinPistonDecay	50(fs)

Table 2.1: Universal parameters of MD simulations presented in this thesis

[a]This refers to which atoms are or are not excluded from calculation of non-bonded interactions.

[b]Interactions between bonded atom pairs or atoms bonded to a same third atom are excluded, and those between atoms connected by two adjacent bonds are scaled. The scaling down ratio is defined by keyword "1-4scaling"

[c]The cut-off distance of van der Waals interactions.

[d]This refers to whether to smooth the force functions before it being truncated at *cutoff* or not. If set to on, the original function between *switchdist* and *cutoff* will be replaced by a smoothing function, which is equivalent at *switchdist*, but decreases faster to zero at *cutoff*, other than the positive value in the original function.

[e]NAMD uses a non-bonded pair list which contained all pairs of atoms for which non-bonded interactions should be calculated. An atom pair located within the given distance is included in the pair list, and the update frequency of the pair list is set by keyword "stepspercycle".

[f]Frequency of non-bonded interaction calculation.

[g]Frequency of full electrostatics calculation.

[h]Wrap the bonded atoms in water/object molecules when the bond crosses a periodic boundary.

[i]The number of time steps between two adjacent rewriting operations of the ".restart" file, same for hereafter "-freq" and "output-" keywords

[j]Initial temperature of the heating process

[k]Temperature increasing gradient of the heating process

Following minimisation, a gradual heating simulation is implemented to optimise the side chains effectively, by immobilising the backbone atoms with a constraint. In this process, the system temperature is initially set to a low value and incrementally increased by a minimal gradient until it reaches the target temperature, identical to the temperature specified for the next step. After each temperature rise and

the following reassignment of atomic velocities based on eq2.18, a short NPT simulation is implemented to equilibrate the conformation. Upon reaching the target temperature, the constraint is lifted, and a NVT simulation is performed to re-equilibrate the entire protein conformation.

Then, the production simulation is implemented with the NPT ensemble. The raw trajectory is divided into segments of equivalent time scales, and one frame is selected from the same position within each segment to form a trajectory for further data analysis. During the production phase, instead of generating a long trajectory, we implement multiple parallel simulations starting from the same initial conformation.

2.3 Analysis of protein conformations and protein-protein interaction in a MD trajectory

MD simulations usually generate two types of output data files, the simulation log file and the trajectory file. The log file contains various sets of information, including a copy of the input parameters, the CPU/GPU time used in the simulations, and a record at each time point when a new conformation is recorded in the trajectory. In addition, for each conformation generated, the log file documents the measured values of the system, and in the product simulations that using the NVT ensemble, it includes values of the potential energy and its contributions, as well as pressure. In MD simulations, these values are commonly regarded as indicators of the normal functioning of the simulation, as an abnormal value suggests that the corresponding conformation violates chemistry principles, signifying the failure of the simulation. Additionally, the kinetic energies of particles can be analysed by the Maxwell-Boltzmann energy distribution as a criterion of the probability of collision, which is an essential condition of chemical reactions or interactions.

In contrast to the log file, the trajectory file is mono-functional, that is to document the coordinates, usually Cartesian coordinates, of all atoms within the protein conformations across all frames. These atomic coordinates can be used to analyse the conformational transformations and biological properties of the

target protein, albeit in a less intuitive manner. Various analysis strategies, as well as a number of software, packages, and scripts performing these strategies, have been published and adapted for the analysis of MD trajectories. We implement several data analysis methods across all simulations in the projects presented in this thesis, while also applying specific methods tailored to the requirements of each individual project.

2.3.1 Structural alignment and analysis of conformational stability

Changes in atomic coordinates arise from two primary factors: the local displacement resulting from the conformational transformations, and the movement of the entire protein, which can be equated to the translation and rotation of the centre of mass of the protein conformation. To examine the conformational changes based on the atomic coordinates, it is necessary to minimise the influence of the overall protein movement on these coordinates. Therefore, the first step in the analysis of a trajectory generated by protein MD simulations is to recentre and overlap the centre of mass of each conformation, and to align the conformations to minimise the root mean square deviation (RMSD) of the atomic coordinates. RMSD is one of the most popular criteria for the evaluation of the structural similarity between two conformations[130]. If R_1 and R_2 are the atomic coordinates of two conformations of the same group of atoms within a simulation, then the RMSD between the two conformations is computed as:

$$RMSD(R_1, R_2) = \sqrt{\sum_{i=1}^{n} \frac{(R_1^i - R_2^i)^2}{n}} \tag{2.20}$$

in which n is the number of atoms within the conformation. By aligning the conformations in a trajectory by minimising the RMSD, it is guaranteed the changes in the atomic coordinates are highly likely to be correlated with the conformational changes.

In addition, RMSD can also be taken as a criterion for the evaluation of the structural stability of protein conformations in a MD trajectory. In practical analysis,

one conformation, usually the initial structure or the structure at frame 0, is selected as the template, and the RMSD between this template and each conformation in the trajectory can be computed as an assessment of their conformational difference, which reveals the conformational stability of the protein structure during the simulations. Commonly, heavy atoms in the backbone are selected as the only source of atomic coordinates for the RMSD calculation, as the hydrogen atoms are more flexible while exhibiting less influence to the protein structure. Besides, the similar evaluation can also be applied to a single residue in the structure, in order to display its conformational changes during the simulations, usually for the investigation of noncovalent interactions concerning this residue[131].

Furthermore, the dispersion of conformations in a trajectory can be examined by calculating the root square mean fluctuation (RMSF) of atomic coordinates[132]. RMSF describes the averaged differences of atomic coordinates of each conformation and the mean values of the coordinates, as:

$$R_{mean} = \frac{\sum_{\tau=1}^{T} R_\tau}{T} \tag{2.21}$$

$$RMSF = \sqrt{\frac{\sum_{\tau=1}^{T} (R_\tau - R_{mean})^2}{T}} \tag{2.22}$$

in which T is the number of conformations in the trajectory. In the analysis of a protein MD simulation, RMSF is often calculated for individual single residue, concerning only the alpha carbon (α-C) atoms of eacj residue. A higher RMSF value for a residue indicates greater flexibility, or reduced conformational stability, during the simulation.

2.3.2 Principal component analysis

Another strategy for examining the correspondence across atomic coordinates of protein conformations in a MD trajectory is to reduce the dimension of the coordinate dataset followed by the examination and visualisation of the dimension-reduced data[133]. Principal Component Analysis (PCA) is one of the most popular algorithms for this examination[133, 134, 135]. The idea of PCA is: given a $T \times 3n$ matrix, X, comprising coordinates of the selected atoms in the trajectory,

in which T is the number of frames in the trajectory and n is the number of selected atoms for PCA computation, as:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,3n} \\ X_{2,1} & X_{2,2} & \dots & X_{2,3n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{T,1} & X_{T,2} & \dots & X_{T,3n} \end{bmatrix} \tag{2.23}$$

First, centring the data to generate a zero-mean matrix, X_c, as:

$$X_c = \begin{bmatrix} X_{1,1} - \overline{X_1} & X_{1,2} - \overline{X_1} & \dots & X_{1,3n} - \overline{X_1} \\ X_{2,1} - \overline{X_2} & X_{2,2} - \overline{X_2} & \dots & X_{2,3n} - \overline{X_2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{T,1} - \overline{X_T} & X_{T,2} - \overline{X_T} & \dots & X_{T,3n} - \overline{X_T} \end{bmatrix} \tag{2.24}$$

Then, calculating the "Principal Components", which are the eigenvalues of $X_c^T X_c$, by performing the singular value decomposition:

$$X_c = U \Sigma V^T \tag{2.25}$$

in which U is a $T * T$ unitary matrix, Σ is a $T * 3n$ diagonal matrix, and V is a $3n * 3n$ unitary matrix, in which the columns are the principle components. In our simulations, we focus primarily on the correspondence across conformations, which can be indicated by the explained variance. The explained variance, λ_i, reveals the relative significance of its corresponding principle component, v_i within all principle components, and is derived from its corresponding singular values, σ_i, as:

$$\lambda_i = \frac{\sigma_i^2}{n-1} \tag{2.26}$$

By sorting the explained variance by a numerical descending sequence and the corresponding principle components, the columns in matrix V can be re-sequenced, and X_c can then be reduced to a matrix with lower dimensions, $X_{reduced}$, with a selected k dimensions ($k \leq 3n$), as:

$$X_{reduced} = X_c V_k \tag{2.27}$$

in which V_k contains the first k columns in V. The explained variance of a principle component is relevant to the variability of the original dataset along the direction

of this PC. In our simulations, if the first two principal components are sufficiently representative, as the sum of the first and second greatest normalised values of the explained variance is greater than 90%, we plot the first two term in each column of the $X_{reduced}$ matrix in a single graph. In the graph, each data point represents a conformation in the trajectory, and the dispersion of the points is related to the conformational stability of the protein structure: if most points are concentrated, the conformations are less likely to diverse during the MD simulation; In contrast, if the data points are dispersed into multiple distinct clusters, it is highly probable that the molecule undergoes a conformational transformation during the MD simulation.

2.3.3 Molecular Mechanics Poisson-Boltzmann (or Generalised Born) Surface Area

In order to study protein-protein interactions in an MD trajectory, we utilise the Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) or Generalised Born Surface Area (MM/GBSA) method to compute the binding energy[136]. For a system comprising a receptor(R)-ligand(L) bound complex(com) solvated in a water solution(s), we can develop a thermodynamic model:

$$\Delta G_{binding,s} = \Delta G_{com,s} - \Delta G_{R,s} - \Delta G_{L,s} \tag{2.28}$$

and a gas(g) phase model:

$$\Delta G_s = \Delta E_g + \Delta G_{solvation} - T\Delta S_{solvation} \tag{2.29}$$

in which:

$$\Delta E_g = U_{all} \tag{2.30}$$

that can be calculated according to eq2.1 and

$$\Delta G_{solvation} = \Delta G_{polar} + \Delta G_{nonpolar} \tag{2.31}$$

in which the non-polar solvation energy can be calculated as either:

$$\Delta G_{nonpolar} = \gamma SASA + b \tag{2.32}$$

in which $SASA$ is the solvent accessible surface area, γ is the surface tension coefficient, and b is the correction term; or:

$$\Delta G_{nonpolar} = \Delta G_{dispersion} + \Delta G_{cavity} = \overbrace{\gamma_1 SASA + b_1}^{dispersion} + \overbrace{\gamma_2 SAV + b_2}^{cavity} \tag{2.33}$$

in which SAV is the solvent accessible volume, which is enclosed by $SASA$; ΔG_{polar} can be calculated by the Poisson-Boltzmann equation, derived from the original Poisson equation:

$$\nabla \cdot \epsilon(\vec{r}) \nabla \phi(\vec{r}) = -4\pi \rho^f(\vec{r}) \tag{2.34}$$

in which $\epsilon(\vec{r})$ is a dielectric distribution function, $\phi(\vec{r})$ is the potential energy function that is equivalent to ΔG_{polar}, and $\rho^f(\vec{r})$ is the charge density. Considering the impact of simple ions in the solute, a salt-related term is integrated with the Poisson Equation:

$$\nabla \cdot \epsilon(\vec{r}) \nabla \phi(\vec{r}) + \lambda(\vec{r}) f(\phi(\vec{r}) = -4\pi \rho^f(\vec{r}) \tag{2.35}$$

in which $\lambda(\vec{r})$ is a ion-exclusion function indicating the accessibility of ions to the space, while $f(\phi(\vec{r})$ is a function related to salt concentration, which can be calculated by the Boltzmann equation:

$$f(\phi(\vec{r})) = 4\pi \sum_i^N z_i q c_i \exp(-\frac{z_i q \phi(\vec{r})}{k_B T}) \tag{2.36}$$

in which z_i is the ion valence, c_i is the concentration at temperature T, and k_B is the Boltzmann constant. The computational of solvation entropy $\Delta S_{solvation}$ is expensive, and the normal mode analysis, which enables the calculation of the entropy term performed by AmberTools[127], may be less accurate than the other term, thereby leading to a significant error to the overall results. We opt to exclude the solvation entropy term from the calculation of mmPB/GBSA in our projects, as we use the algorithm primarily for the comparison between the binding energies of two proteins or ligands binding to the same protein, which generally exhibit comparable values of entropy.

To avoid the expensive computational cost associated with eq2.35 and eq2.36, an alternative algorithm implementing the Generalised Born approximation has been developed to compute the polar solvation energy, by modelling the solute as a series of bi-phase system. Each modelled bi-phase system consists of an internal sphere and an external solvent, which possess different dielectric constant. The polar solvation free energy can then be derived as:

$$\Delta G_{polar} = -\frac{1}{8\pi\epsilon_0}(1-\frac{1}{\epsilon})\sum_{i,j}^{N}\frac{q_i q_j}{f_{GB}} \tag{2.37}$$

$$f_{GB} = \sqrt{r_{ij}^2 + a_i a_j \exp[-(\frac{r_{ij}}{2a_{ij}})^2]} \tag{2.38}$$

in which ϵ_0 is the permittivity of free space, ϵ is the solvent dielectric constant, r_{ij} is the distance between particles i and j, and a_i is defined as the effective Born radius, which can be approximated as the distance from particle i to the surface.

We use a built-in Python script from the AmberTools engine to implement the MM/PBSA and MM/GBSA calculations[137]. Previous research has reported that, while the absolute values of binding energies calculated by the MM/PBSA and MM/GBSA algorithms may not be perfectly accurate, these algorithms are highly effective in elucidating the relative binding energies among complexes with similar conformations[138, 139]. Therefore, in our research, we adopt the method to compare the binding energies of two interactions involving a single protein and its two interactors.

2.4 Steered molecular dynamics: a method to investigate protein behaviour under stress

Classical MD simulations provide an insight into the behaviours and properties of a target molecule as a consequence of its covalent bonds and atomic interactions. However, the ability of classical MD simulations is limited for the prediction of the behaviour of a protein in response to an external force, which is significant for intercellular transportation[140], signalling[141], and drug metabolism[142, 143, 144]. To overcome this limitation, steered molecular dynamics (steered-MD, SMD)

has been developed as a specialized MD simulation technique that examines the behavioural and conformational transformation of a protein structure under the influence of an external force[145, 146]. In computational studies involving protein complexes, SMD is frequently used to simulate the unbinding process of a ligand from a protein[147, 142], as well as the reversible unfolding process of a muscle protein in response to an external force[148, 149]. As discussed in Chapter 1.2, the mechanical sensitivity of various muscle proteins can be regulated by sHSPs, which leads to different structural transformations in sHSP-bound proteins while bearing identical forces over the same time scale, compared to unbound proteins. Therefore, we proposed that SMD simulations can be used in the investigation of interactions between sHSPs and cardiovascular proteins.

A strategy for applying an external force to a protein structure is to fix the atom at a terminal, while pulling the other terminal[149]. This pulling can be executed as either the displacement of a constant velocity of the pulled atom[146, 150], or a constant force applied to the atom[151, 152]. Throughout the stretching process, muscle proteins, as components of the sarcomere, are elongated by the stretching force, which results from the accumulated disruption of the intramolecular interactions. Therefore, to investigate the disruption of intramolecular residue-residue contacts, we select the constant velocity algorithm to conduct SMD simulations, as it enables the alignment between the conformations of bound and unbound protein structures at each time point within the simulations, which facilitates the comparison between the interactions of the same residue pairs in both protein structures.

In constant-velocity SMD simulations, the mobile terminal atom is constrained to a pseudo-atom, which is pulled subsequently for the elongation of the entire protein structure [126, 145]. The connection between the terminal atom and the pseudo-atom is achieved by the introduction of a pseudo covalent bond, which is positioned along the extension of the covalent bond connecting the terminal atom and its adjacent atom. The pseudo-bond is conceptualised as a string, with the harmonic potential determined by a predefined spring constant. At each time step

in a constant-velocity SMD simulation, the pseudo-atom is pulled at a constant velocity, $\boldsymbol{v}$, and the force at a given frame can be calculated as:

$$\boldsymbol{F} = -\nabla U \tag{2.39}$$

$$U = \frac{1}{2}k[\boldsymbol{v}\tau - (\boldsymbol{r} - \boldsymbol{r_0})\boldsymbol{n}]^2 \tag{2.40}$$

in which k is the spring constant of the pseudo bond, τ is the time length of one time step, $\boldsymbol{r}$ is the coordinate vector of the terminal atom at that frame, $\boldsymbol{r_0}$ is the coordinate vector of the terminal atom at the previous frame, and $\boldsymbol{n}$ is the normalised vector indicating the direction of pulling, parallel to the pseudo bond.

Predetermination of several additional parameters, alongside those for classical MD simulations, is essential for SMD simulations, including the spring constant k, the constant velocity $\boldsymbol{v}$, and the simulation timescale, t. The criteria for the selection of $\boldsymbol{v}$ and t for simulating protein unfolding are:

$$\boldsymbol{v}t \geq L_{extension} = L_{full} - L_{initial} \tag{2.41}$$

in which $L_{extension}$ is the extended length of the protein, $L_{initial}$ is the length of the initial structure, and L_{full} is the full length of the protein backbone, which corresponds to the maximum length of the protein while completely extended.

In a SMD simulation, an external force is continuously applied to the protein structure, which leads to concurrent changes of the overall free energy throughout the simulation. In contrast to classical MD simulations, a single SMD simulation generates only one conformation of the structure at a state, forbidding it to deliver a sufficiently unbiased conclusion. Therefore, it is necessary to implement multiple SMD simulations in parallel under identical parameters and conditions, allowing for the application of further data analysis on the basis of the trajectories. Trajectories generated by these parallel simulations should ideally have similar conformations at the same time points in favour of the alignment and further data analysis. Therefore, we anticipate that the protein structures within the parallel simulations will adhere to an identical unfolding pathway.

Various analytical algorithms specifically developed for SMD simulations, along with most data analysis algorithms introduced in 2.3, are performed for the analysis of SMD outputs and trajectories. To study the response of proteins to external stress with and without the presence of binding ligands, we can compute the instantaneous force and the accumulated work for each frame[153]. For a conformation within an SMD trajectory, the instantaneous force, relevant to the overall strength of non-covalent interactions disrupted at that frame, can be determined by eq2.39. For each frame in the simulation, the instantaneous work can be approximately computed as the product of the instantaneous force and the length of the time step. Therefore, for any conformation displayed in the trajectory, the accumulated work required to unfold the protein to that state can be computed as the accumulation of instantaneous work from the beginning of the simulation to the corresponding frame. Both instantaneous force and accumulated work display the mechanical resistance of the protein, and can be reveal the distinct mechanical properties of a specific protein while binding to different ligands.

To explore the correlation between the mechanical resistance of a protein and its non-covalent interactions, we focus on the disruption of each non-covalent interaction during a SMD simulation, through monitoring the corresponding inter-residue distance at each frame. A pair of residues in a protein structure, if coupled by a non-covalent interaction, are typically located within a short distance from each other; while following the disruption of the interaction in a SMD simulation, these two residues is highly likely to separate. To quantify this phenomenon, we define a "breaking point" for a residue pair in a SMD simulation, which refers to the last time point in the simulation that these two residues are located within a specific distance. The breaking points of all residue pairs within a protein structure can be correlated to the instantaneous forces, through chronological sequencing, which may reveal the strength of the non-covalent interactions and their relative contributions to the overall mechanical resistance of the protein.

2.5 Supporting Techniques

Apart from MD simulation, which serves as the primary analytical technique employed in our projects, my colleagues, collaborators, and I have utilised several other analytical methods, including both experimental and computational approaches, to examine the conformation and properties of sHSPs and sHSP-related multi-protein complexes.

2.5.1 X-ray Crystallography

X-ray crystallography is an experimental technique that utilises the phenomenon of diffraction effect to determine the atomic structure of target proteins. To generate a diffraction pattern, the object must be subjected to radiation with a wavelength that is numerically comparable to the dimensions of the object. Therefore, the diffraction pattern of an all-atom protein structure can be obtained through exposure to X-rays, due to the similarity between the lengths of chemical bonds and the wavelength of X-rays[154].

The fundamental principle underlying this technique is that the atomic structure of a target molecule can be derived mathematically from the X-ray diffraction pattern associated with that molecule, which is the Fourier transform of its structure[155]. The diffraction pattern produced by a single molecule is typically too weak to be detected; however, it can be amplified to a detectable level through the accumulation of diffraction signals. To predict the protein structure based on the principle of X-ray diffraction, it is necessary to prepare a crystal of the target protein, which consists of a substantial number of copies of identical protein molecules[154]. It is important to note that the diffraction pattern obtained from a protein crystal does not directly represent the transform of the atomic structure; rather, it is the composite of overlapping patterns from the distanced unit cells within the crystal[156]. As defined in Bragg's law, the variations among the patterns that constitute the overlapping diffraction pattern are determined by the phase of the wave and the dimensions of the unit cell in the crystal[157]. Therefore, the overlapping pattern can be converted into an electron density map in reciprocal

space once the initial phase is determined, either through molecular replacement or the incorporation of heavy atoms.

The resolution of a crystallographic pattern is related to the shortest inter-lattice distance[158]. Typically, the electron density map, and the protein structure derived from this pattern, exhibits low resolution, demanding refinement to yield a reliable protein structure model. During the refinement process, restraints are imposed on the atom positions; notably, regions with low resolution are subjected to greater restraints compared to those of higher resolution, in order to preserve the integrity of bonds and chemical interactions within the structure. The assessment of model refinement is conducted using the R-factor[159], which is calculated from the reflection coefficients of the proposed model and the observed reflections. A low R-factor indicates a high degree of concordance between the model and empirical data. It is customary, when publishing a protein structure obtained through X-ray crystallography or when depositing it into the Protein Data Bank (PDB), to include the R-factor as a measure of the reliability of the structural prediction[160].

Application of protein crystal structure in MD simulations

To conduct protein MD simulations, it is essential to prepare initial structures derived from structural templates generated through either experimental or computational methods. Protein X-ray crystallography, as one of the most widely utilised experimental techniques used for the prediction of protein structures, produces ~70% of all protein structures deposited into the Protein Data Bank, underscoring its significance as a primary source of initial structural templates[125]. In comparison to other protein structural prediction methods, such as Nuclear Magnetic Resonance (NMR), X-ray crystallography shows greater versatility regarding the dimensions of target molecules. It is capable of producing atomic-level structures of large protein molecules as well as biomolecular complexes consisting of protein, peptides, nucleic acids, lipids, and small ligands.

X-ray crystallography, despite being a valuable technique, also has certain limitations, that may impact the adoption of crystal structures as initial structural

templates for MD simulations. The resolution of a crystal structure is indicative of the accuracy of the structural prediction, and should be carefully evaluated prior to the adaptation. For example, X-ray crystallography may fail to collect critical information, such as coordinates of the backbone atoms, particularly when the quality of the crystal is suboptimal. As a result, the deposited structure may retain the sequence of a protein without recording conformational information about some of the residues in the protein. If a structure with missing residue(s) is selected as the template for MD simulation, it is imperative to incorporate the conformational information of the missing residues, usually through structural alignment, followed by additional simulations to stabilise the completed structure.

The resolution of a crystal structure is indicative of the quality of structural predictions and requires careful examination. When the resolution of a crystal structure exceeds 1.5Å, which is frequently observed in structures deposited in PDB, the accuracy of predictions regarding the positions of hydrogen atoms is compromised[161]. Therefore, PDB files for these structures typically do not include hydrogen atoms, which can be added automatically by protein modelling and visualisation software. Furthermore, if the resolution of a protein structure surpasses 3 Å, the structural prediction lacks the precision required for the accurate determination of the location and behaviour of flexible regions, such as side chains, loops, and unstructured regions, within the protein structure. In cases where the resolution of the crystal structure is insufficient or when hydrogen atoms are added automatically to the protein structure, additional simulations are necessary to improve the reliability of the structural predictions.

2.5.2 Native Mass Spectrometry

Mass spectrometry in structural biology

Mass spectrometry (MS) is a widely applied analytical technique in the fields of chemistry and chemical biology for the investigation of the composition and properties of target molecules[162]. In MS experiments, the sample is ionised, and often simultaneously fragmented, by an ion source, and the ions are subsequently

accelerated and deflected by either a magnetic or electric field for detection[163]. The degree of deflection of specific ions, which correlates to their mass-to-charge ratio, is one of the two primary parameters measured by the sensors, alongside the relative abundance of the ions[164]. By analysing the mass-to-charge ratios and relative abundances, a mass spectrum can be generated, thereby elucidating the composition of the sample and the biochemical properties of its constituents[165].

In proteomics studies, MS can be employed to identify protein species by comparing the spectrum with those in protein databases, and to quantify a certain type of protein in a sample, which is related to protein expression and post-translational modification[165, 166]. Various advanced MS techniques have been developed to extract more comprehensive information from samples. These techniques include liquid chromatography-mass spectrometry (LC-MS)[167, 168], tandem mass spectrometry (Tandem-MS or MS/MS)[169], hydrogen-deuterium exchange mass spectrometry (HDX-MS)[170], cross-linking mass spectrometry[171], and native mass spectrometry (native-MS)[172, 173], the latter of which is utilised in our projects.

Native Mass Spectrometry

Native-MS is an advanced technique designed based on mass spectrometry, employed specifically for the structural and functional studies of biochemical macromolecules, including proteins and nucleic acids[173, 47]. In the ionisation phase of native-MS, the sample is ionised without being fragmented into particles, permitting the detection and measurement as intact charged molecules rather than their fragments. Meanwhile, innovations in spectrometer design enable the detection and measurement of intact molecules and provide information about the higher-order structures, the conformations, and intermolecular interactions of macromolecules in the sample[172].

In order to study the intact macromolecules, a native-MS spectrometer utilises specific types of instruments to treat the samples carefully. In the first step, electrospray ioniser is often used for the ionisation of protein complexes, to transfer

molecules into the gas phase while preserving their chemical interactions[172, 174]. Compared to the fragmented ions in the denatured samples, ionised intact molecules produced by electrospray have lower charges, resulting in greater mass-to-charge ratios. Consequently, guiding these ions into detection units is difficult, therefore requiring adjustment of the pressure in the ion source to increase the collision rates of particles, in order to promote the transmission efficiency[175]. Additionally, due to the high mass-to-charge ratios of the analytes, native-MS uses mass analysers that are unbiased towards signals with different mass-to-charge ratios, such as Time-of-Flight[176] or Orbitrap[175, 177] mass analysers.

Application of native-MS

One of the primary applications of native-MS is to investigate the formation of macromolecule complexes that are stabilised by non-covalent interactions[173]. Therefore, this technique has been frequently used in the study of sHSPs, to verify the interaction between sHSPs and their interactors in the sample[173, 47]. Furthermore, native-MS can assess the strength of non-covalent interactions, by partially disrupting these interactions and measuring the relative abundance of both the intact complex and its segments[178]. Collision-induced dissociation[173, 175] and ion-mobility spectrometry[179, 180] are two widely used capabilities affiliated to the instruments to induce and measure the disruption of intermolecular interactions. Ideally, the analysis of a sample containing both intact and dissociated protein ions produces native-MS spectra with peaks corresponding to all components within the sample. The relative concentration of a component within the sample is directly proportional to the cumulative peak area in the native-MS spectra, enabling the calculation of the equilibrium constant and the binding free energy of the complex.

Native-MS is particularly effective for the study of oligomerisation of sHSPs since it can differentiate components based on their oligomerisation states[173]. In the case of a mixture of oligomers formed by individual species of monomers, native-MS can quantify the concentration of oligomers and construct a distribution model of oligomerisation, which can be used to analyse the oligomerisation patterns

and binding free energies. Also, in instances of hetero-oligomerisation, precise mass analysis can differentiate oligomers with different compositions and predict the biases among binding affinities of the monomers.

2.5.3 AlphaFold

AlphaFold is an artificial intelligence programme that aims at protein structure prediction, developed by DeepMind[117]. Since its publication in 2018, it has been widely recognised as one of the most effective and accurate engines for protein structure prediction. In 2020, AlphaFold 2 was released with higher accuracy and efficiency comparing to its predecessor, and the open-source software has been widely used within the field of structural biology. In our projects, we used AlphaFold 2 to generate protein structures in the cases that we were unable to obtain an X-ray crystal structure of a protein or protein-protein complex.

Based on the input sequence, the algorithm applies two searches: one is to search and select an initial structural template from a structural database, usually the Protein Data Bank, and the other one is to search similar sequences across different species and performing multiple sequence alignment (MSA). The algorithm generate two arrays, one consisting of edges the represent the relationships between pairs of residues in the structural template, while the other one includes edges that represent the relationship between residues and the sequences in MSA. These two arrays are refined alternately by introducing the refined parameters from one array into the other, iterating the simulations until convergence is achieved for both arrays. Upon completion of the iterations, most predicted structures align closely with conformations obtained from other experimental and computational techniques, but atomic positions in the structure often violate the stereochemical rules. In AlphaFold 2, an AMBER simulation is implemented as a final step to relax the structure and minimise its energy, in order to reduce stereochemical violations and to rationalise the structure.

In our projects, we use the AlphaFold 2 online server to generate an initial structural model of a protein complex when a structural template could not

be obtained from existing databases or experimental methods, such as X-ray crystallography. For input sequences comprising multiple chains, the AlphaFold online server produces five structural models, which are ranked according to their predicted template modelling (pTM) scores and interface predicted template modelling (ipTM) scores. Both scores are normalized on a scale from 0 to 1 and are derived from the template modelling (TM) score, which assesses the accuracy of the predicted structure. The pTM score indicates the overall accuracy of the model, with a value exceeding 0.5 suggesting a strong correlation with the actual structure. The ipTM score reflects the accuracy of an individual chain within the predicted complex structure, with a reliable prediction typically exhibiting a score above 0.8. In addition to these two scores, the output matrices also include the predicted local distance difference test (pLDDT) scores, which evaluate the concordance between the prediction and an experimental structure, as well as the predicted aligned error (PAE) scores, which assess the confidence regarding the position and orientation of a residue in relation to other residues within the complex. Upon acquiring the structural models and their associated confidence scores and matrices, we first align the structures to determine whether the docking positions across all complexes exhibit significant diversity. If the positions are not significantly diverse, we select the model with the highest pTM score as the initial structure for molecular dynamics (MD) simulations. otherwise, we implement preliminary MD simulations to evaluate the predictions concerning the interacting interface.

3

HspB1 and HspB5 hetero-oligomerise via ACD-CTD interactions

Contents

3.1 Abstract 48
3.2 Introduction 49
3.2.1 Functional properties of HspB1 and B5 49
3.2.2 The co-assembly of HspB1 and B5 influences the bioactivity of sHSPs. 50
3.2.3 Plan of experiments and analysis 52
3.3 Methods 53
3.3.1 Preparation of the initial structures 53
3.3.2 MD simulations 54
3.3.3 Trajectory preprocessing and data analysis 54
3.4 Results and discussions 56
3.4.1 HspB1-HspB5 co-assembly is biased as a result of ACD-CTD interactions. 56
3.4.2 HspB1 CTD forms less intense ACD-CTD interactions than HspB5 CTD. 59
3.4.3 The LCA CTD demonstrates similarity to the HspB5 CTD due to conserved arginines 65
3.5 Conclusions 72

3.1 Abstract

In this chapter, we investigate the co-assembly of HspB1 and HspB5, which predominantly occurs in cardiac muscle tissue *in vivo* and potentially plays a vital role in the regulation of the concentration of HspB1 and B5. Native-MS experiments indicate that interactions between HspB1 and B5 can occur between either two ACDs, or an ACD and a CTD peptide. Notably, only ACD-CTD interactions demonstrate variability in binding affinity within both heterogeneous and homogeneous complexes. To explore the conformational basis of the ACD-CTD interactions and the source of the differences in binding affinity, we construct structural models of homogeneous and CTD-exchanged heterogeneous sHSP dimers, incorporating segments from HspB1, HspB5, and/or their co-ancestor proteins of the two sHSPs. MD simulations of the sHSP dimers reveal that both the ACD and CTD segments in HspB5 are more favourable interactors compared to those in HspB1. A plausible explanation for this observation is that HspB5 contains a higher proportion of charged and polar residues than HspB1, thereby increasing the likelihood of hydrogen bond formation. This hypothesis is further substantiated by simulations involving the co-ancestor proteins.

3.2 Introduction

3.2.1 Functional properties of HspB1 and B5

Human HspB1, also known as Hsp27, is a vertebrate molecular chaperone that is expressed across various cell types, particularly in all forms of muscle cells. Beyond its capability in enhancing the thermal stability of associated proteins, HspB1 is implicated in the regulation of signalling pathways and the inhibition of protein aggregation[79, 81]. Furthermore, is has been observed that the expression levels of HspB1 vary among cells in distinct differentiated states, indicating its significance in cell differentiation, as well as in metabolic processes and morphogenesis[181]. The overexpression of HspB1 may also be a potential consequence of tumour growth[182]. Although functioning as an ATP-independent molecular chaperone, HspB1 has multiple phosphorylation sites, which makes it a regulator for the activity of several muscle proteins, including filamin and actin[183, 184].

HspB5, also know as αB-crystallin, is one of the two sHSPs, alongside HspB4/αA-crystallin, that are highly expressed in the eye lens[55, 185]. In contrast to HspB4, which is localised specifically in the eye lens, HspB5 is expressed systemically and is involved in various biochemical processes, including the prevention of protein aggregation and intercellular signalling[186]. Abnormal expression levels of HspB5 are indicative of neurodegenerative disorders, including Alzheimer's disease and Parkinson's disease[187, 188]. In addition, functional mutations in HspB5 have been implicated as contributing factors to cardiovascular disorders, as will be discussed in Chapter 5.

As detailed in table1.1, both HspB1 and B5 have been found to form large polydisperse homo-oligomers[32, 54, 189]. The fundamental unit of the oligomers of HspB1 and B5 is a dimer, which is associated with two β6+7 strands; consequently, the dominant majority of oligomers observed both *in vivo* and *in vitro* consist of an even number of monomers. Small oligomers are formed through the interconnections of multiple dimers, resulting from the interactions between the ACD and the CTD, while less pronounced interactions between the ACD and the NTD, as well as

multiple entangled NTD regions, contribute to the aggregation of these small oligomers[190]. The flexibility of the terminal regions mitigates the steric effect and facilitates the formation of large oligomers. *In vivo*, both human HspB1 and B5 can form oligomers of considerable size, commonly containing up to 32 to 40 monomers (approximately 800-1000 kDa)[98, 191], and these oligomers exhibit remarkably diversity in terms of ensemble and topological conformation[192]. For instance, the structural models of two 24-mers, HspB1 (PDB: 6DV5)[193] and HspB5 (PDB: 3J07)[52], displayed in fig3.1 are notably distinct.

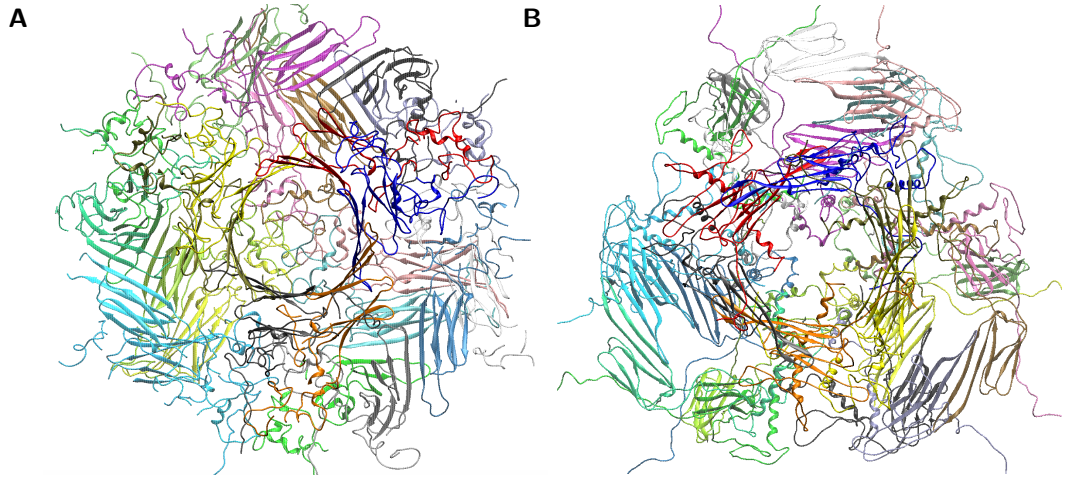


Figure 3.1: **A** Model of a crystal structure of the HspB1 24-mer (PDB: 6DV5). **B** Model of the NMR structure of the HspB5 24-mer (PDB: 3J07). Researchers have demonstrated that both HspB1 and B5 can form large oligomers characterised by specific three-dimensional conformations and potentially distinct functions. In both conformations, interactions occur among ACD-ACD, ACD-CTD, ACD-NTD, and NTD-NTD pairs, resulting in the formation of a substantial cavity or tunnel. In both illustration, six monomers that contributes to the formation of a cavity are highlighted, while the remaining monomers are depicted translucently.

3.2.2 The co-assembly of HspB1 and B5 influences the bioactivity of sHSPs.

The functional regions of heterogeneous proteins encoded by duplicated genes are commonly distinct, allowing them to acquire divergent functions[194]. However, it has been observed that multiple co-localised sHSPs can aggregate within the same human tissues, and hetero-oligomers consisting of multiple species of sHSPs have been identified both in the human body and in laboratory settings, including binary HspB1-B5 oligomers[27, 54]. The co-assembly of multiple sHSPs may occur through

either a precise obligate coassembly [85] or a stochastic process[195, 196]. Hetero-oligomers formed via the first mechanism consist solely of only small hetero-oligomers as basic units, resulting in a fixed quantitative composition ratio. For instance, in the binary HspB2-B3 hetero-oligomer, the proportion of HspB2 to B3 is precisely 3:1[197], as the basic unit of the oligomer is a tetramer composed of three HspB2 monomers and one HspB3 monomers[56]. In contrast, co-assembly following the second mode involved the aggregation of sHSP monomers, and leading to a series of hetero-oligomers with varying dimensions, compositions, and conformations[198]. The ensembles of hetero-oligomers HspB1-HspB5 are reported to be highly polydisperse, with the ratio of HspB1 to B5 in the oligomers not being persistent[199]. Therefore, the oligomerisation of HspB1 and B5 follows the second mechanism, demanding interactions among heterogeneous monomers across multiple domains.

The ensembles of hetero-oligomers composed of HspB1 and B5 have been found to contain a greater maximum number of monomers compared to either the HspB1 or B5 homo-oligomer, with the maximum mass of hetero-oligomers potentially reaching 1200 kDa[199]. The competition between homo- and hetero-interactions among the various domains in sHSP monomers leads to a distinct hetero-oligomerisation process for HspB1 and B5, differing from their homo-oligomerisation in terms of binding affinity and dynamic equilibrium. The presence of HspB1-HspB5 hetero-oligomers regulates the relative chaperoning reactivity of each sHSP, which is influenced by the oligomerisation equilibrium[191]. Consequently, the concentration of either sHSP, along with the ensemble and composition of the hetero-oligomers, can significantly affect the activity of the other sHSP, thereby potentially impacting the biochemical processes involving the protein.

Given the pronounced tendency for co-assembly between HspB1 and B5, we hypothesise that the mutual interaction sites of these two proteins have been conserved throughout their evolution history. Therefore, it is plausible that each protein is capable of forming stable oligomers with their last common ancestor (LCA)(fig3.2), which refers to the most recent ancestral protein that the two proteins commonly evolved from[200]. We propose that certain inter-residue interactions

have remained consistent from the LCA to HspB1 and B5, allowing us to utilise the LCA as a framework to investigate the interactions between HspB1 and B5 at the residue level. We opt to conduct this study with *in silico* methods, as computational simulations of the LCA are considerably more feasible, and can effectively elucidate inter-residue contacts within complex conformations.

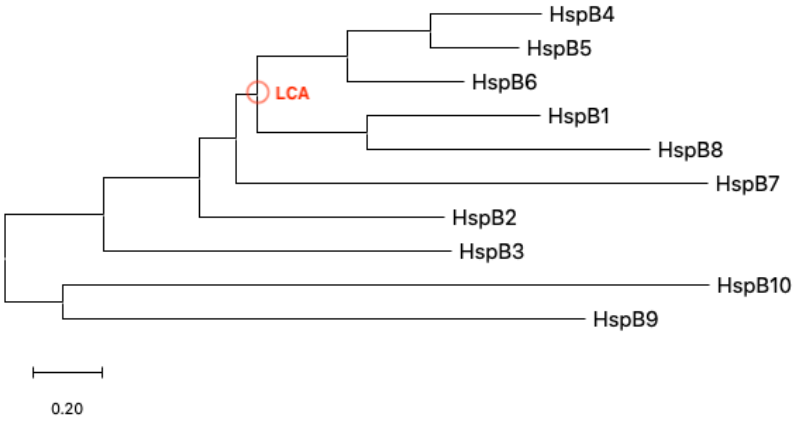


Figure 3.2: The position of the LCA of HspB1 and B5 within the evolution tree of human sHSPs. This image is inspired by the work of G. Wistow[201] and is prepared by MEGA 11[202].

3.2.3 Plan of experiments and analysis

In the computational analysis of the oligomerisation of HspB1, B5, and their LCA, we implemented MD simulations of complex structures formed by two identical ACD-CTD pairs, to investigate the interactions between the ACD and CTD regions. To encompass all possible ACD-CTD combinations of HspB1, B5, and LCA segments, we designed nine distinct structural models, each consisting of two ACD-CTD complexes. Three of the models are composed of ACD and CTD segments from the same sHSP, while the remaining models incorporate ACD and CTD from different sHSPs. In each complex, the two ACD-CTD complexes are interconnected through interactions between the two ACDs, with each ACD associated with a CTD peptide.

In this study, we evaluate the structural stability of the macromolecules and ACD-CTD complexes throughout MD simulations by calculating their RMSD and RMSF. The interaction intensity of each ACD-CTD complex is analysed based

on the binding energy determined using the mmPBSA method. Additionally, we investigate the residue-residue interactions and the formation of hydrogen bonds at the ACD-CTD interfaces to elucidate the reason for the differences in binding affinities among all ACD-CTD interactions.

3.3 Methods

3.3.1 Preparation of the initial structures

Sequence prediction of LCA and sequence alignment

The sequence of the ACD-CTD region of the LCA of HspB1 and B5 is predicted using the ancestral sequence reconstruction (ASR) method, as performed by G. Hochberg[203, 204, 205]. For subsequent simulations, only the highly confident part of the sequence are selected, which include the β3-β9 strands in the ACD and the CTD peptide consisting of nine residues. Following the alignment of the LCA sequence with the ACD-CTD regions of HspB1 and B5 using BLAST+ (fig3.3B)[206], the sequences of HspB1 and B5 are truncated to the same length. Within the LCA sequence, the ACD region is numbered from 1 to 72 and the CTD is numbered from 1 to 9, while the corresponding regions in HspB1 are numbered 95-166 (ACD) and 178-186 (CTD), and those in HspB5 are numbered 75-146 (ACD) and 156-164 (CTD).



Figure 3.3: Sequence alignment of the ACD-CTD sequences of HspB1, B5 and their LCA. The reconstruction of the LCA sequence was performed by G. Hochberg by using the ASR method.

Preparation of the structural models

For the HspB1 dimer, the structure corresponding to the only identifier in the PDB, 4MJH[207], is selected as the structural template after being truncated, as

previously discussed. However, for the HspB5 dimer, two structures corresponding to two identifiers, 4M5S and 4M5T[207], are available in the PDB, in which we select 4M5T for our study, as its ACD-ACD interface is formed following the AP_{II} register, identical to the HspB1 dimer, while the dimer interface in the 4M5S structure is formed following the AP_{III} register. Interfaces formed according to the AP_{III} register typically exhibit smaller overlapping regions, thereby resulting in lower binding intensities compared to AP_{II} register interfaces. In the conformation of 4M5T, coordinates of four residues located in the loop between β5 and β6+7 are missing, which are addressed using Modeller 9.23[118]. The conformations of the remaining macromolecules are predicted using AlphaFold 2[117], and in all predicted conformations, the ACD-ACD interfaces are formed in the AP_{II} register.

3.3.2 MD simulations

Protein solvation and charge neutralisation

The initial structures are parametrised based on the ff19SB force field[128], with the process performed by AmberTools21[127]. During the parametrisation process, each protein structure is solvated in a cubic box of 10 nm dimensions containing 0.15 M NaCl aqueous solution, supplemented with additional ions to neutralise the charge of the protein. The water molecules, as well as the sodium and chloride ions, are modelled by the TIP3P water model[129]. The solvated protein structures are saved in the Amber parameter format, as .prmtop and .inpcrd files, serving as input for MD simulations.

MD simulation process

All MD simulations adhere to the procedure introduced in Chapter2.2.3. For each complex, we implemented five repeating simulations lasting 100 ns.

3.3.3 Trajectory preprocessing and data analysis

Preprocessing MD trajectories, RMSD and RMSF - VMD

The trajectories generated from MD simulations are initially imported into VMD for preprocessing[208]. In simulations involving protein complexes under PBC, the

conformation of a multi-chain protein complex may be divided into segments not to cross the boundaries. These segments must be wrapped, re-centred, and aligned by VMD plugins[209]. VMD is also employed to extract conformations and coordinates of the ACD-CTD complexes, as well as to perform RMSD and RMSF calculations using Tcl scripts. The RMSD and RMSF values are subsequently plotted by scripts that leverage the Matplotlib package in Python[210].

Binding energy calculation - mmPBSA

The mmPBSA calculation is conducted using the Python script mmPBSA.py[137], which is included in AmberTools21. The calculations are performed on the basis of trajectories of solvated four-chain macromolecules, and the binding energies of the two homologous ACD-CTD complexes are computed separately. In this context, the term "complex" refers to the entire complex, "ligand" refers specifically to the CTD peptide within the target ACD-CTD subunit, and the "receptor" is the remainder of the complex, including ACD in the target ACD-CTD subunit and the other subunits.

Trajectory analysis, dimensionality reduction and visualisation

The calculation of residue-residue distances is conducted using the MDAnalysis package in Python[211, 212]. The distance between two residues, A and B, is defined as the shortest distance between any atom in residue A and any atom in residue B, excluding hydrogen atoms. In the analysis of inter-residue distances for a given residue pair, we either compute and record the average inter-residue distance, or determine the frequency with which the two residues are located within a specified distance, requiring a preset cut-off value as a criterion.

Hydrogen bond analysis

The hydrogen bond analysis is processed by the MDAnalysis package in Python[211, 212]. To identify hydrogen bonds in the protein structures, the distances between atom pairs and the angles formed by multiple atoms are calculated and filtered according to specified criteria. The cut-off distance for the donor-hydrogen pair is set as 1.2 Å, while a distance of 3.0 Å is set for the hydrogen-acceptor pair. In addition,

the donor-hydrogen-acceptor angle must not be less than 150°. For each individual residue pair, the frequency of hydrogen bond formation is quantified, and in case that multiple hydrogen bonds are formed between different donor-hydrogen-acceptor pairs, the frequencies of these distinct hydrogen bonds are identified separately, with the final result representing the sum of the frequencies of these hydrogen bonds.

3.4 Results and discussions

3.4.1 HspB1-HspB5 co-assembly is biased as a result of ACD-CTD interactions.

To investigate the interaction mode and co-assembly pattern of the HspB1-B5 hetero-oligomers, M. Collier and D. Šaman performed an experiment involving the mixture of full-length HspB1 and B5 monomers, followed by an examination of the dimension and composition of the resulting oligomers. Through the analysis of the fitted native-MS data, a series of compositions within the mixture was derived. A two-dimensional heat map was generated according to the number of HspB1 and B5 monomers in the oligomers, in which the colour of each cell is linearly correlated with the relative abundance of the corresponding oligomer in the mixture (fig3.4). From the graph, we observed that:

- The majority of the oligomers consist of an even number of sHSP monomers; however, the total number of monomers within the oligomers is not particularly divisible by four. Therefore, we hypothesised that the fundamental units of the oligomers are dimers.

- If HspB1 and B5 monomers are equivalently involved in the oligomerisation process, the distribution of the composition ratio between HspB1 and B5 in the oligomers should reflect the same distribution as the concentration ratio in the sample. However, we noticed that in each subfigure, the distribution of the composition does not precisely align with the HspB1:B5 ratio. Therefore, we assumed that the priority of the self-assembly of HspB1 and B5, as well as the

co-assembly of HspB1 and B5, are varied; in other words, the oligomerisation of HspB1 and B5 exhibits a bias.

- The quantities of each type of monomer present in an oligomer may be either even or odd, suggesting that both homodimers and heterodimers can be formed and are involved in the formation of hetero-oligomers. This implies that heterogeneous interactions can be formed between the ACDs of different sHSPs, thereby supporting the formation of heterodimers, as well as between the ACD and NTD/CTD of different sHSPs, facilitating the formation of large oligomers on the basis of the dimers. Considering the numbers of either of the two monomers in the oligomers, their distributions are approximately normal within a specified range, indicating that the oligomerisation process is stochastic.

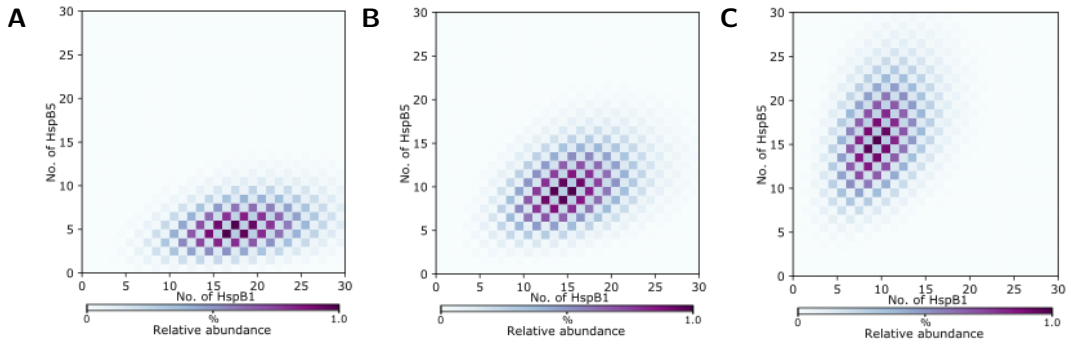


Figure 3.4: Ensembles of HspB1-B5 hetero-oligomers. The three sub-graphs sequentially represent the ensembles obtained with HspB1:HspB5 ratios of 1:3(A), 1:1(B), and 3:1(C). Each marker represents a complex with a distinct composition of HspB1 and B5, and the darkness of its colour correlates to its relative abundance in the mixture. Image prepared by D. Šaman.

Subsequently, my colleagues conducted experiments involving the ACDs of HspB1 and B5 to determine whether the bias attributes to ACD-ACD interactions. The ACDs of HspB1 and B5 were mixed in various ratios (2:3, 1:1, and 3:2), and each mixture was analysed by native-MS (fig3.5). In each mixture, the concentration of the heterodimer and the sum of the concentration of homodimers are approximately equivalent, irrespective of the ratio between the initial concentrations of HspB1

and B5 ACDs. This indicates that the binding intensities of homo- and hetero-dimerisation of HspB1 and/or HspB5 ACDs are not notably varied, suggesting those interactions between ACD and terminus regions of HspB1 and/or B5 are varied.

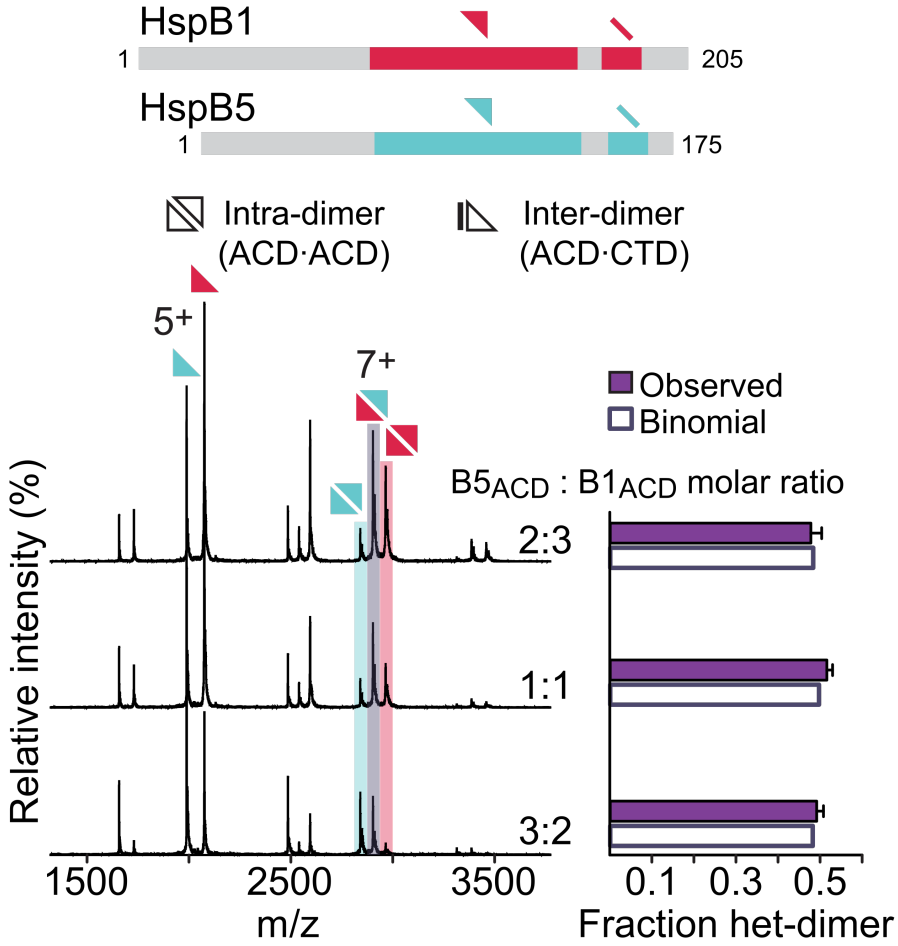


Figure 3.5: Homo- and hetero-dimerisation of HspB1 and HspB5 ACD. Quantitative analysis via native-MS reveals that the concentrations of homo- and hetero-dimers remain approximately equivalent, irrespective of the relative concentrations of HspB1 and HspB5 ACDs in the mixture. This suggests that the ACD-ACD interactions between HspB1 and/or HspB5 are unbiased. This image was prepared by M. Collier and D. Šaman.

In the context of the terminus-involved interactions, our investigation focused on interactions between the ACD and CTD regions, as the NTD regions of both proteins are overwhelmingly flexible and disordered, particularly in small oligomers and dimers. As illustrated in fig3.6, when the ACD segment is the only limiting reactant in a mixture for the ACD-CTD co-assembly, a notable disparity emerges in the binding affinities of the two types of ACDs towards the CTD peptide. In a mixture comprising HspB5 ACD and HspB1 CTD at a 1:8 ratio, the ACD-CTD complex is

the predominant component. Conversely, in the mixture containing HspB1 ACD and HspB5 CTD with the same ingredient ratio, the concentration of the hetero-complex is significantly lower than that of the free ACD. Therefore, the ACD-CTD co-assembly, in contrast to the ACD-ACD dimerisation, is demonstrated to be biased, which contributes to the overall biased co-assembly of HspB1 and HspB5.

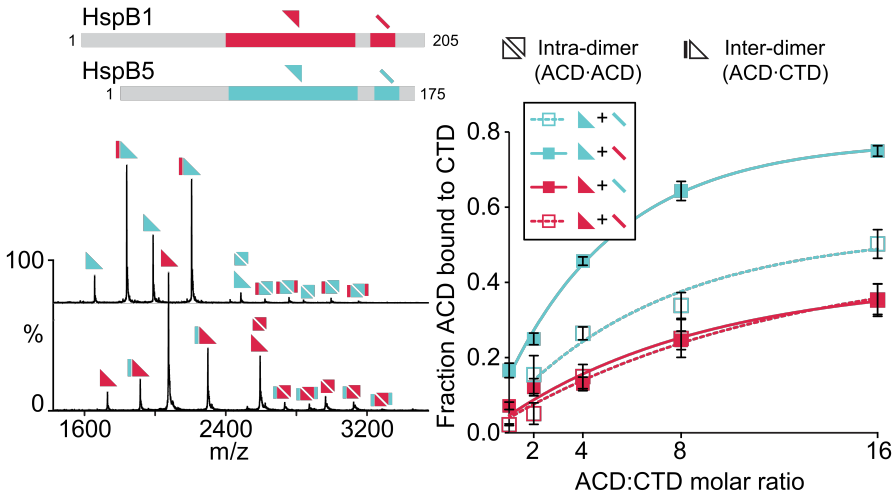


Figure 3.6: Native-MS results of mixtures of heterogeneous ACD and CTD, wherein the CTD peptide plays a role as the excess reactant. In the mixture of HspB5 ACD and HspB1 CTD, the heterogeneous ACD-CTD complex is the predominant component; while in the binary mixture of HspB1 ACD and HspB5 CTD, the hetero-complex is less abundant, with its concentratio lower than that of the free ACD. Image prepared by M. Collier and D. Šaman.

3.4.2 HspB1 CTD forms less intense ACD-CTD interactions than HspB5 CTD.

To investigate the interactions between the ACD and CTD of HspB1 and/or B5, we performed simulations of macromolecules consisting of two homogeneous ACDs and two homogeneous CTD peptides. From the trajectories of these macromolecules, we extracted 1000 conformations, evenly from five 100 ns repeats, for each ACD-CTD pair. Initially, we assessed the conformational stabilities of the complexes throughout the simulations by computing their RMSD and RMSF. For each macromolecule, RMSD values of the "first" complex, of which the initial structure is selected as the template for RMSD computation, are generally lower than those of the other complex(fig3.7A), indicating that the conformations of two ACD-CTD complexes within the same dimers are not precisely identical throughout the simulations. It

implies the occurrence of conformational transformations around the dimerisation site, which break the symmetry of the two ACD-CTD complexes in a macromolecule. As displayed by RMSF values of the residues(fig3.7B), residues located in the C-terminal half of β5 and the loop connecting β5 to β6+7, which are situated adjacent to the ACD-ACD dimerisation site, have the highest RMSF values among all residues in each complex. However, residues in each complex related to the interactions between ACD and CTD, which are located in β4, β8, and the CTD loop, have relatively low RMSF, indicating that the regions remain stable due to the interactions, regardless of the overall structural changes.

We conducted the calculation of the binding energy for each ACD-CTD complex using the mmPBSA algorithm (table3.1) and performed t-tests to compare the observed values among complexes with different compositions(table3.2), two complexes within the same macromolecules, and the complexes from different repeating simulation(fig3.8). Generally, binding energies of the interactions between the CTD of HspB5 and both ACDs are statistically equivalent and significantly greater than those of the interactions involving the CTD peptide of HspB1. The CTD peptide of HspB1, conversely, demonstrates a higher binding affinity towards the HspB5 ACD region, compared to its homogeneous ACD segment. Therefore, we concluded that the binding energies of the ACD-CTD complexes formed by HspB1 and/or B5 follow the sequence: B1ACD-B5CTD $\simeq$ B5ACD-B5CTD $>>$ B5ACD-B1CTD $>$ B1ACD-B1CTD. Within each macromolecule, the binding energies of the two complexes are not statistically identical, which supports our hypothesis that none of the dimers can pertained a perfectly symmetric conformation. In addition, t-tests of the repeats reveal that the conformations of the two homogeneous complexes, compared to those of the heterogeneous complexes, are less notably distinct across the repeating simulations.

To investigate the inter-residue interactions that contribute to variations in binding affinities, we calculated the distances between each pair of residues, and for each pair whose distance ever falls below a preset cut-off of 3 Å, we computed the frequency with which this distance remains below the threshold throughout

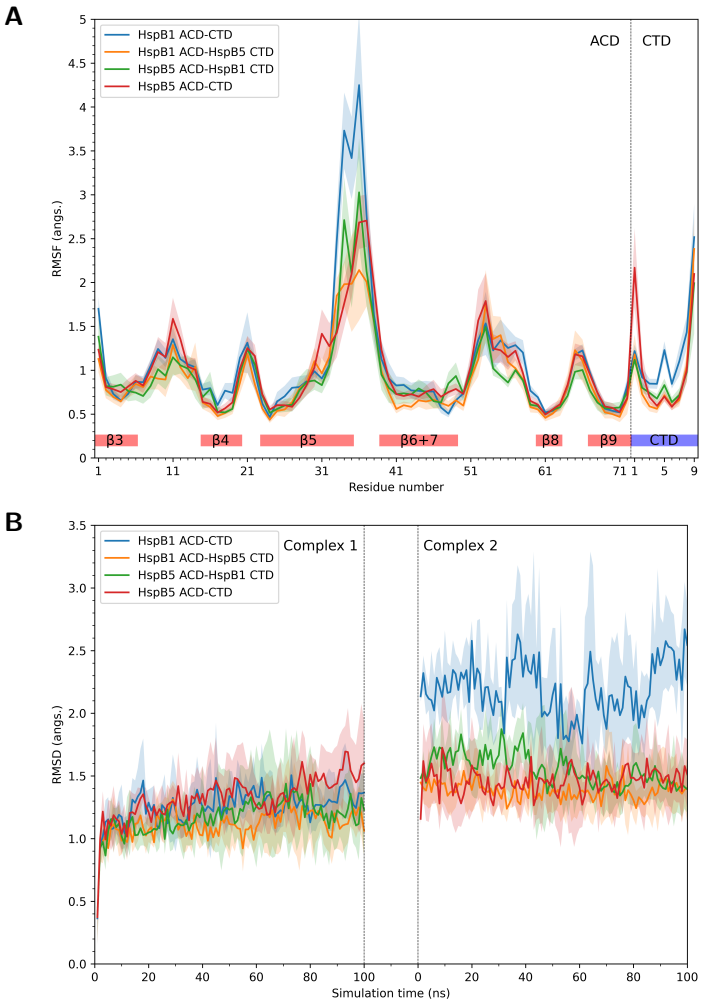


Figure 3.7: **A** Residue RMSF of HspB1/B5 ACD-CTD complexes,calculated based on ten trajectories of the ACD-CTD complexes, which are extracted from all five trajetcories of the macromolecule complex. Standard deviations of residue RMSF are calculated across all ACD-CTD complex traqjetcories and are displayed as the shaded regions. **B** Replica-averaged RMSD of HspB1/B5 ACD-CTD complexes. In the graph, the left half shows mean values of the RMSD of one ACD-CTD complex (complex 1) across five repeating simulations, and the right half represents the other complex (complex 2).

ACD-CTD	Complex 1		Complex 2		Overall	
	Average	Std Dev	Average	Std Dev	Average	Std Dev
B1-B1	-7.2280	8.4579	-7.6613	7.9765	-7.4447	8.2235
B1-B5	-16.6338	8.6195	-18.8900	8.6455	-17.7619	8.7559
B5-B1	-9.5776	7.0289	-10.5396	7.1393	-10.0586	7.1006
B5-B5	-16.2257	9.5507	-19.6700	9.1783	-17.9479	9.5223

Table 3.1: Binding energies of ACD-CTD complexes consisting of HspB1 and B5 segments, calculated by mmPBSA.

the trajectories, which is taken as a criterion for interaction between the two residues. As revealed in fig3.9, the four complexes can be distinctly categorised

ACD-CTD	B1-B1	B1-B5	B5-B1	B5-B5
B1-B1	N/A	~ 0	$4.26 * 10^{-14}$	~ 0
B1-B5	-	N/A	$3.11 * 10^{-93}$	0.65
B5-B1	-	-	N/A	~ 0
B5-B5	-	-	-	N/A

Table 3.2: Double-sided P-values from t-test between values of binding energies of ACD-CTD complexes

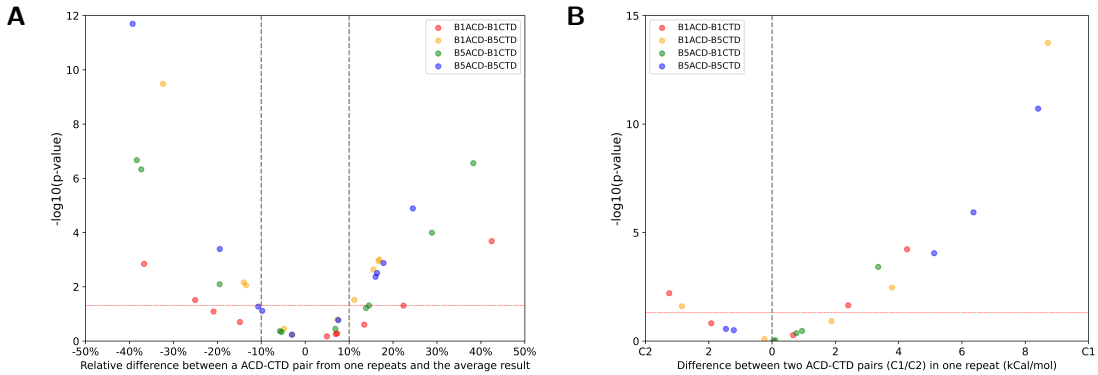


Figure 3.8: **A** Volcano plots of p-values and the relative difference between the binding energy of a ACD-CTD pair in one repeat and the overall average value. **B** Volcano plots of p-values and the difference between the binding energy of two ACD-CTD pair (C1/C2) within a complex structure from one repeat.

into two groups according to the distributions of their inter-residue interactions, which completely correlate with the species of their CTD peptides, rather than their ACD segments. A notable distinction is that R157 and R163 in the HspB5 CTD form more contacts with both ACDs compared to I179 and F185 in the corresponding positions in the HspB1 CTD. In complexes containing the HspB5 CTD, R163 establishes close contacts with HspB1 D107 and HspB5 D87, while R157 intertacts with HspB1 T113 and HspB5 V93 within the β4 strand, as well as with multiple residues in the loop between β6+7 and β8. Conversely, in the HspB1 CTD, F185 cannot maintain a persistent contact with any residues in the ACDs, whereas I179 forms a contact with HspB1 L157 and HspB5 L137, potentially due to their hydrophobicity. Consequently, the interactions involving the two arginines in the HspB5 CTD facilitate the connections across both termini of the peptide and multiple regions in the ACDs, which promote ACD-CTD binding. In contrast, both termini of the HspB1 CTD is occupied by only a single motif in the ACDs, resulting in a significantly lower binding affinity compared to its counterpart.

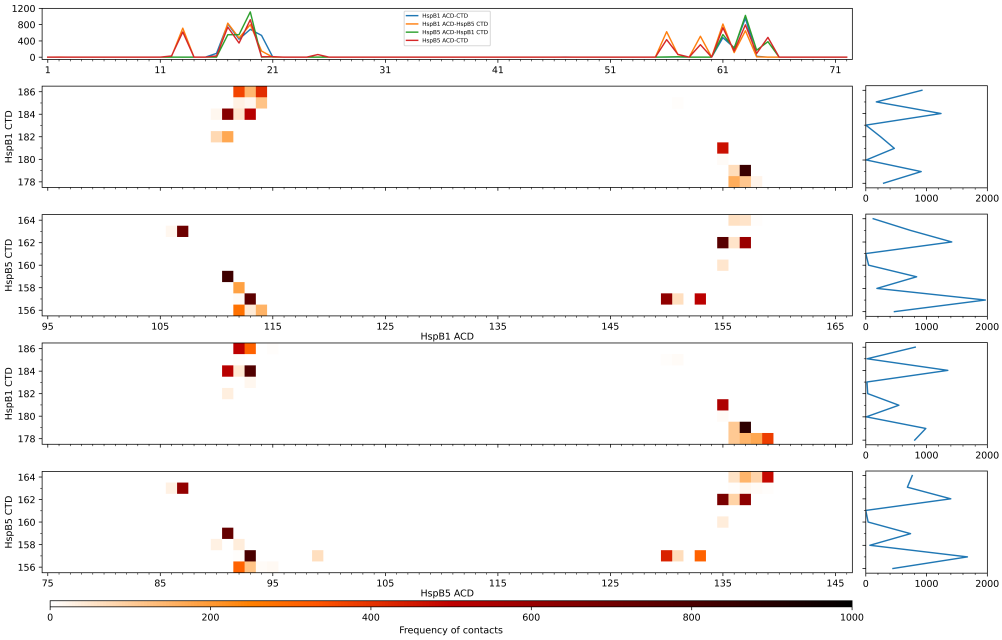


Figure 3.9: Frequencies of close contacts between intermolecular residue pairs in HspB1/B5 ACD-CTD complexes, as well as the cumulative occupancies of each residue in these complexes. The analysis considers only heavy atoms for the calculation of inter-residue distances with a predetermined cut-off distance of 3 Å.

To validate the hypothesis, we identified hydrogen bonds formed between the aforementioned residues and quantified the frequency of each hydrogen bond. As presented in table3.3, the HspB5 CTD forms multiple hydrogen bonds with residues located in the loops between β3 and β4, as well as between β6+7 and β8. Notably, the most frequent hydrogen bonds occur between R163 and HspB1 D107 ($74.2 \pm 12.6\%$) or HspB5 E87 ($83.6 \pm 23.9\%$), as illustrated in fig3.10C/D. In contrast, complexes consisting of the HspB1 CTD have a reduced number of intermolecular residue pairs connected by hydrogen bonds, with these interactions being confined to the β4 and β8 strands (fig3.10A/B), and the overall occupancy is significantly lower. Furthermore, we also noticed that while binding to the HspB1 CTD, two residues located in the β4 strand of the HspB1 ACD, K112 and K114, may form hydrogen bonds with E186 in the CTD, which are potentially exclusive interactions; whereas in the corresponding position in the B5ACD-B1CTD complex, only one lysine, K92, is able to form a hydrogen bond with E186, thereby likely enhancing the interaction between V93 and F185 and stabilising the interface. Additionally, compared to the HspB1 ACD, the HspB5 ACD contains an additional serine located in the β8 strand,

which promotes the formation of hydrogen bonds between this strand and the HspB1 CTD. Therefore, we concluded that the differences in the binding affinities of HspB1 and B5 are primarily attributable to the sequence variations between the two CTD peptides, while the sequence differences between the ACDs play a less important role.

ACD-CTD	**B1-B1**		**B5-B1**	
Location	**H-bond**[a]	**Frequency(%)**[b]	**H-bond**	**Frequency(%)**
β4	V111-T184	18.3 ± 3.1	V91-T184	17.2 ± 7.4
	K112-E186	25.8 ± 4.6	K92-E186	34.8 ± 13.8
	T113-F185	10.4 ± 4.0	V93-F185	24.2 ± 8.9
	K114-E186	29.6 ± 7.3		
β8	S156-E178	10.7 ± 2.7	S138-E178	15.5 ± 4.9
			S139-E178	28.3 ± 8.0
ACD-CTD	**B1-B5**		**B5-B5**	
Location	**H-bond**	**Frequency(%)**	**H-bond**	**Frequency(%)**
β3-β4 loop	D107-R163	74.2 ± 12.6	E87-R163	83.6 ± 23.9
β4	K112-E156	19.4 ± 6.1	K92-E156	18.8 ± 9.1
	K112-T158	10.0 ± 3.6		
	K114-E156	10.2 ± 4.1		
β6+7-β8 loop	P150-R157	41.3 ± 12.0	P130-R157	25.8 ± 11.7
β8	V153-R157	30.8 ± 10.0	I133-R157	21.8 ± 6.7
	S155-T162	26.6 ± 5.3	S135-T162	26.6 ± 6.9
			L137-R163	11.5 ± 4.6
			S139-E164	38.1 ± 5.5

Table 3.3: Frequencies of intermolecular hydrogen bonds in the HspB1/B5 ACD-CTD complexes

[a]A hydrogen bond will only be listed if its overall frequency is greater than 10%.
[b]The errors refer to the standard deviation among the values of the five repeats

In conclusion, our simulations of the HspB1/B5 ACD-CTD pairs demonstrate that the presence of the HspB5 CTD results in the highest binding affinities, irrespective of the species of ACDs. Conversely, the HspB1 CTD forms a complex with the HspB5 ACD, with the binding energy slightly surpassing that of the homogeneous HspB1 ACD-CTD pair. An analysis of inter-residue distances and hydrogen bond distributions indicates that two arginine residues, R157 and R163, located within the HspB5 CTD, engage in interactions with both HspB1 and B5 ACDs through hydrogen bonds formed with multiple residues in the ACDs. In contrast, the hydrophobic residues in the HspB1 CTD that correspond to these arginines do not participate in hydrogen bonding. Additionally, sequence variations

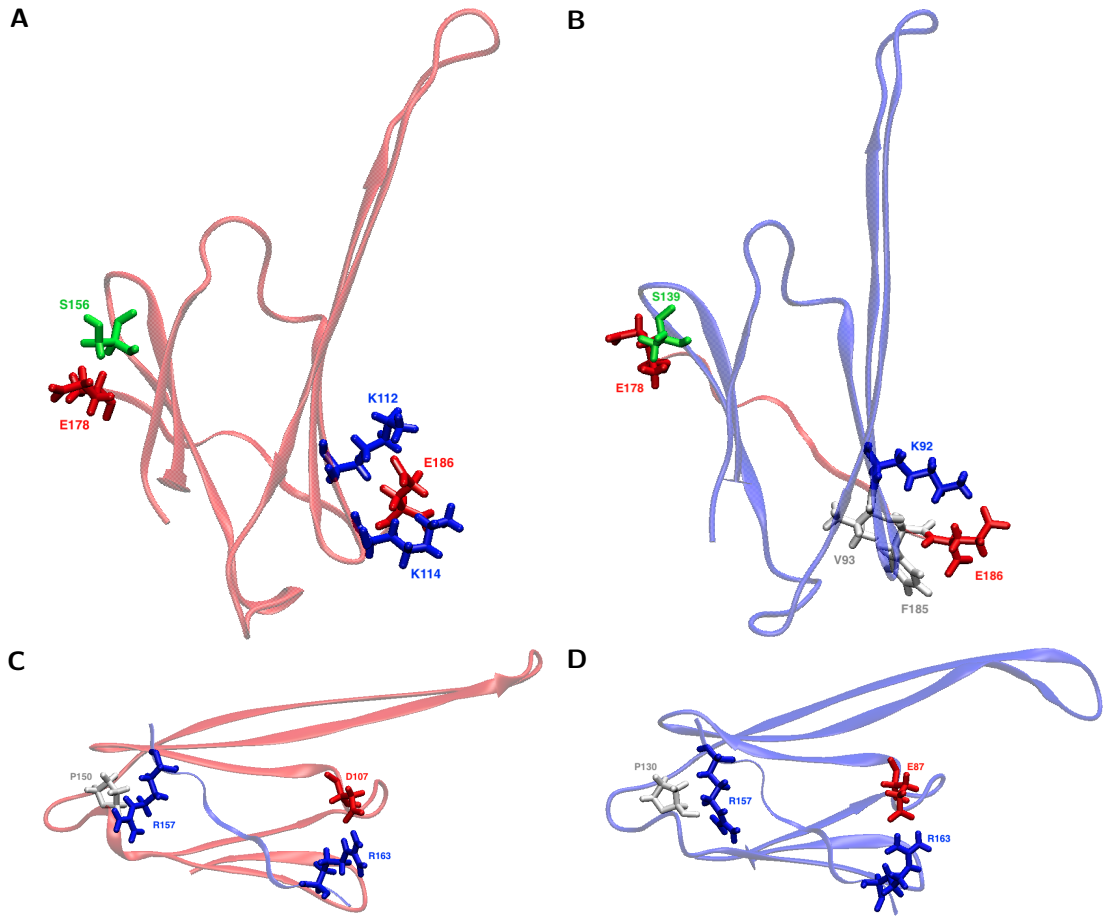


Figure 3.10: Snapshots of the interfaces of HspB1/B5 ACD-CTD complexes, in which interacting intermolecular residue pairs are highlighted. Each interface is extracted from the most representative conformation of the ACD-CTD complex[213, 214]. To identify the most representative conformation of an ACD-CTD complex in the MD simulation, we performed a PCA calculation based on the coordinates of all αC in each ACD-CTD complex individually. The most representative conformation is defined as the conformation with the lowest weighted sum of principal component eigenvalues. **A** HspB1 ACD-CTD complex, in which two hydrogen bonds, K112-E186 and K114-E186, occupy the same residue; **B** B5ACD-B1CTD complex**C** B1ACD-B5CTD complex; **D** HspB5 ACD-CTD complex.

in β4 and β8 contribute to the differences in binding energies observed between the two ACD-CTD pairs involving the HspB1 CTD.

3.4.3 The LCA CTD demonstrates similarity to the HspB5 CTD due to conserved arginines

Analogously to the analysis of HspB1/B5 ACD-CTD complexes, we performed MD simulations of macromolecules comprising LCA segments and extracted an identical number of conformations from the trajectories (fig3.11). Compared to the HspB1/B5

complexes, complexes comprising LCA segments have overall greater RMSD values, ranging between 1 to 3 Å. In terms of the residue RMSF values, the distributions of these values for residues in the ACDs are similar to those in the HspB1/B5 complexes. However, while binding to the LCA ACD, residues located in the HspB5 CTD and the LCA CTD, particularly in their N-termini, are not stabilised while the peptides are attached to the LCA ACD, while RMSF of the residues within the HspB1 CTD are exceedingly low. According to the previous analysis, we supposed that the arginines in HspB5 CTD, including R157 and R163, as well as the arginines in the corresponding positions in the LCA CTD, fail to interact with the LCA ACD, which could be consequent to the sequence variation between the LCA ACD and the post-evolution ACDs or the inaccurate ASR prediction of the sequence.

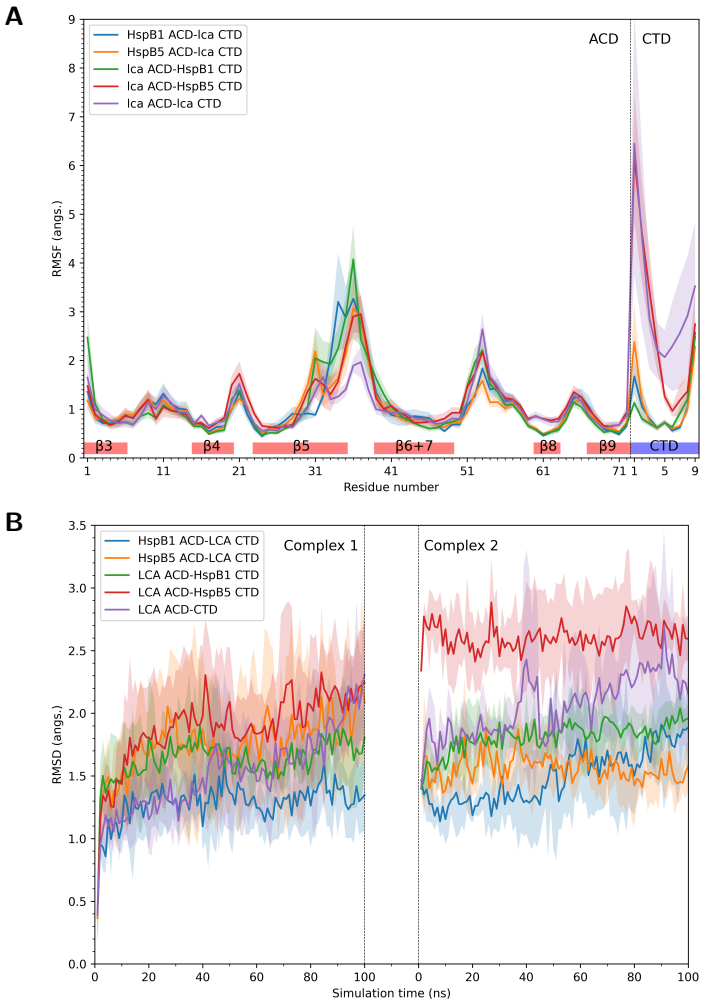


Figure 3.11: **A** Residue RMSF of ACD-CTD complexes containing LCA segment. **B** Replica-average RMSD of ACD-CTD complexes containing LCA segment.

We also computed the binding energies of these complexes using the mmPBSA method (table3.4), and performed t-test on the corresponding values(table3.5 and fig3.12). The complexes comprising the LCA CTD and the ACD from HspB5, or the LCA itself, which have high RMSD/RMSF values, also demonstrate high standard deviations in their binding energies, compared to other complexes. In addition, these two complexes display overall lower binding energies and greater variations among co-existing monomers and throughout repeated simulations, suggesting that the ACD-CTD interfaces remain flexible throughout the simulations. In contrast, the four complexes that include the CTD from HspB5 or LCA, while excluding the LCA ACD, exhibit indistinguishable binding energies, which are higher than those of any other complexes. Among the remaining three complexes containing the HspB1 CTD, the complex with the LCA ACD has the lowest binding energy. Overall, the sequence of binding energies for the complexes is: B1ACD-B5CTD $\simeq$ B5ACD-B5CTD $\simeq$ B1ACD-LCACTD $\simeq$ B5ACD-LCACTD >> B5ACD-B1CTD > B1ACD-B1CTD > LCAACD-B1CTD > LCAACD-B5CTD > LCAACD-LCACTD, in which the binding energies of the last two complexes may not be sufficiently high to stabilise the ACD-CTD interfaces, as indicated by their variations.

ACD-CTD	Complex 1		Complex 2		Overall	
	Average	Std Dev	Average	Std Dev	Average	Std Dev
B1-LCA	-18.5100	7.7066	-17.5398	8.4594	-18.0249	8.1063
B5-LCA	-17.1563	9.1392	-18.5705	9.8055	-17.8634	9.5045
LCA-B1	-4.6029	7.5555	-4.5815	8.4057	-4.5922	7.9919
LCA-B5	-5.2279	17.5782	1.1842	12.0255	-2.0218	15.3975
LCA-LCA	11.1436	24.6141	-2.6817	24.7226	4.2309	25.6187

Table 3.4: Binding energies of ACD-CTD complexes containing LCA segments, calculated by mmPBSA.

ACD-CTD	B1-B1	B1-B5	B5-B1	B5-B5
B1-LCA	~ 0	0.486	~ 0	0.846
B5-LCA	~ 0	0.804	$3.64 * 10^{-87}$	0.843
LCA-B1	$6.00 * 10^{-15}$	~ 0	~ 0	~ 0
LCA-B5	~ 0	~ 0	~ 0	~ 0
LCA-LCA	~ 0	~ 0	~ 0	~ 0

Table 3.5: Double-sided P-values from t-test between values of binding energies of ACD-CTD complexes related and unrelated to LCA segments

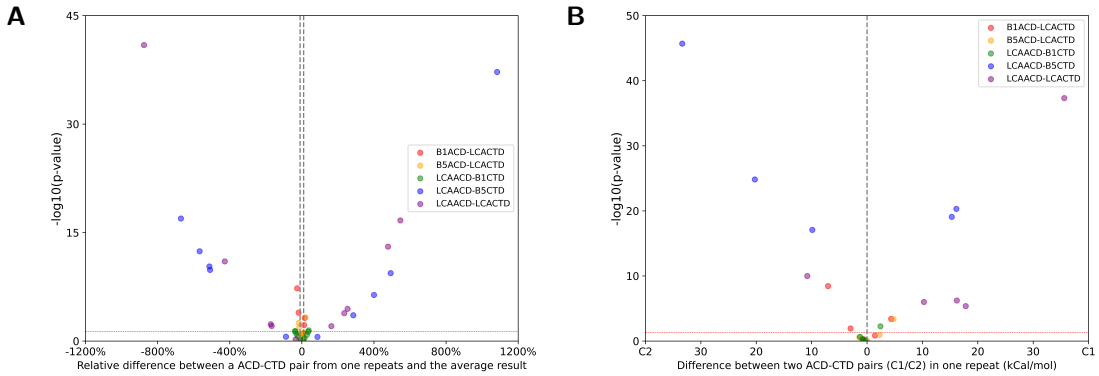


Figure 3.12: **A** Volcano plots of p-values and the relative difference between the binding energy of a ACD-CTD pair in one repeat and the overall average value. **B** Volcano plots of p-values and the difference between the binding energy of two ACD-CTD pair (C1/C2) within a complex structure from one repeat.

According to the distributions of inter-residue contacts in the complexes involving LCA segments, as displayed in fig3.13, along with those displayed in fig3.9, these complexes can be categorised into three columns: one exclusively contains all complexes comprising the HspB1 CTD; one includes all complexes comprising one of the two ACDs from HspB1 or B5 and one of the two CTDs from HspB5 or the LCA, corresponding to those with the highest binding energies; and the last one encompasses the remaining two complexes, which consist of the LCA ACD and the CTD from either HspB5 or the LCA, which are less stable according to their binding energies. Complexes in the first category are conjugated by two groups of inter-residue interactions: one between β4 and the C-terminus of the CTD, while the other occurs between β8 and the N-terminus. Within those complexes in the second group, which exhibit the highest binding energies, the second and penultimate residues in the CTDs engage in inter-residue interactions with residues in multiple regions, including, but not limited to, the β4 and β8 strands in the ACDs, therefore stabilising the ACD-CTD interfaces. Conversely, the complexes in the three column, which are unlikely to be stable according to the positive values of their binding energies, cannot form inter-residue interactions incorporating residues in the β4 and β8 strand. Instead, the peptide is attached to the β3 and β6+7 strand, as well as several loops, within the ACDs, primarily through its C-terminus. The lack of constraints on the N-terminus of the CTD, resulting in

the high RMSF of residues located within this region and the positive values of their binding energies, indicates that these ACD-CTD pairs are unlikely to form stable complexes either *in vivo* or *in vitro*.

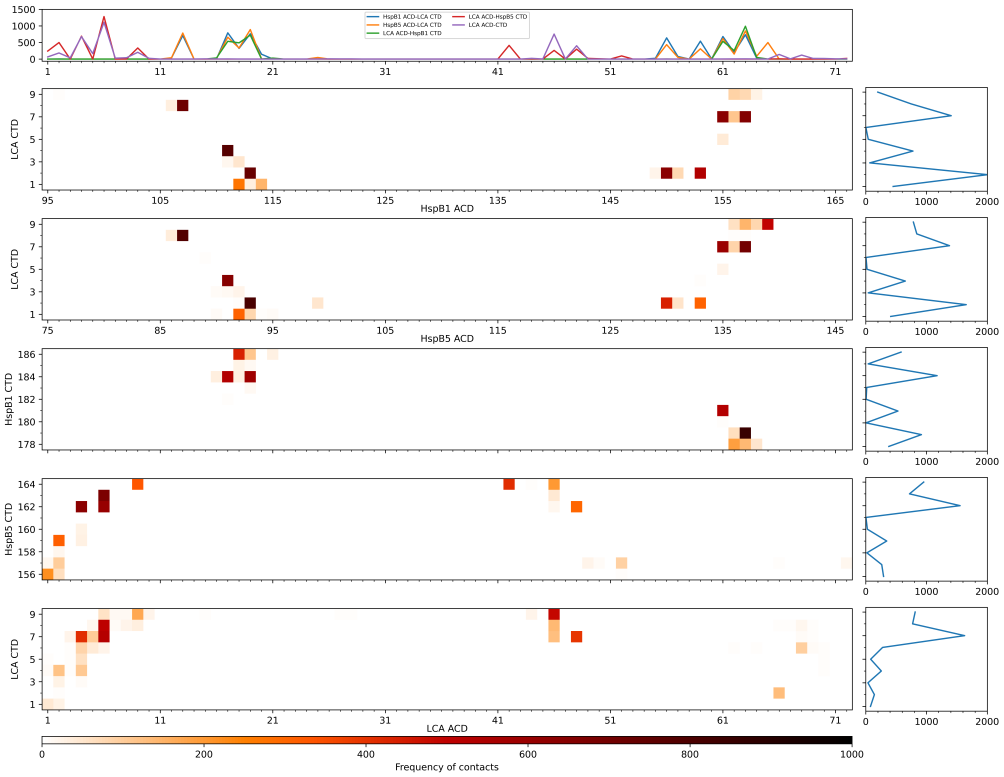


Figure 3.13: Frequencies of close contacts between intermolecular residue pairs in ACD-CTD complexes that include the LCA ACD and/or CTD segments, along with the cumulative occupancies of each residue in these complexes. Only heavy atoms are taken into account in the calculation of inter-residue distances, with a cut-off distance preset as 3 Å.

Similar to the distributions of inter-residue contacts, the hydrogen bonds formed in complexes consisting of the LCA ACD and the HspB5/LCA CTD are located in distinct locations compared to those in the other complexes (table3.6). The frequencies of most hydrogen bonds in these two complexes are significantly varied across different repeats, suggesting that these hydrogen bonds cannot be constantly maintained within the simulations. As noted previously, an arginine in the HspB5 CTD peptide can form a hydrogen bond with a glutamic acid in the loop between β3 and β4 in the HspB5 ACD; However, despite the identity of these amino acids in the LCA sequence, LCA CTD R8 and HspB5 R163 interact with an aspartic acid in β3 (LCA ACD D6), rather than with the glutamic acid in β4 (LCA ACD

E13), as displayed in fig3.14. Furthermore, it is observed that in the HspB5 ACD, the glutamic acid E87, is protected by N85 due to steric effect and the hydrogen bond formed between the two residues. However, in the LCA ACD, D6, which corresponds to HspB5 E87, remains exposed and thus available for interaction with the CTD, as M4, iso-position with HspB5 N85, cannot provide any protective effect.

Despite that the distribution of inter-residue interactions in the three complexes involving the HspB1 CTD is comparable, the occupancy of hydrogen bonds formed between the LCA ACD and the peptide closely resembles that of the B5ACD-B1CTD complex. This similarity arises from the presence of identical lysine and serine residues in the LCA ACD, which are located in the same region as those in the HspB5 ACD, as previously mentioned, and which differ from those in the HspB1 ACD. Consequently, the LCAACD-B1CTD complex exhibits an overall lower occupancy of hydrogen bonds, thereby demonstrating lower binding energies than the other two complexes.

	HspB1 CTD		HspB5 CTD		LCA CTD[a]	
Location	H-bond	Frequency(%)	H-bond	Frequency(%)	H-bond	Frequency(%)
β3			M4-T162	14.6 ± 11.1	D6-R8	28.9 ± 25.5
			D6-R163	51.4 ± 9.2		
β3-β4 loop			H9-E164	24.3 ± 26.3		
β4	V17-T184	17.4 ± 9.1				
	K18-E186	32.0 ± 8.5				
β6+7			R42-E164	42.6 ± 4.5	R46-E9	39.0 ± 39.3
			R46-E164	20.3 ± 32.9	Y48-Q7	30.6 ± 28.7
			Y48-T162	17.7 ± 9.5		
β6+7-β8 loop			E52-R157	11.4 ± 14.8		
β8	S62-I177	11.1 ± 4.6				
β8-β9 loop					D66-R2	18.5 ± 26.4

Table 3.6: Frequencies of hydrogen bonds between LCA ACD and all CTDs

[a]ACD and CTD in LCA are numbered separately, both starting from 1.

In contrast to complexes that incorporate the LCA ACD, the hydrogen bonds formed between the ACD of HspB1/B5 and the LCA CTD, as listed in table3.7, are located in positions analogous to those established between the same ACDs and the HspB5 CTD. Furthermore, the occupancy of the majority of iso-position hydrogen bonds is comparable, which supports our hypothesis that the arginines in the CTD play a critical role in the differences in binding affinity between HspB1 and B5.

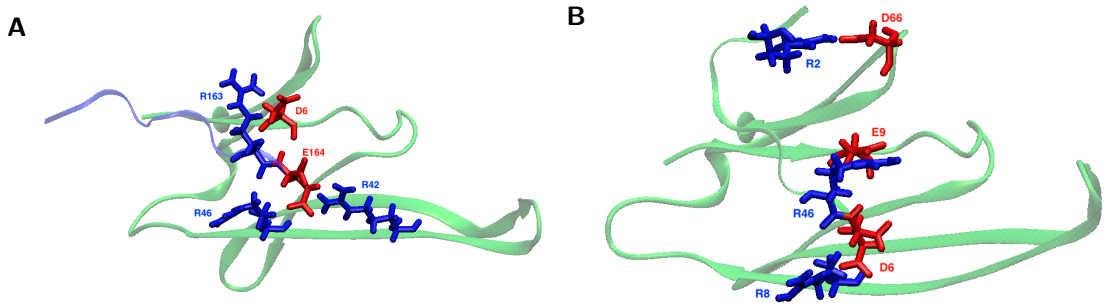


Figure 3.14: Inter-residue contacts in interfaces between LCA ACD and HspB5/LCA CTD. **A**: LCA ACD-HspB5 CTD complex, in which two hydrogen bonds, R42-E164 and R46-E164, are unlikely to co-exist; **B**: LCA ACD-CTD complex.

	HspB1ACD - lcaCTD		HspB5ACD - lcaCTD	
Location	**H-bond**	**Frequency(%)**	**H-bond**	**Frequency(%)**
β3-β4 loop	D107-R8	66.1 ± 9.4	E87-R8	111.3 ± 29.2^a
β4	K112-E1	18.0 ± 5.7	K92-E1	21.9 ± 7.9
	K114-S3	10.2 ± 2.5		
β6+7-β8 loop	P150-R2	41.0 ± 15.2	P130-R2	25.8 ± 8.4
β8	V153-R2	35.4 ± 12.8	I133-R2	25.1 ± 8.7
			S139-E9	39.0 ± 14.1

Table 3.7: Frequencies of hydrogen bonds between the HspB1/B5 ACD and the LCA CTD

[a]This value is greater than 100% because this residue pair can form multiple hydrogen bonds in some conformations, and the hydrogen bonds are counted separately.

In conclusion, the LCA CTD exhibits the capacity to bind to both the ACD regions of HspB1 and HspB5, demonstrating binding affinities that are comparable to those observed with the HspB5 CTD. The inter-residue contacts within these ACD-CTD pairs are remarkably similar. The interactions are primarily attributed to the arginine residues located in analogous positions within both HspB5 and LCA CTD, thereby reinforcing the assertion that these arginines play a crucial role in ACD-CTD interactions. In contrast, the LCA ACD can only form a ACD-CTD complex with the HspB1 CTD peptide, albeit with lower stability compared to the other complexes. Furthermore, the binding energies between this ACD and the other two arginine-enriched peptides are positive, which hinders the formation of the corresponding complexes. The distribution of inter-residue contacts reveals that the residues involved in the interactions between the LCA ACD and the CTD of HspB5 or LCA are distinctly different from those in the other complexes. This

disparity contributes to the conformational transition of the binding interfaces and the reduction in binding affinities.

3.5 Conclusions

In this chapter, we examined the co-assembly of HspB1 and B5, focusing on the hetero-interactions between these two proteins. The hetero-oligomers formed by HspB1 and HspB5 exhibit distinct compositions and conformations, differing in dimensions and chemical properties from the homo-oligomers formed solely by either HspB1 or HspB5. Through laboratory experiments, we confirmed that HspB1 and HspB5 can interact via either ACD-ACD or ACD-CTD interactions, and the binding affinities of the homogeneous and heterogeneous ACD-ACD interactions are not notably diverse. However, native-MS results for ACD-CTD mixtures indicate that the ACD-CTD interactions differ in binding affinity, which consequently affects their formation probabilities.

We implemented MD simulations of dimers consisting of two homogeneous ACD-CTD complexes, comprising HspB1, B5, and their LCA segments, in which the LCA sequence is predicted by the ASR algorithm. The objective of these simulations is to assess the variability in binding affinities among different segments and to elucidate the cause of the differences.

We initially focused on complexes consisting of HspB1 and/or B5 segments. Their RMSD and RMSF indicate that the interaction interface of each ACD-CTD complex is stabilised, despite the asymmetry between the two complexes and the instability observed at the ACD-ACD interfaces in certain macro-molecules. The results from the mmPBSA calculations suggest that the HspB5 CTD peptide can form much stronger interactions with both ACDs, regardless of the ACDs it binds to, compared to the complexes involving the HspB1 CTD; while within the two complexes involving HspB1 CTD, the homogeneous interaction is slightly weaker than the heterogeneous one.

Distributions of inter-molecular inter-residue contacts in these complexes indicate that the HspB5 CTD is affiliated to multiple regions in the ACDs, including the $\beta 4$

and β8 strands and two loops between the β3 and β4 strands and between the β6+7 and β8 strands, while the HspB1 CTD is attached to only the β4 and β8 strands. According to the distributions of occupancies of hydrogen bonds, we noticed that two arginines, R157 and R163, in the HspB5 CTD peptide can form hydrogen bonds with residues in the loops, while the iso-position residues in the HspB1 CTD are both non-polar, therefore unable to stabilise the ACD-CTD interface.

While segments from the predicted LCA are introduced to the simulations, we discovered that, within the five complexes comprising of LCA segments, two consisting of the LCA ACD and the CTDs from HspB5 or the LCA cannot be stabilised, while the other three behave highly similar to the HspB1/B5 complexes. The distributions of the inter-residue distances reveal that within the two unstable complexes, the N-terminus of the LCA CTD peptide is not restrained by any regions in the ACDs, potentially leading to the instability of the whole conformation. The other three complexes, whose values of the ACD-CTD binding energies are negative, can be divided into two categories, along with the HspB1/B5 complexes: one column includes complexes containing the HspB1 CTD, which have their N- and C-termini of the CTD peptide affiliated to the β4 and β8 strands respectively; and the other column consisting the remaining complexes, in which both termini in the peptide cross-links with mutiple regions in the ACDs, therefore exhibiting higher binding energies. According to the inter-residue contacts and occupancy of hydrogen bonds, we discovered that the distinction is predominantly relevant to the arginines located in both termini of the peptide, which play a significant role in the ACD-CTD binding.

4

Phosphorylations at single residues regulate the dimerisation of filamin C domain 24.

Contents

4.1 Abstract 76
4.2 Introduction 76
4.2.1 FLNCd24 plays a crucial role in cardiac muscle tissue . 76
4.2.2 HspB7 is a predominantly monomeric sHSP 78
4.2.3 Residue phosphorylations modify the function of FLNCd24. 81
4.2.4 Plan of experiments and analysis 83
4.3 Methods 83
4.3.1 Preparation of initial structures 83
4.3.2 MD simulations 84
4.3.3 Preprocessing and data analysis 84
4.4 Results and discussions 86
4.4.1 HspB7 is dominantly monomeric, unlike most other sHSPs 86
4.4.2 Phosphorylations of T2677 and Y2683 regulate the protonation state of H2673. 88
4.4.3 The competition between FLNCd24-HspB7 interaction and FLNCd24 dimerisation. 90
4.4.4 The phosphorylation of T2677 and Y2683 results in distinct regulatory effects on the homo- and hetero-interactions 95
4.5 Conclusions 109

4.1 Abstract

In this chapter, we examine the regulatory role of phosphorylations at single residues within filamin C (FLNC) in its dimerisation and its interaction with HspB7. Prior studies have established that FLNC can undergo homo-dimerisation at its domain 24 (FLNCd24), a process that competes with the interaction between FLNCd24 and HspB7. Both interactions are essential for the functionality of FLNC within the sarcomere and are modulated by two phosphorylation sites, T2677 and Y2683, located in FLNCd24. Native MS spectra of phosphorylation-mimicking mutated complexes indicate that phosphorylation at T2677 enhances the homo-dimerisation of FLNCd24 while simultaneously diminishing its interaction with HspB7. Conversely, phosphorylation at Y2683 exerts an opposing effect on both interactions. We implemented MD simulations of wild-type (WT), phosphorylated, and phosphomimicking FLNCd24 homodimers and heterogeneous complexes, focusing on the examination of hydrogen bonds in these conformations. By computing the hydrogen bond formation rates in the WT homo- and hetero-dimer conformations as references, we observed that an increased frequency of intermolecular hydrogen bonds occurs at the interfaces of the T2677-phosphorylated homodimer and the Y2683-phosphorylated heterodimers, whereas hydrogen bond formation is less prevalent in the other complexes. Consequently, our findings suggest that the two types of phosphorylation have opposing effects on FLNCd24 dimerisation and its interactions with HspB7 ACD, which can serve as a pathogenic mechanism or an indicator of various forms of cardiac dysfunction and disorders.

4.2 Introduction

4.2.1 FLNCd24 plays a crucial role in cardiac muscle tissue

Filamins are a class of proteins characterized by a dual-branch structure, with each branch comprising an N-terminal actin-binding domain, 24 immunoglobulin-like domains, and hinge regions located between domains 15-16 and 23-24[215, 216, 217]. The two branches are interconnected through the dimerisation of domain 24[218,

219], and each actin-binding domain can bind an actin protein independently[220]. This configuration enables the whole filamin proteins to function as cross-linkers for actin filaments. The cross-linking capability of filamins facilitates a variety of functions across different biochemical processes, including the formation and operation of muscle tissue[221, 222, 223], signalling[217, 224], and intercellular communication[225]. The filamin protein family is comprised of three distinct types, A, B, and C, in which Filamin A (FLNA) and FLNC are predominantly expressed in cardiac muscle tissue[226, 227, 228], whereas Filamin B (FLNB) functions as a cytoplasmic protein that regulates intracellular signaling and communication between the cell membrane and muscle protein complexes, thereby influencing skeletal development and metabolism[229, 230].

The Z-disc, which connects actin filaments to form the sarcomere—the fundamental unit of cardiac and skeletal muscle fibers[231, 232]—exhibits a remarkably intricate composition and a highly complicated macromolecular structure[233]. It encompasses a diverse array of linker proteins[234, 235], including FLNC[223, 236], which facilitate the binding of adjacent actin filaments (fig4.1).

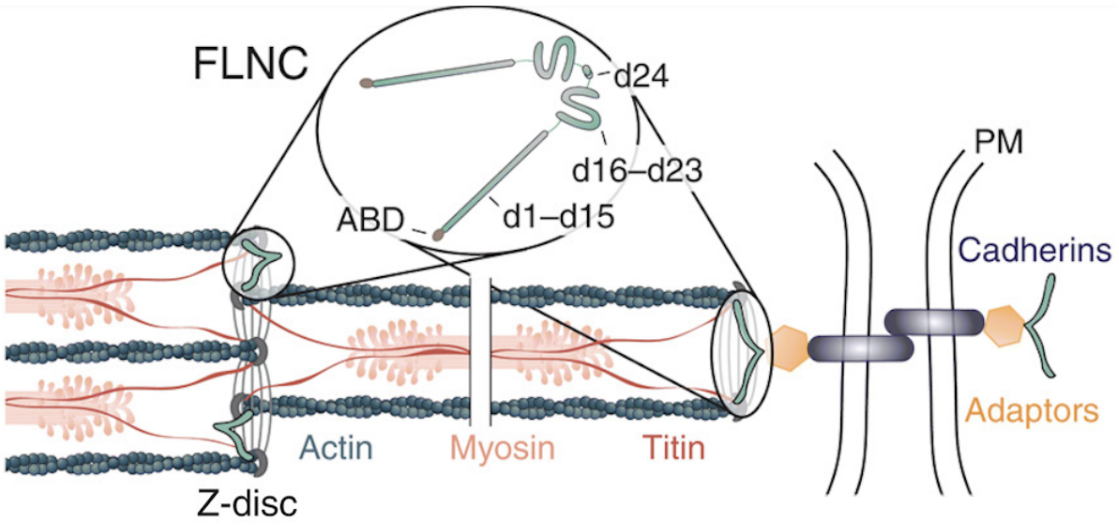


Figure 4.1: FLNC serves as a connector of actin filaments, contributing to the formation of cytoskeletal and cardiac muscle fibres. The FLNC protein is composed of two identical monomers, each containing an N-terminal Actin-binding domain (ABD) and 24 immunoglobulin-like domains (d1-d24), which are organised into three sections separated by hinge regions located between d15-d16 and d23-24[237, 91]. Each of the two N-terminal ABDs interacts with an actin molecule, while two d24 domains dimerise to link the actin molecules associated with the ABDs.

FLNCd24, the homodimerisation site of FLNC, is comprised of two β sheets, respectively containing four strands (A, B, D, and E) and three strands (C, F, and F'/G) respectively, and its conformation resembles the typical structure of sHSP ACD (fig4.2A). The dimerisation interface is predominantly formed by the antiparallel overlap of strands C and D(fig4.2B). As displayed in fig4.2C, FLNCd24, as well as the entire protein, continuously transitions between two ultimate states: in a completely relaxed state, two FLNCd24 monomers are dimerized, and the connecting regions within FLNC remain flexible; while the complex is subjected to shear forces aligned with the direction of the actin filaments, the FLNCd24 dimer experiences partial deformation, which is accompanied by the elongation of the flexible regions; when the shear force surpasses the limit, the FLNCd24 dimer dissociates, resulting in the release of the actin filaments. This reversible deformation, along with the subsequent detachment or recombination of the actin filaments, facilitates the contraction and relaxation of striated muscle, thereby providing resistance to external forces.

4.2.2 HspB7 is a predominantly monomeric sHSP

HspB7 is a cardiovascular sHSP

As the only sHSP expressed predominantly in cardiac muscle tissue[41, 240], HspB7 was initially named as cvHSP, an abbreviation for cardiovascular heat shock protein. One of the primary functions of sHSPs is to inhibit the aggregation of the polyglutamine regions of various proteins[241, 242], which can lead to neurodegenerative disorders such as Huntington's disease[243, 244] and spinocerebellar ataxia[245, 246]. Among all sHSPs, HspB7 has been identified as potentially the most effective suppressor of polyglutamine aggregation[247].

Although HspB7 possesses structural characteristics that are common to other sHSPs, it is unable to self-assemble in any form[247, 248], or to prevent the heat-induced deformation or aggregation of target proteins[195]. However, mutations, truncations, or knockout of HspB7 result in the aggregation of FLNC, which leads to lethal heart failure in murine models[87, 249]. These observations suggest that

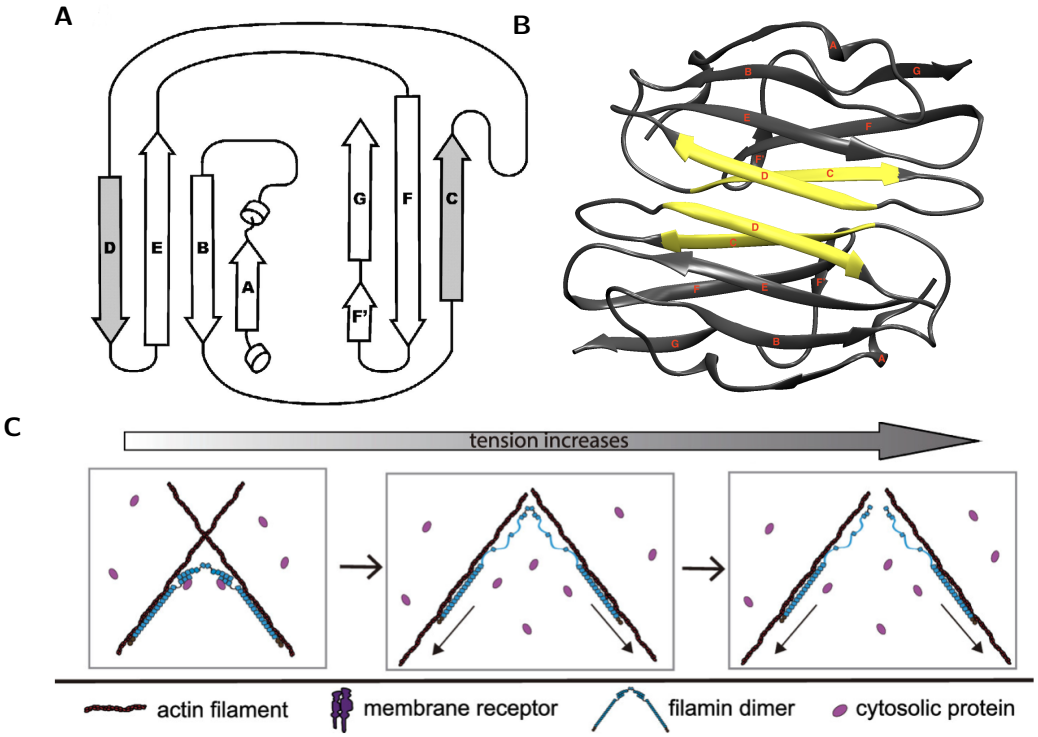


Figure 4.2: **A** A schematic representation of the secondary structure of FLNCd24, adapted from Sellung et al[238]. In FLNCd24, the strands are labelled from A to G, with two short helices located on both sides of strand A, and a short motif, F', situated between strands F and G. The structure contains two β sheets: one formed by strands A, B, Ds and E, and the other consisting of strands C, F, (F',) and G. The grey-filled strands, C and D, play a crucial role in the dimerisation of FLNCd24. **B** The crystal structure of the FLNCd24 dimer (PDB 1v05)[218]. The dimerisation interface is formed by strands C and D, hightlighted in yellow. **C** The deformation process of FLNC-actin complexes under mechanical stress, adapted from Deng et al[239].

HspB7, which is likely a non-promiscuous molecular chaperone, interacts with FLNC to inhibit its aggregation.

The unique property and function of HspB7 is related to its sequence and structure

We utilised the well-studied sHSPs HspB1 and B5 as references for the sequence and structural analysis of HspB7, by aligning their sequences and crystal structures of the three proteins. The sequence alignment reveals that the composition of HspB7 is significantly distinct from those of the other two proteins, particularly in the NTD and CTD regions (fig4.3). An evident characteristic of the HspB7 sequence is the polyserine motif located in the NTD region, which comprises 13 serines within a 15-amino acid motif. This motif is critical for the suppression of

polyglutamine aggregation, and also leads to the lower molecular mass of HspB7 (18.6 kDa) in comparison to HspB1 (22.8 kDa) and B5 (20.1 kDa). Another factor contributing to the reduced mass of HspB7 is the significantly shorter sequence of HspB7, particularly in the CTD. In the truncated HspB7-CTD, the I-X-I motif is positioned at the terminal end of the protein, in contrast to its placement in the middle of the CTD regions of HspB1 and HspB5. Furthermore, the X residue in the I-X-I motif of HspB7 is a lysine, which diminishes the hydrophobicity of this motif. Generally, residues located at the terminal end of a peptide chain exhibit a reduced propensity to engage in stable interactions with other residues due to their relative freedom from constraints imposed by both covalent and non-covalent interactions. Consequently, the short CTD peptide featuring the I-K-I motif at the C-terminus is unlikely to interact with the homogeneous β4-β8 pocket as robustly as the ACD-CTD interactions observed in HspB1 or B5.

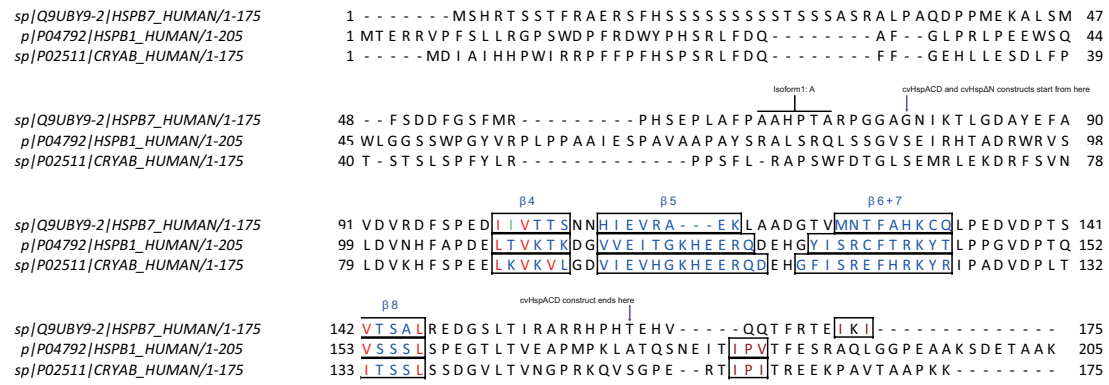


Figure 4.3: The sequence alignment of HspB1, B5 and B7, with the secondary structures of the β4-β8 regions labelled.

In the context of the ACD regions, a notable sequence variation between HspB7 and the other two sHSPs is the absence of a three-residue motif in the β5 strand. Despite involving fewer residues than the differences observed in the NTD and CTD regions, this variation exerts a remarkable impact functional activities of HspB7, as it results in a shorter β5 strand, in turn leading to a correspondingly shorter β6+7 strand, which is the dimerisation site of sHSPs, anti-parallelly engaged to β5 (fig4.4). Given that if two HspB7 ACD monomers are capable to dimerise, the dimerisation

interface formed by the two β6+7 strands is considerably smaller than the interfaces of HspB1 or B5 ACD dimers, indicating that the affinity for dimerisation of HspB7 ACD is less pronounced compared to the other two sHSPs. Due to the dual disadvantage posed by the self-oligomerisation of HspB7, we hypothesised that HspB7 monomers may be incapable of forming stable oligomers.

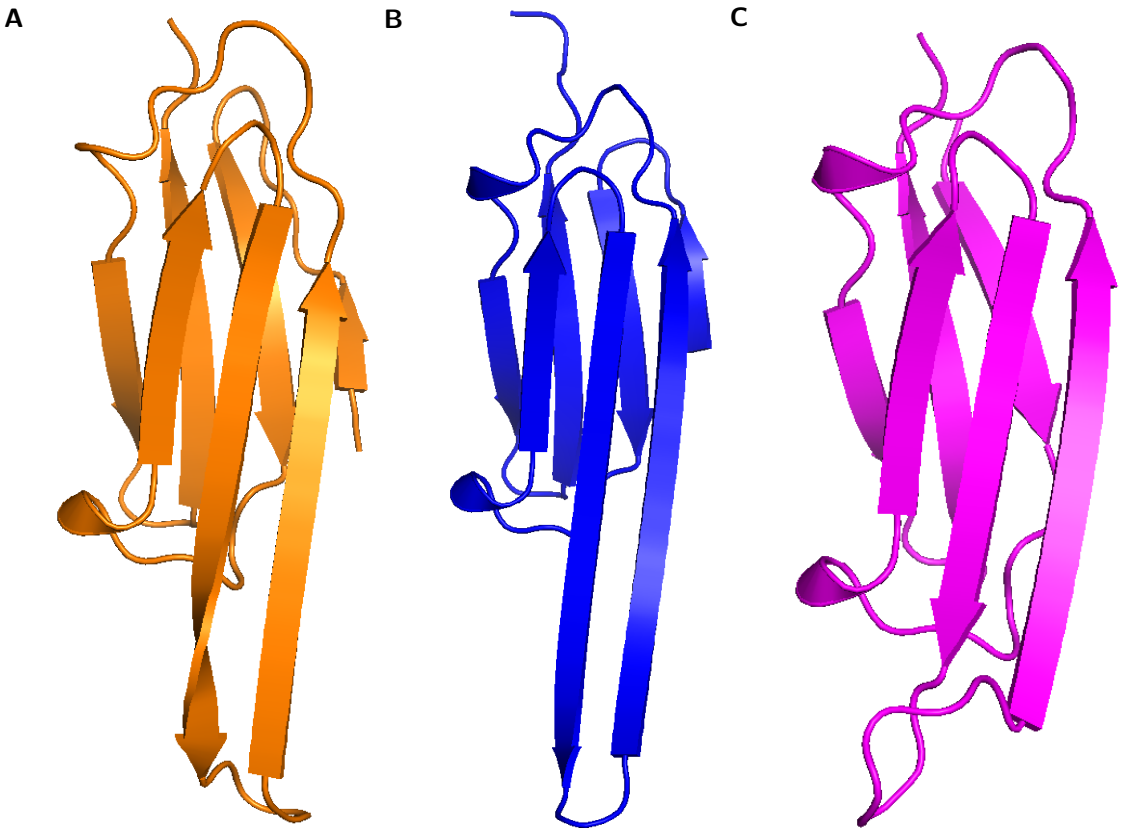


Figure 4.4: **A** The conformation of HspB1 ACD extracted from the crystal structure of the AP_{II} HspB1 ACD-CTD dimer (PDB 4MJH)[207]. **B** The conformation of HspB5 ACD obtained from the crystal structure of the AP_{II} HspB5 ACD dimer (PDB 4M5S)[207]. **C** The conformation of HspB7 ACD in the FLNCd24-HspB7ACD complex determined through crystallography[86]. The structures are aligned by minimising their mutual RMSD. Notably, the β5 and β6+7 strands of HspB7 are significantly shorter than the corresponding strands in HspB1 and B5, impeding the potential of HspB7 to form a stable dimer interface with its β6+7 strand.

4.2.3 Residue phosphorylations modify the function of FLNCd24.

HspB7 ACD binds with the dimerisation site of FLNCd24, competing with FLNCd24 self-dimerisation

HspB7 serves as a functional regulator of FLNC within striated muscle tissue[86, 87]. The expression of HspB7 has been observed to be relevant for the morphogenesis

of cardiac muscle fibre[250] and the development of heart tissue[249], as a low level of HspB7 expression can result in abnormal aggregation of FLNC in mouse heart tissue[251], which contributes to the risk of cardiomyopathy[252]. Previous research have revealed through *in vitro* experiments that HspB7 specifically binds to the FLNC-C section (d19-d24) but not to FLNC-CΔ24 (d19-d23), indicating that FLNCd24 is the sole binding site for HspB7 in the entire FLNC structure[86, 87].

Phosphorylations of residues in FLNCd24 influence the interactions between FLNCd24 and HspB7ACD, as well as the mechano-sensitivity of FLNC.

The phosphorylation of amino acids is a common post-traslation modification (PTM) to regulate the function and behaviour of numerous proteins[253, 254, 255, 256]. In muscle tissue, the phosphorylation of amino acids is frequently associated with reversible conformational transformations of muscle proteins, which are essential for the contraction-relaxation cycle[257, 258]. However, unscheduled phosphorylation of critical amino acids in muscle proteins is a significant contributor to muscle-related diseases and disorders[259, 260]. For example, a study examining the interaction between HspB1 and the multi-domain segment FLNCd18-21 has revealed that the mechano-induced deformation of the multi-domain FLNC region is regulated by the phosphorylation of serines in HspB1[91].

Within the dimerisation interface of FLNCd24, two phosphorylation sites have been identified: T2677, whose phosphorylation is observed in cases of dilated cardiomyopathy[261] and cardiac complications associated with COVID-19 infection[262, 263]; and Y2683, which is implicated in exercise-induced signalling[264, 265] and can be phosphorylated in tumour tissues[266].

In *in vivo* experiments, protein phosphorylation is facilitated by specific kinases, which may exhibit selectivity towards target peptides and necessitate stringent conditions for its optimal activity. Phosphorylation can occur at multiple sites on a protein, complicating the isolation of proteins in the desired phosphorylated state for subsequent analysis. Furthermore, protein phosphorylation is a reversible process, and a phosphorylated protein is unlikely to maintain a stable conformation

during examination with experimental instruments. Consequently, in *in vivo* studies, phosphomimetics are frequently employed as a stable and practical approach to investigate protein phosphorylation[267, 268]. The fundamental concept behind phosphomimetics involves substituting a phosphorylated residue with a negatively charged amino acid that possesses a mass and stoichiometric effect comparable to that of the phosphorylated amino acid. In the experiments conducted in this project, including native-MS and crystallography, pT2677 is substituted with aspartic acid (T2677D), and pY2683 is substituted with glutamic acid (Y2683E).

4.2.4 Plan of experiments and analysis

To ascertain that HspB7 is incapable of forminf a stable homodimer via ACD-ACD interactions, we implemented MD simulations of both the HspB7 ACD monomer and dimer, with their structures elucidated through X-ray crystallography. Additionally, we performed simulations of HspB1 and B5 monomers as controls.

In order to analyse the regulatory effects of phosphorylation at T2677 and Y2683 on the dimerisation of FLNCd24 and its interaction with HspB7, we implemented MD simulations of wild-type (WT), phosphorylated, and phosphomimicking FLNCd24 homodimers, as well as FLNCd24-HspB7 ACD complexes. Based on the trajectories generated from these MD simulations, we assessed the stability and interaction intensity of the interfaces, and compared the inter-residue interactions at iso-positions between the WT and modified structures to understand the regulatory impact of phosphorylation. Additionally, we investigated the relationship between phosphorylation and phosphomimicking mutations to validate the reliability of the results obtained from laboratory experiments.

4.3 Methods

4.3.1 Preparation of initial structures

The structures of the HspB1 and B5 ACD-CTD homodimers are obtained from the Protein Data Bank (HspB1: pdb 4MJH; HspB5: pdb 4M5S)[207], while the monomeric structures of HspB1 and B5 ACD-CTD are extracted from the

corresponding dimeric structures. The structures of the WT FLNCd24 homodimer and the FLNCd24-HspB7ACD complex are determined using X-ray crystallography. Although HspB7 ACD cannot form a stable dimer, we successfully obtained a dimeric structure of HspB7 ACD, as well as a monomeric structure, from a FLNCd24-HspB7ACD multi-oligomer through X-ray crystallography. In laboratory experiments, Cys131 of HspB7 is mutated to serine to prevent the formation of a disulphide bond; this mutation is retained in the protein structure prepared for MD simulations, as the distance from C/S131 to the dimer interface is sufficiently large to avoid interference with any intermolecular interactions. Residue modifications, including the regulation of the protonation state of histidine, T-D and Y-E mutations, and the phosphorylations of threonine and tyrosine, are performed by AmberTools[127] and VMD[208]. All protein molecules are described by the FF14SB force field[128], with the exception of phosphorylated amino acids, which are parametrised by the phosphorylation parameters from the FF19SB force field[104].

4.3.2 MD simulations

The MD simulation process is generally consistent with the procedures outlined in Chapter2.2.3 and Chapter3.3.2. However, that the minimisation steps for the modified FLNCd24 dimers and FLNCd24-HspB7ACD complexes are extended to 10,000 steps to facilitate a more reliable relaxation of the side chains of the modified residues. The production simulations for sHSP monomers and dimers are conducted in five repetitions, each with a duration of 100 ns. Similarly, the production simulations for FLNCd24 dimers and FLNCd24-HspB7ACD complexes are also executed in five repetitions, with each repetition lasting 200 ns.

4.3.3 Preprocessing and data analysis

Preprocessing MD trajectories and numeric quantification: RMSD and RMSF

The preprocessing steps employed are highly consistent with those outlined in Chapter2.2.2. Prior to plotting and comparing the values of the residue RMSD and RMSF, we aligned the sequences of HspB1, B5, and B7 ACD using the BLAST

protein alignment online server[206], to identify the missing residues in the β6+7 strand of HspB7. In the resulting residue RMSD and RMSF plots, the region absent in HspB7 is represented as a vacant space.

Binding energy calculation - mmGBSA

For the computation of binding energies between two FLNCd24 molecules or between FLNCd24 and the ACD of HspB7, mmGBSA calculation is conducted by executing the Python script mmPBSA.py[137]. To investigate the differences in binding energies among the various structures, among WT and modified structures, as well as among the replicates, we performed t-tests on the values obtained from the mmGBSA calculations.

Inter-residue distance calculation

e inter-residue distances are computed invoking the MDAnalysis package in Python[211, 212], as demonstrated in Chapter2.2.3 and Chapter3.3.2. The residue-subunit distance between a residue in chain A and the full-length chain B is defined as the minimum distance between the specified residue in chain A and any residue in chain B.

Hydrogen bond analysis

The analysis of hydrogen bonds is conducted using the MDAnalysis package in Python[211, 212], following the methodology outlined in Chapter3.3.3. To quantitatively assess the accuracy of phosphomimetics, we executed the SciPy and scikit-learn packages in Python to perform the Wilcoxon signed-rank test on the distributions of hydrogen bond frequencies within the complexes [269]. The Wilcoxon algorithm is selected to examine the correlation between the hydrogen bond distributions in phosphorylated and phosphomimicking complexes, as this algorithm is designed for hypothesis testing involving two related data sets. The visualization of data analysis is primarily accomplished through scripts invoking the Matplotlib package in Python[210].

4.4 Results and discussions

4.4.1 HspB7 is dominantly monomeric, unlike most other sHSPs

The ACD-ACD and ACD-CTD interactions within HspB7 were investigated by native-MS, accomplished by mixing ACD with CTD peptide (residue 168-175, TFRTEIKI) in a 1:1 molar ratio and quantifying the concentration of each complexe in the solution, as performed by Z. Wang. Control experiments for HspB1 and B5 were performed under identical conditions. The results of the native-MS analysis (see fig4.5) indicate that in the mixture of HspB7 ACD and the CTD peptide, the ACD monomer is the predominant peak observed in the spectrum, with only minor peaks corresponding to ACD-CTD complexes. In contrast, the spectra for HspB1 and HspB5 reveal a variety of peaks associated with ACD-CTD, ACD-ACD, ACD-ACD-CTD, and CTD-ACD-ACD-CTD complexes, all of which are present in significant proportions. These native-MS spectra provide support for our hypothesis that HspB7 is unlikely to form oligomeric structures.

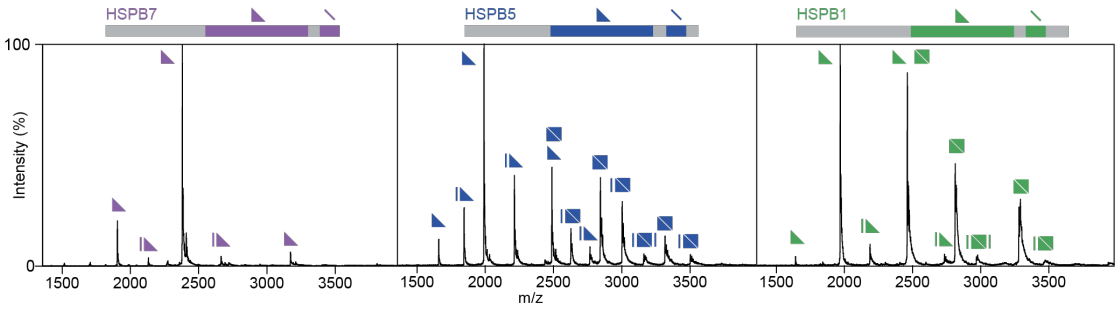


Figure 4.5: Results of Native MS experiment of 1:1 mixed HspB7 ACD and CTD, alongside controlling of HspB1 and B5. The spectra for HspB7 predominantly exhibit free ACD molecules, with trace amounts of ACD-CTD complexes. In contrast, the spectra for HspB1 and HspB5 display multiple peaks corresponding to ACD-CTD and ACD-ACD complexes, as well as complexes composed of ACD-ACD and/or ACD-CTD units. Image prepared by Z. Wang.

To identify the reason that HspB7 is not inclined to form oligomers, we implemented simulations of HspB1, B5, and B7 ACD monomers, each lasting 500 ns. The residue RMSD and RMSF of the monomers are aligned according to the alignment of the sequence displayed in fig4.3 while being plotted, and the missing motif in the β5 strand of HspB7 is displayed as a three-residue gap between

residue 45-47. As showed in fig4.6A, within the monomeric structures of all sHSPs, residues in HspB5 that correspond to the missing motif in HspB7 exhibit the highest RMSD ($\sim$ 13 Å) and RMSF ($\sim$ 4 Å) among all residues, suggesting that the motif is not stabilised when the HspB5 ACD is not bound to any interacting partners. In contrast, the residues in the same region of HspB1 display residue RMSD of approximately 6 Å, which are significantly higher than those of any other residue in HspB1 and B7, including the residues adjacent to the gap in HspB7, D120 and G121($\sim$ 4 Å); however, their RMSF values are not particularly high ($\sim$ 2 Å). This observation implies that the motif adopts different conformations when the HspB1 ACD is either free or dimerised. Therefore, we proposed that the motif present in both β5 strands of the HspB1 and HspB5 ACDs contributes to the stabilisation of the conformation of ACDs in sHSP dimers and oligomers, while the absence of this motif in the HspB7 ACD impedes the oligomerisation of HspB7.

In the analysis of the trajectories generated from the dimeric structures of sHSPs, a notable distinction can be observed between the ACD-ACD interfaces of HspB7 and those of HspB1 or B5 ACD dimers. In the homodimeric structure of the HspB7 ACD (fig4.7A), the two β3 strands converge, a feature not observed in the dimeric structures of HspB1 (fig4.7B) or B5 ACD. The close proximity of these strands suggests that residues in this region may establish an additional interface, which is necessary to compensate for the weak binding affinity between the two short β6+7 strands. The interaction between β3 strands is facilitated by the absence of the β2 strand, and the steric effect posed by the shorter β5 and β6+7 strands is less pronounced compared to that in the ACDs of HspB1 or B5.

We calculated the binding affinity of each sHSP ACD dimer by performing the mmGBSA algorithm (table4.1. In the three sHSP ACD dimers, HspB7 is the only one that has a positive total binding energy, suggesting that the dimerisation of HspB7 ACD is not spontaneous. This clearly reveals that the interactions between the two HspB7 ACDs cannot stabilise the interface, despite the extra contacts between the β3 strands.

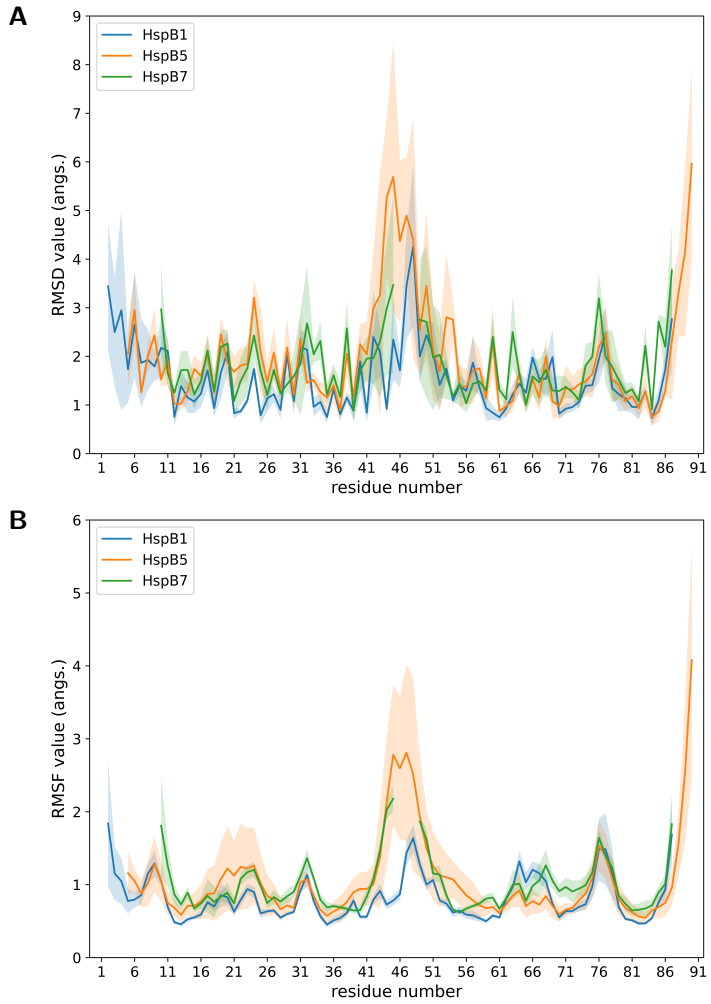


Figure 4.6: **A** RMSD of residues in HspB1 (blue), B5 (orange), and B7 (green) monomers. The curves for HspB7 indicate a gap with null values between residues 45 and 47, which corresponds to the absence of the three-residue motif in the β5 strand. **B** RMSF of residues in HspB1, B5, and B7 monomers.

ACD-ACD	HspB1	HspB5	HspB7
Total	-11.7148	-1.9410	19.1190
Standard deviation	13.9216	8.7464	11.8556

Table 4.1: Binding energies of the sHSP ACD dimers

4.4.2 Phosphorylations of T2677 and Y2683 regulate the protonation state of H2673.

Phosphorylations and phosphomimetic mutations introduce negative charges to the local area within proteins, resulting in a significant Coulombic effect in these confined areas, which subsequently influences neighbouring charged and polar residues. The

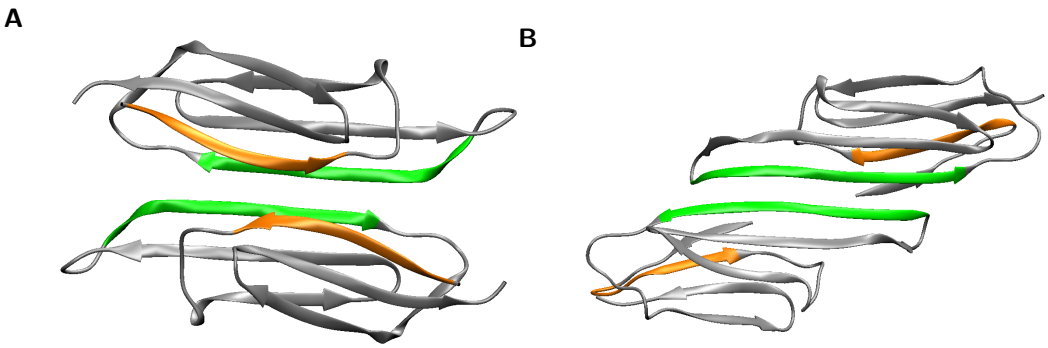


Figure 4.7: **A** Conformation of HspB7 ACD dimer extracted from a tetramer consisting of two FLNCd24 and two HspB7 ACDs. Aside of the interface comprised of β6+7 strands (green), the two β3 strands (orange) also approach to each other, possibly forming another interface. **B** Conformation of HspB1 ACD dimer, collected from PDB: 4MJH. Residues in β3 and β6+7 strands are coloured in the same pattern as those in HspB7 dimer.

pK_{a_2} value of a phosphate group associated with an amino acid is approximately 5.8[270], which is lower than the pH of the system employed in MD simulations that aim to replicate the *in vivo* environment (pH=7) and the pK_a of the histidine side chain (pH=6)[271]. Therefore, we hypothesised that the phosphate group associated with an amino acid may carry a net negative charge of two units, and that the side chain of a histidine adjacent to the phosphate group may be protonated.

While modifying T2677 and Y2683, we assessed the distances from each of these residues to all histidine in the structure, subsequently selecting those histidines that could potentially be influenced by the modifications. Two histidines are in proximity to the modification sites: FLNCd24 H2673, which is adjacent to both T2677 and Y2683 within the FLNCd24 dimer and the FLNCd24-HspB7ACD complex; and HspB7 H109, which is exclusively found in the heterogeneous complex and is situated near T2677 only.

We implemented MD simulations of the pT2677 and pY2683 FLNCd24-HspB7ACD complexes, incorporating protonated H2673 (HIP2673) or both histidines, H2673 and H109, in their protonated forms (HIP2673 and HIP109). The stability of the protonated histidines is assessed through analysis of their RMSD and RMSF (fig4.8). The RMSD values for the two residues, irrespective of their protonation state, did not exhibit significant differences. However, the RMSF values for both protonated histidines in the doubly protonated complexes are notably higher than

those of the corresponding residues in the HIP2673 complexes. Taking into account the proximity to the phosphorylation site and the structural stability, we opted to protonate H2673 while excluding protonation of H109 in the initial structure for the production MD simulations. To maintain consistency across the simulations, H2673 remains protonated in all simulations of the FLNCd24-HspB7ACD complex conformations discussed in the subsequent sections of this chapter.

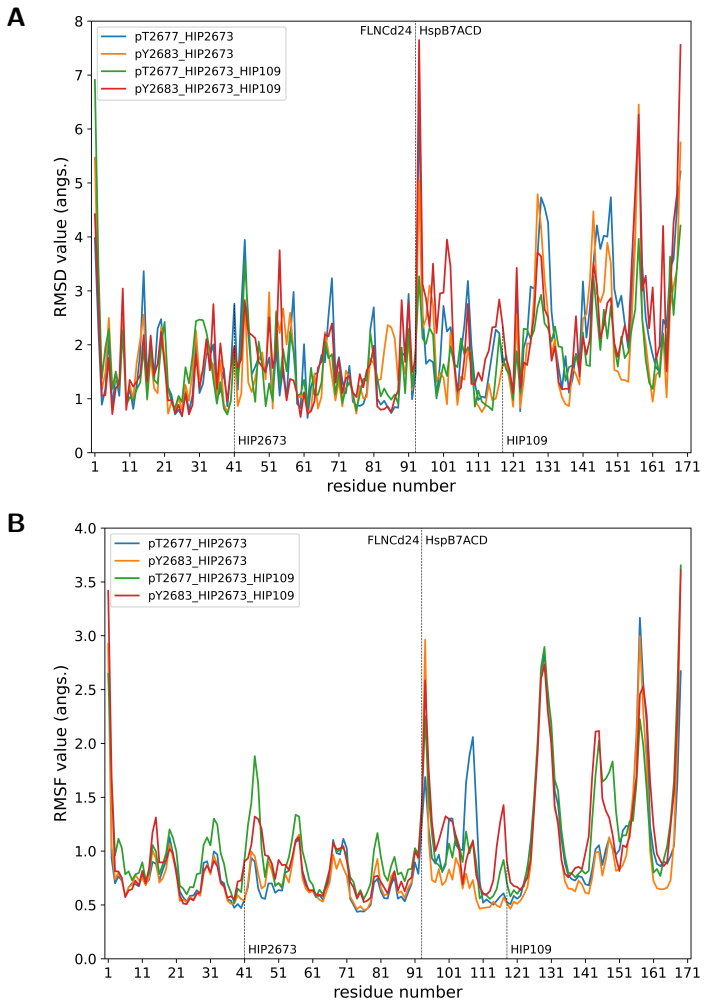


Figure 4.8: Residue RMSD (**A**) and RMSF (**B**) of phosphorylated and protonated FLNCd24-HspB7ACD complexes. The positions of FLNCd24 HIP2673 and HspB7 HIS/HIP109 are labelled.

4.4.3 The competition between FLNCd24-HspB7 interaction and FLNCd24 dimerisation.

Through native-MS experiments, my colleague elucidated the interaction between FLNCd24 and HspB7, demonstrating that FLNCd24 specifically binds to the ACD

region of HspB7 (fig4.9A). The structures of the FLNCd24 dimer and the FLNCd24-HspB7 ACD complex, as determined by crystallography (fig4.9B/C), indicate that the same interface in FLNCd24 is employed for both homo-dimerization and interaction with the HspB7 ACD, suggesting that these two types of interactions are mutually exclusive.

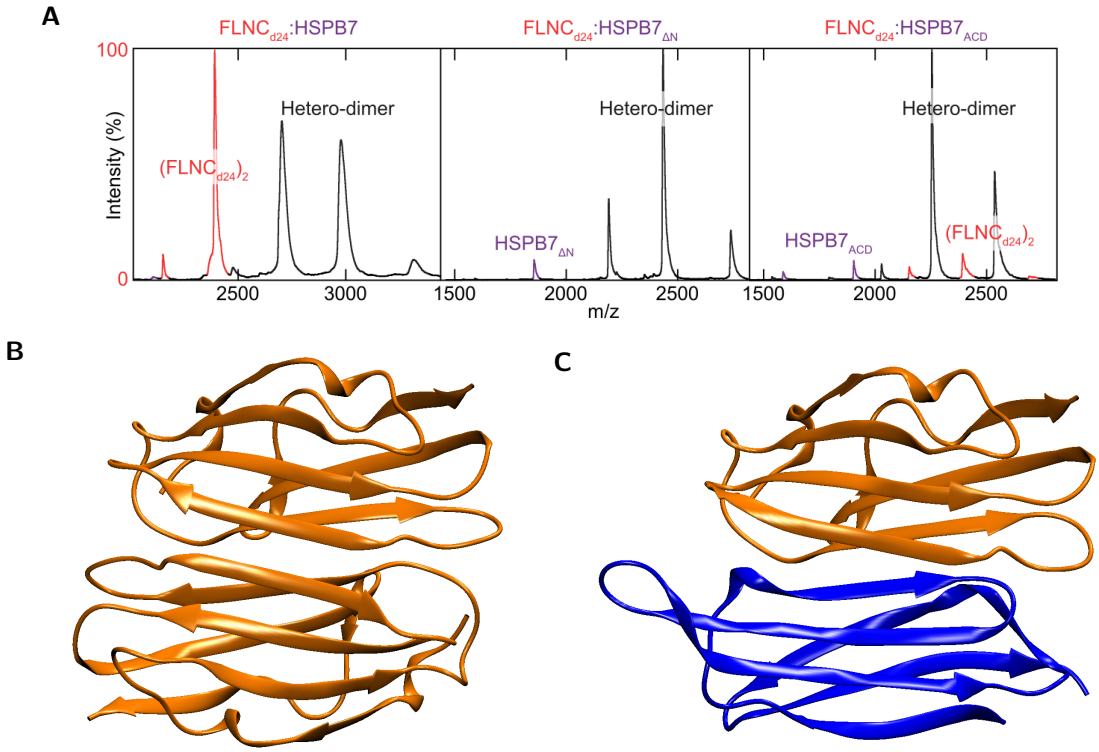


Figure 4.9: Experiments demonstrate the interaction between FLNCd24 and HspB7 ACD. **A** Native-MS data reveal a mixture of FLNCd24 and HspB7 at a constant concentration ratio of 1:1. The consistent presence of FLNCd24-HspB7 peaks suggests that the ACD region of HspB7 functions as the interaction site, and the variations in the FLNCd24 dimer peak across the three spectra imply that these interactions may be influenced by the presence of the NTD of HspB7. **B** The conformation of the FLNCd24 homodimer crystal structure. **C** The conformation of the FLNCd24-HspB7ACD complex crystal structure. In both crystal structures, FLNCd24 is represented in orange, HspB7 in purple, and the two structures are aligned based on the coordinates of one FLNCd24 monomer. The interaction with HspB7 and the homo-dimerisation occur at the same interface of FLNCd24, highlighting the exclusivity of these two types of interactions. Both crystal structures were resolved by Z. Wang.

We conducted production simulations for both bimolecular complexes and calculated the binding energy of each complex using the mmGBSA algorithm. As indicated in table4.2, the average binding energies of the two complexes are closely comparable. A t-test performed on the binding energies derived from individual

trajectories generated by repeated simulations revealed that the binding energies of the hetero-complex exhibit greater variability. However, it is noteworthy that none of the variations observed in any of the repeated simulations exceed 10%.

Energy (kcal/mol)	FLNCd24 dimer		FLNCd24-HspB7ACD	
	Mean	Std Dev	Mean	Std Dev
Overall	-55.9956	9.9278	-56.4983	7.8476

Repeat	1	2	3	4	5
Homodimer	↓↓[a]	=	=	=	↑↑
Hetero-complex	↓↓	↑↑	↑↑	↑	↓↓

Table 4.2: Binding energies of the WT FLNCd24 dimer and the WT FLNCd24-HspB7ACD complex, and p-values from t-tests for the variation of the repeating simulations

[a]The upward arrows indicate that the binding energy calculated from the repeat is greater than that derived from the conjugated trajectory, and vice versa. The number of arrows signifies the statistical difference between the two samples: one arrow denotes a p-value between 0.01 and 0.05; two arrows indicate a p-value less than 0.01, with the difference between the two mean values being less than 10% of the mean value of the reference sample; and three arrows correspond to a p-value less than 0.01, with the difference between the two mean values exceeding 10% of the mean value of the reference sample.

To identify the inter-molecular residue pairs that may engage in non-covalent interactions through simulations, and to quantify the intensities of these interactions, we measured the inter-residue distances between all intermolecular residue pairs within the two complexes, calculated the average values and filtered out those that are smaller than a preset cut-off, which is set at 4 Å for this analysis. As displayed in fig4.10, a greater number of residues in FLNCd24 participate in the formation of intermolecular interactions when the FLNCd24 monomer is bound to the HspB7 ACD. In the dimerised FLNCd24, the majority of inter-residue contacts are located in two regions: residue 2665-2674, primarily strand C, and residue 2678-2685, mainly strand D. The red bars representing these contacts form two symmetric zigzag lines along the diagonal, indicating that the interfaces are strictly anti-parallel and consist solely of strands C and D. In the hetero-complex, most residues between positions 2665 and 2692, which include strands C, D, and three loops, are situated adjacent to the β4 strand of HspB7 ACD (residues 98-109). Additionally, other

groups of close contacts are formed between the β8 strand of HspB7 ACD (residues 139-144) and strand F of FLNCd24 (residues 2707-2712).

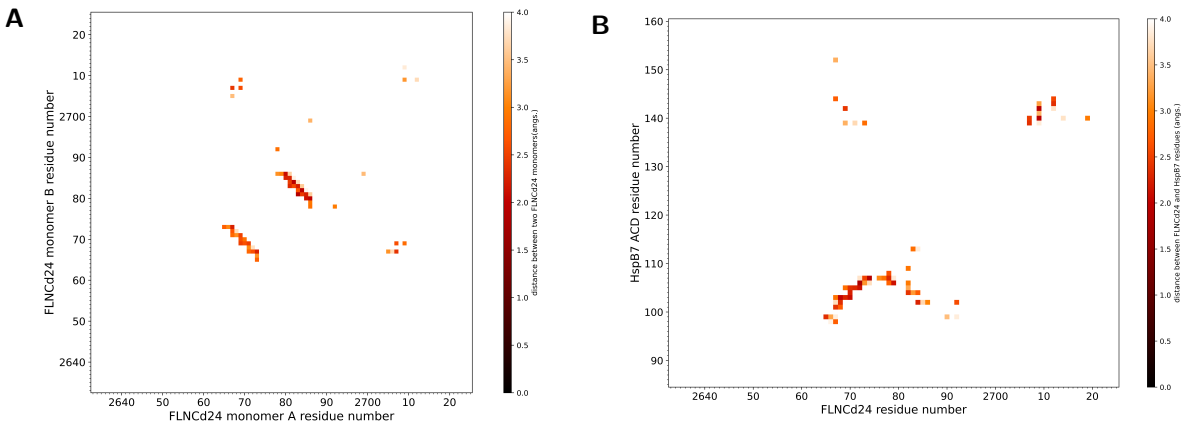


Figure 4.10: **A** Distances between intermolecular residue pairs in the FLNCd24 homodimer. The distance between two residues is defined as the minimum value from the set of distances of any inter-residue atom pairs. Distances exceeding 4 Å are excluded from the plotting. **B** Distances of intermolecular residue pairs in the FLNCd24-HspB7ACD complex.

Based on the inter-residue distances, we calculated the average and minimum residue-ligand distances between each residue within a FLNCd24 monomer and its interacting partner, which may be either another FLNCd24 monomer or the HspB7 ACD, in order to quantify the transitions of the binding interfaces (fig4.11). Throughout the simulations of both FLNCd24 homodimer and the hetero-complex, it is observed that the majority of residues within FLNCd24 that participate in intermolecular interactions are concentrated in two specific regions: residues 2665-2692 and 2705-2714. The former region, which includes strands C and D along with the adjacent loops, constitutes the primary components of the binding interface, whereas most residues in the latter region are found in strand F. However, the conformations of the two binding interfaces exhibit variability, as indicated by the average distances. In the homodimer, residues located in the loop between strands C and D or in strand F are less engaged in the formation of the interface, as evidenced by their average distances from the interactors being significantly greater than the corresponding minimum values. In contrast, in the hetero-complex, strand D appears to play a less critical role in the interactions for similar reasons. Furthermore, within the homodimer, two lysine residues, K2653 and K2699, which are not included in

either of the aforementioned regions, may approach the other FLNCd24 monomer within approximately 2 Å, despite not exhibiting any activity in the hetero-complex.

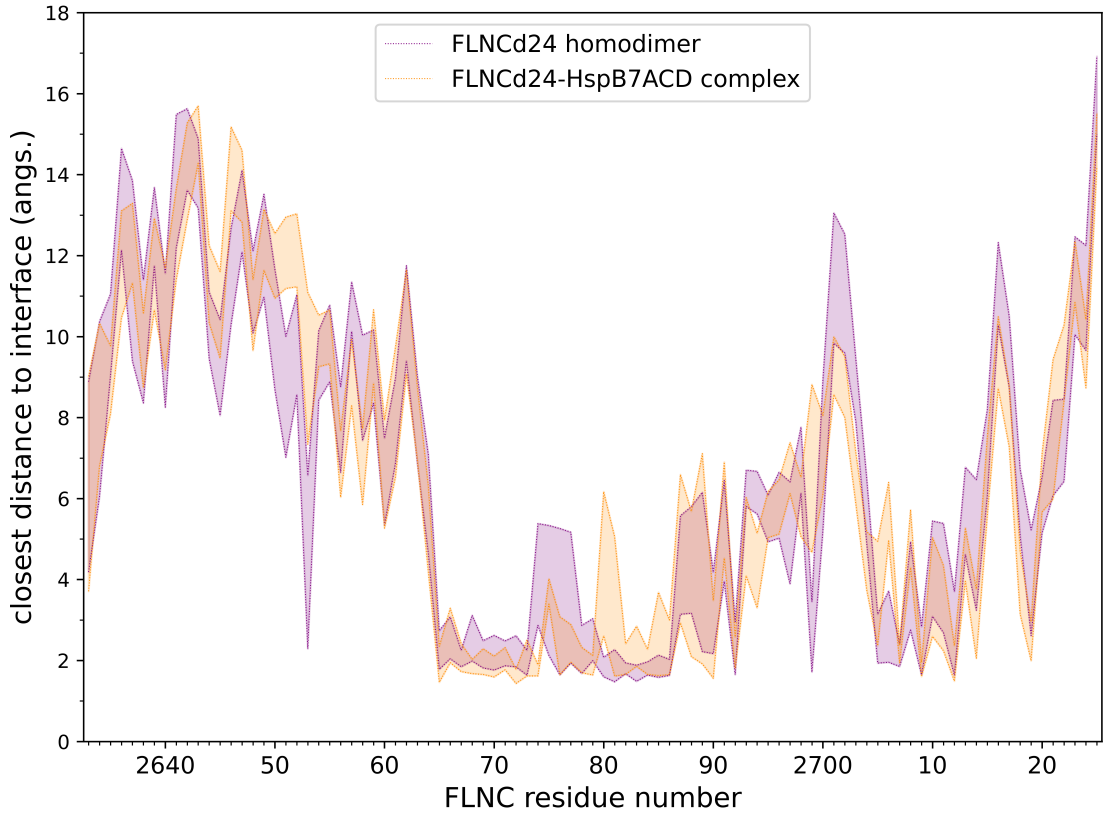


Figure 4.11: The distances between all residues in FLNCd24 and its interacting partners. The lower boundary of the single-colour shaded region represents the minimum distances observed during the simulation, while the upper boundary indicates the average distances.

Given that the interface of the complexes comprises multiple charged and polar residues, we hypothesised that hydrogen bonds play a significant role in mediating intermolecular interactions. In both instances, hydrogen bonds exert a notable influence on these interactions; consequently, we calculated the intermolecular hydrogen bonds formed in both complexes (fig4.12). In the FLNCd24 dimer, the hydrogen bond between E2681 and Y2683, both located in strand D, is the most frequently observed interaction, occurring at a rate of $70.70 \pm 3.87\%$, and is the only hydrogen bond detected in more than 50% of all frames. Additional hydrogen bonds include the interactions of K2685 with neighbouring residues E2680 and E2681($34.50 \pm 1.16\%$ and $18.8 \pm 0.51\%$), all located in strand D, an interaction

between two strands F, K2709-E2712 ($14.60 \pm 0.38\%$), as well as a hydrogen bond between H2686 and K2699 ($15.35 \pm 0.66\%$), formed between the loops.

In comparison to the FLNCd24 dimer, the hydrogen bond interactions within the hetero-complex are significantly more intense, as a greater number of residues in FLNCd24 are involved. Similar to the distribution of inter-residue contacts, the residues in FLNCd24 that contribute to the formation of hydrogen bonds are more widely dispersed. This includes interactions in strand C (V2672-N107, $91.6 \pm 1.21\%$ and G2674-N108, $62.7 \pm 1.61\%$), in the loop between strands D and E (R2690-E99/D100, $22.1 \pm 3.50\%/26.8 \pm 6.52\%$), and in strand F (K2709-T140/V142, $56.9 \pm 0.79\%/67.9 \pm 3.25\%$), and the cumulative occupancy of hydrogen bond within the complexes are notably high. Although hydrogen bonds associated with residues in strand D are also observed, they occur with relatively low probabilities (<20%), which supports our hypothesis that strand D is less intensely engaged in the interactions.

In accordance with the analysis of binding energy, inter-residue distances, and hydrogen bond distribution, the interaction between FLNCd24 and HspB7 ACD is more favorable compared to the FLNCd24 dimerisation. Furthermore, the structure of the heterogeneous complex exhibits increased rigidity.

4.4.4 The phosphorylation of T2677 and Y2683 results in distinct regulatory effects on the homo- and hetero-interactions

My colleague employed Native-MS to the T2677D or Y2683E mutated FLNCd24 homodimer and FLNCd24-HspB7ACD complex to investigate the impact of phosphorylation on homo- and hetero-interactions (fig4.13). In the solution of WT FLNCd24, the concentrations of FLNCd24 monomers and homodimers are both remarkable, while the dominant majority of FLNCd24 are dimerised following the occurrence of the pT2677-mimicking T2677D mutation, and a notable amount of proteins remain monomeric as a result of the pY2683-mimicking Y2683E mutations. The effects are retained in the solution of mixture of FLNCd24 and HspB7 ACD, as the T2677D mutation leads to an increased intensity of the FLNCd24 dimer peaks, along with

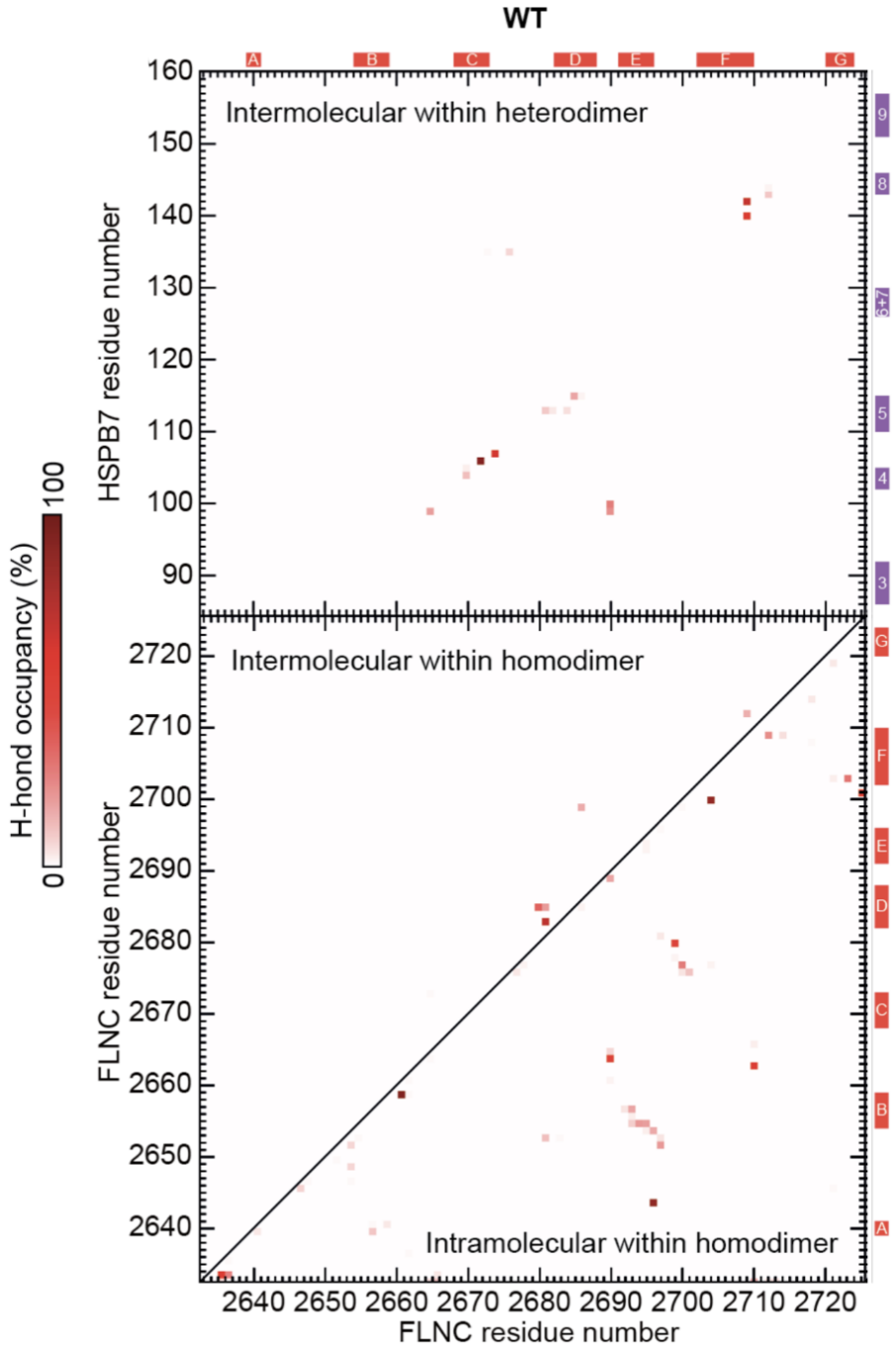


Figure 4.12: The probabilities of H-bond formed between residue pairs in the FLNCd24 dimer (upper) and between intermolecular residue pairs in the FLNCd24-HspB7ACD complex (lower). The probability of a hydrogen bond between a residue pair is defined as the ratio of the number of frames in which the residue pair can form a hydrogen bond to the total number of frames in the simulation. In the upper graph, which depicts homodimer hydrogen bond occupancies, the upper-left half represents intermolecular H-bonds and the lower-right half corresponds to intramolecular H-bonds.

multiple peaks corresponding to free HspB7 ACD, while Y2683E decreases the probabilities of forming FLNCd24 homodimers. These findings indicate that the

pT2677-mimicking T2677D mutation promotes the formation of FLNCd24 dimer and inhibiting the interactions between FLNCd24 and HspB7 ACD, while the pY2683-mimicking Y2683E mutation exhibits the opposite effects on both interactions.

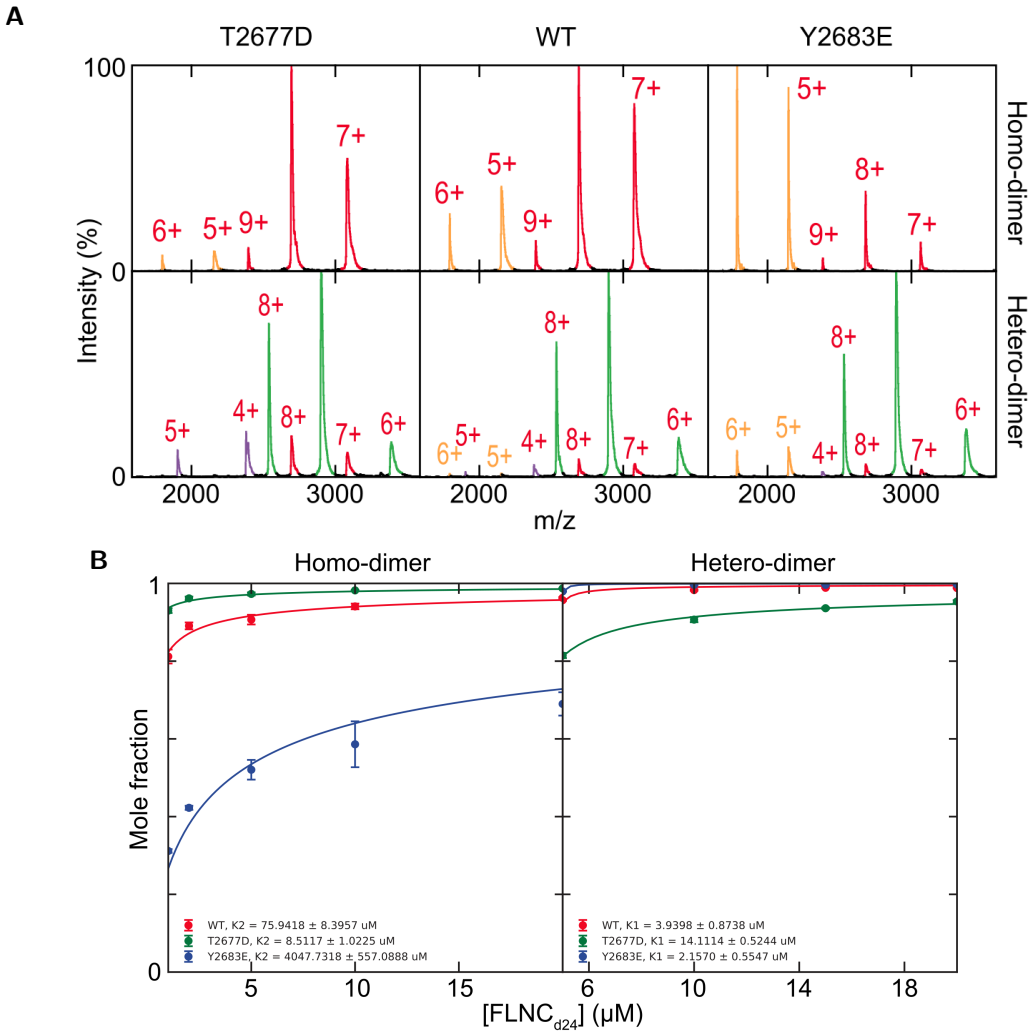


Figure 4.13: **A** Native-MS spectra of samples containing WT and mutated FLNCd24 and HspB7 ACD. The subfigures in the top respectively refers to the samples containing T2677D, WT, and Y2683E FLNCd24, and those in the bottom correspond to samples comprising mixture of HspB7 ACD and T2677D, WT, and Y2683E FLNCd24, all at a ratio of 1:1. The colour of peaks refer to the components in the mixture as follows: red: FLNCd24 homodimer; yellow: FLNCd24 monomer; purple: HspB7 ACD monomer; green: FLNCd24-HspB7ACD complex. **B** Native-MS data of different concentration of WT, T2677D and Y2683E FLNCd24 individually (left) and with 5 μM HspB7 ACD (right). In low FLNCd24 concentration solution, the impact of phosphomimicking mutation is more prominent, especially for the FLNCd24-HspB7ACD interaction.

Phosphorylation or T-D mutation of FLNCd24 T2677 strengthens the FLNCd24 dimerisation and weakens the FLNCd24-HspB7ACD interaction

To elucidate if the phosphorylations exert an impact on the dimerisation of FLNCd24, the interaction between FLNCd24 and the ACD of HspB7, or both interactions simultaneously, we conducted MD simulations of both phosphorylated and phosphomimicking mutated FLNCd24 dimers and FLNCd24-HspB7ACD complexes. The binding energies, calculated using the mmGBSA method (table4.3), indicate that phosphorylation or phosphomimetic mutation of T2677 enhances the interaction within the FLNCd24 homodimer, while concurrently diminishing the interaction between FLNCd24 and HspB7 ACD. Notably, the binding energies for both homo- and hetero-dimers of the T2677-mutated and phosphorylated structures are relatively comparable, thereby reinforcing the validity of our phosphomimetic approach.

Energy (kcal/mol)	Homodimer		Hetero-complex	
	pT2677	T2677D	pT2677	T2677D
Total	-71.8673	-69.7707	-48.1495	-47.6290
Std Dev	12.2803	9.8085	8.1400	11.4522
T-test result	↑↑↑[a]	↑↑↑	↓↓↓	↓↓↓
T-test between overall result and that from each repeat				
Repeat 1	↑↑↑	↑↑	↓↓	↓
Repeat 2	↓↓	↑↑↑	=	↑↑↑
Repeat 3	↓	↓↓	↑	↓↓↓
Repeat 4	↓↓	↓↓	=	↓↓
Repeat 5	↓↓↓	↓↓	↑	↑↑

Table 4.3: Binding energies of the pT2677 and T2677D FLNCd24 dimer and FLNCd24-HspB7ACD complex, p-values from t-tests between binding energies of WT and modified structures, and from t-tests for the variation of repeating simulations.

[a]The upward arrows indicate that the modified dimer exhibits more intense interactions compared to the reference (the corresponding WT dimer, in this case), and vice versa.

In accordance with the identical method utilised in the previous section, we computed the average closest distances between all residues within a FLNCd24 molecule and the binding interactors of the FLNCd24 chain, subsequently plotting the results for both homodimers and heterodimers separately in fig4.14. Upon phosphorylation of T2677, the residues at the interfaces exhibit differential responses:

in the homodimers, K2676 moves closer to the other monomer, whereas in the heterodimers, H2673 and G2674 depart from the HspB7 ACD. However, when T2677 is mutated to aspartic acid, the proximity of K2676 to the other FLNCd24 monomer becomes less pronounced, and the distances between H2673/G2674 and the HspB7 ACD in the hetero-complex do not significantly decrease compared to those in the WT complex. The values of most residues that are excluded from the interface in the pT2677 FLNCd24 variant are less distinguishable from the residues in the WT chain when compared to the T2677D residues. This suggests that the effect of phosphorylation at T2677 is confined to the interface residues, whereas the T-D mutation influences the conformation and interactions of residues in other regions. Consequently, the phosphomimetic T-D mutation does not fully replicate the effects of phosphorylation.

We quantified the hydrogen bonds present in the T2677 modified structures and assessed the probabilities of all residue pairs being linked by hydrogen bonds. To elucidate the structural differences induced by the T2677 modification, we computed the variations in hydrogen bond probabilities for corresponding residue pairs between the WT and the pT2677/T2677D complexes (fig4.15). In the pT2677 FLNCd24 homodimer, the primary alteration observed at the intermolecular interface is an increased frequency of the hydrogen bond between pT2677 and R2690 ($26.4 \pm 4.60\%$). Conversely, in the T2677D homodimer, in addition to this change, there is a heightened likelihood of E2680 forming a hydrogen bond with H2686 ($44.3 \pm 4.92\%$), as compared to the WT or pT2677 homodimer, which features K2685. Given that E2680, K2685, and H2686 are located at the periphery of the dimer interface, it is plausible that the T2677D mutation has resulted in a slight displacement of the homodimer interface, which is not observed in the pT2677 homodimer.

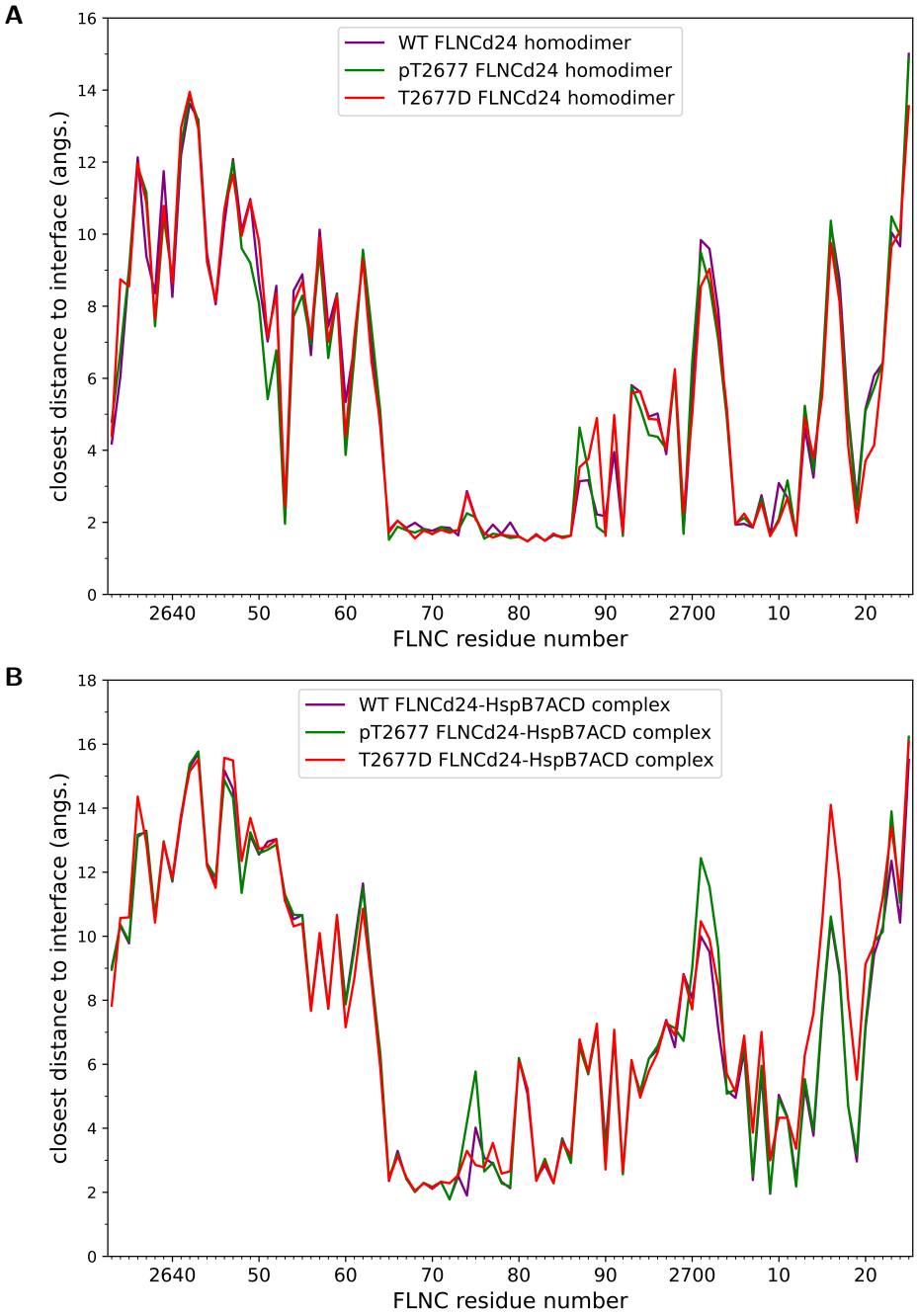


Figure 4.14: **A** The distances from each residue in one FLNCd24 chain to the other FLNCd24 molecule in the WT (purple), pT2677 (green), and T2677D (red) FLNCd24 homodimers. **B** The distances from each residue in the FLNCd24 chain to the HspB7 ACD molecule in the WT, pT2677, and T2677D FLNCd24-HspB7ACD complexes.

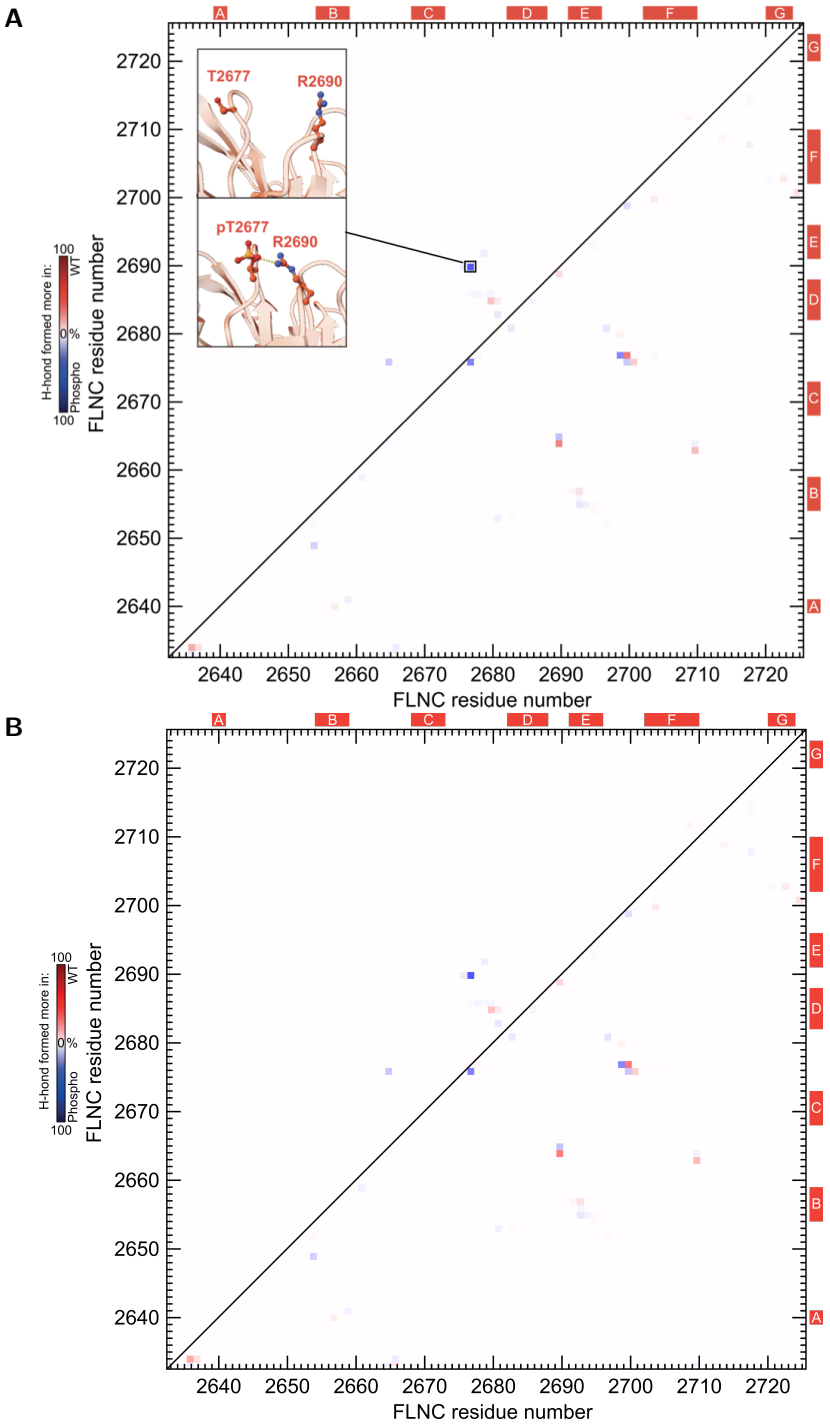


Figure 4.15: **A** The differences in hydrogen bond probabilities between intermolecular (upper left) and intramolecular (lower right) residue pairs in WT and pT2677 FLNCd24 homodimers. The difference in hydrogen bond probability for a given residue pair is determined by calculating the disparity between the H-bond probabilities in the WT and modified structures. A positive value (indicated in red) signifies that the probability in the WT structure is greater, while a negative value (indicated in blue) indicates the opposite. In the hydrogen bond probability difference map for the homodimer, the upper left represents intermolecular hydrogen bonds, whereas the lower right half pertains to intramolecular hydrogen bonds. The number of coloured bars corresponds to the number of residue pairs influenced by the respective residue modification. **B** The differences in hydrogen bond probabilities between inter- and intra-molecular residue pairs in WT and T2677D FLNCd24 homodimers.

We also assessed the differences in the probability of hydrogen bond formation between the WT and T2677 modified heterogeneous complexes(fig4.16). In the FLNCd24-HspB7ACD complexes, two hydrogen bonds are altered as a result of the phosphorylation of T2677: the hydrogen bond between G2674 and N107 is diminished, potentially due to the spatial effects and the negative charge of pT2677, while the hydrogen bond between D2712 and T143 is strengthened ($38.0 \pm 7.09\%$), possibly a consequence of the proximity of these two residues, which is related to the spatial effect of the phosphate group. Conversely, in the T2677D complex, a greater number of hydrogen bonds are impacted by the mutation, including the decline of the V2672-S106, K2709-T140, and K2709-V142 hydrogen bonds. These alterations are likely attributable to backbone hydrogen bonds resulting from structural changes rather than the Coulomb effect.

The phosphorylation or Y-E mutation of FLNCd24 at Y2683, conversely, diminishes the dimersation of FLNCd24 while enhancing the interaction between FLNCd24 and HspB7ACD

Taking into account the energy differences between the WT and T2677 modified conformations, which range from approximately 10~20 kcal/mol, it is observed that the binding energies of the Y2683 modified FLNCd24-HspB7ACD complexes are significantly closer to those of the WT complex, approximately 3~5 kcal/mol(table4.4). However, the binding energy of the pY2683 homodimer is notably low; conversely, the Y2683E mutant exhibits greater stability, with a binding energy that is less than 10 kcal/mol lower than that of the WT homodimer. Furthermore, the standard deviation of the pY2683 complex is considerably higher than that of the other structures, along with a remarkable variation across repeated simulations. Based on the positive electrostatic contribution to the solvation free energy of this complex, we hypothesised that the negative charges introduced by the phosphorylation of Y2683 lead to considerable electrostatic repulsion, thereby hindering the formation of a stable binding interface.

Unlike the T2677 modified structures, the complexes modified at Y2683 do not exhibit significant differences from the WT complexes regarding the distances

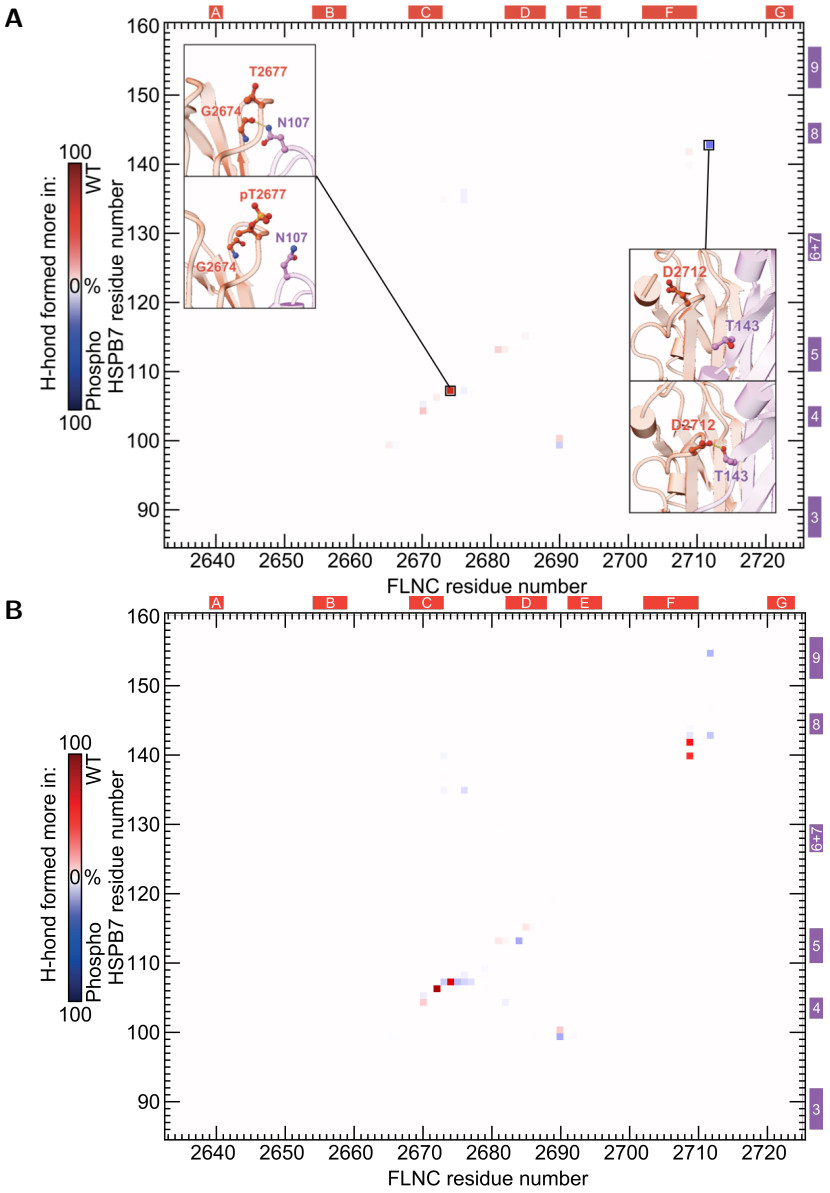


Figure 4.16: **A** The differences in hydrogen bond probabilities between intermolecular residue pairs in WT and pT2677 FLNCd24-HspB7ACD complexes. **B** The differences in hydrogen bond probabilities between intermolecular residue pairs in WT and T2677D FLNCd24-HspB7ACD complexes.

between the segments (see fig4.17. The most notable distinction is that the G2688-N2689-R2690 motif in pY2683 FLNCd24 is positioned closer to the HspB7 ACD compared to the WT FLNCd24, while the phosphomimetic Y2683E mutation does not influence the positioning of that motif.

In the pY2683 FLNCd24 homodimer (fig4.18A), only one intermolecular residue pair, E2681-pY2683, loses its hydrogen bond connection. This bond is characterised

Energy (kcal/mol)	Homodimer		Hetero-complex	
	pY2683	Y2683E	pY2683	Y2683E
Total	-2.0867	-48.2426	-62.8252	-59.5157
Std Dev	22.1104	10.1663	9.4639	11.9639
T-test result	↓↓↓	↓↓↓	↑↑↑	↑↑
T-test between overall result and that from each repeat				
Repeat 1	↑↑↑	↓↓↓	=	↑
Repeat 2	↓↓↓	↑	↑↑↑	↓↓↓
Repeat 3	↑↑↑	=	=	↓↓
Repeat 4	↓↓↓	↑↑	=	↓↓↓
Repeat 5	↓↓↓	=	↓↓↓	↓↓

Table 4.4: Binding energies of the pY2683 and Y2683E FLNCd24 dimer and FLNCd24-HspB7ACD complex, p-values from t-tests between binding energies of WT and modified structures, and from t-tests for the variation of repeating simulations.

as an significant interaction in the WT homodimer, arising from the association between positively and negatively charged residues. In contrast, pY2683 establishes an intramolecular hydrogen bond with K2653 ($35.7 \pm 9.12\%$); however, this bond is less intense than the salt bridge in the WT homodimer. Furthermore, the phosphorylation of Y2683 introduces two negative charges in proximity to E2681, resulting in substantial electrostatic repulsion. Consequently, the conformational transition caused by the phosphorylation of Y2683 significantly hinders the dimerisation of FLNCd24. In the Y2683E mutant, the formation of a hydrogen bond between E2681 and E2683 is also improbable (fig4.18B). Nevertheless, there is no clear evidence of an intramolecular hydrogen bond between E2683 and K2653, and the mutation introduces only a single negative charge. These factors contribute to the dimerisation of Y2683E FLNCd24 being less severely compromised than that of pY2683 FLNCd24.

The phosphorylation of Y2683, in contrast to pT2677, induces several modifications in hydrogen bonding (fig4.19A), including the formation of a hydrogen bond between pY2683 and R113 ($32.0 \pm 12.07\%$), as well as an increase in hydrogen bonds involving E99 within the HspB7 ACD, which interacts with H2686($27.5 \pm 1.57\%$), N2689($21.3 \pm 1.02\%$), or R2690($45.2 \pm 7.95\%$). In the Y2683E complex (fig4.19B), E2683 also demonstrates a signifi enhanced tendency to form hydrogen bonds with R113($82.9 \pm 14.68\%$), attributed to the Coulombic effect. However, this

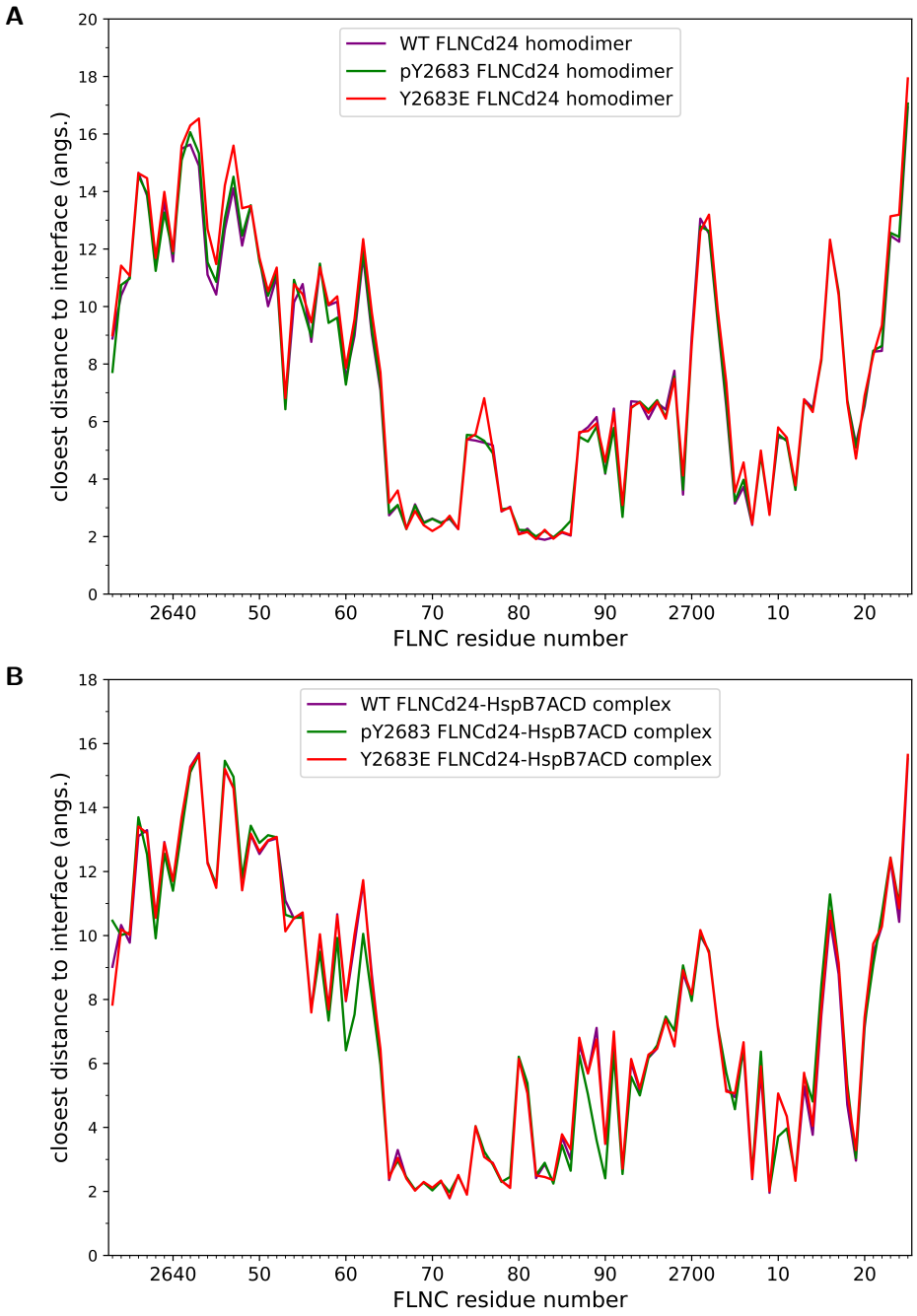


Figure 4.17: **A** The distances from each residue in one FLNCd24 chain to the other FLNCd24 molecule in the WT (purple), pY2683 (green), and Y2683E (red) FLNCd24 homodimers. **B** The distances from each residue in the FLNCd24 chain to the HspB7 ACD molecule in the WT, pY2683, and Y2683E FLNCd24-HspB7ACD complexes.

represents the only substantial alteration in hydrogen bonding, as the changes associated with E99 are not observed.

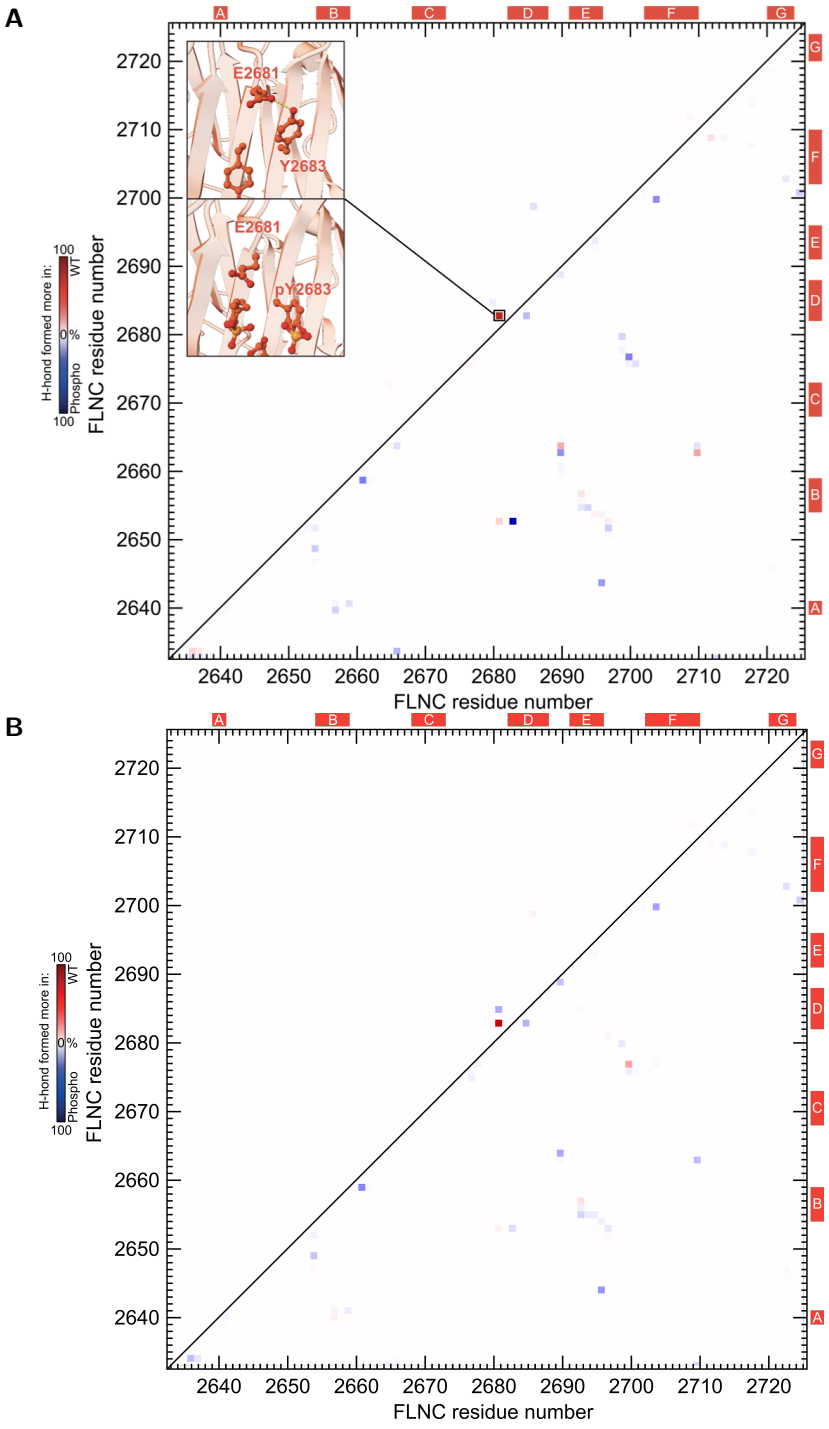


Figure 4.18: **A** The differences in hydrogen bond probabilities between intermolecular (upper left) and intramolecular (lower right) residue pairs in WT and pY2683 FLNCd24-HspB7ACD complexes. **B** The differences in hydrogen bond probabilities between intermolecular (upper left) and intramolecular (lower right) residue pairs in WT and Y2683E FLNCd24-HspB7ACD complexes.

Examination of the accuracy of phosphomimetic mutations

In laboratory experiments, phosphomimicking mutated FLNCd24 is utilized as a substitute for phosphorylated proteins. Through MD simulations, we investigated

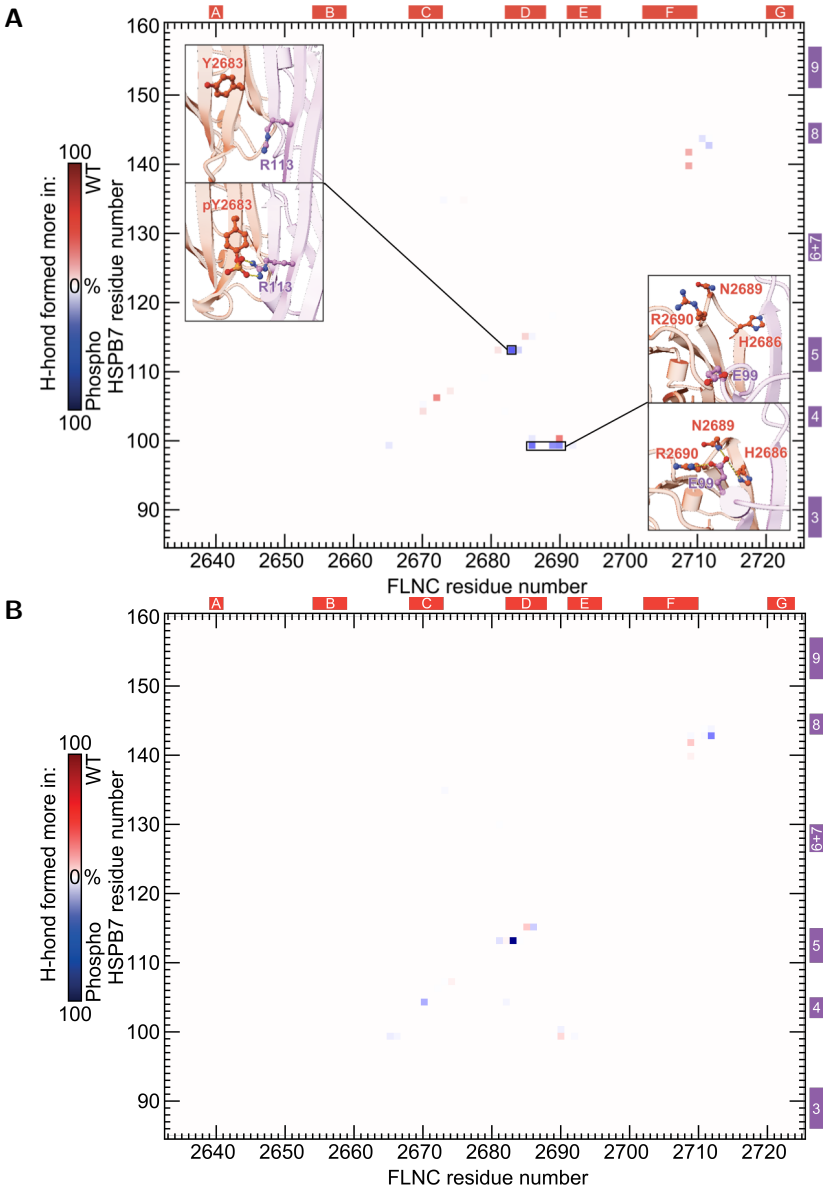


Figure 4.19: **A** The differences in hydrogen bond probabilities between intermolecular residue pairs in WT and pY2683 FLNCd24-HspB7ACD complexes. **B** The differences in hydrogen bond probabilities between intermolecular residue pairs in WT and Y2683E FLNCd24-HspB7ACD complexes.

both phosphomimicking and phosphorylated protein complexes, facilitating a comparative analysis of the effects of these two types of modifications. Based on the binding energies calculated by the mmGBSA method (table4.5), we observed that both the T2677D and Y2683E phosphomimicking mutations exhibit regulatory effects that are analogous to those of the corresponding phosphorylations in both the FLNCd24 dimer and the FLNCd24-HspB7ACD complex. However, these effects

manifest with varying intensities. For both homo- and hetero-dimers, the two types of T2677 modifications induce negligible or minor variations in binding energies. In contrast, the two modifications of Y2683 demonstrate significant differences, particularly within the homodimer. As illustrated in the Y2683 section, the observed variation suggests that the binding interface within the pY2683 homodimer may be unstable, whereas the Y2683E homodimer exhibits greater stability throughout the simulations and potentially in laboratory experiments as well.

T-test	Homodimer		Hetero-complex	
	Complex[a]	P-value	Complex	P-value
T2677	pT2677	$8.45 * 10^{-3}$	=	0.540
Y2683	Y2683E	~0	pY2683	$2.37 * 10^{-4}$

Table 4.5: T-test of binding energies of T2677 and Y2683 modified complexes

[a]This refers to the complex with higher binding energy within the two phosphorylated and phosphomimicking mutated complexes

To further explore the relationship between the phosphomimicking mutation and phosphorylation, we analysed the h-bond occupancy datasets using the Wilcoxon signed-rank test[272] and mapped the H-bond occupancies in fig4.20. In the context of the Wilcoxon signed-rank tests, the null hypothesis posits that the difference between the two datasets follows a distribution that is symmetric about zero, which is, in this case, equivalent to asserting that the corresponding hydrogen bond occupancies in the phosphorylated and phosphomimicking complexes are not biased. As demonstrated in table4.6 the p-values for inter-molecular H-bond occupancies in T2677 modified hetero-complexes and intra-molecular occupancies in T2677 modified homo-dimers are both below 0.05. According to one-sided tests, the accepted hypotheses indicate that the pT2677 hetero-complex exhibits greater inter-molecular H-bond occupancies than the T2677D hetero-complex, while the intra-molecular hydrogen bond occupancies in the pT2677 homo-dimer are lower than those in its subsititute. In the T2677 modified hetero-complexes, the observed differences are associated with hydrogen bonds between three specific residue pairs: V2672-S106, K2709-T140, and K2709-V142, which are all highly occupied in both the pT2677 complex and WT complex, but are weakened in the

T2677D complex. This observation correlates with the three dark red indicators, excluding the G2674-N107 hydrogen bond, as illustrated in fig4.16B. Conversely, in the T2677 modified homo-dimers, the differences arise from a series of hydrogen bonds that are distributed dispersedly in the complex structure. Therefore, we proposed that the T-D mutation can accurately mimic the phosphorylation of T2677 in the formation of FLNCd24 homodimer, while in other complexes, both phosphomimetic mutations may exhibit effects to the interactions that differ from those of the corresponding phosphorylations.

Complex	**FLNCd24 homodimer**		**FLNCd24-HspB7ACD complex**	
P-values	T2677	Y2683	T2677	Y2683
inter-molecular	0.4329	0.1005	0.03043[a]	0.3340
intra-molecular	0.001314[b]	0.5194	0.7292	0.8096
full-complex	0.009086[c]	0.8194	0.4324	0.8290

Table 4.6: Correlation of h-bond occupancies between phosphorylated and phosphomimicking protein complexes, revealed by p-values calculated by the Wilcoxon signed-rank test

[a] $p = 0.01522$ for the alternative hypothesis that h-bond occupancies in the phosphomimicking complex are less than those in the phosphorylated complex

[b] $p = 0.0006568$ for the alternative hypothesis that h-bond occupancies in the phosphomimicking complex are less than those in the phosphorylated complex

[c] $p = 0.004543$ for the alternative hypothesis that h-bond occupancies in the phosphomimicking complex are less than those in the phosphorylated complex

4.5 Conclusions

In this chapter, we examined the characteristics of the HspB7 ACD, its influence on the dimerisation of FLNCd24, and the regulatory role of phosphorylation at the FLNCd24 residues T2677 and Y2683 on the interaction affinities associated with the dimerisation of FLNCd24, as well as the interactions between FLNCd24 and HspB7 ACD.

Initially, we performed MD simulations of the HspB7 ACD monomers and dimers, utilising their crystal structures as well as monomeric and dimeric HspB1 and HspB5 as references. The analysis of RMSD and RMSF for both HspB7 ACD monomer and dimer supports our hypothesis that dimerisation of HspB7 ACD does not enhance structural stability and is therefore unlikely to occur. The differences

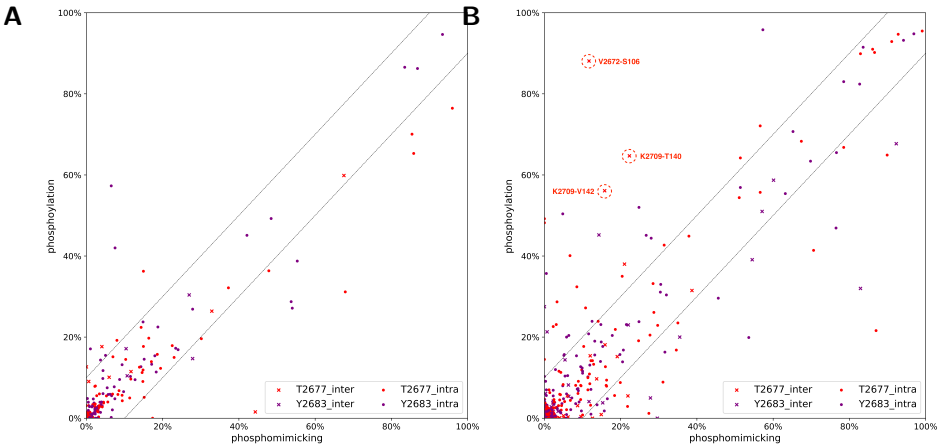


Figure 4.20: **A** The distribution of occupancies for intermolecular (×) and intramolecular (•) hydrogen bonds in the T2677 (red) and Y2683 (purple) modified FLNCd24 homodimers. Each symbol represents the probability of hydrogen bond formation within the residue pair at the same position across the two structures. The parallel lines delineate the hydrogen bonds for which the differences in probabilities between the phosphorylated and phosphomimicking mutated structures are less than 20%. **B** The distribution of occupancies for hydrogen bonds in modified FLNCd24-HspB7ACD complexes.

in residue RMSD and RMSF among the three homologous monomers, along with their correlation to the respective conformations, indicate that the relatively weaker dimerisation interaction of HspB7 ACD, in comparison to the other two proteins, is attributable to the absence of a three-residue motif in the $\beta 5$ strand.

We then conducted MD simulations of the WT HspB7ACD-FLNCd24 complex and the FLNCd24 homodimer, analysing the two trajectories from multiple perspectives, including binding energies, inter-residue distances, and hydrogen bonding interactions. The comparable binding energies indicate that for a FLNCd24 monomer, the binding affinities with a homogeneous protein and an HspB7 ACD are approximately equivalent. Given the overlapping interaction surfaces and the closer proximity of residue pairs in the hetero-complex, we proposed that the interaction between FLNCd24 and HspB7 ACD is slightly more favourable than the homo-dimerisation of FLNCd24.

Before modifying the phosphorylatable residues, we assessed the protonation states of adjacent histidines, given the comparable pK_a of the histidine side chain and the substituted phosphate group. Based on their relative positions, we selected H2673 in FLNCd24, which is close to both T2677 and Y2683, as well as H109 in HspB7

ACD, which is adjacent to T2677 only. We conducted a series of simulation tests on hetero-complexes with H2673-only or doubly protonated state. Analysis of RMSF of the histidines and neighbouring residues indicates that the protonation of H109, but not H2673, disrupts the conformational stability of the complexes. Consequently, we decided to conduct subsequent simulation of structures with protonated H2673.

Subsequently, we implemented MD simulations of FLNCd24 homodimers and FLNCd24-HspB7ACD complexes, incorporating either T2677 or Y2683 in a phosphorylated state or with phosphomimetic mutations. The binding energies were computed using the mmGBSA algorithm to quantify the interaction intensities. To elucidate the source of the intensity differences observed between the WT and modified structures, we employed H-bond as a criterion and calculated the probabilities of H-bond formation for all intermolecular residue pairs within the structures, using the WT H-bond probability as a benchmark. Our findings indicate that:

- pT2677 or T2677D:
  - Strengthen the FLNCd24 homodimer, primarily through the formation of an intermolecular hydrogen bond between pT/D2677 and R2690;
  - Weaken the formation of FLNCd24-HspB7ACD complex through the inhibition of the hydrogen bond between G2674 and N107;
  - A significantly greater number of residue pairs are influenced by the T2677D mutation compared to pT2677, suggesting that the conformations associated with pT2677 exhibit greater stability.
- pY2683 or Y2683E:
  - Diminish the dimerisation of FLNCd24, as the modified residue can no longer form an intermolecular H-bond with E2681. The dimerisation of pY2683 FLNCd24 is extremely weak, possibly due to the formation of an intramolecular pY2683-K2653 H-bond.
  - Promote the formation of the FLNCd24-HspB7ACD complexthrough the formation of a hydrogen bond between pY/E2683-R113.

Furthermore, we analysed the reliability of the phosphomimetic method employed in laboratory experiments by evaluating the correlation between the binding energies and the H-bond occupancies of the phosphomimicking mutated structures and their corresponding phosphorylated complexes. Based on the statistical analysis, we hypothesised that the T2677D mutation has a comparable effect, particularly on the dimerisation of FLNCd24. In contrast, the Y2683E mutation is not an appropriate substitute for pY2683, as it results in an overestimation of the stability of the Y2683-modified FLNCd24 homodimer.

5

The R157H mutation of HspB5 CTD regulates the mechano-sensitivity of the titin immunoglobulin-like domain 26

Contents

5.1 Abstract 114
5.2 Introduction 115
5.2.1 The role of titin in regulating the mechanical properties of cardiac muscle 115
5.2.2 HspB5 modulates the retraction of titin through its interaction with titin Ig26 117
5.2.3 The mutation point: R157 or R163 118
5.2.4 Plan of experiments and analysis 118
5.3 Methods 119
5.3.1 MD simulations 119
5.3.2 Steered MD simulations 120
5.4 Results and discussions 121
5.4.1 Native-MS experiments verifies the interactions between titin Ig26 domain and HspB5 CTD peptide, which is weakened by the R157H mutation of the CTD peptide. . 121
5.4.2 Determination of pulling velocity and identification of R-H mutation point 123
5.4.3 HspB5 CTD peptide binds titin Ig26 domain, and the interaction is weakened by R157H mutation. 126
5.4.4 Interaction with the HspB5 CTD peptide protects titin Ig26 against unfolding under mechanical force, which is diminished by the R157H mutation. 131
5.5 Conclusions 140

5.1 Abstract

In this chapter, we examined the interaction between the 26 immunoglobulin-like domain (Ig26) of titin and the CTD peptide of HspB5, as well as the regulation of the mechano-sensitivity of the titin domain resulting from this interaction. Experimental results obtained through *in vitro* methods confirm that the interaction, particularly between the CTD peptide of HspB5 and titin Ig26, enhances the mechanical resistance of the Ig domain. Furthermore, a mutation in the HspB5 CTD, specifically R157H, is found to partially mitigate this enhancement effect. Based on computationally simulated structures of titin Ig26 and the titin Ig26-HspB5 CTD complexes, we implemented regular and steered MD simulations to investigate the mechanical resistance of titin Ig26 under various conditions. Analysis of the MD conformational snapshots and the mechanical properties of the titin Ig26 domains revealed that, in comparison to WT, the R157H-mutated HspB5 CTD exhibits a significantly smaller interaction interface with titin Ig26, leading to weaker interactions with the Ig domain. Consequently, the strengthening effect of the R157H HspB5 CTD on the mechanical resistance of titin Ig26 is considerably less pronounced than that of the WT, which may contribute to the pathogenesis of dilated cardiomyopathy.

5.2 Introduction

5.2.1 The role of titin in regulating the mechanical properties of cardiac muscle

Titin is recognised as the largest protein in the human body, located within the sarcomere, and serves as a critical structural component alongside actin and myosin[273, 274]. It plays a pivotal role in providing elasticity to the cardiovascular system[66, 275], as it exists in a loosely coiled state within relaxed cardiac fibres. This configuration enables titin to be passively stretched or compressed in response to the movement of the thin filaments to which it is attached[276, 277]. During the cardiac contraction cycle, the passive force generated by the elastic potential energy of titin facilitates the repositioning of the thin filaments, thereby ensuring the structural integrity of the cardiac sarcomere complex[278, 279]. Consequently, dysfunction of titin is identified as a significant pathogenic factor contributing to heart failure[280, 281] and various forms of cardiomyopathy[282, 283, 284, 285].

Titin is comprised of 244 distinct domains[65], and is typically categorised into two primary regions: the fixed region and the elastic region[67]. The fixed region interacts with myosin to form the thick filament, anchoring to the M-band via its tail[286, 287]This region is classified as part of the A-band (A stands for this region being anistropic) and is primarily composed of two types of domains: immunoglobulin (Ig) domains and fibronectin-3 (Fn3) domains [288, 289]. In contrast, the elastic region, which is coiled, constitutes the I-band (I stands for isotropic), with the head of the elastic region connecting to the thick filament at the Z-disc[290], of which the head connects to the thick filament in the Z-disc[291, 292]. Unlike the fixed region, the elastic portion is predominantly composed of Ig domains[66].

The titin protein is encoded by 363 exons located on chromosome 2 [293], and various isoforms of titin can be produced through alternative splicing[294, 295] (see fig5.1). Within the I-band, the PEVK region, which is rich in proline, glutamic acid, valine, and lysine residues, is located between two extensive segments that contain tandem Ig domains[296]. Additionally, the N2A and N2B regions are positioned

in the segments preceding the PEVK region [297, 298]. These two regions are among the first segments of titin to be identified in laboratory experiments and are designated as N2A/N2B due to their predominant representation on the N2 line in electron microscopic images[299].The N2A and N2B regions can be included or excluded individually in different isoforms; for instance, the isoforms that incorporate both regions (N2BA isoform) or solely the N2B region (N2B isoform) are the most prevalent in cardiac muscle[273]. Conversely, the isoform that contains only the N2A region, situated between extensive tandem Ig domains and the PEVK region (N2A isoform), is expressed in skeletal muscle[273, 300]. Both the N2A and N2B regions are significantly associated with the elasticity of the entire titin protein.

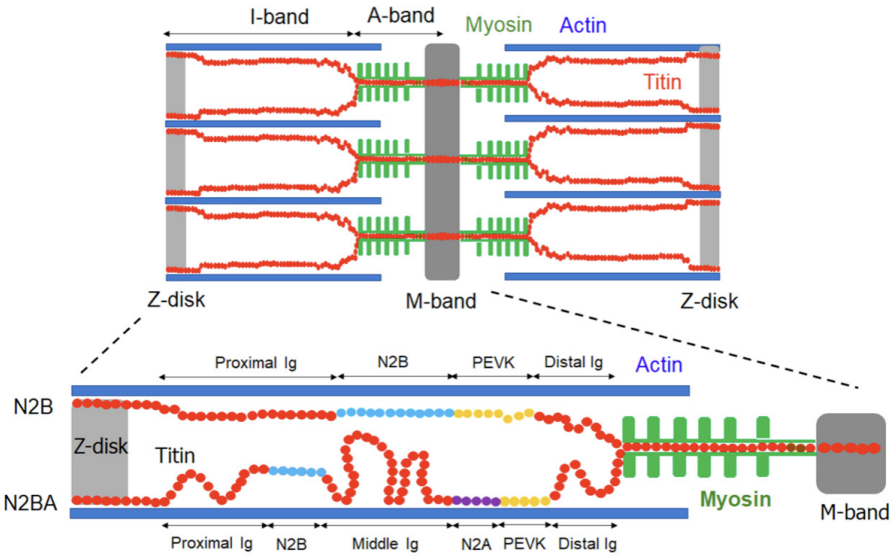


Figure 5.1: Titin within the sarcomere is composed of several components, including the Z-disc, I-band, A-band, and M-band. The elastic regions, which encompass the N2A, N2B, tandem Ig region, and PEVK region, are predominantly situated within the I-band. In human cardiac muscle, two principal isoforms of titin are expressed, designated based on the presence of the N2A and N2B regions: N2BA, which includes both the N2B and N2A regions along with the middle-Ig segment located between the two N2 regions, and N2B, which contains solely the N2B region. The image is adapted from Adewale et al.[301].

Both N2 regions are composed of Ig domains and unique sequences, with the proportion of Ig domains in these regions being lower than the average observed in the I-band section[302]. In the N2B region, a notable unique sequence, referred to as N2B-Us [303], plays a critical role in facilitating the extension of titin to its maximum length (fig5.2). The Ig domain that follows N2B-Us is named Ig

domain 26 (Ig26), which functions as an active interaction site for regulators of titin function[304]. Both N2B-Us and Ig26 exhibit retractable properties, contributing to the retraction of the full-length titin protein. Furthermore, ligand binding within these regions has the potential to alter the mechanical resistance of the titin protein by modulating the elasticity of these regions[305, 306].

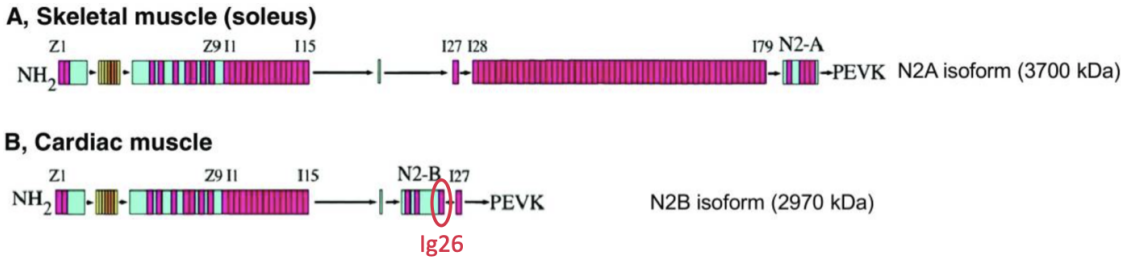


Figure 5.2: N2B region represents a unique segment of cardiac titin isoforms, as it is absent in the N2A isoform (**A**), which is the predominant isoform of titin in skeletal muscle. Ig26, located in the tail of the N2B region (**B**), serves as one of the binding sites for HspB5, and its length is responsive to external mechanical forces. This image is adapted from Bang et al.[293].

5.2.2 HspB5 modulates the retraction of titin through its interaction with titin Ig26

HspB5 is expressed in cardiac muscle and is localised within the sarcomere, where it binds to cardiac myofibrils[307]. This interaction is associated with external stress, with one of the binding sites identified in the titin I-band[93]. In a relaxed sarcomere, HspB5 predominantly resides in the Z-disc, but it translocates to the N2B region of the I-band in response to significant myofibril extension[42, 93]. Experimental findings indicate that HspB5 can bind to either the N2B-Us or the adjacent Ig26 domain, but not to Ig24 or Ig25, which are also located within the N2B region[304].The N2B region bound by HspB5 exhibits greater resistance to external stress compared to the unbound state, as evidenced by a greater maximum extended length and a higher threshold of external force required to stretch the region to its maximum length[93, 304, 308].

Furthermore, residue mutations in HspB5 have been shown to influence its regulation of N2B extension, thereby contributing to pathogenic cardiac muscle disorders, such as dilated cardiomyopathy[309, 310]. The R120G mutation, which

occurs within the ACD of HspB5, significantly alters the structure of the protein. Notably, the presence of the R120G HspB5 variant does not induce changes in the mechanical properties of the N2B-U region or other Ig domains[311]. Consequently, it is posited that the R120G HspB5 variant is unable to effectively bind to these regions in titin. In contrast, the R157H mutation, which occurs in the CTD peptide, inhibits interactions between HspB5 and the N2B-U region, while allowing for interactions with the Ig26 domain[312]. The interactions between the Ig26 domain and the R157H peptide result in a decrease in the mechanical resistance of the Ig domain, as the unfolding force in this scenario is significantly lower than that observed for Ig26 bound to WT HspB5, although it remains higher than that of the unbound Ig26 domain.

5.2.3 The mutation point: R157 or R163

HspB5 CTD peptide, which includes the R157 mutation site, displays a palindromic sequence, ERTIPITRE, in which the R157 is symmetrically positioned relative to R163. Consequently, it remains ambiguous whether the mutated peptide is oriented in the same manner as in the *in vivo* state, or whether the mutation site in the structure utilised for MD simulations corresponds to R157 or R163. For the purposes of this study, we have provisionally designated the mutated complex that is isomorphic to the R157H mutated complex present *in vivo* as the "R157H complex", while the alternative is referred to as the “R163H complex”.

Prior to conducting MD simulations of the R157H mutated complex, it is essential to verify the binding orientation of the peptide. Therefore, we intended to predict the structures of the R157H and R163H complexes using AlphaFold. Additionally, we perform both classical and steered MD simulations on these structures to compare the mechanical properties of the two complexes.

5.2.4 Plan of experiments and analysis

In laboratory experiments, my colleagues examined the interaction affinities between titin Ig26 and various variants of the HspB5 CTD. This was accomplished by

preparing mixtures of the two components at different concentration ratios and performing native-MS experiments to evaluate the composition of the mixtures.

For performing *in silico* simulations, we generated a structural model of the titin Ig26 domain, as well as a model of the complex formed by the titin Ig26 domain and the WT HspB5 CTD, utilising AlphaFold. Modifications are made to the HspB5 CTD sequences, including the introduction of the R157H and R163H mutations and the preparation of a scrambled control peptide. Based on the predicted conformations, we performed classical MD simulations to analyse the stability of both the Ig26 domain and the complexes. Subsequently, we employed steered MD simulations, applying external forces to the Ig26 domain in order to obtain data pertinent to the mechanical properties of the domains, including the force and work required to unfold the structure, as well as the degrees of deformation of different Ig26 domains at corresponding time points. The steered MD simulations for both the unbound and bound Ig26 domains are repeated and supplemented with control tests to minimise error. Our objective is to elucidate the regulatory effect of the HspB5 CTD and the attenuation of this effect resulting from the R157H mutation by investigating and comparing inter-residue interactions between the two chains.

5.3 Methods

5.3.1 MD simulations

Preparation of the initial structures

The structures of the free titin Ig26 domain and the WT titin Ig26-HspB5 CTD complex are simulated by AlphaFold 2[117], performed by N. Sen. For each structure, the prediction is repeated five times, resulting in the generation of five templates each time, which are evaluated using the internal scoring functions of AlphaFold. For the complexes containing mutated HspB5 CTD, structural models can be generated either through individual predictions using AlphaFold or by introducing mutations to the residue within the model of the WT complex.

Procedure and parameters of MD simulations

The settings and parameters of MD simulations, as well as the analytical methods and procedures, adhere to the protocols outlined in Chapter2.2.3 and Chapter3.3.2. Classical MD simulations, encompassing both the WT and R157H complexes and the free Ig26 domain, are conducted in five replicates, with each replicate lasting 200 ns. The MD simulations for the R163H and CTD scrambled complexes, aimed at structural equilibration and template sampling, are conducted for a duration of 10 ns each.

5.3.2 Steered MD simulations

SMD procedure and settings

In a constant-velocity SMD simulation, one terminal of the stretched polymer chain is fixed while the other is compelled to move a specified distance in a single direction over a designated time interval. In this study, the backbone atoms of the final residue in the Ig26 domain are immobilised, and following each application of force to the structure, a short simulation is conducted to equilibrate the distorted configuration. In this process, we employed the Nosé-Hoover Langevin piston method to regulate the pressure, consistent with the approach utilised during the constraining phase in classical MD simulations.

Data analysis

The instantaneous force values can be derived from the diagonally decomposed force values presented in the output file of the SMD simulations. The work at a specific time point is defined as the cumulative force from the initiation of the simulation to that particular time point. While plotting the values to visualise their transitions, we employed a smoothing technique, whereby the exact value at each frame is replaced with the average of the values within a short time period ranging 200 frames, centred on that frame. This approach is implemented to mitigate local fluctuations and to more effectively illustrate the overall trend of the values throughout the simulations. Furthermore, the standard deviations of

the values obtained from repeated simulations are calculated and represented in the plotted figures as shaded regions.

Our analysis of inter-residue distances indicated that certain pairs of residues demonstrated transient proximity at specific time points, followed by rapid separation. This behavior is atypical, as the residues in question do not exhibit close proximity in classical MD simulations; therefore, the observed approach is likely a consequence of external forces. Consequently, in determining breaking points, we focused exclusively on residue pairs that are classified as being in close contact, defined as having an average distance of less than a predetermined cutoff, which is set at 4 Å in this study. We recorded only the frame number in which the residue pair is last observed to be within this distance threshold.

5.4 Results and discussions

5.4.1 Native-MS experiments verifies the interactions between titin Ig26 domain and HspB5 CTD peptide, which is weakened by the R157H mutation of the CTD peptide.

In native-MS experiments, we mixed the titin Ig26 domain with WT, R157H, or a scrambled controlling HspB5 CTD peptide without the I-P-I motif and examined the abundance of each component by Native-MS. As displayed in fig5.3, a greater portion of the WT peptide is bound to the titin domain, compared to any of the other two peptides, while a gap exists between the mutated peptide and non-specific binding. The native-MS results suggest that the mutation leads to a decrease in binding affinity but does not completely eliminate the effect of the I-P-I motif.

According to prior research, which has demonstrated through *in vivo* experiments that the R157H mutated peptide is capable of binding to titin Ig26 and can subsequently result in a reduction of its mechanical resistance, and the observation of the *in vitro* experiments, we proposed that the complexes present in the mixed sample containing the mutated peptide may consist of either R157H, R163H, or a combination of both, in accordance with one of the proposed hypotheses:

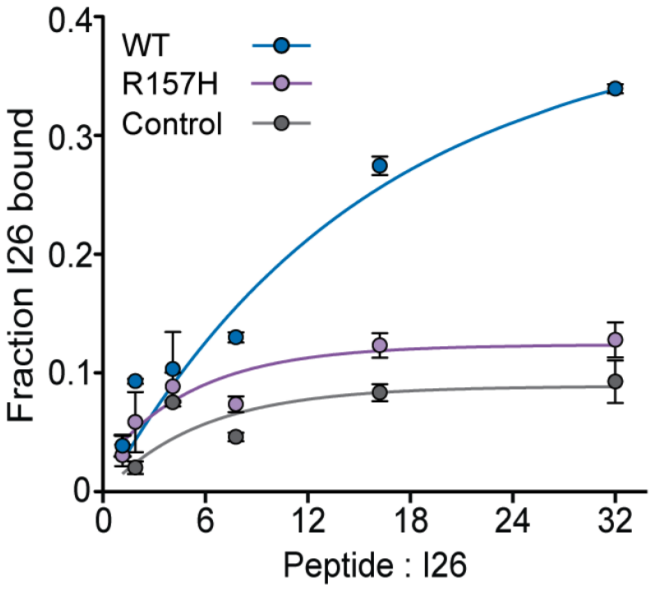


Figure 5.3: The ratio of concentration between domain-peptide complex and the initial Ig domain concentration, obtained from Native-MS. Each mixture is prepared with 5μM titin Ig26 and WT (blue), R157H (violet) or scrambled (grey, the control) HspB5 CTD of which the concentration is 1, 2, 4, 8, 16 and 32 times of titin Ig26. Image prepared by M. Collier.

- Complexes containing the R-H mutated peptide are predominantly R157H complexes. It suggests that the interaction within the R157H complex is much stronger than the R163H complex, and the R157H peptide is much less effective in terms of promoting the mechanical resistance of titin Ig26. In this circumstance, the R157H complex causes the weakening of mechanical resistance observed in the native-MS test, despite the fact that the effect of R163H remains unclear.

- Complexes containing the R-H mutated peptide are predominantly R163H complexes. It suggests that the interaction within the R163H complex is much stronger than the R157H complex, and the R163H peptide is much less effective in promoting the mechanical resistance of titin Ig26. In this circumstance, the R163H complex causes the weakening of mechanical resistance observed in the native-MS test, regardless that the effect of R157H remains unclear.

- Complexes containing the R-H mutated peptide include both R157H and R163H complexes, both at considerable concentration. In this circumstance, it is possible for the mutated peptide to bind to the Ig domain from both

directions, and the Ig domain in the R157H complex is less resistant to mechanical stress. However, the effect of R163H remains unclear, identical to the cause of native-MS results.

5.4.2 Determination of pulling velocity and identification of R-H mutation point

We initially employed AlphaFold to generate structural models of the free titin Ig26 domain and the WT, R157H and R163H complexes. The structural predictions for both the free Ig domain and the WT complex are compelling, as evidenced by the high degree of coincidence and the high scores of the templates for both structures. In contrast, the templates for the modified complexes are deemed unreliable, as they yield low scores and exhibit significantly varied conformations, particularly at the binding interfaces. Furthermore, within all templates associated with the R-H mutation, the orientations of the peptide, revealed by the relative position of the mutated histidine and the corresponding arginine, are identical. It indicates that the binding orientation of the mutated peptide in a complex is highly biased, and therefore disallows the hypothesis that both considerable concentrations of R157H and R163H complexes can co-exist in the mixture. Consequently, we alternatively implemented mutations to the HspB5 CTD peptide by VMD[208], which necessitate an extended minimisation phase (10,000 steps).

Within SMD simulations aimed at elucidating protein unfolding mechanisms, the selection of pulling velocity is a critical factor influencing the unfolding pathway of proteins. It is generally advised that the pulling velocity should not exceed 0.01 m/s (100 Å/ns) to prevent the occurrence of abrupt force spikes that may compromise the structural integrity of proteins. Additionally, employing a lower pulling velocity facilitates the preservation of conformations during simulations conducted in quasi-equilibrated states. However, as discussed in the literature regarding the dependence of the unfolding pathway on pulling velocity[313], when the pulling velocity is relatively high (>10 Å/ns), the outer regions of a protein tend to unfold prior to the inner regions, whereas this preference is neutralised at lower

pulling velocities. Given our focus on the conformations of the binding interfaces throughout the unfolding process, we anticipated that, across repeated simulations under identical conditions, the Ig domains would undergo unfolding following a consistent pathway. Consequently, we selected a pulling velocity of 25 Å/ns, which is supported by studies that have employed SMD simulations with constant velocities of 10 and 50 Å/ns to investigate the unfolding process of other titin segments[314].

To validate the rationale and benefits of our selected pulling velocity, we implemented testing SMD simulations of conformations of the WT complex, which are sampled from short production simulations of the structure generated by AlphaFold, utilising pulling velocities of 1, 5, and 25 Å/ns, each velocity being repeated ten times. To evaluate the consistency of the pulling pathway of the titin Ig26 domains, we analysed the transitions in the distances between β strands throughout the simulations. For each pair of β strands, we recorded the time points at which the intermediate distance exceeded 5 Å/ns beyond the distance observed in the corresponding conformational template sampled from equilibrated simulations. This threshold marked the departure point for that pair of β strands, allowing us to subsequently calculate the corresponding extension dimension. By examining the relationship between the breakage of a specific β strand pair and the associated extension length, we can delineate the sequence of the β strands being unfolded and assess the consistency of the unfolding pathway.

As illustrated in fig5.4, protein conformations observed in simulations conducted at velocities of 5 and 25 Å/ns generally follow a similar unfolding pathway, as the sequence of values remains largely consistent. However, the lower pulling speed results in increased sensitivity and variation, as evidenced by the overall lower values and higher standard deviations. In contrast, the conformations subjected to a pulling speed of 1 Å/ns exhibit a distinct unfolding pathway. The interacting β strands pairs, mainly the consecutive β strands pairs, demonstrate greater resistance to the applied pulling force, therefore leading to a distinct unfolding pathway and greater variations of the break time points of the peptide-associated β strands (C~D) among the repeats. In order to better analyse the impact of the HspB5 CTD

to the Ig domain, we proposed that a velocity of 25 Å/ns is suitable for subsequent simulations, as it results in the lowest chronological variation of the departure of the β strand pairs, while requiring the least computational cost. Meanwhile, the rationality of the unfolding pathway it results in can be validated by the simulations conducted at a lower velocity. As indicated by the extension length, the majority of conformations are fully unfolded at 300 Å; therefore, for simulations conducted at 25 Å/ns, we have selected a simulation time limit of 15 ns, which is slightly longer than the time range corresponding to the extension (12 ns).

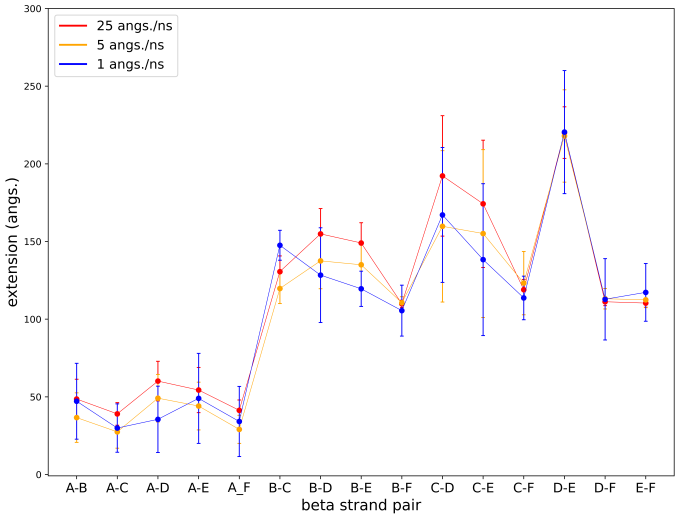


Figure 5.4: Time points of separation of beta strand pairs within Ig domains unfolded at the speed of 1 (blue), 5 (yellow) and 25 (red) Å/ns. For a beta strand pair, this time point is defined as the last frame in the SMD trajetcory that the distance between their centres of mass (CoMs) is not 5 Å greater than the average distance between the CoMs in the classical MD simulation.

Next, we performed SMD simulations to investigate the mechanical resistance of the Ig domains within the complexes. For each complex, a short production simulation lasting 10 ns is implemented, and ten conformations are sampled as initial structures for the subsequent SMD simulations. As illustrated in Chapter2.4, greater force and accumulated work are indicative of higher mechanical resistance, and the curves presented in fig5.5 demonstrate the differences in averaged accumulated work. During the initial 6 ns, the resistances of all structures are indistinguishable. Subsequently, the R157H complex exhibits a lower work requirement to unfold its higher-order structure compared to the other two complexes, and by the end of the simulations, the difference in averaged work between the WT and R157H complexes

reaches $1 * 10^{-21}$ J. Conversely, the R163H complex displayed a resistance to pulling stress that is not only comparable to the WT complex but is, in fact, slightly greater, particularly after 8 ns, with a gap of approximately $0.5 * 10^{-21}$ J at the end of the simulations. Consequently, we proposed that the R157H mutation, rather than the R163H mutation, is the primary factor contributing to the reduction in the mechano-resistance of the Ig domain. Therefore, the mutated complexes generated both *in vitro* and *in silico* are predominantly or exclusively R157H complexes.

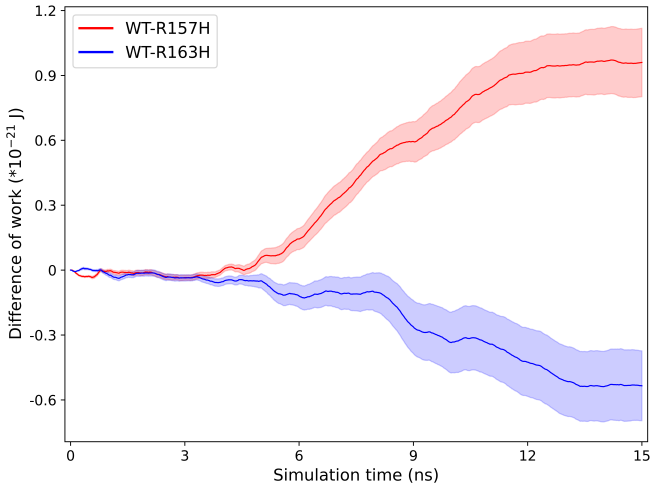


Figure 5.5: Difference of replica-averaged accumulated work between WT and R157H complexes (red) or between WT and R163H complexes (blue). Positive values refer to the fact that WT complex requires higher work than the R-H complex to be unfolded to the same level, and vice versa.

5.4.3 HspB5 CTD peptide binds titin Ig26 domain, and the interaction is weakened by R157H mutation.

In the classical MD simulations for both the WT and R157H mutated complexes, we executed five repeated simulations, each with a duration of 200 ns, with initial structural templates selected as outlined in the preceding section. Additionally, simulations following the same protocol are performed for the free titin Ig26 domain, serving as a reference for the structural transformation of the Ig domain. Upon aligning the trajectories by minimising the RMSD of the Ig domain, we observed that, by the end of the simulations, the WT peptide exhibits a significant displacement

from its initial docking position, whereas the mutated peptide maintained its original position, using the position of the Ig domain as a reference(fig5.6).

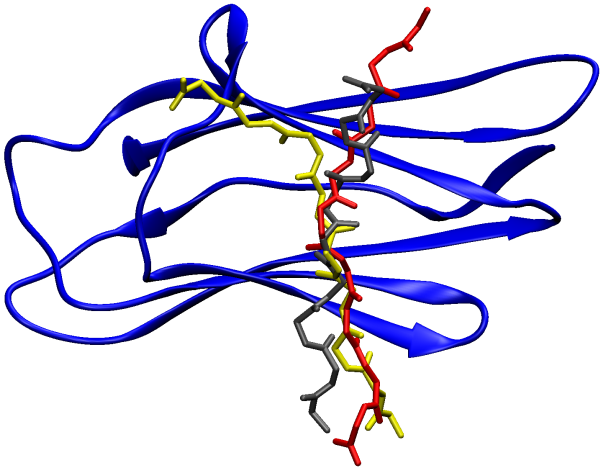


Figure 5.6: The positions of HspB5 CTD peptides, either WT or R157H-mutated, within the complexes. The grey chain represents the original position of the peptide as generated by AlphaFold, while the yellow chain denotes the WT peptide in the final frame of the simulation, and the red chain indicates the R157H peptide in the final frame of the simulation. The conformations have been aligned based on the coordinates of Ig26 domains (depicted in blue), and two of the three Ig26 domains omitted from the display for clarity.

In the WT complex, upon transformation of the docking position, the residues located at the C terminal of the peptide are forced to move, therefore obtaining high residue RMSD (fig5.7). Meanwhile, RMSF of the residues are not excessively high compared to any other residues in the complex, suggesting that the C-terminal of the peptide is intensely restrained. However, in the R157H complex, both RMSD and RMSF of residues located in either terminal of the mutated peptide are extremely high. The instability of the peptide suggests that the peptide is not fixed in the initial position, as displayed in fig5.6, since the interactions are weakened by mutation.

The mutation also leads to a greater structural transformation of its interacting region in the Ig domain compared to that of the rest of the protein, as revealed by their higher RMSD. Most affected regions, including strand A and A' (residue 4615-4622), C-D loop (residue 4644-4648), and E-F loop (residue 4667-4672), experience less structural fluctuation when the domain binds to WT peptide, compared to those in the free domain or the domain interacting with mutated peptide, while the reverse tendency only occurs in the F-G loop (residue 4680-4684)(5.7B). It suggests

that the stabilisation effect of the WT peptide applies significantly to multiple regions in the Ig domain via their interactions. Meanwhile, RMSD and RMSF of the forementioned regions have similar RMSD and RMSF values, suggesting that the interaction between the mutated HspB5 peptide and the Ig domain is not necessarily intense to stabilise the conformation.

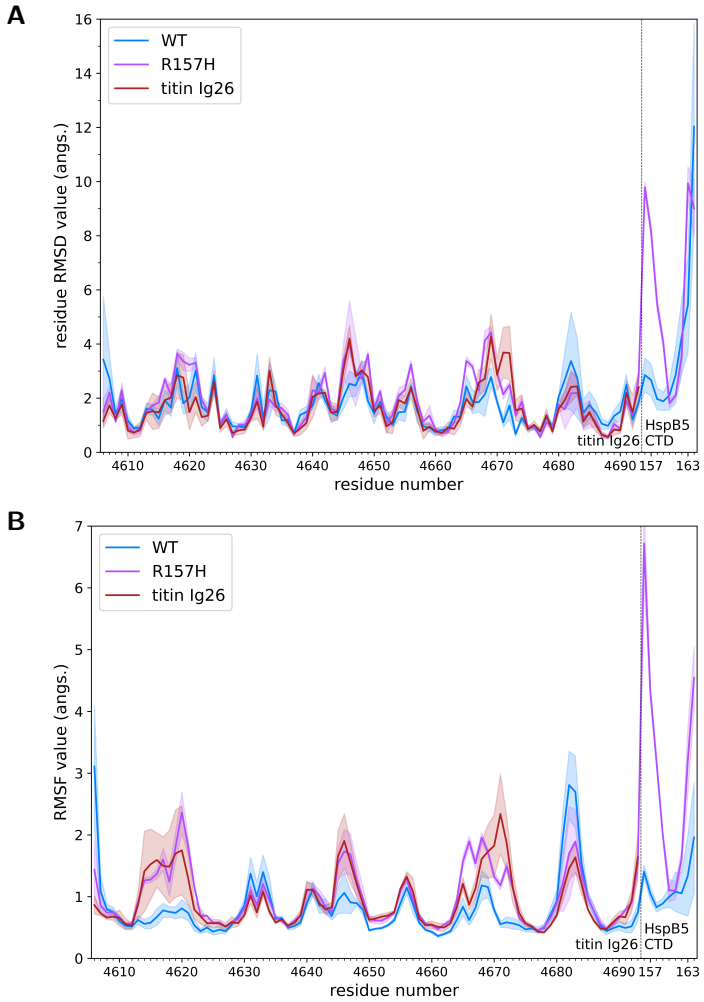


Figure 5.7: **A** Average residue RMSD of WT (blue) and R157H (purple) complexes and free titin Ig26 (red). For the complexes, sequences of titin Ig26 and HspB5 CTD are divided by a vertical line, and R/H157 in HspB5 CTD is labelled. **B** Residue RMSF of WT and R157H complexes and free titin Ig26.

We conducted a mmPBSA calculation to determine the binding energy of the two complexes[137], and applied a t-test to assess the variation across the repeating simulations as a criterion for evaluating the conformational stability of the structures throughout the simulations. The average binding energy between the WT peptide

and the Ig domain is approximately -14.0 kcal/mol, whereas the binding energy in the R157H mutated complex is merely -1.2 kcal/mol (table5.1). Furthermore, the repeated simulations of the R157H complex exhibit significantly greater variations, as evidenced by their higher p-values. Based on the differences in the absolute values and the variations of the binding energies, we concluded that the R157H mutation diminishes the interactions between the peptide and the titin Ig26 domain, which may lead to the distinct conformations observed across the repeating simulations.

Energy (kcal/mol)	**WT**		**R157H**	
	Mean	**Std Dev**	**Mean**	**Std Dev**
Overall	-14.0443	10.2176	-1.2240	10.1208

Repeat	**1**	**2**	**3**	**4**	**5**
WT	$5.44 * 10^{-3}(\downarrow)$	0.0808	$0.0261(\uparrow)$	$0.0439(\downarrow)$	0.222
R157H	$9.81 * 10^{-5}(\downarrow)$	$1.07 * 10^{-3}(\uparrow)$	0.0685	$4.90 * 10^{-13}(\downarrow)$	$2.09 * 10^{-10}(\uparrow)$

Table 5.1: The binding energies of the WT and R157H titin Ig26-HspB5 CTD complex, along with the p-values obtained from t-tests utilised to evaluate the variations across the repeating simulations.

To illustrate the phenomena at the residue level, we calculated the distances between inter-molecular residue pairs in WT and mutated complexes, and a residue pair is defined as a contact if the inter-residue distance is shorter than 4 Å. For each inter-residue contact, we calculated its occupancy ratio as the ratio between the number of frames in which the contact occurs and the number of frames in the trajectory. In the WT complex (fig5.8A), the HspB5 peptide is strictly immobilised by two regions of the titin Ig26 domain: C terminal of the peptide is intensely affiliated with the C-D loop of the Ig domain (residue 4643-4647), and the N-terminal is bound to strand D (residue 4651-4655). In the R157H complex, the same regions in titin Ig26 are occupied, but both regions reach only the midsection, rather than the two termini, of the R157H peptide. This shows that the topologies of the interfaces in the WT and R157H mutated complexes are significantly distinct, and the difference leads to the divergences in terms of the docking positions and the stabilities of interactions in the two complexes.

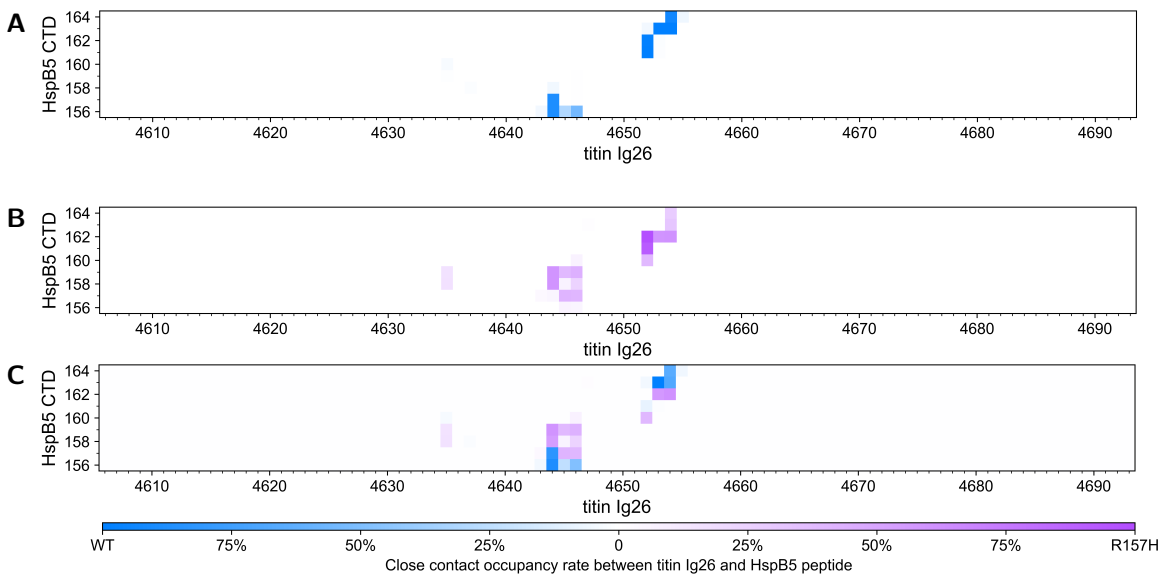


Figure 5.8: **A** The occupancy ratios of inter-residue contacts in the WT complex, with a distance cut-off value set at 4 Å. **B** The occupancy ratios of inter-molecular contacts in the R157H complex. **C** The differences between the occupancy ratios of inter-residue contacts in the WT and R157H complexes.

To investigate how the interaction affects the structure of the Ig26 domain, we isolated the coordinates of the titin Ig26 domains from the trajectories of the WT and R157H complexes, aligned them together with the free titin Ig26 domain, and applied PCA calculation of the coordinates of the backbone atoms (fig5.9). The eigenvalues of the two highest-weighted PCs show that all conformations of the Ig domain generated by MD simulations can be divided into two columns: one of them is indistinguishable from the initial topology, and the majority of conformations of the Ig26 domain in the WT complex are columned in this group; However, both the free and the R157H-peptide bounded Ig26 domains transform to similar conformations, clearly different from the original ones.

We also calculated the distances between intramolecular residue pairs to investigate the inter-residue contacts influenced by the peptide. In fig5.10, the differences in distances between the intramolecular residue pairs of the free and ligand-binding Ig26 domains are presented, with a cutoff distance of 4 Å. Excluding the minor colour bars adjacent to the diagonal lines, which indicate local residue-residue contacts that are not relevant to across-strand interactions, the majority of the

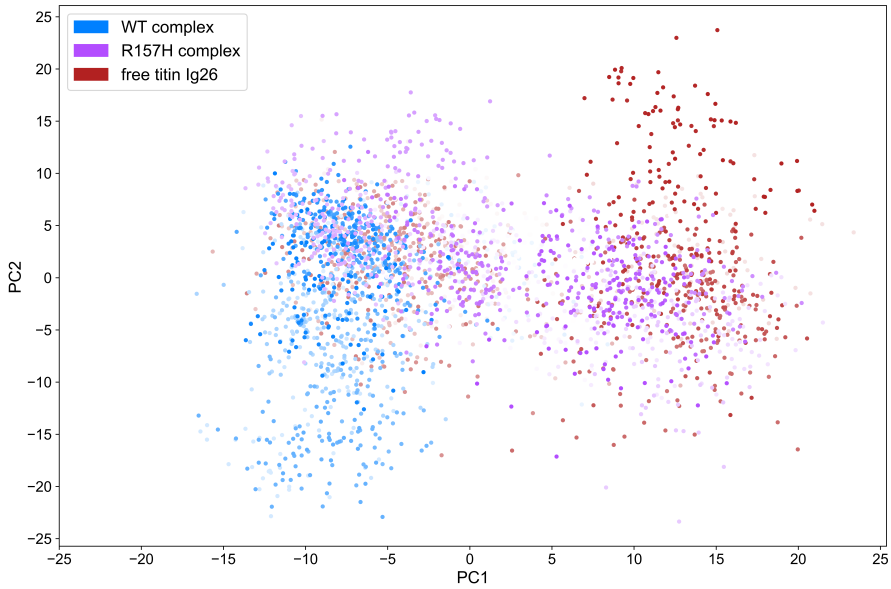


Figure 5.9: PCA of atomic coordinates of titin Ig26 in WT complex (blue), R157H complex (purple) or independent (red). Transparency of a dot is reversely linearly correlated to the number of frame in the simulation.

differences in residue-residue contacts between the free and R157H-HspB5CTD binding Ig26 domains can be attributed to the reshaping of the contacting A-B and E-F loops, as well as the enhanced interactions between the C-D and E-F loops, which compete with the ligand-binding interactions. Notably, the Ig26 domain in the WT complex exhibits a greater number of differing residue pairs compared to the free domain: interactions involving the E-F loops intensify; the conformation of N-terminus undergoes significant changes, inhibiting its binding with the C-terminal binding or the B-C loop.

5.4.4 Interaction with the HspB5 CTD peptide protects titin Ig26 against unfolding under mechanical force, which is diminished by the R157H mutation.

To investigate the impact of the mutation on the mechanical sensitivity of the titin Ig26 domain at the residue level, we sampled 40 conformations from each trajectory to serve as the initial structures for the SMD simulations. The SMD parameters employed for these simulations are consistent with those utilised in Chapter5.4.2.

According to the force data extracted from the output file, we calculated the

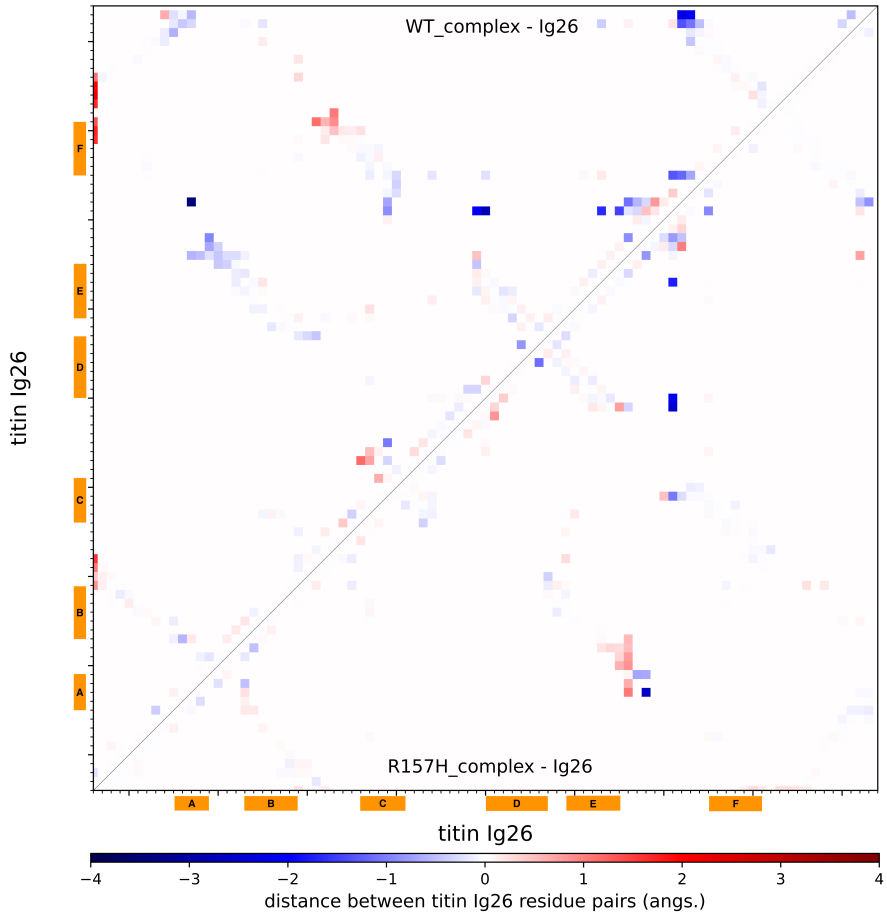


Figure 5.10: The differences in intramolecular residue-residue distances within titin Ig26 molecules. The dots in the upper left triangle indicate the distance differences between titin Ig26 in the WT complex and free titin Ig26, while the lower right triangle illustrates the differences between titin Ig26 in the R157H complex and free titin Ig26. The distance difference for a given residue pair is calculated only if the average distance of that pair in at least one structural configuration is less than 4 Å, and the cutoff boundary for the distance difference is established as -4~4 Å.

averaged instantaneous force applied to the titin Ig26 domain during the simulations, and performed t-tests between the values of the force applied to peptide-bound and free Ig domains(fig5.11). The overall trends of the instantaneous force for all three types of Ig domain pertain generally identical, as all requiring a notably great force to unfold the structures at approximately 0.5 ns and 4 ns. The most notable difference between the WT complex and the free Ig26 domain located between 6 and 8 ns, as most p-values within this period are below 0.05, indicating that the instantaneous force for the WT complex are persistently greater. The purple line illustrates that the force exerted on the Ig domain in the R157H complex during

specific short time intervals, such as around 8 ns and 10 ns, is lower than that of the free Ig domain. However, between 6 and 8 ns, the accumulated work of the R157H complex remains significantly lower than that of the WT complex, and is not significantly different from the free Ig domain. This suggests that the R157H peptide can enhance the mechanical resistance of the Ig26 domain to which it binds, albeit less effectively than the WT peptide.

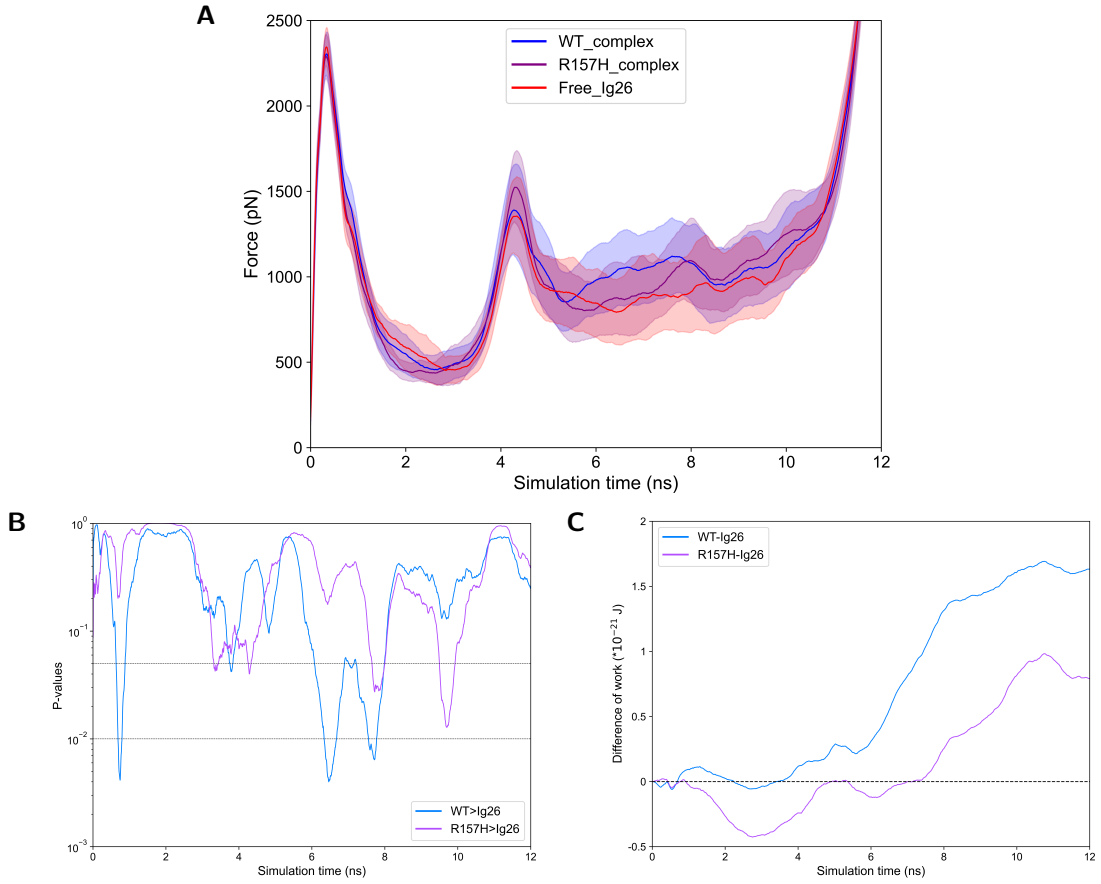


Figure 5.11: **A** The instantaneous force of the WT complex (depicted in blue), the R157H complex (purple), and the free titin Ig26 (red).The shaded regions indicate the error bars, with the single-sided width representing the standard deviation. **B** One sided p-values from t tests between the values of instantaneous force of the WT complex and the free titin Ig26 (depicted in blue), and between the R157H complex and the free titin Ig26 (purple). The alternative hypothesis is the instantaneous force applied to the peptide-bound Ig26 domain is greater than that for the free Ig domain. **C** The differences in accumulated work between the WT complex and the free titin Ig26 domain (blue) and between the R157H complex and the free titin Ig26 domain (purple). The calculations of work, which incorporate extension in their derivation, are not applicable beyond the time point at which the structure reaches its maximum length. Additionally, the instantaneous force is not associated with the mechano-resistance beyond this time point. Consequently, we only plot the values of instantaneous force and accumulate work within the initial 12 ns.

To investigate the underlying causes of the significant divergences observed between 6 and 8 ns, we calculated the average inter-residue distances across all configurations at each time point. Fig5.12 illustrates the intramolecular inter-residue contacts with a cutoff distance of 4 Å in all three variants of the Ig26 domains. At both time points, the majority of contacts, primarily located at strand C and the C-D loop, remain consistent irrespective of the binding state. The sole exception is the presence of a group of interacting residue pairs that is observed in both peptide-bound Ig26 domains at 6 ns, but not in the unbound variant. The overall high degree of correspondence among the residue contacts at 6 and 8 ns indicates that the titin Ig26, regardless of its interaction state, adheres to a conserved unfolding pathway between these time points in the simulations. Consequently, the differences in mechanical resistance among the various Ig26 variants cannot be attributed to distinct unfolding pathways; rather, they are likely a result of the protective effect exerted by the HspB5 CTD peptide.

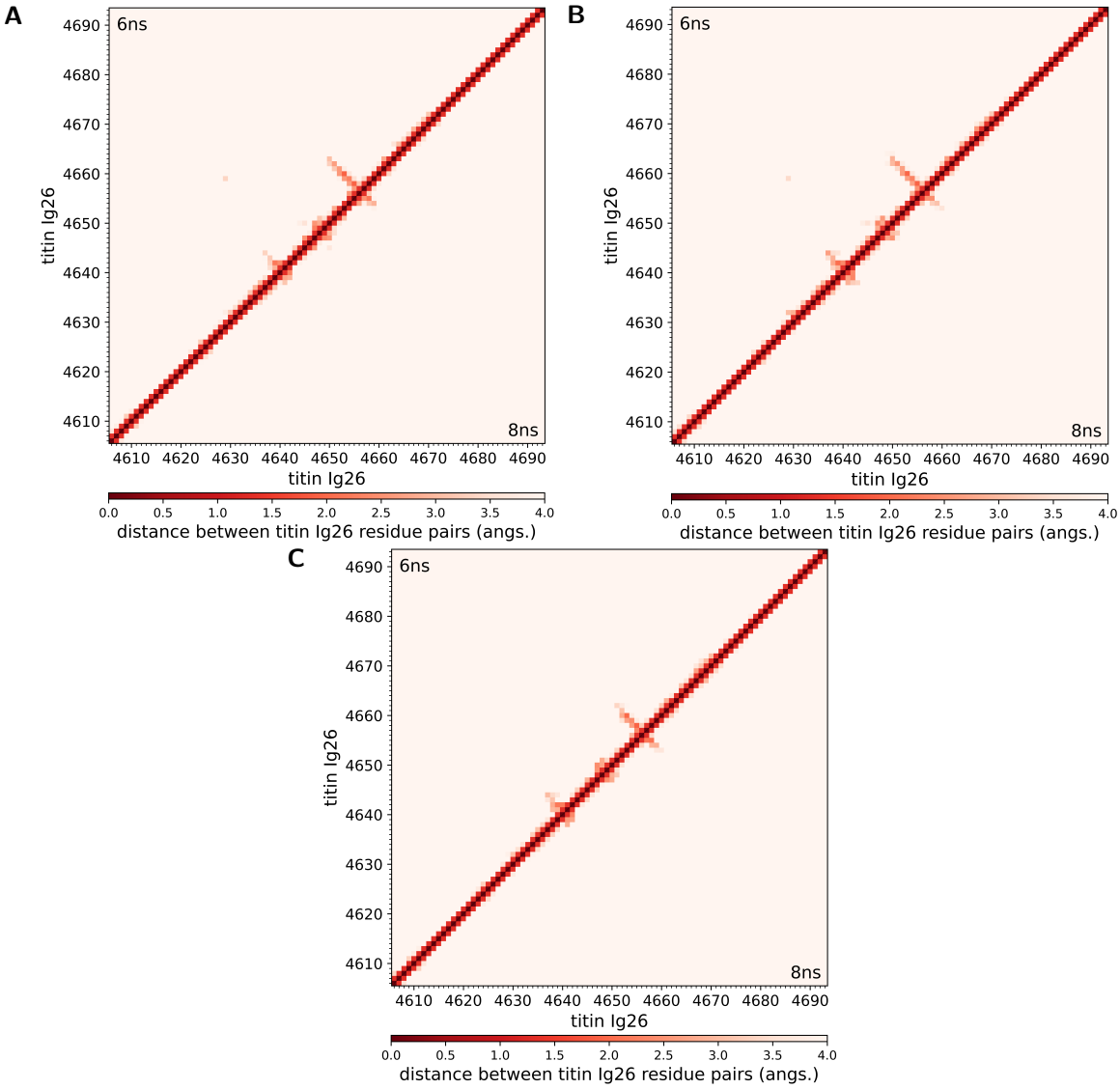


Figure 5.12: **A** The replicate-averaged intramolecular inter-residue distances in the titin Ig26 domain of the WT complex at 6 ns (upper left) and 8 ns (lower right) during the SMD simulations. **B** The corresponding inter-residue distances of the R157H complex. **C** The corresponding inter-residue distances of the free titin Ig26 domain.

Conformations of the binding interfaces in the WT and R157H complexes exhibit significant variability, thereby influencing the unfolding of the Ig domain within 6~8 ns during the simulations. At 6 ns, the interaction between Ig26 and the WT HspB5 CTD is notably strong, as the residues at both termini of the peptide are distinctly associated with the C-D loop and strand D of Ig26 (fig5.13A). These contacts are relatively weakened by 8 ns, nevertheless without displaying a sign of utterly disruption. In the R157H mutated complex, the interactions between titin Ig26 and the peptide at 6 ns are considerably weaker compared to the WT complex at the same time point, with only a few residues in the central region of the peptide binding to strand D of Ig26, while no intense interactions with the C-D loop (fig5.13B). Furthermore, at 8 ns, the residue contacts between the peptide and strand D remain relatively weak. The notable differences in residue interactions observed during the 6~8 ns interval between the WT and R157H complexes, which can be directly visualised through conformational snapshots of the complexes at corresponding time points, as illustrated in fig5.14, contribute to the variations in force and work exerted during this period.

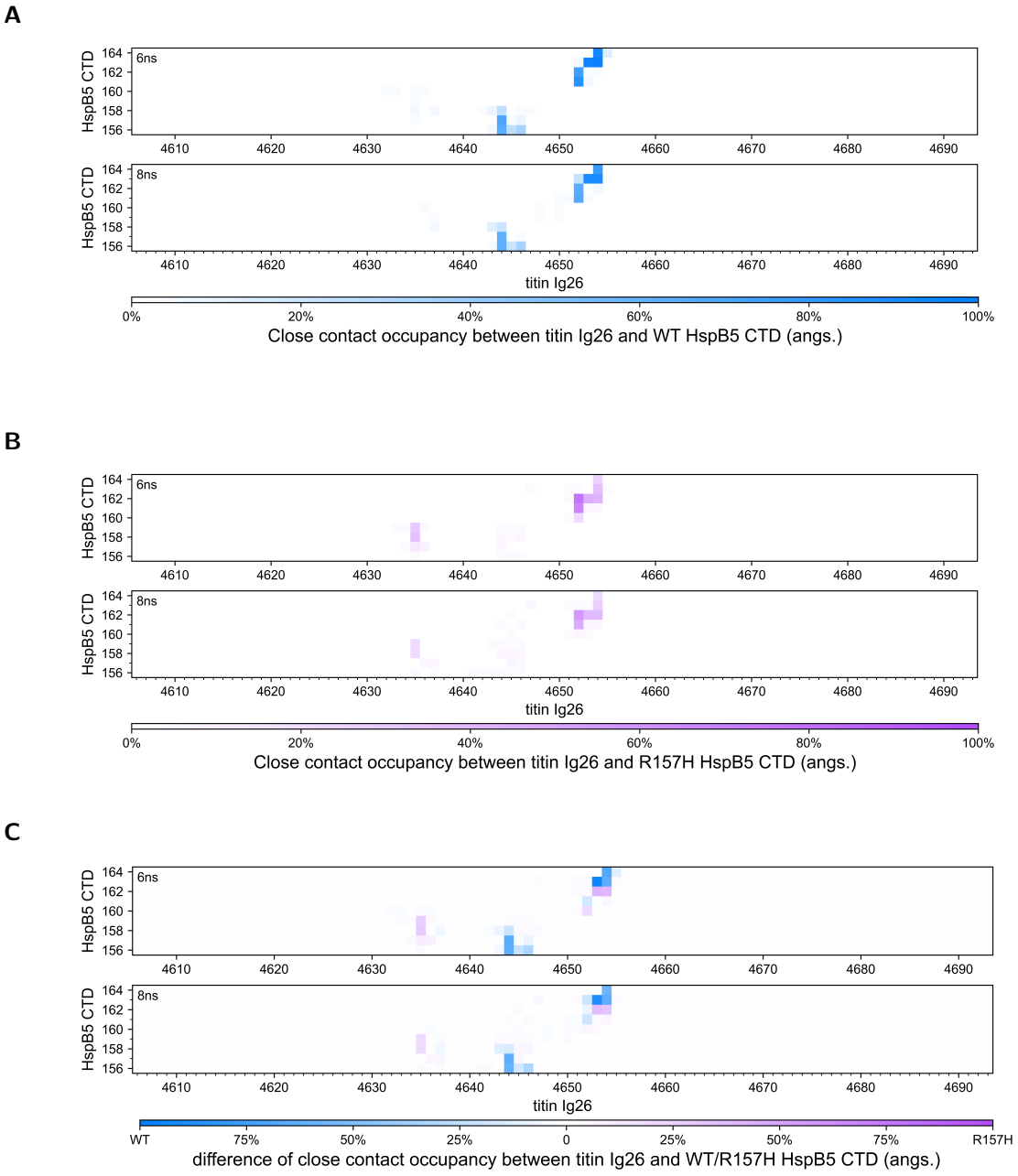


Figure 5.13: **A** The replicate-averaged intermolecular inter-residue distances in the WT complex at 6 ns (upper) and 8 ns (lower) within the simulations. **B** Corresponding intermolecular inter-residue distances in the R157H complex. **C** The differences between the replicate-averaged intermolecular residue-residue distances in the WT and R157H complexes at 6 ns (upper) and 8 ns (lower) within the simulations.

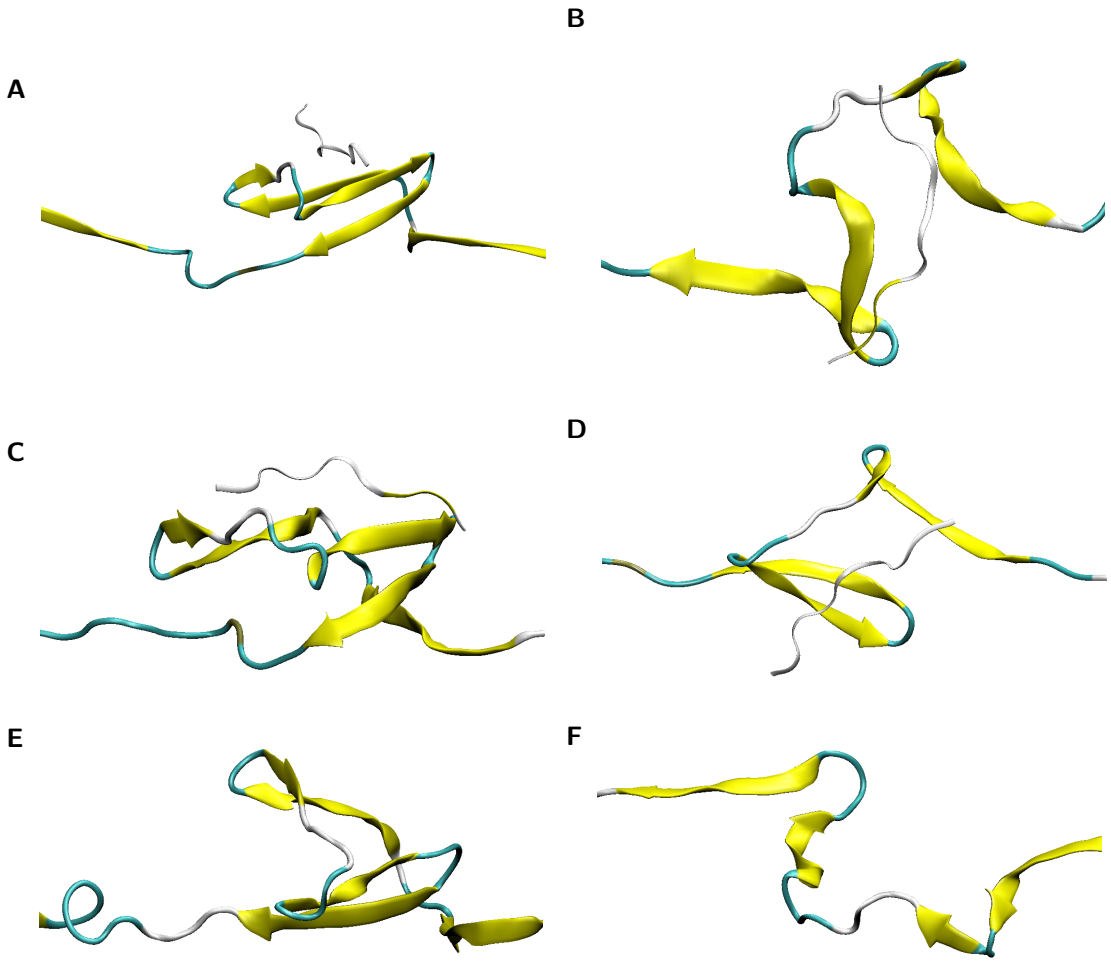


Figure 5.14: Conformational snapshots of the folded regions within the titin Ig26 domain, derived from the most representative conformations obtained during the SMD simulations. The snapshots correspond from top to bottom to the WT complex, the R157H complex, and the free Ig26 domain, while the images on the left refer to conformations at 6 ns and those on the right illustrate conformations at 8 ns.

In addition, we assessed the damage to inter-residue contacts during the steered MD simulations by calculating instantaneous inter-residue distances and comparing these distances with those observed in static states, as described in Chapter2.4. For each residue pair exhibiting close contact in the equilibrated MD simulations, as illustrated in fig5.8, we recorded the final frame in which a contact is detectable within the steered MD trajectory, designating this time point as the "breaking point" for the residue pair. Fig5.15 demonstrates that, as a consequence of the R157H mutation, inter-residue contacts involving H157, as well as adjacent residues, and titin Ig26 display reduced resistance to pulling forces and sustain damage at an earlier stage of the simulations compared to interactions at analogous positions in the WT complexes.

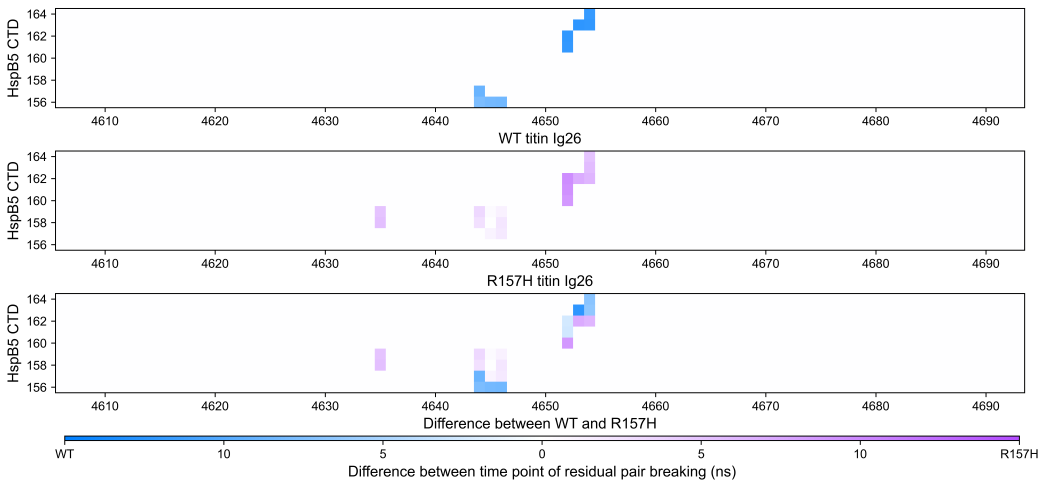


Figure 5.15: The replicate-averaged breaking points of intermolecular inter-residue contacts in the WT complex (top), the R157H complex (centre), and the differences between the two complexes (bottom). The breaking point for a given residue pair is calculated only if the corresponding distance between the residues observed in the equilibrated state is less than 4 Å; otherwise, the breaking point is recorded as 0 ns in the graph. Additionally, if a inter-residue contact remains intact at the end of the simulation, it is denoted as 15 ns (the duration of the simulation) in the graph.

The breaking points of intramolecular residue contacts in titin Ig26 (displayed in fig5.16), distinct from intermolecular contacts, display highly similar shapes and colour distributions, regardless of whether the protein is unbound or complexed with either the WT or the R157H mutated peptide. This observation suggests that the

titin Ig26 domains adhere to a comparable unfolding pathway, independent of their ligand-binding state. Furthermore, it implies that the affinities of intramolecular interactions within titin Ig26 are not significantly affected by the presence of the HspB5 CTD peptide. Consequently, the enhancement of mechanical resistance in titin Ig26 upon binding to HspB5 CTD, as well as the reduction of this effect in the presence of the R157H mutation, can be primarily attributed to intermolecular interactions rather than the modulation of non-covalent interactions or conformational transformations within the Ig domain.

5.5 Conclusions

In this chapter, we examined the regulatory effect of HspB5 on the mechanical properties of the immunoglobulin-like domain Ig26 of titin, a colossal cardiovascular protein that is essential for sarcomeric function. Titin is present in all types of muscle tissue; however, the Ig26 domain is situated within the N2B region, which is characterised by its high elasticity and contributes significantly to the mechanical resistance of the entire protein. Notably, this domain is exclusively expressed in striated muscle. Prior research has demonstrated that HspB5 can bind to the N2B region of titin, and this interaction enhances the mechanical resistance of the region. Consequently, dysregulation of HspB5 expression is identified as a primary factor contributing to titin-related cardiomyopathy. Furthermore, the findings indicate that the binding site for HspB5 is located within the CTD peptide, and the R157H mutation in this peptide diminishes the enhancing effect on the mechanical resistance of the Ig26 domain.

Laboratory native-MS experiments conducted by my colleagues indicate that the Ig26 domain, in conjunction with the N2B-U region, constitutes one of the binding sites within the N2B region. Furthermore, the interaction with the HspB5 CTD peptide remains largely unaffected by the R157H mutation. The results of the pulling test demonstrate that the mechanical resistance of the Ig26 domain when bound to the R157H HspB5 CTD peptide is lower than that observed with WT peptide, yet higher than that of an unbound Ig26 domain.

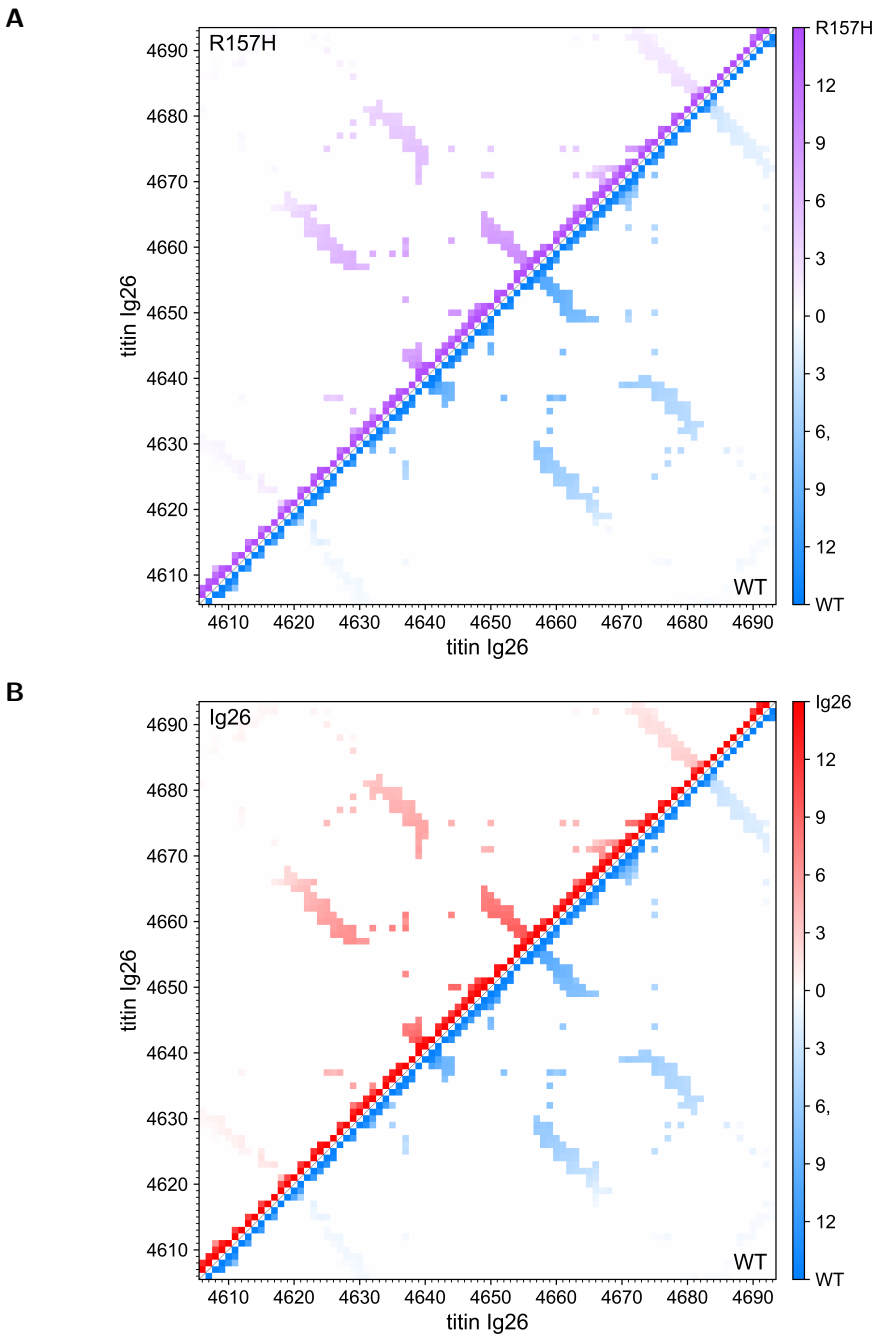


Figure 5.16: The replicate-averaged breaking points of intramolecular inter-residue contacts in the WT complex (top), the R157H complex (centre), and the free titin Ig26 domain(bottom).

Due to the palindromic sequence of the HspB5 CTD peptide, both *in vitro* and *in silico* experiments face challenges in distinguishing the orientation of the peptide. Consequently, it remains uncertain whether the R-H mutation observed *in vitro* or simulated *in silico* corresponds to the R157H mutation that occurs *in vivo*. To address this issue, we modelled two mutated peptides, R157H and R163H,

and conducted separate docking experiments with the Ig26 domain. For both models, we employed steered MD simulations to assess the impact of each mutation and to determine which one aligns with the *in vivo* findings. The results from ten repeat SMD tests indicate that the R157H mutation significantly diminishes the mechanical resistance of the Ig26 domain. In contrast, the domain associated with the R163H-mutated peptide exhibits mechanical resistance comparable to that of the WT peptide when subjected to external forces. Therefore, it can be concluded that the positioning of the R-H mutated residue in the R157H complexes generated *in vitro* or *in silico* is consistent with the *in vivo* complex, as the R157H peptide consistently serves as the source of reduced mechano-resistance in the Ig domain across all experimental conditions.

Following the confirmation of the mutation site, we conducted classical MD simulations for complexes involving either the WT or R157H mutated peptide, as well as the free Ig26 domain, to investigate the conformations and residue-residue contacts. The RMSF and binding energy analyses indicate that the interaction between Ig26 and the R157H peptide is less pronounced compared to that of the WT complex. The simulation trajectories demonstrate that the interaction with the WT HspB5 CTD peptide induces a substantial structural transformation, whereas the conformation of the Ig26 domain bound to the R157H peptide remains highly similar to that of the free Ig26 domain. Furthermore, the binding positions of the WT and R157H HspB5 CTD peptides diverge significantly by the end of the simulations, with variations observed in the regions constituting the interaction interface. Consequently, we have determined that the interaction mechanisms of the titin Ig26 domain with the WT and R157H mutated peptides are distinct.

We prepared a sampling of models from three classical MD trajectories and subsequently performed SMD simulations using these models as initial structures. The analysis of instantaneous force and accumulated work as a function of time reveals that the Ig26 domain in the R157H complex exhibits lower mechanical resistance compared to the WT complex, yet demonstrates greater resistance than the free Ig26 domain. The distinctions between the WT and the other two

protein variants are most pronounced during the time interval of 6~8 ns within the simulations. At both 6 and 8 ns, the close contacts between Ig26 and the WT peptide are sufficient to stabilise the structure; conversely, in the R157H complex, there are fewer contacts at 6 ns, and minimal strong contacts are maintained at 8 ns. Consequently, the protective effect on the higher-order structure of the HspB5 CTD peptide is diminished due to the R157H mutation. Furthermore, a greater number of inter-residue contacts between titin Ig26 and the WT HspB5 CTD persist longer than those observed between the Ig26 domain and the R157H peptide. Nevertheless, despite the differences in various aspects between titin Ig26 and the respective complexes, the time series depicting the departure of interactions within any Ig26 domain are notably similar, suggesting that the unfolding process of the Ig26 domain, in contrast to the mechanical resistance, remains unaffected by the peptide to which it is bound.

6
Concluding remarks

Contents

6.1 Conclusion of the work 145
6.2 Future potential applications of our technique 148
6.2.1 Co-assembly of HspB1 and B6 148
6.2.2 The structure of HspB8 and its function in the sarcomere 148

6.1 Conclusion of the work

The primary inspiration of our research was to investigate the significance of sHSPs in relation to the normal functioning and dysfunction, specifically cardiomyopathy, of cardiac muscle proteins. Previous studies have demonstrated that multiple sHSPs are expressed in cardiac muscle tissue, thereby playing a crucial role in regulating the functions and behaviours of various types of sarcomeric proteins through protein-protein interactions. After reviewing the literature on projects examining these interactions, we focused on three specific types of interactions: the coassembly of HspB1 and B5, the interaction between FLNCd24 and HspB7, and the interaction between titin Ig26 and HspB5.

Inspired by various published projects[315, 316, 317], we set out to implement MD simulations of proteins and multi-protein complexes in collaboration with

laboratory experiments conducted by my colleagues and collaborators. This approach aimed to investigate three specific types of interactions involving sHSPs and/or sarcomeric proteins. Generally, computational simulations of protein structures and protein-protein complexes provide insights into conformations at the atomic and residue levels.

In Chapter 3, we presented our research on the heterogeneous interactions between HspB1 and HspB5, which facilitate the co-assembly of these two sHSPs. Our simulations indicate that the CTD regions of HspB5 can form stronger interactions with the ACDs of both sHSPs compared to the CTD of HspB1. This disparity results in unequal participation of the two sHSPs in the formation of hetero-oligomers, as observed in native-MS experiments. In our study, we have also introduced a sequence prediction for the LCA of the two proteins and employed MD simulations involving this hypothetical protein. These simulations, which validate our conclusions and help identify potential objectives for subsequent experiments, suggest that the concept of introducing the LCA may be applicable to the investigation of interactions across other sHSPs, as well as among other homologous proteins.

In Chapter 4, we investigated the interaction between the ACD region of HspB7 and FLNCd24, as well as its regulatory effects on the dimerisation of FLNCd24. Our simulations of phosphorylated structures revealed that the phosphorylation of T2677 shifts the equilibrium towards favouring the dimerisation of FLNCd24, which is essential for the normal functioning of FLNC as actin-linking proteins and is implicated in stress-induced cardiomyopathy. Conversely, the phosphorylation of Y2683 exhibits an opposing effect, suggesting a detrimental impact on cardiovascular tissue and potentially contributing to the development of cancerous tissue. The simulations of these phosphorylated protein-protein complexes, which are substituted with phosphomimetic mutants in native-MS experiments, facilitate the assessment of the accuracy of phosphomimetics. Furthermore, according to a massive amount of published research regarding the phosphorylation of sarcomeric proteins, we proposed that the simulations of phosphorylated structures can be broadly applied in future studies investigating the interactions between sHSPs and sarcomeric proteins.

In Chapter 5, we examined the mechano-sensitivity of titin Ig26 and its regulation by the CTD of HspB5. Titin Ig26 is an elastic region within the titin protein, undergoing reversible elongation in response to external forces, and demonstrates increased mechanical resistance when interacting with the CTD region of HspB5. However, the R157H mutation within the HspB5 CTD leads to a reduction in this mechanical resistance. MD simulations revealed that the R157H mutation reduces the binding interfaces within the complex, resulting in a lower force requirement to disrupt the interactions compared to the WT complex. This study builds upon previous research that employed steered MD simulations to investigate the unfolding of sarcomeric proteins, thereby broadening the scope of research to include the computational simulation of the mechano-sensitivity of sHSPs-bound sarcomeric proteins, which may be explored in future investigations. Furthermore, classical MD simulations of both the WT and R157H complexes indicate that the docking positions predicted by the AlphaFold algorithm may not represent the optimal configurations, as the R157H complex exhibits a high degree of similarity with the AlphaFold predictions for both structures, whereas the WT complex demonstrates a larger binding interface and higher binding energy throughout the simulations.

In summary, we conducted MD simulations to investigate various types of protein-protein interactions involving sHSPs and/or sarcomeric proteins, which are essential for the normal functioning of cardiac muscle tissue. Our findings elucidate how modifications at the residue and atomic levels of interacting proteins influence these interactions, thereby affecting the function and behaviour of cardiac muscle in both beneficial and pathological contexts. Furthermore, our research confirms the effectiveness and distinctiveness of MD simulations in the study of sHSPs and cardiovascular proteins, particularly in simulating protein structures and protein-protein interactions that are modulated by protein modifications. The multiple methods and strategies involving MD simulations, as well as the scripts and documentation employed in these projects presented in this thesis, have the potential to be effective in future research on the interactions of sHSPs and cardiovascular proteins, as well as other related topics.

6.2 Future potential applications of our technique

6.2.1 Co-assembly of HspB1 and B6

Following the established pattern of co-assembly between HspB1 and B5, it has been documented that HspB6 can also co-assemble with HspB1 in human striated muscle tissue[196]. However, in contrast to HspB1 and B5, human HspB6 has not yet been shown to form large homo-oligomers [318]. A plausible explanation for this phenomenon is the evolutionary loss of the I-X-I motif in HspB6, which may diminish the interactions between its ACD and CTD. Consequently, we hypothesise that the co-assembly of HspB1 and B6 is unlikely to be unbiased, as the interaction between the HspB1 ACD and the HspB6 CTD is also expected to be weak. According to the MS and Mass photometry (MP) results for the mixture of ACD and CTD chains of HspB1 and B6, D. Saman has demonstrated that the co-assembly of HspB1 and B6 favours the heterogeneous ACD-ACD interaction, as the ensemble predominantly exhibits a 1:1 ratio in the mixture of the two sHSPs. However, quantifying the preference for interaction in in vitro experiments presents challenges, as the concentration measurements of mixed HspB1 and B6 are susceptible to inaccuracies due to their differing molar absorption coefficients. Therefore, we propose that it is essential to further investigate the ACD-CTD and ACD-ACD complexes formed by HspB1 and B6 segments through MD simulations to accurately quantify the binding affinity between the two proteins.

6.2.2 The structure of HspB8 and its function in the sarcomere

HspB8 is expressed in cardiac muscle tissue[307, 319, 320] and is involved in CASA[88], which is responsible for the degradation of irreversibly deformed and aggregated sarcomeric proteins. My colleague, Z. Wang, has observed that HspB8 predominantly exists in a monomeric form, despite the comparable lengths of its β5 and β6+7 strands to those of HspB1 and B5, which are significantly longer than those of HspB7. Consequently, we hypothesise that the monomeric state of HspB8 may arise from other sequential or structural characteristics, which could

be further explored through MD simulations. Additionally, Z. Wang has identified that despite that neither the ACD nor the ACD-CTD pair exhibits a stabilised conformation *in vitro*, HspB8 is capable of binding to other sHSPs, including HspB5 and B7, through the interaction involving the β4-β8 hydrophobic groove within the ACD of HspB8 and heterogeneous CTD peptides. Furthermore, as a component of the CASA protein complexes, HspB8 interacts with Bcl-2-associated athanogene 3 (BAG3) and serves as a sensor for unfolded filamin proteins [321]. Given the low conformational stability of HspB8 and the structural complexity of CASA protein complexes, we propose to utilise MD simulations to analyse the structural dynamics of protein complexes and the protein-protein interactions involving HspB8.

A
Appendix

ACD-CTD		HspB1-HspB1		HspB1-HspB5	
Binding energy (kcal/mol)		Average	Std Dev	Average	Std Dev
Complex 1	repeat 1	-7.8124	7.8430	-20.7611	7.5368
	repeat 2	-8.4450	8.5678	-15.2780	8.6271
	repeat 3	-6.3393	7.9682	-19.7484	8.4602
	repeat 4	-7.9634	9.2244	-12.0117	7.3851
	repeat 5	-5.5801	8.2648	-15.3699	7.8860
	overall	-7.2280	8.4579	-16.6338	8.6195
Complex 2	repeat 1	-5.8916	10.8019	-20.5199	6.8460
	repeat 2	-9.1099	6.3207	-19.0625	9.4401
	repeat 3	-10.6042	6.6627	-16.9003	9.3322
	repeat 4	-4.7139	7.2951	-20.7244	7.4968
	repeat 5	-7.9871	6.4182	-17.2428	9.0403
	overall	-7.6613	7.9765	-18.8900	8.6455
overall		-7.4447	8.2235	-17.7619	8.7559
ACD-CTD		HspB5-HspB1		HspB5-HspB5	
Binding energy (kcal/mol)		Average	Std Dev	Average	Std Dev
Complex 1	repeat 1	-10.7499	7.2029	-22.3453	10.0274
	repeat 2	-8.0980	6.4078	-10.8979	8.6513
	repeat 3	-12.9596	6.9700	-14.4561	7.7996
	repeat 4	-6.2026	6.4219	-17.4021	8.2575
	repeat 5	-9.4781	6.3564	-16.0273	9.0151
	overall	-9.5776	7.0289	-16.2257	9.5507
Complex 2	repeat 1	-11.5192	6.6346	-20.8864	8.7840
	repeat 2	-11.4516	6.7071	-19.3026	8.0285
	repeat 3	-13.9069	7.0531	-20.8214	10.0118
	repeat 4	-6.3039	6.6399	-16.1926	8.7622
	repeat 5	-9.5165	6.3159	-21.1472	9.0797
	overall	-10.5396	7.1393	-19.6700	9.1783
overall		-10.0586	7.1006	-17.9479	9.5223

Table A.1: Raw data of binding energies of ACD-CTD complexes consisting of HspB1 and B5 segments, calculated by mmPBSA.

ACD-CTD		B1-LCA		B5-LCA		LCA-B1		LCA-B5		LCA-LCA	
Binding energy (kcal/mol)		Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev
Complex 1	**repeat 1**	-20.9165	7.2360	-17.0696	7.6758	-3.2026	8.8751	-23.9222	15.7172	24.4619	13.2768
	repeat 2	-18.4527	6.8112	-16.1873	9.3034	-6.2114	7.1743	8.3444	10.4906	2.7878	12.2771
	repeat 3	-18.0521	6.6393	-14.8985	9.6868	-6.3580	6.3574	-11.9844	10.9829	-13.8540	11.8115
	repeat 4	-14.7734	8.9663	-19.1119	9.7789	-2.8307	5.1406	-10.1183	7.2798	14.9640	28.3910
	repeat 5	-20.3554	7.0966	-18.5140	8.4250	-4.4117	8.8059	11.5408	11.3938	27.3582	24.6398
	overall	-18.5100	7.7066	-16.6338	8.6195	-4.6029	7.5555	-5.2279	17.5782	11.1436	24.6141
Complex 2	**repeat 1**	-17.9412	9.1738	-17.1563	9.1392	-2.9451	7.3417	9.4450	7.8950	14.1904	15.3902
	repeat 2	-17.8294	7.3168	-15.3659	9.6503	-4.9485	7.3974	-7.7869	11.0576	-32.8226	18.2825
	repeat 3	-19.4876	7.1426	-19.6933	9.2813	-5.8703	10.4296	8.2745	12.6593	-3.0740	10.4645
	repeat 4	-19.1245	8.0757	-21.3577	10.3494	-5.2594	6.9523	-0.2424	7.4586	-2.8495	24.7664
	repeat 5	-13.3164	8.9159	-17.1177	9.0133	-3.6144	8.4340	-3.7692	9.8494	11.1470	19.4293
	overall	-17.5398	8.4594	-18.5705	9.8055	-4.5815	8.4057	1.1842	12.0255	-2.6817	24.7226
overall		-18.0249	8.1063	-17.8634	9.5045	-4.5922	7.9919	-2.0218	15.3975	4.2309	25.6187

Table A.2: Raw data of binding energies of ACD-CTD complexes consisting of LCA segments, calculated by mmPBSA.

Homodimer	WT		T2677D		pT2677		Y2683E		pY2683	
Repeat	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev
1	-52.2154	9.4828	-74.1533	9.3471	-79.4300	13.6293	-43.1721	10.3133	-22.1407	21.2308
2	-56.1555	9.2661	-76.4238	7.4727	-68.8612	11.5564	-50.0198	9.5996	8.4995	12.5827
3	-57.3198	9.3404	-65.6013	7.9915	-69.8602	9.5395	-48.4174	9.0697	-18.0476	21.3384
4	-55.7601	9.2133	-66.7805	9.0596	-69.3131	10.7671	-50.6013	9.4532	8.4579	11.3283
5	-58.5311	11.0643	-65.8948	9.3332	-64.6038	9.2469	-48.9994	10.5643	12.8076	12.2465
Overall	-55.9956	9.9278	-69.7707	9.8085	-71.8673	12.2803	-48.2426	10.1663	-2.0867	22.1104

Heterodimer	WT		T2677D		pT2677		Y2683E		pY2683	
Repeat	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev
1	-55.1136	8.4017	-45.7737	9.4489	-46.0025	8.4395	-64.3398	8.9920	-58.1488	12.0275
2	-59.5546	7.1746	-56.2708	10.9980	-47.6311	7.1208	-59.2673	8.5537	-66.9061	12.6452
3	-58.3464	6.7824	-42.0625	11.2447	-49.6951	8.1130	-59.6668	11.0068	-58.7774	9.1547
4	-57.9109	6.9521	-43.6163	8.8551	-47.8669	8.5014	-52.5055	10.0212	-60.7799	9.6592
5	-51.5689	7.1133	-50.4165	10.3672	-49.5492	7.8736	-59.3356	9.9693	-52.9679	11.5154
Overall	-56.4983	7.8476	-47.6290	11.4522	-48.1495	8.1400	-62.8252	9.4639	-59.5157	11.9639

Table A.3: Binding energies of WT and modified FLNCd24 homodimers and FLNCd24-HspB7ACD complexes

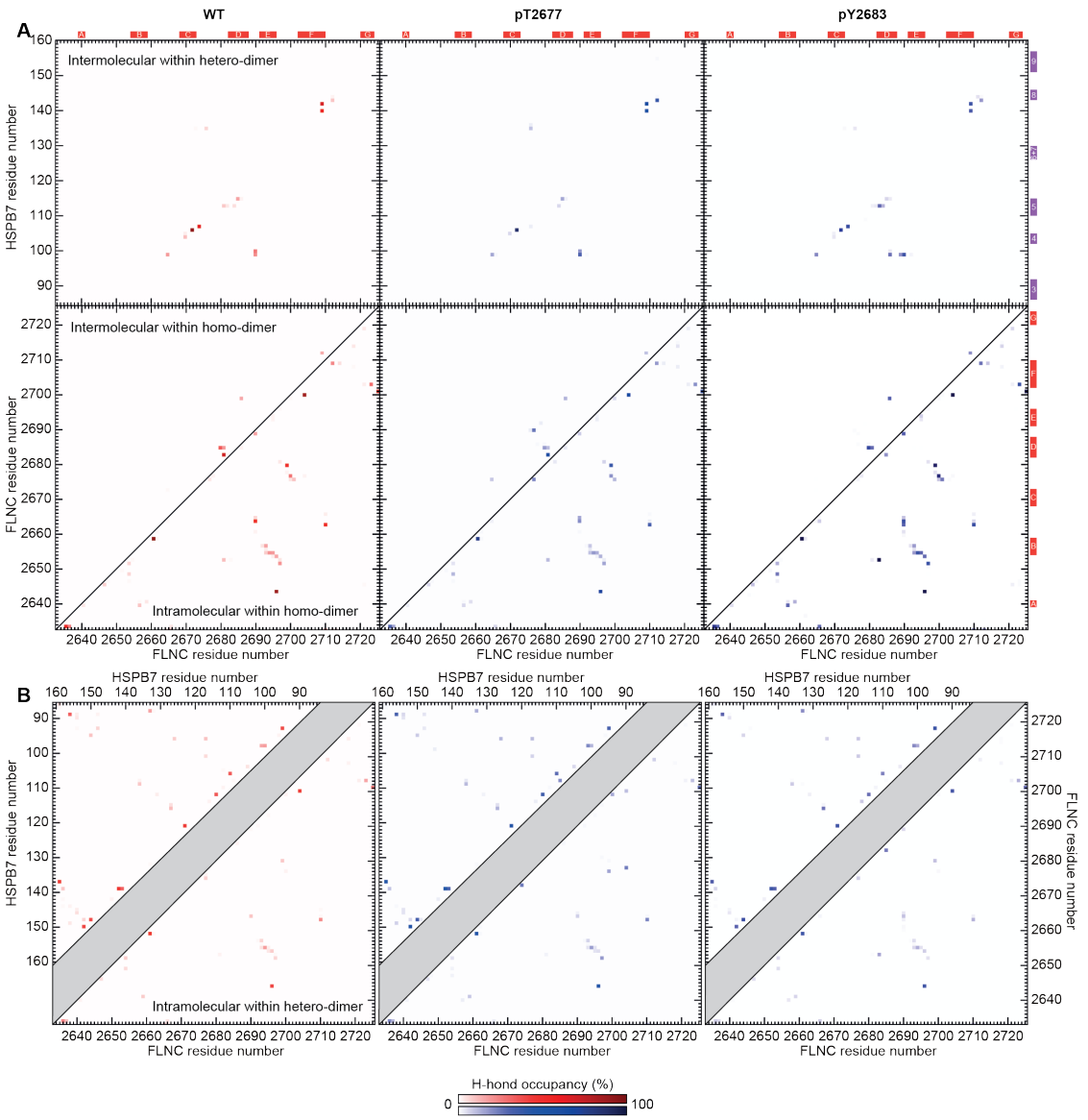


Figure A.1: **A** Distributions of probabilities of intermolecular hydrogen bonds in WT and phosphorylated FLNCd24-HspB7 ACD complexes (upper), and distributions of probabilities of intermolecular and intramolecular hydrogen bonds in WT and phosphorylated FLNCd24 homodimers (lower). **B** Distributions of probabilities of intramolecular hydrogen bonds in WT and phosphorylated FLNCd24-HspB7 ACD complexes.

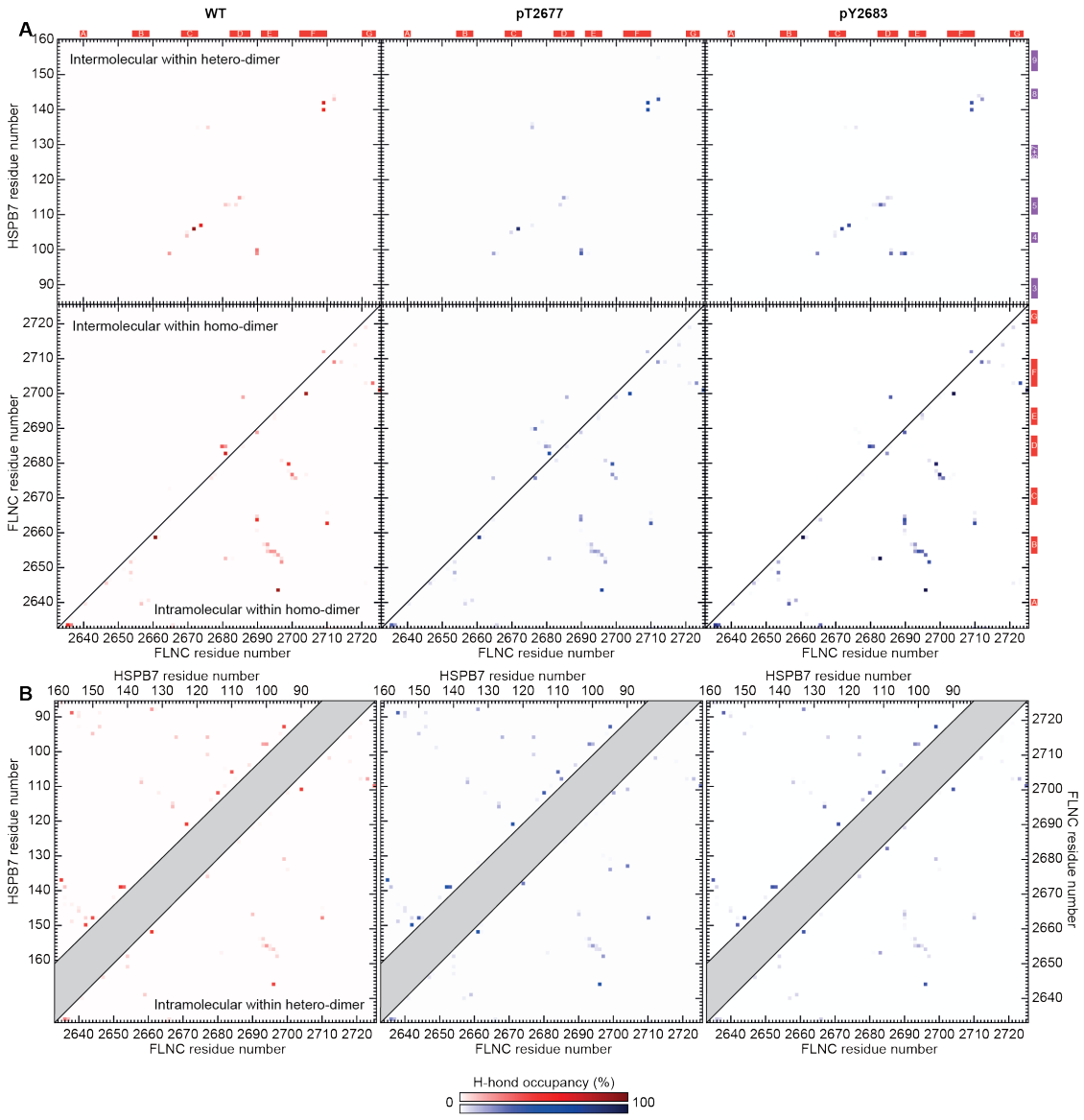


Figure A.2: **A** Differences between distributions of probabilities of intramolecular hydrogen bonds in WT and phosphorylated FLNCd24-HspB7 ACD complexes. **B** Differences between distributions of probabilities of intermolecular and intramolecular hydrogen bonds in WT and phosphorylated FLNCd24 homodimers.

ACD-CTD	WT		R157H	
Binding energy (kcal/mol)	**Average**	**Std Dev**	**Average**	**Std Dev**
repeat 1	-11.8987	8.4611	1.8218	9.7539
repeat 2	-15.4078	9.3218	-3.7492	8.9747
repeat 3	-15.8643	12.0598	-2.6268	8.9185
repeat 4	-12.4314	10.8301	4.5384	10.4610
repeat 5	-14.9907	8.8030	-6.1227	8.4709
overall	-14.0443	10.2176	-1.2240	10.1208

Table A.4: Raw data of binding energies between titin Ig26 and HspB5 CTD peptides, calculated by mmPBSA.

References

1. Brändén, C.-I. *Introduction to protein structure* Second Edition. eng (ed Tooze, J.) (Garland Publishing, New York, 2009).
2. Aguzzi, A. & O'Connor, T. Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nature Reviews Drug Discovery* **9,** 237–248. doi:10.1038/nrd3050 (2010).
3. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *Journal of Molecular Medicine* **81,** 678–699. doi:10.1007/s00109-003-0464-5 (2003).
4. Gregersen, N., Bolund, L. & Bross, P. Protein misfolding, aggregation, and degradation in disease. *Molecular Biotechnology* **31,** 141–150. doi:10.1385/MB:31:2:141 (2005).
5. Soto, C., Estrada, L. & Castilla, J. Amyloids, prions and the inherent infectious nature of misfolded protein aggregates. *Trends in Biochemical Sciences* **31,** 150–155. doi:10.1016/j.tibs.2006.01.002 (Mar. 2006).
6. Bross, P. *et al.* Protein misfolding and degradation in genetic diseases. *Human Mutation* **14,** 186–198. doi:https://doi.org/10.1002/(SICI)1098-1004(1999)14:3<186::AID-HUMU2>3.0.CO;2-J (Jan. 1999).
7. Del Monte, F. & Agnetti, G. Protein post-translational modifications and misfolding: New concepts in heart failure. *PROTEOMICS – Clinical Applications* **8,** 534–542. doi:https://doi.org/10.1002/prca.201400037 (Aug. 2014).
8. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. Forces contributing to the conformational stability of proteins. *The FASEB Journal* **10,** 75–83. doi:https://doi.org/10.1096/fasebj.10.1.8566551 (1996).
9. Alberts, B. *Molecular biology of the cell.* Sixth edition / B... eng (eds Wilson, J. H. & Hunt, T.) (Garland Science, New York, NY, 2015).
10. Anfinsen, C. B. The formation and stabilization of protein structure. *Biochemical Journal* **128,** 737–749. doi:10.1042/bj1280737 (July 1972).
11. Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426,** 900–904. doi:10.1038/nature02264 (2003).
12. Ellis, J. Proteins as molecular chaperones. *Nature* **328,** 378–379. doi:10.1038/328378a0 (1987).
13. Mashaghi, A., Kramer, G., Lamb, D. C., Mayer, M. P. & Tans, S. J. Chaperone Action at the Single-Molecule Level. *Chemical Reviews* **114,** 660–676. doi:10.1021/cr400326k (Jan. 2014).

14. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342,** 884–889. doi:10.1038/342884a0 (1989).

15. Ritossa, F. A new puffing pattern induced by temperature shock and DNP in drosophila. *Experientia* **18,** 571–573. doi:10.1007/BF02172188 (1962).

16. Lindquist, S. & Craig, E. A. The Heat-Shock Proteins. *Annual Review of Genetics* **22,** 631–677. doi:10.1146/annurev.ge.22.120188.003215 (Dec. 1988).

17. Burton, V., Mitchell, H. K., Young, P. & Petersen, N. S. Heat Shock Protection Against Cold Stress of Drosophila melanogaster. *Molecular and Cellular Biology* **8,** 3550–3552. doi:10.1128/mcb.8.8.3550-3552.1988 (Aug. 1988).

18. Matz, J. M., Blake, M. J., Tatelman, H. M., Lavoi, K. P. & Holbrook, N. J. Characterization and regulation of cold-induced heat shock protein expression in mouse brown adipose tissue. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **269,** R38–R47. doi:10.1152/ajpregu.1995.269.1.R38 (July 1995).

19. Simon, M. M. *et al.* Heat shock protein 70 overexpression affects the response to ultraviolet light in murine fibroblasts. Evidence for increased cell viability and suppression of cytokine release. *The Journal of Clinical Investigation* **95,** 926–933. doi:10.1172/JCI117800 (Mar. 1995).

20. Laplante, A. F. *et al.* Expression of Heat Shock Proteins in Mouse Skin During Wound Healing. *Journal of Histochemistry & Cytochemistry* **46,** 1291–1301. doi:10.1177/002215549804601109 (Nov. 1998).

21. Santoro, M. Heat shock factors and the control of the stress response. *Biochemical Pharmacology* **59,** 55–63. doi:https://doi.org/10.1016/S0006-2952(99)00299-3 (2000).

22. Murshid, A., Gong, J. & Calderwood, S. The Role of Heat Shock Proteins in Antigen Cross Presentation. *Frontiers in Immunology* **3** (2012).

23. Binder, R. J. Functions of Heat Shock Proteins in Pathways of the Innate and Adaptive Immune System. *The Journal of Immunology* **193,** 5765–5771. doi:10.4049/jimmunol.1401417 (Dec. 2014).

24. Benjamin, I. J. & McMillan, D. R. Stress (Heat Shock) Proteins. *Circulation Research* **83,** 117–132. doi:10.1161/01.RES.83.2.117 (July 1998).

25. Schlesinger, M. J. Heat shock proteins. *Journal of Biological Chemistry* **265,** 12111–12114. doi:10.1016/S0021-9258(19)38314-0 (July 1990).

26. Kampinga, H. H. *et al.* Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress and Chaperones* **14,** 105–111. doi:https://doi.org/10.1007/s12192-008-0068-7 (2009).

27. Collier, M. P. & Benesch, J. L. P. Small heat-shock proteins and their role in mechanical stress. *Cell Stress and Chaperones* **25,** 601–613. doi:https://doi.org/10.1007/s12192-020-01095-z (2020).

28. Haslbeck, M., Weinkauf, S. & Buchner, J. Small heat shock proteins: Simplicity meets complexity. *Journal of Biological Chemistry* **294,** 2121–2132. doi:https://doi.org/10.1074/jbc.REV118.002809 (2019).

29. Mogk, A. *et al.* Refolding of Substrates Bound to Small Hsps Relies on a Disaggregation Reaction Mediated Most Efficiently by ClpB/DnaK. *Journal of Biological Chemistry* **278,** 31033–31042. doi:https://doi.org/10.1074/jbc.M303587200 (2003).

30. Latchman, D. S. Heat shock proteins and cardiac protection. *Cardiovascular Research* **51,** 637–646. doi:10.1016/S0008-6363(01)00354-6 (Sept. 2001).

31. Snoeckx, L. H. E. H., Cornelussen, R. N., Van Nieuwenhoven, F. A., Reneman, R. S. & Van der Vusse, G. J. Heat Shock Proteins and Cardiovascular Pathophysiology. *Physiological Reviews* **81,** 1461–1497. doi:10.1152/physrev.2001.81.4.1461 (Jan. 2001).

32. Benesch, J. L. P., Ayoub, M., Robinson, C. V. & Aquilina, J. A. Small Heat Shock Protein Activity Is Regulated by Variable Oligomeric Substructure. *Journal of Biological Chemistry* **283,** 28513–28517. doi:10.1074/jbc.M804729200 (Oct. 2008).

33. Haslbeck, M. sHsps and their role in the chaperone network. *Cellular and Molecular Life Sciences CMLS* **59,** 1649–1657. doi:10.1007/PL00012492 (2002).

34. Taylor, R. P. & Benjamin, I. J. Small heat shock proteins: a new classification scheme in mammals. *Journal of Molecular and Cellular Cardiology* **38,** 433–444. doi:https://doi.org/10.1016/j.yjmcc.2004.12.014 (2005).

35. Chowdary, T. K., Raman, B., Ramahrishna, T. & Rao, C. M. Mammalian Hsp22 is a heat-inducible small heat-shock protein with chaperone-like activity. *Biochemical Journal* **381,** 379–387. doi:10.1042/BJ20031958 (July 2004).

36. Acunzo, J., Katsogiannou, M. & Rocchi, P. Small heat shock proteins HSP27 (HspB1), αB-crystallin (HspB5) and HSP22 (HspB8) as regulators of cell death. *The International Journal of Biochemistry & Cell Biology* **44,** 1622–1631. doi:https://doi.org/10.1016/j.biocel.2012.04.002 (2012).

37. Penke, B. *et al.* Heat Shock Proteins and Autophagy Pathways in Neuroprotection: From Molecular Bases to Pharmacological Interventions. *International Journal of Molecular Sciences* **19.** doi:10.3390/ijms19010325 (2018).

38. Wilhelmus, M. M. M. *et al.* Specific association of small heat shock proteins with the pathological hallmarks of Alzheimer's disease brains. *Neuropathology and Applied Neurobiology* **32,** 119–130. doi:https://doi.org/10.1111/j.1365-2990.2006.00689.x (Apr. 2006).

39. Bruinsma, I. B. *et al.* Inhibition of α-synuclein aggregation by small heat shock proteins. *Proteins: Structure, Function, and Bioinformatics* **79,** 2956–2967. doi:https://doi.org/10.1002/prot.23152 (Oct. 2011).

40. Rothbard, J. B. *et al.* Chaperone Activity of α B-Crystallin Is Responsible for Its Incorrect Assignment as an Autoantigen in Multiple Sclerosis. *The Journal of Immunology* **186,** 4263–4268. doi:10.4049/jimmunol.1003934 (Apr. 2011).

41. Golenhofen, N., Perng, M. D., 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. doi:10.1007/s00418-004-0711-z (2000).

42. Dubińska-Magiera, M. *et al.* Contribution of small heat shock proteins to muscle development and function. *FEBS Letters* **588,** 517–530. doi:https://doi.org/10.1016/j.febslet.2014.01.005 (2014).

43. Janowska, M. K., Baughman, H. E., Woods, C. N. & Klevit, R. E. Mechanisms of Small Heat Shock Proteins. *Cold Spring Harbor Perspectives in Biology* **11,** a034025. doi:10.1101/cshperspect.a034025 (Oct. 2019).

44. Waters, E. R. & Vierling, E. Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist* **227,** 24–37. doi:https://doi.org/10.1111/nph.16536 (July 2020).

45. Sun, Y. & MacRae, T. H. Small heat shock proteins: molecular structure and chaperone function. *Cellular and Molecular Life Sciences CMLS* **62,** 2460–2476. doi:10.1007/s00018-005-5190-4 (2005).

46. Giese, K. C. & Vierling, E. Changes in Oligomerization Are Essential for the Chaperone Activity of a Small Heat Shock Protein in-vivo and in-vitro. *Journal of Biological Chemistry* **277,** 46310–46318. doi:10.1074/jbc.M208926200 (Nov. 2002).

47. Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Science* **19,** 1031–1043. doi:https://doi.org/10.1002/pro.380 (May 2010).

48. Jaya, N., Garcia, V. & Vierling, E. Substrate binding site flexibility of the small heat shock protein molecular chaperones. *Proceedings of the National Academy of Sciences* **106,** 15604–15609. doi:10.1073/pnas.0902177106 (Sept. 2009).

49. Gusev, N. B., Bogatcheva, N. V. & Marston, S. B. Structure and Properties of Small Heat Shock Proteins (sHsp) and Their Interaction with Cytoskeleton Proteins. *Biochemistry (Moscow)* **67,** 511–519. doi:10.1023/A:1015549725819 (2002).

50. Delbecq, S. P., Jehle, S. & Klevit, R. Binding determinants of the small heat shock protein, αB-crystallin: recognition of the 'IxI' motif. *The EMBO Journal* **31,** 4587–4594. doi:https://doi.org/10.1038/emboj.2012.318 (Dec. 2012).

51. Bagnéris, C. *et al.* Crystal Structures of α-Crystallin Domain Dimers of αB-Crystallin and Hsp20. *Journal of Molecular Biology* **392,** 1242–1252. doi:https://doi.org/10.1016/j.jmb.2009.07.069 (2009).

52. 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* **108,** 6409–6414. doi:10.1073/pnas.1014656108 (Apr. 2011).

53. Basha, E., Friedrich, K. L. & Vierling, E. The N-terminal Arm of Small Heat Shock Proteins Is Important for Both Chaperone Activity and Substrate Specificity. *Journal of Biological Chemistry* **281,** 39943–39952. doi:10.1074/jbc.M607677200 (Dec. 2006).

54. Arrigo, A.-P. Human small heat shock proteins: Protein interactomes of homo- and hetero-oligomeric complexes: An update. *FEBS Letters* **587,** 1959–1969. doi:https://doi.org/10.1016/j.febslet.2013.05.011 (2013).

55. Mymrikov, E. V., Seit-Nebi, A. S. & Gusev, N. B. Heterooligomeric complexes of human small heat shock proteins. *Cell Stress and Chaperones* **17,** 157–169. doi:10.1007/s12192-011-0296-0 (2012).

56. Den Engelsman, J. *et al.* The Small Heat-Shock Proteins HSPB2 and HSPB3 Form Well-defined Heterooligomers in a Unique 3 to 1 Subunit Ratio. *Journal of Molecular Biology* **393,** 1022–1032. doi:https://doi.org/10.1016/j.jmb.2009.08.052 (2009).

57. Campbell, N. A., Reece, J. B., Taylor, M. R., Simon, E. J. & Dickey, J. L. *Biology: Concepts and Connections* (Benjamin-Cummings Publishing Company, 2009).

58. Van der Velden, J. & Stienen, G. J. M. Cardiac disorders and pathophysiology of sarcomeric proteins. *Physiological reviews* **99,** 381–426 (2019).

59. Gresham, K. S., Mamidi, R., Li, J., Kwak, H. & Stelzer, J. E. Sarcomeric protein modification during adrenergic stress enhances cross-bridge kinetics and cardiac output. *Journal of applied physiology* **122,** 520–530 (2017).

60. Walker, S. M. & Schrodt, G. R. I segment lengths and thin filament periods in skeletal muscle fibers of the rhesus monkey and the human. *The Anatomical Record* **178,** 63–81. doi:https://doi.org/10.1002/ar.1091780107 (Jan. 1974).

61. Rassier, D. E., MacIntosh, B. R. & Herzog, W. Length dependence of active force production in skeletal muscle. *Journal of Applied Physiology* **86,** 1445–1457. doi:10.1152/jappl.1999.86.5.1445 (May 1999).

62. Squire, J. M. Architecture and function in the muscle sarcomere. *Current Opinion in Structural Biology* **7,** 247–257. doi:https://doi.org/10.1016/S0959-440X(97)80033-4 (1997).

63. Huxley, H. & Hanson, J. Changes in the Cross-Striations of Muscle during Contraction and Stretch and their Structural Interpretation. *Nature* **173,** 973–976. doi:10.1038/173973a0 (1954).

64. Spudich, J. A. The myosin swinging cross-bridge model. *Nature Reviews Molecular Cell Biology* **2,** 387–392. doi:10.1038/35073086 (2001).

65. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science* **276,** 1109–1112. doi:10.1126/science.276.5315.1109 (May 1997).

66. Linke, W. A. *et al.* Towards a Molecular Understanding of the Elasticity of Titin. *Journal of Molecular Biology* **261,** 62–71. doi:https://doi.org/10.1006/jmbi.1996.0441 (1996).

67. Herzog, W. The role of titin in eccentric muscle contraction. *Journal of Experimental Biology* **217,** 2825–2833. doi:10.1242/jeb.099127 (Aug. 2014).

68. Richfield, D. *Medical gallery of David Richfield 2014* 2014. doi:10.15347/wjm/2014.009.

69. Lange, S., Ehler, E. & Gautel, M. From A to Z and back? Multicompartment proteins in the sarcomere. *Trends in Cell Biology* **16,** 11–18. doi:10.1016/j.tcb.2005.11.007 (Jan. 2006).

70. Kamisago, M. *et al.* Mutations in Sarcomere Protein Genes as a Cause of Dilated Cardiomyopathy. *New England Journal of Medicine* **343,** 1688–1696. doi:10.1056/NEJM200012073432304 (Dec. 2000).

71. Thierfelder, L. *et al.* α-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: A disease of the sarcomere. *Cell* **77,** 701–712. doi:10.1016/0092-8674(94)90054-X (June 1994).

72. Fatkin, D. & Graham, R. M. Molecular Mechanisms of Inherited Cardiomyopathies. *Physiological Reviews* **82,** 945–980. doi:10.1152/physrev.00012.2002 (Jan. 2002).

73. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry* **82,** 323–355. doi:10.1146/annurev-biochem-060208-092442 (June 2013).

74. Smith, D. A., Carland, C. R., Guo, Y. & Bernstein, S. I. Getting Folded: Chaperone Proteins in Muscle Development, Maintenance and Disease. *The Anatomical Record* **297,** 1637–1649. doi:https://doi.org/10.1002/ar.22980 (Sept. 2014).

75. Li, F., Xiao, H., Zhou, F., Hu, Z. & Yang, B. Study of HSPB6: Insights into the Properties of the Multifunctional Protective Agent. *Cellular Physiology and Biochemistry* **44,** 314–332. doi:10.1159/000484889 (Nov. 2017).

76. Mounier, N. & Arrigo, A.-P. Actin cytoskeleton and small heat shock proteins: how do they interact? *Cell Stress & Chaperones* **7,** 167. doi:10.1379/1466-1268(2002)007<0167:ACASHS>2.0.CO;2 (2002).

77. Golenhofen, N., Arbeiter, A., Koob, R. & Drenckhahn, D. Ischemia-induced association of the stress protein αB-crystallin with I-band portion of cardiac titin. *Journal of Molecular and Cellular Cardiology* **34,** 309–319. doi:10.1006/jmcc.2001.1513 (2002).

78. Van der Ven, P. F. M. *et al.* Unusual splicing events result in distinct Xin isoforms that associate differentially with filamin c and Mena/VASP. *Experimental Cell Research* **312,** 2154–2167. doi:https://doi.org/10.1016/j.yexcr.2006.03.015 (2006).

79. Hoffman, L., Jensen, C. C., Yoshigi, M. & Beckerle, M. Mechanical signals activate p38 MAPK pathway-dependent reinforcement of actin via mechanosensitive HspB1. *Molecular Biology of the Cell* **28,** 2661–2675. doi:10.1091/mbc.e17-02-0087 (Aug. 2017).

80. Yin, B. *et al.* CRYAB protects cardiomyocytes against heat stress by preventing caspase-mediated apoptosis and reducing F-actin aggregation. *Cell Stress and Chaperones* **24,** 59–68. doi:10.1007/s12192-018-0941-y (2019).

81. Koh, T. J. & Escobedo, J. Cytoskeletal disruption and small heat shock protein translocation immediately after lengthening contractions. *American Journal of Physiology-Cell Physiology* **286,** C713–C722. doi:10.1152/ajpcell.00341.2003 (Mar. 2004).

82. Unger, A. *et al.* Translocation of molecular chaperones to the titin springs is common in skeletal myopathy patients and affects sarcomere function. *Acta Neuropathologica Communications* **5,** 72. doi:10.1186/s40478-017-0474-0 (2017).

83. Seit-Nebi, A. S. & Gusev, N. B. Versatility of the small heat shock protein HSPB6 (Hsp20). *Cell Stress and Chaperones* **15,** 233–236. doi:10.1007/s12192-009-0141-x (2010).

84. Fan, G.-C. & Kranias, E. G. Small heat shock protein 20 (HspB6) in cardiac hypertrophy and failure. *Journal of Molecular and Cellular Cardiology* **51,** 574–577. doi:https://doi.org/10.1016/j.yjmcc.2010.09.013 (2011).

85. Sugiyama, Y. *et al.* Muscle Develops a Specific Form of Small Heat Shock Protein Complex Composed of MKBP/HSPB2 and HSPB3 during Myogenic Differentiation. *Journal of Biological Chemistry* **275,** 1095–1104. doi:https://doi.org/10.1074/jbc.275.2.1095 (2000).

86. Wang, Z. *et al.* Cardiac stress leads to regulation of Filamin C dimerisation via an ancient phosphorylation-modulated interaction with HSPB7. *bioRxiv,* 2024.01.05.574393. doi:10.1101/2024.01.05.574393 (Jan. 2024).

87. Juo, L.-Y. *et al.* HSPB7 interacts with dimerized FLNC and its absence results in progressive myopathy in skeletal muscles. *Journal of Cell Science* **129,** 1661–1670. doi:10.1242/jcs.179887 (Apr. 2016).

88. Cristofani, R. *et al.* The Role of HSPB8, a Component of the Chaperone-Assisted Selective Autophagy Machinery, in Cancer. *Cells* **10.** doi:10.3390/cells10020335 (2021).

89. Treweek, T. M., Meehan, S., Ecroyd, H. & Carver, J. A. Small heat-shock proteins: important players in regulating cellular proteostasis. *Cellular and Molecular Life Sciences* **72,** 429–451. doi:10.1007/s00018-014-1754-5 (2015).

90. Gaestel, M. in *Small Stress Proteins* (eds Arrigo, A.-P. & Müller, W. E. G.) 151–169 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2002). doi:10.1007/978-3-642-56348-5{_}8.

91. Collier, M. P. *et al.* HspB1 phosphorylation regulates its intramolecular dynamics and mechanosensitive molecular chaperone interaction with filamin C. *Science Advances* **5,** eaav8421 (2019).

92. Simon, S. *et al.* Analysis of the Dominant Effects Mediated by Wild Type or R120G Mutant of αB-crystallin (HspB5) towards Hsp27 (HspB1). *PLOS ONE* **8,** e70545– (Aug. 2013).

93. Dimauro, I., Antonioni, A., Mercatelli, N. & Caporossi, D. The role of αB-crystallin in skeletal and cardiac muscle tissues. *Cell Stress and Chaperones* **23,** 491–505. doi:10.1007/s12192-017-0866-x (July 2018).

94. Inagaki, N. *et al.* αB-crystallin mutation in dilated cardiomyopathy. *Biochemical and Biophysical Research Communications* **342,** 379–386. doi:https://doi.org/10.1016/j.bbrc.2006.01.154 (2006).

95. Kley, R. A. *et al.* A Combined Laser Microdissection and Mass Spectrometry Approach Reveals New Disease Relevant Proteins Accumulating in Aggregates of Filaminopathy Patients. *Molecular & Cellular Proteomics* **12,** 215–227. doi:https://doi.org/10.1074/mcp.M112.023176 (2013).

96. Ortiz-Genga, M. F. *et al.* Truncating FLNC Mutations Are Associated With High-Risk Dilated and Arrhythmogenic Cardiomyopathies. *Journal of the American College of Cardiology* **68,** 2440–2451. doi:https://doi.org/10.1016/j.jacc.2016.09.927 (2016).

97. Morelli, F. F. *et al.* Aberrant Compartment Formation by HSPB2 Mislocalizes Lamin A and Compromises Nuclear Integrity and Function. *Cell Reports* **20,** 2100–2115. doi:https://doi.org/10.1016/j.celrep.2017.08.018 (2017).

98. Weeks, S. D. *et al.* Characterization of human small heat shock protein HSPB1 α-crystallin domain localized mutants associated with hereditary motor neuron diseases. *Scientific Reports* **8,** 688. doi:10.1038/s41598-017-18874-x (2018).

99. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9,** 646–652. doi:10.1038/nsb0902-646 (2002).

100. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* doi:10.1093/oso/9780198803195.001.0001 (Oxford University Press, June 2017).

101. Van Gunsteren, W. F. & Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angewandte Chemie International Edition in English* **29,** 992–1023. doi:https://doi.org/10.1002/anie.199009921 (Sept. 1990).

102. MacKerell, A. D. J. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **102,** 3586–3616. doi:10.1021/jp973084f (Apr. 1998).

103. Lindahl, E. R. in *Molecular Modeling of Proteins* (ed Kukol, A.) 3–23 (Humana Press, Totowa, NJ, 2008). doi:10.1007/978-1-59745-177-2{_}1.

104. Tian, C. *et al.* ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **16,** 528–552. doi:10.1021/acs.jctc.9b00591 (Jan. 2020).

105. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4,** 187–217. doi:https://doi.org/10.1002/jcc.540040211 (June 1983).

106. Jones, J. E. & Chapman, S. On the determination of molecular fields. I. From the variation of the viscosity of a gas with temperature. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **106,** 441–462. doi:10.1098/rspa.1924.0081 (Jan. 1997).

107. Jones, J. E. & Chapman, S. On the determination of molecular fields. II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **106,** 463–477. doi:10.1098/rspa.1924.0082 (Jan. 1997).

108. J E Lennard-Jones. Cohesion. *Proceedings of the Physical Society* **43,** 461. doi:10.1088/0959-5309/43/5/301 (1931).

109. De Leeuw, S. W., Perram, J. W., Smith, E. R. & Rowlinson, J. S. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **373,** 27–56. doi:10.1098/rspa.1980.0135 (Jan. 1997).

110. De Leeuw, S. W., Perram, J. W., Smith, E. R. & Rowlinson, J. S. Simulation of electrostatic systems in periodic boundary conditions. II. Equivalence of boundary conditions. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **373,** 57–66. doi:10.1098/rspa.1980.0136 (Jan. 1997).

111. De Leeuw, S. W., Perram, J. W., Smith, E. R. & Rowlinson, J. S. Simulation of electrostatic systems in periodic boundary conditions. III. Further theory and applications. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **388,** 177–193. doi:10.1098/rspa.1983.0077 (Jan. 1997).

112. Perera, L., Essmann, U. & Berkowitz, M. L. Effect of the treatment of long-range forces on the dynamics of ions in aqueous solutions. *The Journal of Chemical Physics* **102,** 450–456. doi:10.1063/1.469422 (Jan. 1995).

113. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159,** 98–103. doi:10.1103/PhysRev.159.98 (July 1967).

114. Verlet, L. Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Physical Review* **165,** 201–214. doi:10.1103/PhysRev.165.201 (Jan. 1968).

115. Levesque, D. & Verlet, L. Computer "Experiments" on Classical Fluids. III. Time-Dependent Self-Correlation Functions. *Physical Review A* **2,** 2514–2528. doi:10.1103/PhysRevA.2.2514 (Dec. 1970).

116. Brünger, A. T. *X-PLOR, Version 3.1 : a system for X-ray crystallography and NMR* (Yale University Press, New Haven ; 1992).

117. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589. doi:10.1038/s41586-021-03819-2 (2021).

118. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234,** 779–815. doi:https://doi.org/10.1006/jmbi.1993.1626 (1993).

119. Simons, K. T., Bonneau, R., Ruczinski, I. & Baker, D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins: Structure, Function, and Bioinformatics* **37,** 171–176. doi:https://doi.org/10.1002/(SICI)1097-0134(1999)37:3+<171::AID-PROT21>3.0.CO;2-Z (Jan. 1999).

120. Rowlinson, J. S. The maxwell–boltzmann distribution. *Molecular Physics* **103,** 2821–2828 (2005).

121. Lutz, H. D., Eckers, W. & Haeuseler, H. OH stretching frequencies of solid hydroxides and of free OH ions. *Journal of Molecular Structure* **80,** 221–224. doi:https://doi.org/10.1016/0022-2860(82)87236-0 (1982).

122. Langevin, P. Sur la théorie du mouvement brownien [On the Theory of Brownian Motion]. *Comptes rendus de l'Académie des Sciences* **146,** 530–533 (1908).

123. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81,** 511–519. doi:10.1063/1.447334 (July 1984).

124. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31,** 1695–1697. doi:10.1103/PhysRevA.31.1695 (Mar. 1985).

125. Burley, S. K. *et al.* Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Research* **53,** D564–D574. doi:10.1093/nar/gkae1091 (Jan. 2025).

126. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26,** 1781–1802. doi:https://doi.org/10.1002/jcc.20289 (Dec. 2005).

127. Case, D. A. *et al.* AmberTools. *Journal of Chemical Information and Modeling* **63,** 6183–6191. doi:10.1021/acs.jcim.3c01153 (Oct. 2023).

128. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11,** 3696–3713. doi:10.1021/acs.jctc.5b00255 (Aug. 2015).

129. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **105,** 9954–9960. doi:10.1021/jp003020w (Nov. 2001).

130. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **32,** 922–923. doi:10.1107/S0567739476001873 (Sept. 1976).

131. Mathews, N. VMD User's Guide (2016).

132. Martínez, L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLOS ONE* **10,** e0119264– (Mar. 2015).

133. Stein, S. A. M., Loccisano, A. E., Firestine, S. M. & Evanseck, J. D. in *Annual Reports in Computational Chemistry* (ed Spellmeyer, D. C.) 233–261 (Elsevier, 2006). doi:https://doi.org/10.1016/S1574-1400(06)02013-5.

134. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2,** 559–572. doi:10.1080/14786440109462720 (Nov. 1901).

135. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374,** 20150202. doi:10.1098/rsta.2015.0202 (Apr. 2016).

136. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **10,** 449–461. doi:10.1517/17460441.2015.1032936 (May 2015).

137. Miller, B. R. I. I. I. *et al.* MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8,** 3314–3321. doi:10.1021/ct300418h (Sept. 2012).

138. Wang, E. *et al.* End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical Reviews* **119,** 9478–9508. doi:10.1021/acs.chemrev.9b00055 (Aug. 2019).

139. Guvench, O. & MacKerell, A. D. Computational evaluation of protein–small molecule binding. *Current Opinion in Structural Biology* **19,** 56–61. doi:https://doi.org/10.1016/j.sbi.2008.11.009 (2009).

140. Wells, D. B., Abramkina, V. & Aksimentiev, A. Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics. *The Journal of Chemical Physics* **127,** 125101. doi:10.1063/1.2770738 (Sept. 2007).

141. Genchev, G. Z. *et al.* Mechanical Signaling on the Single Protein Level Studied Using Steered Molecular Dynamics. *Cell Biochemistry and Biophysics* **55,** 141–152. doi:10.1007/s12013-009-9064-5 (2009).

142. Do, P.-C., Lee, E. H. & Le, L. Steered Molecular Dynamics Simulation in Rational Drug Design. *Journal of Chemical Information and Modeling* **58,** 1473–1482. doi:10.1021/acs.jcim.8b00261 (Aug. 2018).

143. Suan Li, M. & Khanh Mai, B. Steered molecular dynamics-a promising tool for drug design. *Current Bioinformatics* **7,** 342–351 (2012).

144. Decherchi, S. & Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chemical Reviews* **120,** 12788–12833. doi:10.1021/acs.chemrev.0c00534 (Dec. 2020).

145. Izrailev, S. *et al. Steered Molecular Dynamics* in *Computational Molecular Dynamics: Challenges, Methods, Ideas* (eds Deuflhard, P. *et al.*) (Springer Berlin Heidelberg, Berlin, Heidelberg, 1999), 39–65.

146. Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* **11,** 224–230. doi:https://doi.org/10.1016/S0959-440X(00)00194-9 (2001).

147. Patel, J. S., Berteotti, A., Ronsisvalle, S., Rocchia, W. & Cavalli, A. Steered Molecular Dynamics Simulations for Studying Protein–Ligand Interaction in Cyclin-Dependent Kinase 5. *Journal of Chemical Information and Modeling* **54,** 470–480. doi:10.1021/ci4003574 (Feb. 2014).

148. Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation. *Biophysical Journal* **75,** 662–671. doi:https://doi.org/10.1016/S0006-3495(98)77556-3 (1998).

149. Gao, M., Craig, D., Vogel, V. & Schulten, K. Identifying Unfolding Intermediates of FN-III10 by Steered Molecular Dynamics. *Journal of Molecular Biology* **323,** 939–950. doi:https://doi.org/10.1016/S0022-2836(02)01001-X (2002).

150. Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. & Schulten, K. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics and Modelling* **19,** 13–25. doi:https://doi.org/10.1016/S1093-3263(00)00133-9 (2001).

151. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophysical Journal* **72,** 1541–1555. doi:10.1016/S0006-3495(97)78802-7 (Apr. 1997).

152. Iida, S. & Kameda, T. Dissociation Rate Calculation via Constant-Force Steered Molecular Dynamics Simulation. *Journal of Chemical Information and Modeling* **63,** 3369–3376. doi:10.1021/acs.jcim.2c01529 (June 2023).

153. Park, S. & Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *The Journal of Chemical Physics* **120,** 5946–5961. doi:10.1063/1.1651473 (Apr. 2004).

154. Drenth, J. *Principles of Protein X-Ray Crystallography* (Springer New York, 2007).

155. Ilari, A. & Savino, C. in *Bioinformatics: Data, Sequence Analysis and Evolution* (ed Keith, J. M.) 63–87 (Humana Press, Totowa, NJ, 2008). doi:10.1007/978-1-60327-159-2{_}3.

156. Bragg, L. X-RAY CRYSTALLOGRAPHY. *Scientific American* **219,** 58–74 (1968).

157. Bragg, W. L. X-Ray Analysis of Proteins. *Proceedings of the Physical Society. Section B* **65,** 833. doi:10.1088/0370-1301/65/11/301 (1952).

158. Carter, C. W. & Sweet, R. M. *Macromolecular Crystallography* (Academic Press, 1997).

159. Smyth, M. S. & Martin, J. H. J. x Ray crystallography. *Molecular Pathology* **53,** 8 (2000).

160. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–475. doi:10.1038/355472a0 (1992).

161. Maveyraud, L. & Mourey, L. Protein X-ray crystallography and drug discovery. *Molecules* **25,** 1030 (2020).

162. Sparkman, O. D. *Mass Spectrometry Desk Reference* (Global View Pub., 2000).

163. Ghosh, P. K. in *Introduction to Protein Mass Spectrometry* (ed Ghosh, P. K.) 9–67 (Academic Press, Boston, 2016). doi:https://doi.org/10.1016/B978-0-12-805123-8.50002-2.

164. Ghosh, P. K. in *Introduction to Protein Mass Spectrometry* (ed Ghosh, P. K.) 68–134 (Academic Press, Boston, 2016). doi:https://doi.org/10.1016/B978-0-12-805123-8.50003-4.

165. Ghosh, P. K. in *Introduction to Protein Mass Spectrometry* (ed Ghosh, P. K.) 135–219 (Academic Press, Boston, 2016). doi:https://doi.org/10.1016/B978-0-12-805123-8.50004-6.

166. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207. doi:10.1038/nature01511 (2003).

167. McMaster, M. C. *LC/MS: A Practical User's Guide* (Wiley, 2005).

168. Niessen, W. M. A. *Liquid Chromatography-Mass Spectrometry* (CRC Press, 2006).

169. De Hoffmann, E. Tandem mass spectrometry: a primer. *Journal of mass spectrometry* **31,** 129–137 (1996).

170. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nature Methods* **16,** 595–602. doi:10.1038/s41592-019-0459-y (2019).

171. Leitner, A. *et al.* Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics*. *Molecular & Cellular Proteomics* **9,** 1634–1649. doi:https://doi.org/10.1074/mcp.R000001-MCP201 (2010).

172. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods* **5,** 927–933. doi:10.1038/nmeth.1265 (2008).

173. Chandler, S. A. & Benesch, J. L. P. Mass spectrometry beyond the native state. *Current opinion in chemical biology* **42,** 130–137 (2018).

174. Barth, M. & Schmidt, C. Native mass spectrometry—A valuable tool in structural biology. *Journal of Mass Spectrometry* **55,** e4578 (2020).

175. Tamara, S., den Boer, M. A. & Heck, A. J. R. High-resolution native mass spectrometry. *Chemical reviews* **122,** 7269–7326 (2021).

176. Leney, A. C. & Heck, A. J. R. Native mass spectrometry: what is in the name? *Journal of the American Society for Mass Spectrometry* **28,** 5–13 (2016).

177. Dyachenko, A. *et al.* Tandem native mass-spectrometry on antibody–drug conjugates and submillion Da antibody–antigen protein assemblies on an Orbitrap EMR equipped with a high-mass quadrupole mass selector. *Analytical chemistry* **87,** 6095–6102 (2015).

178. Sobott, F., Benesch, J. L. P., Vierling, E. & Robinson, C. V. Subunit exchange of multimeric protein complexes: real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *Journal of Biological Chemistry* **277,** 38921–38929 (2002).

179. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. & Heck, A. J. R. Ion mobility mass spectrometry of proteins and protein assemblies. *Chemical Society Reviews* **39,** 1633–1655 (2010).

180. Konijnenberg, A., Butterer, A. & Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1834,** 1239–1256 (2013).

181. Arrigo, A.-P. *et al.* Hsp27 (HspB1) and αB-crystallin (HspB5) as therapeutic targets. *FEBS Letters* **581,** 3665–3674. doi:https://doi.org/10.1016/j.febslet.2007.04.033 (2007).

182. Sun, X. *et al.* HSPB1 as a novel regulator of ferroptotic cancer cell death. *Oncogene* **34,** 5617–5625. doi:10.1038/onc.2015.32 (2015).

183. McDonald, E. T., Bortolus, M., Koteiche, H. A. & Mchaourab, H. S. Sequence, Structure, and Dynamic Determinants of Hsp27 (HspB1) Equilibrium Dissociation Are Encoded by the N-Terminal Domain. *Biochemistry* **51,** 1257–1268. doi:10.1021/bi2017624 (Feb. 2012).

184. Jovcevski, B. *et al.* Phosphomimics Destabilize Hsp27 Oligomeric Assemblies and Enhance Chaperone Activity. *Chemistry & Biology* **22,** 186–195. doi:https://doi.org/10.1016/j.chembiol.2015.01.001 (2015).

185. Clark, A. R., Lubsen, N. H. & Slingsby, C. sHSP in the eye lens: Crystallin mutations, cataract and proteostasis. *The International Journal of Biochemistry & Cell Biology* **44,** 1687–1697. doi:https://doi.org/10.1016/j.biocel.2012.02.015 (2012).

186. Arrigo, A.-P. & Gibert, B. HspB1, HspB5 and HspB4 in Human Cancers: Potent Oncogenic Role of Some of Their Client Proteins. *Cancers* **6,** 333–365. doi:10.3390/cancers6010333 (2014).

187. Webster, J. M., Darling, A. L., Uversky, V. N. & Blair, L. J. Small Heat Shock Proteins, Big Impact on Protein Aggregation in Neurodegenerative Disease. *Frontiers in Pharmacology* **10** (2019).

188. Zhu, Z. & Reiser, G. The small heat shock proteins, especially HspB4 and HspB5 are promising protectants in neurodegenerative diseases. *Neurochemistry International* **115,** 69–79. doi:https://doi.org/10.1016/j.neuint.2018.02.006 (2018).

189. Delbecq, S. P., Rosenbaum, J. C. & Klevit, R. E. A Mechanism of Subunit Recruitment in Human Small Heat Shock Protein Oligomers. *Biochemistry* **54,** 4276–4284. doi:10.1021/acs.biochem.5b00490 (July 2015).

190. Heirbaut, M. *et al.* The preferential heterodimerization of human small heat shock proteins HSPB1 and HSPB6 is dictated by the N-terminal domain. *Archives of Biochemistry and Biophysics* **610,** 41–50. doi:https://doi.org/10.1016/j.abb.2016.10.002 (2016).

191. Mymrikov, E. V. *et al.* Regulation of small heat-shock proteins by hetero-oligomer formation. *Journal of Biological Chemistry* **295,** 158–169. doi:10.1074/jbc.RA119.011143 (Jan. 2020).

192. Kurokawa, N. *et al.* Oligomeric Structural Transition of HspB1 from Chinese Hamster. *International Journal of Molecular Sciences* **22.** doi:10.3390/ijms221910797 (2021).

193. Nappi, L. *et al.* Ivermectin inhibits HSP27 and potentiates efficacy of oncogene targeting in tumor models. *The Journal of Clinical Investigation* **130,** 699–714. doi:10.1172/JCI130819 (Feb. 2020).

194. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **256,** 119–124. doi:10.1098/rspb.1994.0058 (Jan. 1997).

195. Vos, M. J., Kanon, B. & Kampinga, H. H. HSPB7 is a SC35 speckle resident small heat shock protein. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1793,** 1343–1353. doi:https://doi.org/10.1016/j.bbamcr.2009.05.005 (2009).

196. Heirbaut, M. *et al.* Specific sequences in the N-terminal domain of human small heat-shock protein HSPB6 dictate preferential hetero-oligomerization with the orthologue HSPB1. *Journal of Biological Chemistry* **292,** 9944–9957. doi:10.1074/jbc.M116.773515 (June 2017).

197. 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. doi:https://doi.org/10.1007/s12192-017-0769-x (2017).

198. 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* **100,** 10611–10616. doi:10.1073/pnas.1932958100 (Sept. 2003).

199. Aquilina, J. A., Shrestha, S., Morris, A. M. & Ecroyd, H. Structural and Functional Aspects of Hetero-oligomers Formed by the Small Heat Shock Proteins αB-Crystallin and HSP27. *Journal of Biological Chemistry* **288,** 13602–13609. doi:https://doi.org/10.1074/jbc.M112.443812 (2013).

200. Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annual Review of Biophysics* **46,** 247–269. doi:10.1146/annurev-biophys-070816-033631 (May 2017).

201. Wistow, G. The human crystallin gene families. *Human Genomics* **6,** 26. doi:10.1186/1479-7364-6-26 (2012).

202. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* **38,** 3022–3027. doi:10.1093/molbev/msab120 (June 2021).

203. Pauling, L. C., Zuckerkandl, E., Henriksen, T. & Lövstad, R. Chemical Paleogenetics. Molecular "Restoration Studies" of Extinct Forms of Life. *Acta Chemica Scandinavica,* 9–16 (1963).

204. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Reviews Genetics* **5,** 366–375. doi:10.1038/nrg1324 (2004).

205. Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nature Reviews Genetics* **14,** 559–571. doi:10.1038/nrg3540 (2013).

206. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421. doi:10.1186/1471-2105-10-421 (2009).

207. Hochberg, G. K. *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.** doi:10.1073/pnas.1322673111 (2014).

208. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14,** 33–38. doi:https://doi.org/10.1016/0263-7855(96)00018-5 (1996).

209. Eargle, J., Wright, D. & Luthey-Schulten, Z. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* **22,** 504–506. doi:10.1093/bioinformatics/bti825 (Feb. 2006).

210. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9,** 90–95. doi:10.1109/MCSE.2007.55 (2007).

211. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32,** 2319–2327. doi:https://doi.org/10.1002/jcc.21787 (July 2011).

212. Gowers, R. *et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations* in (2016), 98–105. doi:10.25080/Majora-629e541a-00e.

213. Stone, J. *An Efficient Library for Parallel Ray Tracing and Animation* PhD thesis (Mar. 1998).

214. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics* **23,** 566–579. doi:https://doi.org/10.1002/prot.340230412 (Dec. 1995).

215. Wang, K., Ash, J. F. & Singer, S. J. Filamin, a new high-molecular-weight protein found in smooth muscle and non-muscle cells. *Proceedings of the National Academy of Sciences* **72,** 4483–4486. doi:10.1073/pnas.72.11.4483 (Nov. 1975).

216. Wang, K. & Singer, S. J. Interaction of filamin with f-actin in solution. *Proceedings of the National Academy of Sciences* **74,** 2021–2025. doi:10.1073/pnas.74.5.2021 (May 1977).

217. Feng, Y. & Walsh, C. A. The many faces of filamin: A versatile molecular scaffold for cell motility and signalling. *Nature Cell Biology* **6,** 1034–1038. doi:10.1038/ncb1104-1034 (2004).

218. Pudas, R., Kiema, T.-R., Butler, P. J. G., Stewart, M. & Ylänne, J. Structural Basis for Vertebrate Filamin Dimerization. *Structure* **13,** 111–119. doi:https://doi.org/10.1016/j.str.2004.10.014 (2005).

219. 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. doi:10.1083/jcb.200707073 (Dec. 2007).

220. Vadlamudi, R. K. *et al.* Filamin is essential in actin cytoskeletal assembly mediated by p21-activated kinase 1. *Nature Cell Biology* **4,** 681–690. doi:10.1038/ncb838 (2002).

221. Small, J. V., Fürst, D. O. & De Mey, J. Localization of filamin in smooth muscle. *Journal of Cell Biology* **102,** 210–220. doi:10.1083/jcb.102.1.210 (Jan. 1986).

222. 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* **45,** 149–162. doi:https://doi.org/10.1002/(SICI)1097-0169(200002)45:2<149::AID-CM6>3.0.CO;2-G (Feb. 2000).

223. Mao, Z. & Nakamura, F. Structure and Function of Filamin C in the Muscle Z-Disc. *International Journal of Molecular Sciences* **21.** doi:10.3390/ijms21082696 (2020).

224. Kajita, M. *et al.* Filamin acts as a key regulator in epithelial defence against transformed cells. *Nature Communications* **5,** 4428. doi:10.1038/ncomms5428 (2014).

225. Nakamura, F. *et al.* Amino- and carboxyl-terminal domains of Filamin-A interact with CRMP1 to mediate Sema3A signalling. *Nature Communications* **5,** 5325. doi:10.1038/ncomms6325 (2014).

226. Zhou, X., Borén, J. & Akyürek, L. M. Filamins in Cardiovascular Development. *Trends in Cardiovascular Medicine* **17,** 222–229. doi:https://doi.org/10.1016/j.tcm.2007.08.001 (2007).

227. Kyndt, F. *et al.* Mutations in the Gene Encoding Filamin A as a Cause for Familial Cardiac Valvular Dystrophy. *Circulation* **115,** 40–49. doi:10.1161/CIRCULATIONAHA.106.622621 (Jan. 2007).

228. Zhou, X. *et al.* Interaction of Filamin C With Actin Is Essential for Cardiac Development and Function. *Circulation Research* **133,** 400–411. doi:10.1161/CIRCRESAHA.123.322750 (Aug. 2023).

229. Razinia, Z., Mäkelä, T., Ylänne, J. & Calderwood, D. A. Filamins in Mechanosensing and Signaling. *Annual Review of Biophysics* **41,** 227–246. doi:10.1146/annurev-biophys-050511-102252 (May 2012).

230. Del Valle-Pérez, B. *et al.* Filamin B Plays a Key Role in Vascular Endothelial Growth Factor-induced Endothelial Cell Motility through Its Interaction with Rac-1 and Vav-2. *Journal of Biological Chemistry* **285,** 10748–10760. doi:10.1074/jbc.M109.062984 (Apr. 2010).

231. Frank, D., Kuhn, C., Katus, H. A. & Frey, N. The sarcomeric Z-disc: a nodal point in signalling and disease. *Journal of Molecular Medicine* **84,** 446–468. doi:10.1007/s00109-005-0033-1 (2006).

232. Luther, P. K. The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. *Journal of Muscle Research and Cell Motility* **30,** 171–185. doi:10.1007/s10974-009-9189-6 (2009).

233. Knappeis, G. G. & Carlsen, F. The Ultrastructure of the Z Disc in Skeletal Muscle. *Journal of Cell Biology* **13,** 323–335. doi:10.1083/jcb.13.2.323 (May 1962).

234. Frank, D. & Frey, N. Cardiac Z-disc Signaling Network. *Journal of Biological Chemistry* **286,** 9897–9904. doi:10.1074/jbc.R110.174268 (Mar. 2011).

235. Knöll, R., Buyandelger, B. & Lab, M. The Sarcomeric Z-Disc and Z-Discopathies. *Journal of Biomedicine and Biotechnology* **2011** (ed Granzier, H.) 569628. doi:10.1155/2011/569628 (2011).

236. Leber, Y. *et al.* Filamin C is a highly dynamic protein associated with fast repair of myofibrillar microdamage. *Human Molecular Genetics* **25,** 2776–2788. doi:10.1093/hmg/ddw135 (July 2016).

237. Collier, M. *Small heat shock protein interactions with in vivo partners* eng. 2018.

238. Sellung, D. *et al.* Novel Filamin C Myofibrillar Myopathy Variants Cause Different Pathomechanisms and Alterations in Protein Quality Systems. *Cells* **12.** doi:10.3390/cells12091321 (2023).

239. Deng, Y. & Yan, J. Force-Dependent Structural Changes of Filamin C Rod Domains Regulated by Filamin C Dimer. *Journal of the American Chemical Society* **145,** 14670–14678. doi:10.1021/jacs.3c02303 (July 2023).

240. Krief, S. *et al.* Identification and Characterization of cvHsp: a Novel Human Small Stress Protein Selectively Expressed in Cardiovascular and Insulin-sensitive Tissues. *Journal of Biological Chemistry* **274,** 36592–36600. doi:10.1074/jbc.274.51.36592 (Dec. 1999).

241. Robertson, A. L. *et al.* Small heat-shock proteins interact with a flanking domain to suppress polyglutamine aggregation. *Proceedings of the National Academy of Sciences* **107,** 10424–10429. doi:10.1073/pnas.0914773107 (June 2010).

242. Healy, E. F., Little, C. & King, P. J. A Model for Small Heat Shock Protein Inhibition of Polyglutamine Aggregation. *Cell Biochemistry and Biophysics* **69,** 275–281. doi:10.1007/s12013-013-9795-1 (2014).

243. Trottier, Y. *et al.* Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* **378,** 403–406. doi:10.1038/378403a0 (1995).

244. Ross, C. A. Polyglutamine Pathogenesis: Emergence of Unifying Mechanisms for Huntington's Disease and Related Disorders. *Neuron* **35,** 819–822. doi:10.1016/S0896-6273(02)00872-3 (Aug. 2002).

245. Chen, S., Ferrone, F. A. & Wetzel, R. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proceedings of the National Academy of Sciences* **99,** 11884–11889. doi:10.1073/pnas.182276099 (Sept. 2002).

246. Rüb, U. *et al.* Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. *Progress in Neurobiology* **104,** 38–66. doi:https://doi.org/10.1016/j.pneurobio.2013.01.001 (2013).

247. 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. doi:10.1093/hmg/ddq398 (Dec. 2010).

248. 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. doi:https://doi.org/10.1074/jbc.M116.760413 (2017).

249. Wu, T. *et al.* HSPB7 is indispensable for heart development by modulating actin filament assembly. *Proceedings of the National Academy of Sciences* **114,** 11956–11961. doi:10.1073/pnas.1713763114 (Nov. 2017).

250. 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. doi:https://doi.org/10.1016/j.ydbio.2013.06.025 (2013).

251. Mercer, E. J., Lin, Y.-F., Cohen-Gould, L. & Evans, T. Hspb7 is a cardioprotective chaperone facilitating sarcomeric proteostasis. *Developmental Biology* **435,** 41–55. doi:https://doi.org/10.1016/j.ydbio.2018.01.005 (2018).

252. 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– (Aug. 2017).

253. Rubin, C. S. & Rosen, O. M. Protein phosphorylation. *Annual review of biochemistry* **44,** 831–887 (1975).

254. Krebs, E. G. & Beavo, J. A. Phosphorylation-dephosphorylation of enzymes. *Annual review of biochemistry* **48,** 923–959 (1979).

255. Cohen, P. The origins of protein phosphorylation. *Nature cell biology* **4,** E127–E130 (2002).

256. Ubersax, J. A. & Ferrell Jr, J. E. Mechanisms of specificity in protein phosphorylation. *Nature reviews Molecular cell biology* **8,** 530–541 (2007).

257. Korzeniewski, B. & Zoladz, J. A. A model of oxidative phosphorylation in mammalian skeletal muscle. *Biophysical Chemistry* **92,** 17–34. doi:https://doi.org/10.1016/S0301-4622(01)00184-3 (2001).

258. Gnaiger, E. Capacity of oxidative phosphorylation in human skeletal muscle: New perspectives of mitochondrial physiology. *The International Journal of Biochemistry & Cell Biology* **41,** 1837–1845. doi:https://doi.org/10.1016/j.biocel.2009.03.013 (2009).

259. Colson, B. A., Gruber, S. J. & Thomas, D. D. Structural dynamics of muscle protein phosphorylation. *Journal of Muscle Research and Cell Motility* **33,** 419–429. doi:10.1007/s10974-012-9317-6 (2012).

260. Lemieux, H. & Warren, B. E. An animal model to study human muscular diseases involving mitochondrial oxidative phosphorylation. *Journal of Bioenergetics and Biomembranes* **44,** 503–512. doi:10.1007/s10863-012-9451-2 (2012).

261. Reitz, C. J. *et al.* Proteomics and phosphoproteomics of failing human left ventricle identifies dilated cardiomyopathy-associated phosphorylation of CTNNA3. *Proceedings of the National Academy of Sciences* **120,** e2212118120. doi:10.1073/pnas.2212118120 (May 2023).

262. Bouhaddou, M. *et al.* The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **182,** 685–712. doi:10.1016/j.cell.2020.06.034 (Aug. 2020).

263. Bonow, R. O., Fonarow, G. C., O'Gara, P. T. & Yancy, C. W. Association of Coronavirus Disease 2019 (COVID-19) With Myocardial Injury and Mortality. *JAMA Cardiology* **5,** 751–753. doi:10.1001/jamacardio.2020.1105 (July 2020).

264. Hoffman, N. J. *et al.* Global Phosphoproteomic Analysis of Human Skeletal Muscle Reveals a Network of Exercise-Regulated Kinases and AMPK Substrates. *Cell Metabolism* **22,** 922–935. doi:https://doi.org/10.1016/j.cmet.2015.09.001 (2015).

265. Nelson, M. E. *et al.* Phosphoproteomics reveals conserved exercise-stimulated signaling and AMPK regulation of store-operated calcium entry. *The EMBO Journal* **38,** e102578. doi:https://doi.org/10.15252/embj.2019102578 (Dec. 2019).

266. Bai, Y. *et al.* Phosphoproteomics Identifies Driver Tyrosine Kinases in Sarcoma Cell Lines and Tumors. *Cancer Research* **72,** 2501–2511. doi:10.1158/0008-5472.CAN-11-3015 (May 2012).

267. Hiscott, J. *et al.* Triggering the Interferon Response: The Role of IRF-3 Transcription Factor. *Journal of Interferon & Cytokine Research* **19,** 1–13. doi:10.1089/107999099314360 (Jan. 1999).

268. Paleologou, K. E. *et al.* Phosphorylation at Ser-129 but Not the Phosphomimics S129E/D Inhibits the Fibrillation of α-Synuclein. *Journal of Biological Chemistry* **283,** 16895–16905. doi:10.1074/jbc.M800747200 (June 2008).

269. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12,** 2825–2830 (2011).

270. Hunter, T. Why nature chose phosphate to modify proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367,** 2513–2516. doi:10.1098/rstb.2012.0013 (Sept. 2012).

271. Liu, T., Ryan, M., Dahlquist, F. W. & Griffith, O. H. Determination of pKa values of the histidine side chains of phosphatidylinositol-specific phospholipase C from Bacillus cereus by NMR spectroscopy and site-directed mutagenesis. *Protein Science* **6,** 1937–1944. doi:https://doi.org/10.1002/pro.5560060914 (Sept. 1997).

272. McDonald, J. H. *Handbook of Biological Statistics* 3rd (2014).

273. LeWinter, M. M., Wu, Y., Labeit, S. & Granzier, H. Cardiac titin: Structure, functions and role in disease. *Clinica Chimica Acta* **375,** 1–9. doi:https://doi.org/10.1016/j.cca.2006.06.035 (2007).

274. Granzier, H. L. & Labeit, S. in *Advances in Protein Chemistry* 89–119 (Academic Press, 2005). doi:https://doi.org/10.1016/S0065-3233(04)71003-7.

275. Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387,** 308–312. doi:10.1038/387308a0 (1997).

276. Horowits, R. Passive force generation and titin isoforms in mammalian skeletal muscle. *Biophysical Journal* **61,** 392–398. doi:10.1016/S0006-3495(92)81845-3 (Feb. 1992).

277. Herzog, W. The multiple roles of titin in muscle contraction and force production. *Biophysical Reviews* **10,** 1187–1199. doi:10.1007/s12551-017-0395-y (2018).

278. Nishikawa, K. C. *et al.* Is titin a 'winding filament'? A new twist on muscle contraction. *Proceedings of the Royal Society B: Biological Sciences* **279,** 981–990. doi:10.1098/rspb.2011.1304 (Sept. 2011).

279. Tskhovrebova, L. & Trinick, J. in *Fibrous Proteins: Structures and Mechanisms* (eds Parry, D. A. D. & Squire, J. M.) 285–318 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-49674-0{_}10.

280. Nagueh, S. F. *et al.* Altered Titin Expression, Myocardial Stiffness, and Left Ventricular Function in Patients With Dilated Cardiomyopathy. *Circulation* **110,** 155–162. doi:10.1161/01.CIR.0000135591.37759.AF (July 2004).

281. LeWinter, M. M. & Granzier, H. L. Titin Is a Major Human Disease Gene. *Circulation* **127,** 938–944. doi:10.1161/CIRCULATIONAHA.112.139717 (Feb. 2013).

282. Gerull, B. *et al.* Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nature Genetics* **30,** 201–204. doi:10.1038/ng815 (2002).

283. Itoh-Satoh, M. *et al.* Titin Mutations as the Molecular Basis for Dilated Cardiomyopathy. *Biochemical and Biophysical Research Communications* **291,** 385–393. doi:https://doi.org/10.1006/bbrc.2002.6448 (2002).

284. Herman, D. S. *et al.* Truncations of Titin Causing Dilated Cardiomyopathy. *New England Journal of Medicine* **366,** 619–628. doi:10.1056/NEJMoa1110186 (Feb. 2012).

285. Zile, M. R. *et al.* Myocardial Stiffness in Patients With Heart Failure and a Preserved Ejection Fraction. *Circulation* **131,** 1247–1259. doi:10.1161/CIRCULATIONAHA.114.013215 (Apr. 2015).

286. Labeit, S., Gautel, M., Lakey, A. & Trinick, J. Towards a molecular understanding of titin. *The EMBO Journal* **11,** 1711–1716. doi:https://doi.org/10.1002/j.1460-2075.1992.tb05222.x (May 1992).

287. Granzier, H. L. *et al.* Deleting titin's I-band/A-band junction reveals critical roles for titin in biomechanical sensing and cardiac function. *Proceedings of the National Academy of Sciences* **111,** 14589–14594. doi:10.1073/pnas.1411493111 (Oct. 2014).

288. Trinick, J. Cytoskeleton: Titin as a scaffold and spring. *Current Biology* **6,** 258–260. doi:https://doi.org/10.1016/S0960-9822(02)00472-4 (1996).

289. LeWinter, M. M. & Granzier, H. Cardiac Titin. *Circulation* **121,** 2137–2145. doi:10.1161/CIRCULATIONAHA.109.860171 (May 2010).

290. Krüger, M. *et al.* Protein Kinase G Modulates Human Myocardial Passive Stiffness by Phosphorylation of the Titin Springs. *Circulation Research* **104,** 87–94. doi:10.1161/CIRCRESAHA.108.184408 (Jan. 2009).

291. Sorimachi, H. *et al.* Tissue-specific expression and α-actinin binding properties of the Z-disc titin: implications for the nature of vertebrate Z-discs. *Journal of Molecular Biology* **270,** 688–695. doi:https://doi.org/10.1006/jmbi.1997.1145 (1997).

292. Mues, A., van der Ven, P. F. M., Young, P., Fürst, D. O. & Gautel, M. Two immunoglobulin-like domains of the Z-disc portion of titin interact in a conformation-dependent way with telethonin. *FEBS Letters* **428,** 111–114. doi:https://doi.org/10.1016/S0014-5793(98)00501-8 (1998).

293. Bang, M.-L. *et al.* The Complete Gene Sequence of Titin, Expression of an Unusual 700-kDa Titin Isoform, and Its Interaction With Obscurin Identify a Novel Z-Line to I-Band Linking System. *Circulation Research* **89,** 1065–1072. doi:10.1161/hh2301.100981 (Nov. 2001).

294. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402,** 100–103. doi:10.1038/47083 (1999).

295. Kellermayer, D., Smith, J. E. & Granzier, H. Titin mutations and muscle disease. *Pflügers Archiv - European Journal of Physiology* **471,** 673–682. doi:10.1007/s00424-019-02272-5 (2019).

296. Linke, W. A., Ivemeyer, M., Mundel, P., Stockmeier, M. R. & Kolmerer, B. Nature of PEVK-titin elasticity in skeletal muscle. *Proceedings of the National Academy of Sciences* **95,** 8052–8057. doi:10.1073/pnas.95.14.8052 (July 1998).

297. Cazorla, O. *et al.* Differential Expression of Cardiac Titin Isoforms and Modulation of Cellular Stiffness. *Circulation Research* **86,** 59–67. doi:10.1161/01.RES.86.1.59 (Jan. 2000).

298. Van der Pijl, R. J. *et al.* The titin N2B and N2A regions: biomechanical and metabolic signaling hubs in cross-striated muscles. *Biophysical Reviews* **13,** 653–677. doi:10.1007/s12551-021-00836-3 (2021).

299. Wang, K. & Williamson, C. L. Identification of an N2 line protein of striated muscle. *Proceedings of the National Academy of Sciences* **77,** 3254–3258. doi:10.1073/pnas.77.6.3254 (June 1980).

300. Trombitás, K., Wu, Y., Labeit, D., Labeit, S. & Granzier, H. Cardiac titin isoforms are coexpressed in the half-sarcomere and extend independently. *American Journal of Physiology-Heart and Circulatory Physiology* **281,** H1793–H1799. doi:10.1152/ajpheart.2001.281.4.H1793 (Oct. 2001).

301. Adewale, A. O. & Ahn, Y.-H. Titin N2A Domain and Its Interactions at the Sarcomere. *International Journal of Molecular Sciences* **22.** doi:10.3390/ijms22147563 (2021).

302. Linke, W. A. *et al.* I-Band Titin in Cardiac Muscle Is a Three-Element Molecular Spring and Is Critical for Maintaining Thin Filament Structure. *Journal of Cell Biology* **146,** 631–644. doi:10.1083/jcb.146.3.631 (Aug. 1999).

303. Patel, J. R., Pleitner, J. M., Moss, R. L. & Greaser, M. L. Magnitude of length-dependent changes in contractile properties varies with titin isoform in rat ventricles. *American Journal of Physiology-Heart and Circulatory Physiology* **302,** H697–H708. doi:10.1152/ajpheart.00800.2011 (Dec. 2011).

304. Zhu, Y., Bogomolovas, J., Labeit, S. & Granzier, H. Single Molecule Force Spectroscopy of the Cardiac Titin N2B Element: Effects of The Molecular Chaperone αB-Crystallin with Disease-causing Mutations. *Journal of Biological Chemistry* **284,** 13914–13923. doi:10.1074/jbc.M809743200 (May 2009).

305. Watanabe, K. *et al.* Molecular Mechanics of Cardiac Titin's PEVK and N2B Spring Elements. *Journal of Biological Chemistry* **277,** 11549–11558. doi:https://doi.org/10.1074/jbc.M200356200 (2002).

306. Kontrogianni-Konstantopoulos, A., Ackermann, M. A., Bowman, A. L., Yap, S. V. & Bloch, R. J. Muscle Giants: Molecular Scaffolds in Sarcomerogenesis. *Physiological Reviews* **89,** 1217–1267. doi:10.1152/physrev.00017.2009 (Oct. 2009).

307. Marunouchi, T. *et al.* Changes in Small Heat Shock Proteins HSPB1, HSPB5 and HSPB8 in Mitochondria of the Failing Heart Following Myocardial Infarction in Rats. *Biological and Pharmaceutical Bulletin* **36,** 529–539. doi:10.1248/bpb.b12-00796 (2013).

308. Dimauro, I. & Caporossi, D. Alpha B-Crystallin in Muscle Disease Prevention: The Role of Physical Activity. *Molecules* **27.** doi:10.3390/molecules27031147 (2022).

309. S Christians, E., B Mustafi, S. & J Benjamin, I. Chaperones and cardiac misfolding protein diseases. *Current Protein and Peptide Science* **15,** 189–204 (2014).

310. Benndorf, R., Martin, J. L., Kosakovsky Pond, S. L. & Wertheim, J. O. Neuropathy- and myopathy-associated mutations in human small heat shock proteins: Characteristics and evolutionary history of the mutation sites. *Mutation Research/Reviews in Mutation Research* **761,** 15–30. doi:https://doi.org/10.1016/j.mrrev.2014.02.004 (2014).

311. Mitzelfelt, K. A. & Benjamin, I. J. in *The Big Book on Small Heat Shock Proteins* (eds Tanguay, R. M. & Hightower, L. E.) 269–299 (Springer International Publishing, Cham, 2015). doi:10.1007/978-3-319-16077-1{_}11.

312. Nasiri, P. *et al.* The biochemical association between R157H mutation in human αB-crystallin and development of cardiomyopathy: Structural and functional analyses of the mutant protein. *Biochimie* **190,** 36–49. doi:https://doi.org/10.1016/j.biochi.2021.06.019 (2021).

313. Sheridan, S., Gräter, F. & Daday, C. How Fast Is Too Fast in Force-Probe Molecular Dynamics Simulations? *The Journal of Physical Chemistry B* **123,** 3658–3664. doi:10.1021/acs.jpcb.9b01251 (May 2019).

314. Lee, E. H., Gao, M., Pinotsis, N., Wilmanns, M. & Schulten, K. Mechanical Strength of the Titin Z1Z2-Telethonin Complex. *Structure* **14,** 497–509. doi:https://doi.org/10.1016/j.str.2005.12.005 (2006).

315. Patel, S., Vierling, E. & Tama, F. Replica Exchange Molecular Dynamics Simulations Provide Insight into Substrate Recognition by Small Heat Shock Proteins. *Biophysical Journal* **106,** 2644–2655. doi:https://doi.org/10.1016/j.bpj.2014.04.048 (2014).

316. Fossa, P. & Cichero, E. In silico evaluation of human small heat shock protein HSP27: Homology modeling, mutation analyses and docking studies. *Bioorganic & Medicinal Chemistry* **23,** 3215–3220. doi:https://doi.org/10.1016/j.bmc.2015.04.070 (2015).

317. Chiappori, F., Mattiazzi, L., Milanesi, L. & Merelli, I. A novel molecular dynamics approach to evaluate the effect of phosphorylation on multimeric protein interface: the αB-Crystallin case study. *BMC Bioinformatics* **17,** 57. doi:10.1186/s12859-016-0909-9 (2016).

318. Bukach, O. V., Glukhova, A. E., Seit-Nebi, A. S. & Gusev, N. B. Heterooligomeric complexes formed by human small heat shock proteins HspB1 (Hsp27) and HspB6 (Hsp20). *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1794,** 486–495. doi:https://doi.org/10.1016/j.bbapap.2008.11.010 (2009).

319. Ke, L. *et al.* HSPB1, HSPB6, HSPB7 and HSPB8 Protect against RhoA GTPase-Induced Remodeling in Tachypaced Atrial Myocytes. *PLoS ONE* **6,** e20395. doi:10.1371/journal.pone.0020395 (June 2011).

320. Sanbe, A. *et al.* Protective Effect of Geranylgeranylacetone via Enhancement of HSPB8 Induction in Desmin-Related Cardiomyopathy. *PLoS ONE* **4,** e5351. doi:10.1371/journal.pone.0005351 (Apr. 2009).

321. Carra, S., Seguin, S. J. & Landry, J. HspB8 and Bag3: A new chaperone complex targeting misfolded proteins to macroautophagy. *Autophagy* **4,** 237–239. doi:10.4161/auto.5407 (Feb. 2008).