A Dynamics Based Analysis of Allosteric Modulation in Heat Shock Proteins

A thesis submitted in fulfilment of the requirement for the degree
of

DOCTOR OF PHILOSOPHY
IN BIOINFORMATICS

Rhodes University, South Africa
Department of Biochemistry and Microbiology
Faculty of Science

By

David Lawrence Penkler

November 2018

Abstract

The 70 kDa and 90 kDa heat shock proteins (Hsp70 and Hsp90) are molecular chaperones that play central roles in maintaining cellular homeostasis in all organisms of life with the exception of archaea. In addition to their general chaperone function in protein quality control, Hsp70 and Hsp90 cooperate in the regulation and activity of some 200 known natively folded protein clients which include protein kinases, transcription factors and receptors, many of which are implicated as key regulators of essential signal transduction pathways. Both chaperones are considered to be large multi-domain proteins that rely on ATPase activity and co-chaperone interactions to regulate their conformational cycles for peptide binding and release. The unique positioning of Hsp90 at the crossroads of several fundamental cellular pathways coupled with its known association with diverse oncogenic peptide clients has brought the molecular chaperone under increasing interest as a potential anti-cancer target that is crucially implicated with all eight hallmarks of the disease. Current orthosteric drug discovery efforts aimed at the inhibition of the ATPase domain of Hsp90 have been limited due to high levels of associated toxicity. In an effort to circumnavigate this, the combined focus of research efforts is shifting toward alternative approaches such as interference with co-chaperone binding and the allosteric inhibition/activation of the molecular chaperone.

The overriding aim of this thesis was to demonstrate how the computational technique of Perturbation response scanning (PRS) coupled with all-atom molecular dynamics simulations (MD) and dynamic residue interaction network (DRN) analysis can be used as a viable strategy to efficiently scan and accurately identify allosteric control element capable of modulating the functional dynamics of a protein. In pursuit of this goal, this thesis also contributes to the current understanding of the nucleotide dependent allosteric mechanisms at play in cellular functionality of both Hsp70 and Hsp90.

All-atom MD simulations of *E. coli* DnaK provided evidence of nucleotide driven modulation of conformational dynamics in both the catalytically active and inactive states. PRS analysis employed on these trajectories demonstrated sensitivity toward bound nucleotide and peptide substrate, and provided evidence of a putative allosterically active intermediate state between the ATPase active and inactive conformational states. Simultaneous binding of ATP and peptide substrate was found to allosterically prime the chaperone for interstate conversion regardless of the transition direction. Detailed analysis of these allosterically primed states revealed select residue sites capable of

selecting a coordinate shift towards the opposite conformational state. In an effort to validate these results, the predicted allosteric hot spot sites were cross-validated with known experimental works and found to overlap with functional sites implicated in allosteric signal propagation and ATPase activation in Hsp70. This study presented for the first time, the application of PRS as a suitable diagnostic tool for the elucidation and quantification of the allosteric potential of select residues to effect functionally relevant global conformational rearrangements. The PRS methodology described in this study was packaged within the Python programming environment in the MD-TASK software suite for command-line ease of use and made freely available.

Homology modelling techniques were used to address the lack of experimental structural data for the human cytosolic isoform of Hsp90 and for the first time provided accurate full-length structural models of human Hsp90α in fully-closed and partially-open conformations. Long-range all-atom MD simulations of these structures revealed nucleotide driven modulation of conformational dynamics in Hsp90. Subsequent DRN and PRS analysis of these MD trajectories allowed for the quantification and elucidation of nucleotide driven allosteric modulation in the molecular chaperone. A detailed PRS analysis revealed allosteric inter-domain coupling between the extreme terminals of the chaperone in response to external force perturbations at either domain. Furthermore PRS also identified several individual residue sites that are capable of selecting conformational rearrangements towards functionally relevant states which may be considered to be putative allosteric target sites for future drug discovery efforts.

Molecular docking techniques were employed to investigate the modulation of conformational dynamics of human Hsp90α in response to ligand binding interactions at two identified allosteric sites at the C-terminal. High throughput screening of a small library of natural compounds indigenous to South Africa revealed three hit compounds at these sites: Cephalostatin 17, 20(29)-Lupene-3β-isoferulate and 3'-Bromorubrolide F. All-atom MD simulations on these protein-ligand complexes coupled with DRN analysis and several advanced trajectory based analysis techniques provided evidence of selective allosteric modulation of Hsp90α conformational dynamics in response to the identity and location of the bound ligands. Ligands bound at the four-helix bundle presented as putative allosteric inhibitors of Hsp90α, driving conformational dynamics in favour of dimer opening and possibly dimer separation. Meanwhile, ligand interactions at an adjacent sub-pocket located near the interface between the middle and C-terminal domains demonstrated allosteric activation of the chaperone, modulating conformational dynamics in favour of the fully-closed catalytically active conformational state.

Taken together, the data presented in this thesis contributes to the understanding of allosteric modulation of conformational dynamics in Hsp70 and Hsp90, and provides a suitable platform for future biochemical and drug discovery studies. Furthermore, the molecular docking and computational identification of allosteric compounds with suitable binding affinity for allosteric sites at the CTD of human Hsp90α provide for the first time "proof-of-principle" for the use of PRS in conjunction with MD simulations and DRN analysis as a suitable method for the rapid identification of allosteric sites in proteins that can be probed by small molecule interaction. The data presented in this section could pave the way for future allosteric drug discovery studies for the treatment of Hsp90 associated pathologies.

Declaration

I David Lawrence Penkler declare that this thesis is my own, unaided work, unless otherwise stated. It is being submitted for the degree of Doctor of Philosophy at Rhodes University. It has not been submitted before for any degree or examination at any other university.

David Lawrence Penkler

Acknowledgements

My gratitude and appreciation are expressed to the following, without which, this contribution would not have been possible:

My Supervisor Prof. Özlem Tastan Bishop, for her academic and financial support. Her guidance and motivation were crucial to the completion of this work. Thank you for recognizing my potential and helping me achieve it.

Dr. David Brown, for his technical support on the RUBi compute cluster and for the numerous collaborations, each of which helped hone and develop my programming skills.

Dr. Rowan Hatherley, for his friendship and moral support and always being available to discuss an idea or thought.

My colleagues of the RUBi lab past and present with whom I have collaborated and interacted with on a day-to-day basis.

Prof. Canan Atilgan and her lab at Sabanci University, Istanbul, with whom I collaborated with on several publications. Thank you for hosting me and for sharing your combined academic knowledge.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. The opinions expressed and derived conclusions are mine alone and are not necessarily to be attributed to the NRF.

Rhodes University, for awarding me the Prestigious Henderson Scholarship and for financial support to attend several conferences.

The Center for High Performance Computing (CHPC), South Africa, for providing computing resources.

My precious wife Jo-Anne Penkler for her love, support and enduring patience for which I am ever grateful. Your industrious spirit has kept me afloat whenever I needed motivation.

My dearest mother Carol Penkler, for providing me with the platform to succeed and for instilling me with the fundamentals of scientific reasoning from an early age. My achievements are a credit to all your sacrifices for me.

Stephanie and Michaela Penkler, my best sisters, thank you for all your support and understanding over the years.

Special thank you to Michaela who assisted with the final printing and compilation of this thesis.

~ This thesis is dedicated to my beloved Father ~

Lawrence John Penkler

(1964-2003)

"Medicinal chemistry has entered a new age of rational approaches, based on an understanding of the principles underlying molecular recognition. Conformational analysis of receptor selective ligands and computer based methods for structure representation and comparison provide useful techniques for the determination of receptor bound conformation of drugs, essential as basis for understanding molecular recognition" – *Lawrence John Penkler (1998)*

Table of Contents

Chapter 4: **111**

List of Figures

List of Tables

List of Abbreviations

3D	Three-dimensional
ATP	Adenosine tri-phosphate
ADP	Adenosine di-phosphate
C_{α}	Carbon-alpha
$C_{ß}$	Carbon-beta
CP	Communication propensity
CTD	C-terminal domain
DRN	Dynamic residue network
EAM	Ensemble allosteric model
ED	Essential dynamics
EM	Electron micrographs
ENM	Elastic network model
HOP	Hsp70-Hsp90 organizing protein
Hsp	Heat shock protein
Hsp70	Heat shock protein 70 KDa
Hsp90	Heat shock protein 90 KDa
Hsp40	Heat shock protein 40 kDa
HSF	Heat shock transcription factor
HtpG	High temperature protein G
JDP	J-domain protein
kDa	Kilodalton
KNF	Koshland-Nemethy-Filmer (model)
MiD	Middle domain
MD	Molecular dynamics
MWC	Monod-Wyman-Changeux (model)
NBD	Nucleotide binding domain
NTD	N-terminal domain
NMA	Normal mode analysis
NMR	Nuclear magnetic resonance
PC	Principle component
PCA	Principle component analysis
PDB	Protein data bank

PRS	Perturbation response scanning
PTM	Post-translational modification
RMSD	Root-mean-square-deviation
RMSF	Root-mean-square-fluctuation
SANCDB	South African natural compounds database
SANC	South African natural compound
SBD	Substrate binding domain
SHR	Steroid hormone receptors
TPR	Tetracopetide repeat domain

Research Outputs

Primary Publications

1. Penkler, D., Sensoy, Ö., Atilgan, C., and Tastan Bishop, Ö. **Perturbation-Response Scanning Reveals Key Residues for Allosteric Control in Hsp70**. *Journal of Chemical Information and Modeling*, 57: 1359-1374, (2017).
2. Brown, DK., Penkler, D., Amamuddy, OS., Ross, C., Atilgan, AR., Atilgan, C., and Tastan Bishop, Ö. **MD-TASK: a software suite for analyzing molecular dynamics trajectories**. *Bioinformatics*, 33 (17): 2768-2771, (2017).
3. Penkler, D., Atilgan, C., and Tastan Bishop, Ö. **Allosteric Modulation of Human Hsp90α Conformational Dynamics**. *Journal of Chemical Information and Modeling*, 58: 383-404, (2018).
4. Penkler, D., and Tastan Bishop, Ö. **Modulation of Human Hsp90α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain**. *Scientific Reports* (2018). (Accepted, In press)

Secondary Publications

5. Hatherley, R., Brown, DK., Musyoka, T., Penkler, D., Faya, N., Lobb, KA., and Tastan Bishop, Ö. **SANCDB: A South African Natural Compound Database**. *Journal of Cheminformatics*, 7:29, (2015).
6. Brown, DK., Penkler, DL., Musyoka, T., and Tastan Bishop, Ö. **JMS: An Open Source Workflow Management System and Web-Based Cluster Front-End for High Performance Computing**. *PLoS One*, 10 (8): e0134273, (2015).
7. Faya, N., Penkler, DL., and Tastan Bishop, Ö. **Human, vector and parasite Hsp90 proteins: A comparative bioinformatics analysis**. *FEBS openbio*, 5: 916-927, (2015).

Conference Presentations

1. Penkler, D.L, Tastan Bishop, O.A. **Targeted drug discovery for human heat shock protein 90: an in silico approach**. Joint SASBi-SAGS Congress, (2014). Pretoria, South Africa.
2. D.L Penkler and O. Tastan Bishop. **Understanding the biophysics involved in human HSP90 flexibility**. Biophysics in the Understanding, Diagnosis, and Treatment of Infectious Diseases, (2015). Stellenbosch, South Africa
3. D.L. Penkler, Ö. Sensoy, G. Güven, C. Atilgan, Ö. Tastan Bishop. **Investigation of Hsp70 allosteric control through Perturbation response scanning analysis**. 25th SASBMB Congress, (2016). East London, South Africa.
4. D.L. Penkler, Ö. Sensoy, G. Güven, C. Atilgan, Ö. Tastan Bishop. **Determining Allosteric Hot Spots in Hsp70 using Perturbation response scanning**. ISMB/ECCB, (2017). Prague, Czech Republic.

Overview of this thesis

The aim of this thesis is to contribute to the growing body of evidence for the allosteric regulation of the 70 kDa and 90 kDa heat shock proteins (Hsps), by demonstrating the use of Perturbation response scanning (PRS) coupled with all-atom molecular dynamics (MD) simulation analysis techniques and dynamic residue interaction network (DRN) analysis as a viable strategy for the efficient and accurate identification of allosteric control elements in large multi-domain proteins that undergo complex conformational restructuring that involves both rotational and translational motions. The research presented in this thesis is comprised and based on three separate studies that have been published in reputable peer-reviewed journals. The thesis is written as an all-encompassing account of these studies under the common theme of dynamics based investigation of allostery in molecular heat shock proteins for allosteric drug design using computational techniques.

In *Chapter 1* the concept and origins of protein allostery is briefly reviewed together with its application as a biological regulatory mechanism in drug discovery, and the structural computational methods for investigating allostery in dynamic proteins briefly summarized. The molecular chaperones are introduced and the biological role and mechanism of action for 70 kDa and90 kDa family members reviewed in detail. Finally the stage is set for therapeutic targeting of Hsp90 for the treatment of cancer.

Chapter 2 introduces the computational technique of PRS and discusses the advantages of using MD trajectories to build the underlying kernel. PRS is applied to the model system of Hsp70 to establish the accuracy of the technique for use in large multi-domain proteins that undergo complex structural rearrangements. This was the first time PRS has been applied to such a system using this particular methodology. PRS is used to probe Hsp70's regulatory allosteric mechanism with respect to chaperones conformational transition between the substrate free inactive state and the substrate bound ATPase active state.

Chapter 3 builds on the PRS methodology presented in Chapter 2 to investigate nucleotide dependent allosteric modulation of conformational dynamics in the human cytosolic isoform of Hsp90. The absence of suitable experimental structural data for Hsp90α is addressed by homology modelling and novel insights regarding the nucleotide dependent conformational dynamics and underlying allosteric regulation of human Hsp90 in fully-closed, partially-open and fully-open conformational states explored using all-atom MD simulation analysis techniques.

Chapter 4 is a follow up study to Chapter 3, in which allosteric control elements identified at the C-terminal domain of the fully-open conformation of human Hsp90α are targeted in a series of computational molecular docking experiments, to probe allosteric modulation of conformational dynamics in the chaperone by small molecule interaction. This works provides for the first time “proof-of-concept” for the use of PRS to accurately identify allosteric sites on a protein that can be targeted by small molecule allosteric inhibitors/activators for drug discovery purposes.

Chapter 1:

Introduction

This chapter outlines several key biological concepts and methodologies pertinent to the thesis. Protein allostery is defined and its history and biotechnological application briefly described. The 70 kDa and 90 kDa molecular chaperones are introduced, detailing their biological function, structure, and allosteric modulation. Finally the aims of this thesis are laid out in context of allosteric regulation in molecular heat shock proteins.

1.1 Protein Allostery

The term 'allostery' was first introduced more than 50 years ago to describe regulation of enzymatic activity by small molecules that bind a site distinct from the substrate binding site [1]. These binding sites have since been termed 'allosteric' sites from the Greek word άλλα meaning other/different. Since its induction into the biological community in 1963, the definition for allostery has been refined to describe any type of long-range communication ('allosteric signaling') in proteins, in which excess energy stemming from perturbations at one site manifest as structural and/or dynamical changes elsewhere on the protein [2,3]. This phenomenon has since been reported for a plethora of biological molecules; including hemoglobin [4], aspartic-trans-carbamylase [5], homoserine-dehydrogenase [6], aspartokinase [7] and L-theronine deaminase [1]. The phenomenon of allostery remains a central focus in biology being implicated most cellular processes and its quantitative description is widely considered to be a fundamental requirement to understanding of most biological processes beyond the molecular level [2,8].

1.1.1 Brief history of allostery

Early experimental rationalization of allostery led to the development of two influential phenomenological models that formed an initial foundation for describing allostery in the absence of adequate structural data. In their classical paper, Monod, Wyman and Changeux introduced the concept of allostery to describe the cooperative O_2 binding phenomenon in the hemoglobin tetramer, where O_2 molecules binding at one monomer increases the binding affinity of additional molecules in the adjacent monomer [9]. This landmark work led to the development of the symmetric Monod-Wyman-Changeux (MWC) (Figure 1.1 – green) model, which proposed the idea of there being tensed (*T*) and relaxed (*R*) quaternary end point states whose equilibrium is shifted upon effector binding [9]. Later, Koshland, Nemethy and Filmer (KNF) proposed a sequential model for allostery (Figure 1.1 – red dash), positing an 'induced-fit' mechanism formalized on inherent protein flexibility, where effector binding at one domain triggers a conformational change in another, essentially describing allosteric transmission as the sequential propagation of structural deformations from the effector site to the active site [10]. Although both the MWC and KNF models successfully explained allosteric regulation for several biological molecules, their phenomenological origins and lack of sufficient structural data meant neither model was able to provide detailed insight as to the physical mechanism for allosteric communication between two sites.

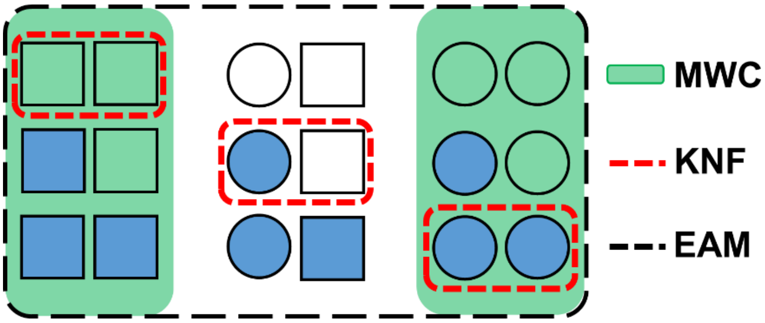


Figure 1.1: Schematic depiction of the founding allosteric models. A hypothetical two-domain protein is represented as an ensemble of conformational ligand bound states. The squares and circles denote the *T* and *R* states for each domain – blue shading represents the presence of bound effector. The possible conformational states accessible by each model are shown according the key. The MWC model is restricted to matching conformational states in both domains, while the KNF model demonstrates an 'induced-fit' mechanism. The more recent EAM model incorporates combinations of domain conformational states with/without bound effector. (Adapted from [11])

It was not until two decades later that the description of allostery was extended, particularly through revolutionary work by Cooper and Dryden in 1984 which introduced the idea of "dynamic allostery" [12]. The authors proposed that allostery could be manifest without structural change through entropy changes that alter the width of a conformational distribution without changing the average atomic coordinates and thus for the first time challenged the requirement for observable conformational change, which until this point was considered to be a fundamental characteristic signature of allostery [12]. The conclusions raised by Cooper and Dryden were established and supported by several theoretical [13], experimental [14,15], and computational [16–18] studies which highlighted the importance of conformational dynamics in protein function. Indeed, "dynamic allostery" and the concept of entropy driven allostery paved the way for the ensuing ensemble description of allostery [8,11,19–27].

Hilser and co-workers posited the ensemble allosteric model (EAM) which treats allostery as a macroscopic effect that arises from ligand-driven stabilization of cooperative elements [11] (Figure 1.1). Unlike the aforementioned MWC and KNF models, the EAM can be utilized to investigate how the thermodynamic architecture of a protein facilitates allosteric coupling [11]. Importantly, the EAM applies to all allosteric systems irrespective of their structural complexity (monomers, multi-domain, and multi-subunit proteins), as well as the type and magnitude of structural changes associated with allosteric transitions. The Nussinov group proposed a unified view of allostery from three separate standpoints: *thermodynamics*, *free energy landscape of population shift*, and *structure* using the same allosteric descriptors using the two-state allosteric model [2]. The thermodynamic view provides experimental quantification, emphasizing that allostery reflects preferred ligand binding to one of the two active/inactive states. The free energy landscape representation encapsulates preferred ligand binding and transforms population shift into an amount of energy that is proportional to the stabilization/destabilization of the active/inactive conformations.

In summary, allostery regulates biological molecules through a remote action-at-a-distance mode, and is a crucial requirement for function in cell life. Gaining an understanding of allostery is thus essential for the comprehension of a vast array of complex biological systems.

1.1.2 Allosteric regulation and its application

In a broad sense, allostery is the phenomenon whereby two physically distant sites on a biological molecule are dynamically coupled through structural excitation events at either site. The external nature of these excitation events (i.e. ligand binding, post-translational modifications, or structural deformations) present allostery as a tunable regulatory mechanism for many, if not all, enzymes and signaling proteins. In most cases, allosteric excitation at a site distant from the active site propagates via a precise allosteric pathway leading to functionally relevant conformational changes or concerted motions that modulate protein activity/function (Figure 1.2). Effector molecules that enhance protein activity are referred to as *allosteric activators*, while those that engender decreased activity termed *allosteric inhibitors*.

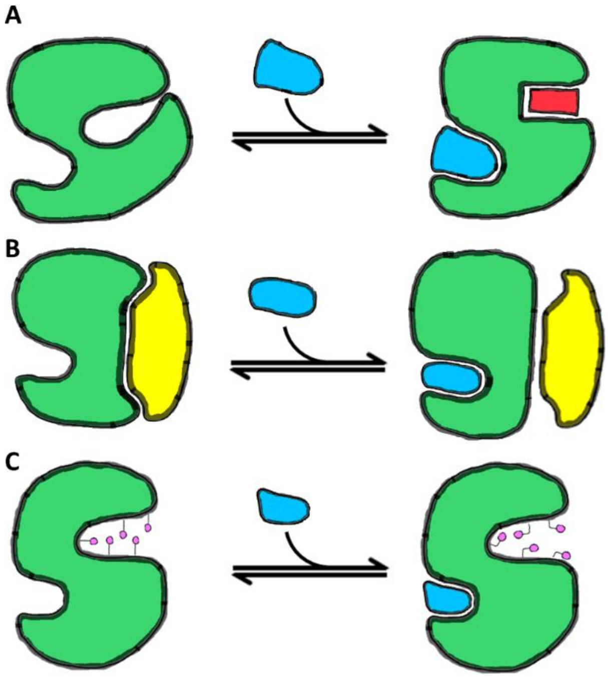


Figure 1.2: Illustration of the different allostery types. (A) Ligand-driven substrate binding: where ligand binding (blue) induces a conformational change that enables substrate (red) access to an otherwise inaccessible binding site. **(B)** Ligand-driven disruption of inter-protein interactions: where ligand binding (blue) results in disassociation of the of bound protein partner. **(C)** Ligand-driven dynamics: where ligand binding triggers dynamical changes at the active site (pink) ultimately affecting its activity. (Adapted from [28])

In recent years allosteric regulation has gained popularity as an alternative mechanism for the therapeutic regulation of proteins [2]. Unlike orthosteric drugs which perturb the active site by competing with the substrates of target proteins, allosteric regulators provide a “noninvasive” mechanism for the control of proteins through an action-at-a-distance mode meaning they (i) do not interfere with endogenous regulators, (ii) are often more accessible compared to buried active sites, (iii) can function as activatable modules, and (iv) can be utilized in concert with traditional active site regulators. Therapeutic allosteric regulation can also be exploited to improve compound selectivity [29–33]. Drug candidate selectivity is best demonstrated by the kinases which share structurally similar ATP binding sites. Drugs that compete with ATP for binding in the kinases are by nature highly promiscuous and their promiscuity results in poor specificity and inevitable undesirable drug site effects. In addition, allosteric sites do not appear to be evolutionarily conserved and thus offer unique ligand binding sites with high specificity [34]. Amino acid conservation is largely associated to protein folding kinetics, stability, or particular protein functions [35,36], they have also been shown to be capable of evolving to bind diverse ligands [37]. Taken together, allosteric sites by nature provide a unique opportunity to target select isoforms from homologous protein families [38].

The uniqueness of allosteric drugs is evidently one of their most appealing characteristics, however this aspect also brings several challenges from the standpoint of conventional drug discovery and in particular rational drug design [39]. These design programs have traditionally relied on static high throughput X-ray crystallographic structures of targets in complex with bound ligand, and in most cases biochemical and biophysical assays have been developed around these known orthosteric sites. While these efforts bring efficiencies to drug design programs, but are not routinely applicable for the rational design of allosteric drugs. As previously mentioned, the key to unlocking allosteric mechanisms is centered on the rationalization of the target’s dynamic conformational landscape – information that is not accessible through static protein imaging. Computational methods and techniques provide an answer to these limitations as they enable researches the ability to explore the time dependent dynamics of putative allosteric drug targets at an atomic level [39,40].

1.1.3 Structure based computational methods for investigating allostery

From an experimental standpoint, nuclear magnetic resonance (NMR) is an extremely powerful approach for measuring allostery in proteins given the ability of the technique to sample time scales in the order of picoseconds to nanoseconds for side chain dynamics and microsecond to millisecond for backbone dynamics [26,27,41–43]. Despite the experimental advantages of NMR, the amount of time and the relative cost of limit its extensive application for dynamics based resolution of protein structures and allostery. For this reason many researcher have turned to computational techniques to unearth dynamical data from static protein structures.

- *Molecular dynamics simulations*

From a dynamics view, allosteric communication arises through effector induced alteration of protein dynamics either through large-scale structural changes or via more subtle changes such as correlated residue motions. Given the intimate link between allostery and the dynamics of the protein, suggests a full understanding of the underlying dynamic conformational flexibility and energetic landscape are fundamental to properly understand allosteric mechanisms. Molecular dynamics (MD) simulations is a popular method for accessing dynamic atomic coordinate information for each particle in the system in three-dimensional (3D) space over time. Molecular systems are treated as beads (atoms) connected by springs (bonds) and Newtonian mechanics used to predict the time dependent position of every atom which is described by potential energy functions (known as force fields) that are used to approximate the underlying physics [44]. MD simulations have proven invaluable for investigating allostery in that they can provide an all-atom description of protein dynamics with high resolution in both time (femtoseconds) and space (sub-angstrom). The simulation time is discretized into steps of 1 to 2 fs to accurately simulate the fastest atomic motions such as hydrogen-bond stretching. The potential energy of the system is quantified at each simulation time step by considering the energies associated with bonded atoms (angle bending, torsion rotation etc.) and non-bonded atom pairs (van der Waal and electrostatic forces). The resultant forces on each atom are calculated from a negative gradient potential energy and used to update the atomic coordinates of each residue. While the all-atom description allows for the inclusion of specific interactions and a holistically more accurate description of the mechanics at play, it limits conformational sampling of the molecule by (i) having a more complex conformational energy landscape which enables the possibility of the molecule becoming trapped in an energy minimum, and (ii) because all degrees of freedom are accounted for, the computational cost is increased impacting total accessible simulation times.

- *Protein residue network graphs*

Protein network graphs are routinely used to visualize protein structures as a network of interacting residues. This method is centered on the concept of real world networks and thus requires selection of nodes and defining suitable criteria for connecting node pairs. Typically, 3D protein structures are represented as weighted graphs by representing each residue as a node in the network and edges established based on pre-defined residue-residue interactions (Figure 1.3). Graphs such as these demonstrate “small-world” network properties [45–48] and reflect the inherent cooperative behavior of macromolecular systems, by characterizing dynamic and thermodynamic properties [45,46,49], which ultimately facilitate allosteric signal propagation [28,50].

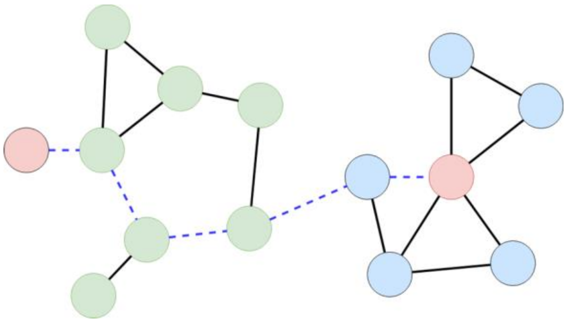


Figure 1.3: Protein residue network graphs. Featuring residue nodes with various decrees of connectivity ranging from one to four edges. The shortest possible path between the green and red nodes is denoted by a dashed blue edges. The neighbourhood of the orange node is represented by the surrounding yellow nodes.

In general, graph-based analyses primarily focus on partitioning the network into highly interconnected sub-graphs referred to as 'communities', and mapping the optimal communication routes between communities (Figure 1.3). Central to these graph-partitioning methods has been the definition of 'centrality' measures for the identification of critical residues that mediate intra-protein communication [45,49,51,52]. Several studies have used protein network graphs to identify functional and allosteric hot spots, domain motions, and protein- ligand/protein binding sites (see reviews [28,49,52]).

- *Normal mode analysis and elastic network models*

Normal mode analysis (NMA) uses statistical mechanics to examine protein vibrations around an energy minimum through the quadratic approximation of the energy potential [53–57]. This energy approximation requires diagonalization of the 3*N* x 3*N* Hessian matrix, where *N* is the total number of atoms in the system. This matrix operation maybe a limiting factor for larger proteins due to the higher matrix rank. The NMA methodology is simplified by the elastic network model, in which atoms that fall within a specified cut-off are considered to be connected by 'springs' and modeled using the Hookean potential [58,59]. This model was extended in terms of polymeric networks as the Gaussian network model by Bahar and colleagues where connected atoms are connected with a uniform spring constant [60,61]. The Gaussian model enables the sampling of individual modes, providing information on the mean-square-fluctuation and cross-correlations between atoms. Unlike the 3*N* x 3*N* Hessian, the Gaussian model is based on an *N* x *N* Kirchhoff matrix and thus cannot provide information on 3D motions. In contrast, the anisotropic model allows for the construction of the Hessian matrix based on the Cartesian coordinates of C_α, providing a coarse-grained NMA approach. The introduction of normal-mode analysis had a major impact on structural biology in that it enabled the dynamics based studies of large-scale macromolecular machinery, and has since been extensively used for the detection of allosteric sites [62,63] and the mapping of allosteric communication pathways [28,39,64–66].

1.2 Molecular Chaperones

Protein folding in living cells is a much more complex process than the controlled *in vitro* environment. Varying temperatures and the high concentration of proteins in the cell (ca. 200 mg/ml) contribute to a molecular "crowding" effect that leads to fierce competition between folding processes often resulting in aggregation and misfolding. Cellular protein aggregation is a fatal outcome for the cell as it interferes with important subcellular processes and in many instances is the cause of neurodegenerative disease such as Alzheimer's disease, prion diseases and Parkinson's disease [67,68]. To circumnavigate the risk of protein aggregate formation, cells maintain a set of specially designed proteins called molecular chaperones, which play a crucial role in assisting the *de novo* folding process of newly synthesized proteins and the re-folding of misfolded proteins under natural and environmental stress conditions such as high temperatures [69,70]. The realization that chaperones are upregulated under heat shock conditions led to the descriptive term of "heat shock protein" (Hsp) [71]. As a class the Hsps are highly conserved existing in most kingdoms of life and are named/grouped according to their respective molecular weights: e.g. Hsp110, Hsp100, Hsp90, Hsp70, Hsp60, Hsp40, and the small Hsps [72,73].

The fundamental feature of chaperone function is the transient binding of their peptide/protein substrates. Hsps are classed into three functional groups: *holder* chaperones, *folder* chaperones, and *disaggregases*. The "holder" chaperones are responsible for preventing unfolded peptides from interacting with each other by binding them and thus prevent formation of protein aggregates. Member of this group include the smaller Hsps (e.g. Hsp26, and Hsp40/DnaJ). The "folder" chaperones are unique in that they actively assist in the protein folding process. Due to the energy requirements for this function, members of this group typically harbor an ATPase domain. Two well-studied examples include the Hsp60 chaperonins (e.g. *E. coli* GroEL, human TRiC) and the Hsp70 chaperone family (e.g. *E. coli* DnaK and human Hsp70) [74]. As expected from their group name, the "disaggregases" have disaggregating activity and are thus able to dissolve already formed protein aggregates and release polypeptides for refolding/degradation. Examples of this group include the Hsp100 family members, *E. coli* ClpB, and *S. cerevisiae* Hsp104. In collaboration with members from the Hsp70 family the disaggregases are able to solubilize aggregated proteins [75,76].

1.2.1 The Hsp70 chaperone machinery

– *Cellular functions of Hsp70*

The 70-kDa heat shock proteins (Hsp70) form a central part of the ubiquitous chaperone system and are present in most cellular compartments, where they assist in a wide range of folding processes [77,78] (Figure 1.4). Participation in these processes uniquely positions Hsp70 at the center of several quality control functions which can be divided up into two sets: (i) general housekeeping functions under normal conditions and (ii) stress related functions when the cell is under environmental insult such as heat shock. Under normal conditions Hsp70's primary function is assisting in the *de novo* folding of native proteins, in addition to several other responsibilities including: the disassembly of protein complexes such as clathrin coats [79,80], translocation of proteins from the cytosol into the mitochondrial matrix [81] or endoplasmic reticulum (ER) [82,83], and the regulation of the heat-shock response and control of regulatory proteins in tandem with the Hsp90 system. Under extreme conditions such as heat-shock, hypoxia, increased pH or salt etc. Hsp70 play a crucial part in the rescue of misfolded, partially denatured, or aggregated proteins [76,84–87].

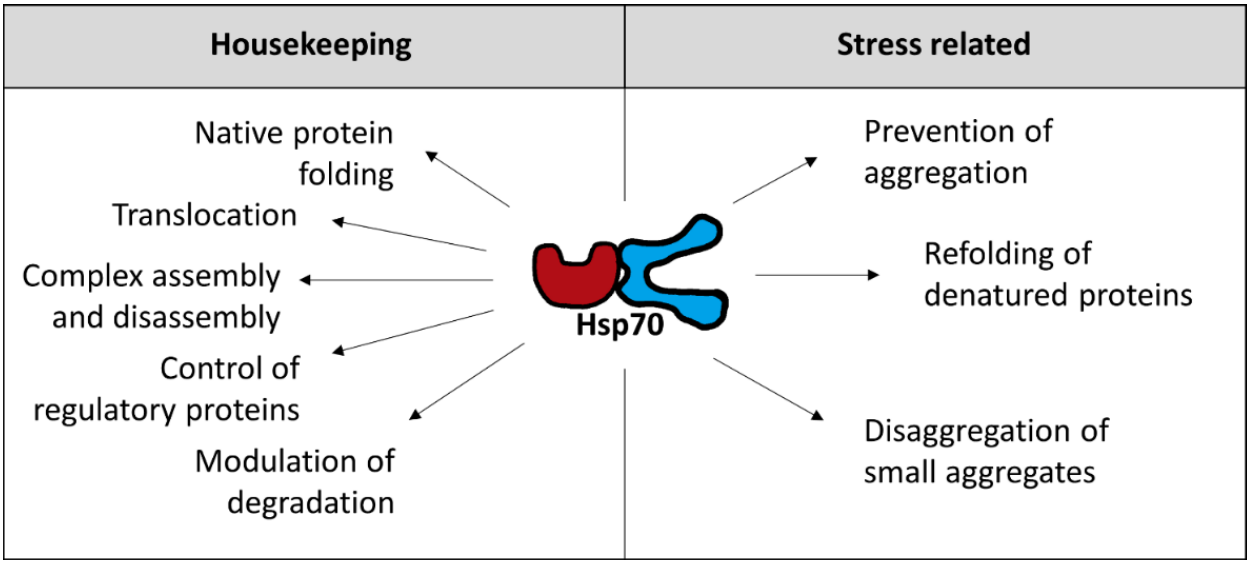


Figure 1.4: Cellular functions of the Hsp70 chaperones.

– *Hsp70 structure*

Biochemical and structural studies have revealed that the structure of Hsp70 is highly conserved and consists of two well-defined domains, a 44-kDa nucleotide binding domain (NBD) at the N-terminal and a 25-kDa substrate binding domain (SBD) at the C-terminal (Figure 1.5-A). The NBD and SBD are connected via a short 10-12 residues hydrophobic linker (Figure 1.5 B) [74]. The structure of Hsp70's NBD is homologous to hexokinase, actin, and glycerokinase. The domain is made up of two lobes (I and II) that together form a V-like cleft. Each lobe is further divided into two subdomains: IA, IB, IIA and IIB that together form a nucleotide binding pocket located at the base of the V-like cleft (Figure 1.5). Cross-connecting helices from each lobe position ß- and γ-binding loops that provide contact points adenosine nucleotide binding [88]. X-ray crystal structures of the ATPase domain of human, bovine, and bacterial Hsp70 have indicated minor conformational restructuring in response to bound nucleotide [88,89], while NMR studies have revealed enhanced NBD flexibility with predominant shearing and tilting motions of the subdomains relative to each other that engender opening and closing of the nucleotide binding cleft [90,91]. The opening frequency has been reported to be in the order of apo > ADP > ATP [92–94]. The SBD of Hsp70 is composed of a two-layered hydrophobic ß-sandwich (SBDß) which forms a well-defined peptide binding cavity that contains a central pocket that binds a single hydrophobic residue of the peptide substrate. A five member α-helical lid segment (SBDα) acts as a mechanical arm that encloses bound peptide in the SBDß [95]. The extreme C-terminal is composed of an intrinsically disordered 30 residue segment of an as yet uncertain function. The presence of an EEVD motif at the C-terminus in the cytosolic eukaryote isoform suggests involvement in chaperone activity and interaction with the co-chaperones Hop/Sti1 and chip [95–100].

Over the past two decades, experimental structural studies of full-length Hsp70 have revealed two well-defined nucleotide-dependent conformations of Hsp70. In particular, X-ray crystallography structures of ATP-bound DnaK (ATP-DnaK) broadened our knowledge of the nucleotide dependent mechanism of inter-state conversion and the ATP-driven opening of the substrate biding cavity at the SBD [101,102]. In the ATP bound state the chaperone has a compact structure in which the SBDα and SBDß are completely detached from each other and docked onto two sides of the NBD. This inter-domain association is initiated and guided by the precise packing of the inter-domain linker into a binding cleft on the NBD that arises due to ATP-driven conformational repositioning of the IA and IIA subdomains relative to each other [101]. The association and interaction of the linker with the NBD brings the SBD into close contact with the NBD, allowing stabilizing inter-domain interactions to form with both the SBDα and SBDß subdomains [101,102]. In this compact state, the

SBDα and SBDß are completely detached from each other such that the SBDß binding cavity is completely exposed and thus more accessible to peptide substrates (Figure 1.5-A). Solution NMR studies show that the ADP bound and nucleotide-free states engender dramatic structural rearrangements in which the NBD and SBD undock from each other releasing the inter-domain linker [91,103,104]. In this state, the NBD and SBD are largely independent allowing the SBDα to fold over the SBDß binding cavity in a lid-like fashion to entrap bound substrate (Figure 1.5-B).

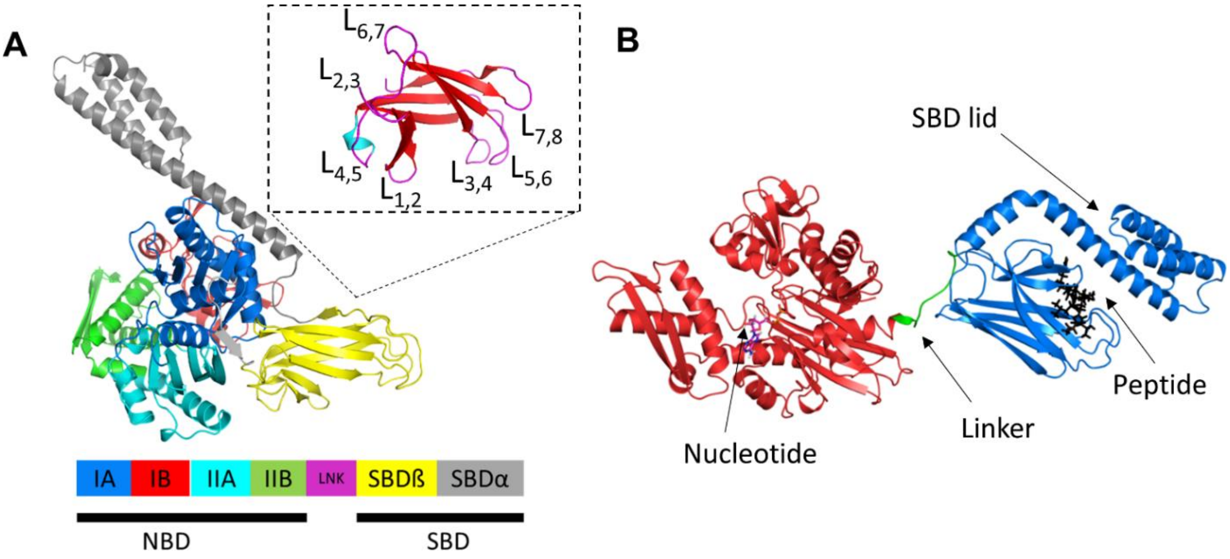


Figure 1.5: Experimentally determined structures of *E. coli* Hsp70 DnaK. (A) X-ray crystal structure of the open ATP bound conformation (PDB ID 4B9Q) colored by subdomain. The inset hyphenated box describes the loop regions of the SBD indicating: $L_{1,2}$ (res 404-406), $L_{2,3}$ (res 412-420), $L_{3,4}$ (res 428-434), $L_{4,5}$ (res 439-457), $L_{5,6}$ (res 458-476), $L_{6,7}$ (res 479-482), and $L_{7,8}$ (res 490-496). **(B)** NMR structure of the closed ADP and peptide bound conformation, showing ADP (magenta sticks) bound to the NBD (red) and peptide substrate (black sticks) bound to the SBD (blue). Reproduced with permission form Penkler et. al 2017 [105]

Several studies have demonstrated that the Hsp70 chaperone system is highly dynamic and that the two aforementioned conformations resemble mere snapshots of the chaperone's conformational ensemble. Nevertheless, these conformations are widely thought of as structural endpoints in the chaperone's functional conformational cycle and are commonly referred to as the ATP-bound "open" and ADP-bound "closed" states, due to concomitant opening and closing of the SBD in response to the NBD nucleotide bound configuration. This is largely due to the understanding that while Hsp70s are able to transition between the open and closed states irrespective of the nucleotide status [101,106–108], binding of ATP/ADP constrain the diverse conformational ensemble into two different sets, and thus modulate the frequency of structural transitions and therefore the equilibrium between the open and closed conformers [109].

– *Allosteric modulation of Hsp70's conformational cycle*

Central to the Hsp70-mediated folding process is the chaperones ability to bind and release unfolded polypeptide substrate segments via a nucleotide dependent conformational cycle in which the chaperone transitions between the open and closed states (Figure 1.6) [70,74,110,111]. The binding and release of clientele from the SBD is intricately modulated through the nucleotide dependent allosteric coupling of the NBD and SBD, allowing for allosteric tuning of Hsp70's conformational dynamics though ATP binding and hydrolysis. This functional conformational cycle can be summarized by four key stages/steps (Figure 1.6 1-4).

(**1**) In the ATP-bound state, Hsp70s are in equilibrium between the open and closed state favoring the open conformation in which the SBD is suitably exposed for peptide loading (Figure 1.5-A). In this state, Hsp70 has a low substrate binding affinity which is coupled to high association/dissociation rates giving rise to the so called low-affinity state. Several studies have shown that Hsp70 proteins have a higher affinity for ATP (~1 nM) than for ADP/P_i (~10 nM), which implies that under most *in vivo* conditions (~3 mM ATP) they exists in an ATP bound state.

(**2**) Due to the low affinity state of the open conformation, substrate loading is facilitated the Hsp40 chaperone family (J-domain proteins; JDP) (Figure 1.6-2) [112]. Substrates associate with JDPs before being transferred to the Hsp70 with high transfer rates and the physical action of substrate loading triggers ATP hydrolysis at the NBD. The basal rate for ATP hydrolysis in DnaK has been reported to be ~19 min, and thus too low to be physiologically relevant [113–119]. To circumnavigate this, it appears that JDPs such as DnaJ in *E. coli*, not only facilitate substrate binding but also accelerate Hsp70's ATPase activity 1000-fold [111,120]. The JDPs thus play a crucial role in facilitating the efficient transfer of substrate to Hsp70, by enhancing the coupling of ATP hydrolysis with substrate loading. Cells encode a broad spectrum of JDPs which differ in their domain composition some of which display varying substrate specificities [121,122].

(**3**) ATP hydrolysis and subsequent binding of ADP at the NBD gives rise to a stable closed conformational state (Figure 1.5 -B). Under these conditions the Hsp70 protein is only able to access the open conformation transiently, leading to the high-affinity state with low dissociation rates (Figure 1.6-3). The acceleration of ATPase activity by Hsp40 co-chaperones, leaves this equilibrium state as the rate-limiting step In Hsp70's functional cycle, whereby ADP release together with inorganic phosphate must occur before subsequent ATP binding can return the chaperone back to the low-affinity open conformation.

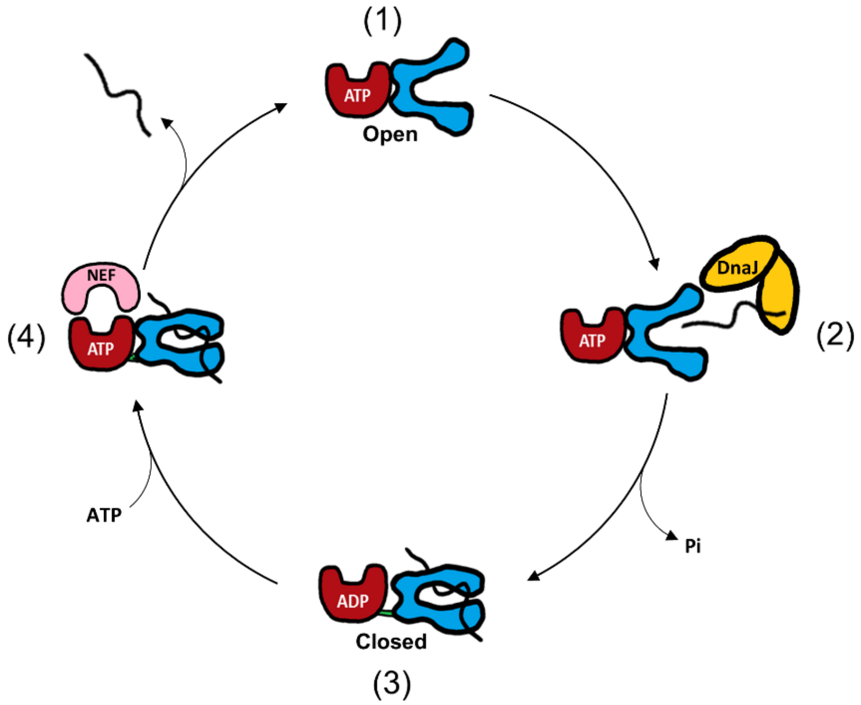


Figure 1.6: Schematic representation of Hsp70s nucleotide dependent conformational cycle. (1) Bound ATP at the NBD engenders an open conformation with low substrate affinity but high transfer rates. **(2)** The co-chaperone DnaJ facilitates substrate binding to the SBD, triggering ATP hydrolysis and inter-domain undocking. **(3)** Bound ADP at the NBD engenders conformational changes at the SBD entrapping the substrate resulting in a high substrate affinity state with low on/off transfer rates. **(4)** Intervention by NEF facilitated ADP exchange for ATP at the NBD, triggering conformational restructuring resulting in inter-domain docking and the opening of the SBD and eventual release of the peptide substrate. Reproduced with permission from Penkler et. al 2017 [105]

(**4**) As with JDP facilitation of ATPase activity, several Hsp70 proteins acquire the services of *in vivo* co-factors to accelerate nucleotide exchange. Most bacteria and eukaryotic organellar Hsp70 homologs interact with the GrpE family of nucleotide exchange factors (NEF) [123–127], which lower ADP affinity and increase the rate of nucleotide exchange by 5000-fold. This aspect of Hsp70's conformational cycle was confirmed by the X-ray crystal structure of GrpE in complex with the nucleotide-free NBD of DnaK [89]. In the proposed functional model of GrpE, the co-factor inserts a helical bundle into the ATP binding cleft, mechanically widening the binding site and thus accelerating the dissociation of ADP [128]. Release of ADP from the NBD and the subsequent binding of ATP allosterically triggers a conformational shift back to the ATP-bound open conformation and the eventual client release.

Hsp70's ability to allosterically communicate the relative nucleotide/substrate bound state of the NBD and SBD in a bi-directional manner, formalizes the underlying mechanism for functional regulation. Allosteric coupling between the two functional domains is absolutely crucial to this process: the physical action of binding at the SBD in synergism with JDP, allosterically triggers ATP hydrolysis at the NBD which in turn triggers conformational transition towards the high-affinity state.

Conversely, exchange of ADP for ATP in synergism with NEF at the NBD initiates an allosteric trigger signal for conformational restructuring at both the NBD an SBD that ultimately induces transition from the high-affinity state to the open low-affinity state. The exact mechanism behind this intricate bi-directional allosteric communication on the structural level remained elusive and poorly understood for many years due to the dynamic nature and size of the multi-domain chaperone. However, the availability of improved structural data coupled with detailed biochemical studies and the development of computational methods for predicting allostery in protein structures has led to a more unified understanding of the mechanism in recent years [129–131]. The details surrounding Hsp70's allosteric mechanism and the techniques used for its elucidation are discussed at length in Chapter 2 of this thesis, in addition to Perturbation response scanning as a suitable computational approach for predicting allosteric hotspots in large multi-domain proteins.

1.2.2 The Hsp90 chaperone machinery

The 90 kDa heat shock protein (Hsp90) is an essential homodimeric chaperone in all eukaryotic cells and make up about 1 – 2% of all cellular proteins under normal biological conditions. The Hsp90 chaperone family is highly conserved across all kingdoms of life with the exception of archea, sharing a sequence identity of 60% between yeast and human, and 40% between *E. coli* and human homologs. Eukaryotic Hsp90 is an essential player in signal transduction networks in the cell, interacting with a broad range of client substrates in complex with Hsp70 and an array of co-chaperones. Two cytosolic isoforms of Hsp90 are present in the cytosol of humans and yeast, sharing 99% sequence identity. The first isoform Hsp90ß in humans and Hsc82 in yeast is constitutively expressed at high levels under normal conditions. The second isoform Hsp90α in humans and Hsp82 in yeast, is mass produced under heat shock conditions and regulated by the heat-shock transcription factor HSF-1 [132–134]. In addition, other compartmentalized isoforms of Hsp90 exist, the endoplasmic reticulum inhabitant Grp94 in higher eukaryotes [135,136], and the mitochondrial isoform TRAP1 [137], which is a descendent of the bacterial homolog HtpG [134].

– *Cellular functions of Hsp90*

While Hsp90 chaperones are also able to bind misfolded proteins [138], their main cellular function is to interact with and stabilize the activity of native or near-native client proteins. To date several hundred proteins have been implicated as clients of Hsp90, placing the molecular chaperone as a central modulator of many different cellular processes ranging from protein folding and cellular homeostasis at both physiological and stress conditions, to DNA repair, cell cycle control, proliferation, differentiation, and neuronal signaling, to name a few (Table 1.1) [139].

Table 1.1 | Select Hsp90 clients

Client	Function	Related disease
E3 ubiquitin ligases		
MDM2	P53 degradation	Cancer
UHRF1	DNA methylation	Cancer
Kinases		
AKT (PKB)	Mitogen signaling	Cancer
CDK4	Cell cycle regulation	Cancer
ERBB2	EGF receptor	Cancer
HCK	Immune response	Cancer
JAK1/JAK2	Cytokine signaling	Cancer
SRC	Constitutively active Tyr kinase	Cancer
BRAF	Mitogen signaling	Cancer
BCR-ABL	Constitutively active Tyr kinase	Cancer
Steroid hormone receptors		
Androgen receptor	Androgen response	Spinobulbar muscular atrophy
Glucocorticoid receptor	Glucocorticoid response	Cushing syndrome
Mineralocorticoid receptor	Mineralocorticoid response	Chronic kidney disease
Oestrogen receptor	Oestrogen response	Cancer
Progesterone receptor	Progesterone response	Cancer
Transcription factors		
HIF1α	Angiogenesis	Cancer
p53	Tumor suppressor protein	Cancer
OCT4	Embryonic development	Cancer
PPARα, PPARß, PPARγ	Fast and glucose metabolism	Diabetes mellitus
STAT2, STAT3, STAT5	Cytokine signaling	Cancer
Other		
Argonaute 1, Argonaute 2	RNA interference	
Calcineurin	Immune response	Rheumatoid arthritis
Calmodulin	Signaling pathways	
CFTR	Chloride channels	Cystic fibrosis
HSF1	Regulation of heat shock response	
NLR proteins	Innate immune response	Inflammation
TERT	Telomere maintenance	Cancer
eNOS	Nitric oxide synthesis	
RAD51/ RAD52	DNA repair	Cancer
Tau	Microtubule stabilization	Alzheimer disease

See https://www.picard.ch/downloads for a comprehensive list of Hsp90 clients

Of these client proteins, the SHRs and protein kinases have been the most extensively studied [140–144]. A quantitative study of Hsp90 clients in humans revealed that approximately 60% of the known kinome and 30% of E3 ubiquitin ligases associate with Hsp90, while only 7% of transcription factors bind Hsp90 [145]. Despite this promiscuity of Hsp90 binding, it is apparent that Hsp90 clients are limited compared to other chaperones such as Hsp60 or Hsp70, which interact with most if not all unfolded proteins. Hsp90s client interactions can classified into three broad categories: (i) formation of a specific active client conformation, typically seen in the case of kinases [146,147]; (ii) facilitate the assembly of multi-member protein complexes, such as the kinetochore [148] and Rvb1-Rvb2-Tah1-Pih1 [149] complexes; and (iii) promote protein-ligand interactions by stabilizing a binding-competent receptor conformation. Binding of steroid hormones to SHR [150–152], telomerase association with DNA [153,154], and Agronaute 2 loading of siRNA [155] are all examples of the third category. In addition to these categorical functions, regulation of protein dynamics and the ensemble of conformational states as a function of Hsp90 interactions has also been reported [156]. Given that Hsp90 facilitates a wide range of proteins that act as central ‘hubs’ in cellular networks and signaling cascades, positions the chaperone as a key regulator for a large subset of the proteome by integrating a plethora of signaling pathways.

– *Hsp90 structure*

Hsp90 functions as a homodimer, and dimerization has been shown to be essential for its *in vivo* function [157,158]. Several biochemical and structural studies have revealed that each protomer consists of three well characterized, highly conserved domains [158–160]: the N-terminal domain (NTD) is responsible for ATP binding and hydrolysis as well as facilitating protomer dimerization [161]; the middle domain (MiD) is implicated in ATPase activation [162] and provides a large surface area for client interactions; the C-terminal domain (CTD) is the primary site for inter-protomer dimerization [163,164]. In eukaryotes the NTD and MiD are connected via a charged highly disordered flexible inter-domain linker region of about 50 residue in length, that modulates NTD-MiD contacts and affects Hsp90 function [165–167]. Eukaryotic Hsp90’s contain a C-terminal Met-Glu-Glu-Val-Asp (MEEVD) motif that is not present in prokaryotes, and functions as a binding site for several cytosolic co-chaperones that contain tetracopetide repeat TPD domains [168].

The first crystal structure of the NTD was solved for yeast Hsp82 in 1997 [161] revealed a unique ATP binding site made up of an unusual binding pocket known as the “Bergerat fold” [169]. Nucleotides bind to a cleft formed by α-helices positioned above an 8-member ß-sheet, and assumes a special kinked orientation in which the γ-phosphate is completely exposed. The homology of this particular fold is shared across the GHKL-ATPase superfamily, including DNA-modifying protein GyraseB, the histidine kinases CheA, and EnvZ, as well as MutL. A common structural feature of the ATPase domain of these proteins is a flexible ATP-lid, formed by two α-helices that when folded, over entrap bound nucleotide in the binding pocket (Figure 3.6). This lid structure has been heavily implicated in modulating conformational dynamics acting as a conformational switch for the ATP hydrolysis function of Hsp90 and also forms a binding site for several co-chaperones such as p23/Sba and Cdc37 [163,170,171]. The NTD of Hsp90 has also been crystalized in complex with several small molecule inhibitors [172]. Structural data of the MiD in isolation was resolved for yeast [162], and also in complex with the co-chaperone Aha1 [173], while a two domain construct of the NTD and MiD was resolved for *E. coli* HtpG [174]. The MiD is comprised of a large and small α-ß-α-domain that is interconnected by a helical coil. The MiD contains a catalytic loop with an exposed arginine (R380 in yeast Hsp82) that crucially interacts with the γ-phosphate of bound ATP at the NTD to initiate ATP hydrolysis. The MiD is heavily implicated in protein-client interactions via to structural elements involving an exposed hydrophobic patch surrounding Trp300 and an amphipathic loop (res 329-339). Mutations to the latter region was shown to severely affect v-Src kinase maturation [162]. The CTD plays an essential role for the functional dimerization of Hsp90 [175]. Crystal structures of CTD of *E. coli* HtpG revealed a mixed α/ß domain, the dimer interface being formed by a pair of helices at the C-terminal end of the domain where they form a four-helix bundle [164] (Figure 1.7). A flexible amphipathic helix (HtpG res 549-558) has been suggested to be involved in protein interactions [164]. Several independent research groups have provided evidence of a second putative nucleotide binding site at the CTD – a sight that has also been proposed to be a possible binding site for CTD drugs such as Novobiocin and *csi*-platin [176–178]. Due to the disordered nature of the last 30 residues of Eukaryotic Hsp90, the MEEVD TPR domain has not been structurally resolved.

Efforts to analyse full-length Hsp90 at an atomic resolution have been largely hindered by the dynamic nature of the dimer under normal conditions and for many years, structural data could only be determined for each of Hsp90's three stable domains in isolation. More recently, extensive crystallization efforts by Ali *et al*. (2006) provided a first glimpse of full-length yeast Hsp82 in complex with the non-hydrolysable ATP analogue AMPPNP as well as the co-chaperone p23 [163]. This structure of Hsp90 revealed dimerization at both the NTD and CTD, with two p23 molecules bound symmetrically either side of the NTDs (Figure 1.7). Interestingly, this "closed state" structure could only be achieved when the flexible NTD-MiD was deleted, and resembles the closed "molecular clamp" proposed by Prodromou and colleagues for stable client binding [179]. The two Hsp90 protomers twist around the central axis and interact through N-terminal strand swapping.

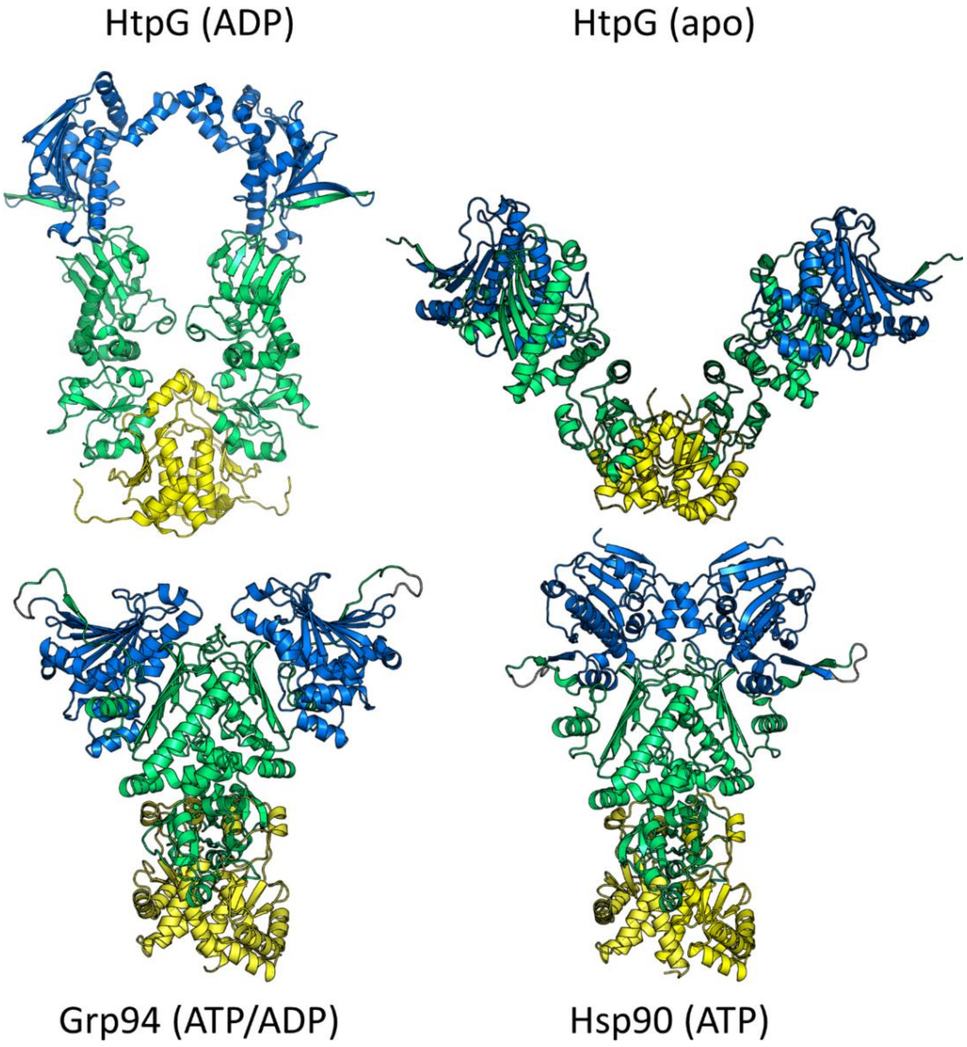


Figure 1.7: Full-length experimental structures of Hsp90. The bacterial Hsp90 homologue HtpG was resolved in a flexible open apo-state (PDB ID 2IOQ) and in a compact ADP-state (PDB ID 2IOP). The ATP and ADP bound complexes of mammalian Grp94 was resolved in identical partially-open conformations (PDB ID 2O1U). Yeast Hsp90 was co-crystalized with ATP, revealing a closed N-terminal dimerized state (PDB ID 2CG9). The three functional domains are coloured: NTD-blue, MiD-green, and CTD-yellow.

Crystal structures and electron micrographs (EM) of *E.coli* HtpG in the absence of co-chaperones indicated similar domain structuring to that observed for the yeast Hsp90-p23 complex, but noted that the arrangement of the domains relative to each other differs drastically in a nucleotide dependent manner. In these structures, the NTDs were observed to be either far apart in an “open state” conformation, or form small contact surfaces on the opposite face under specific crystallization conditions [180]. Electron micrographs (EM) of full-length Hsp90 in the absence of co-chaperones, revealed multiple distinct conformational states consistent with hinge bending motions between the domains, and provided further evidence of nucleotide modulated conformational dynamics, demonstrating that bound ATP analogues induce a conformational shift in favour of the closed state in a similar orientation to that observed for the crystal structure of the Hsp90-p23 complex [181]. Collectively these structural studies formed a basis for establishing a working hypothesis for Hsp90's mechanism of action, in which Hsp90's conformational ensemble of can be influenced by nucleotide and/or co-chaperone binding.

- *The nucleotide driven allosteric conformational cycle of Hsp90*

Hsp90's ability to bind and release client proteins revolves around a complex nucleotide dependent conformational cycle in which the dimer transitions between an open catalytically inactive state, and an ATPase active closed state (Figure 1.8).

(1) Crystal structures of HtpG in the nucleotide free state revealed CTD dimerization and an open "v-like" conformation in which the MiDs are exposed in preparation for client binding [164,180,182] (Figure 1.7 and Figure 1.8-1). Binding of ATP at the NTD triggers conformational rearrangements that lead to NTD dimerization and a conformational transition towards the closed catalytically active state. The molecular switch for this nucleotide dependent conformational restructuring has been attributed to the ATP-lid, which has to first close over the ATP binding pocket to entrap bound nucleotide – a conformational change that exposes essential hydrophobic surface residues that are important facilitators of NTD dimerization (Figure 1.7) [163,183].

(2) Conformational transition to the closed state and full ATPase activation is an inherently slow process, recording time constants in the order of minutes [161,184,185], and is likely to occur via structural intermediates which may present energy barriers that must be overcome [186,187]. Such structural intermediates have been tentatively reported for the ER resident mammalian Hsp90, Grp94, crystal structures of which revealed structurally identical partially open/closed intermediate structures in the presence of either non-hydrolysable ATP or ADP [188] (Figure 1.7 and Figure 1.8 2 & 4). In addition to nucleotide conformational regulation, co-chaperones have been shown to play an important role in the modulation of ATPase activity in Hsp90 [183,189]. Interactions with Aha1 at the NTD-MiD interface has been shown it increase ATPase activity 10-fold by promoting the closing transition by facilitating ATP-lid closure [187] and reorienting the MiD loop towards the ATP binding site such that Arg-300 is able to coordinate with the γ-phosphate of ATP and initiate ATPase activity [162,190].

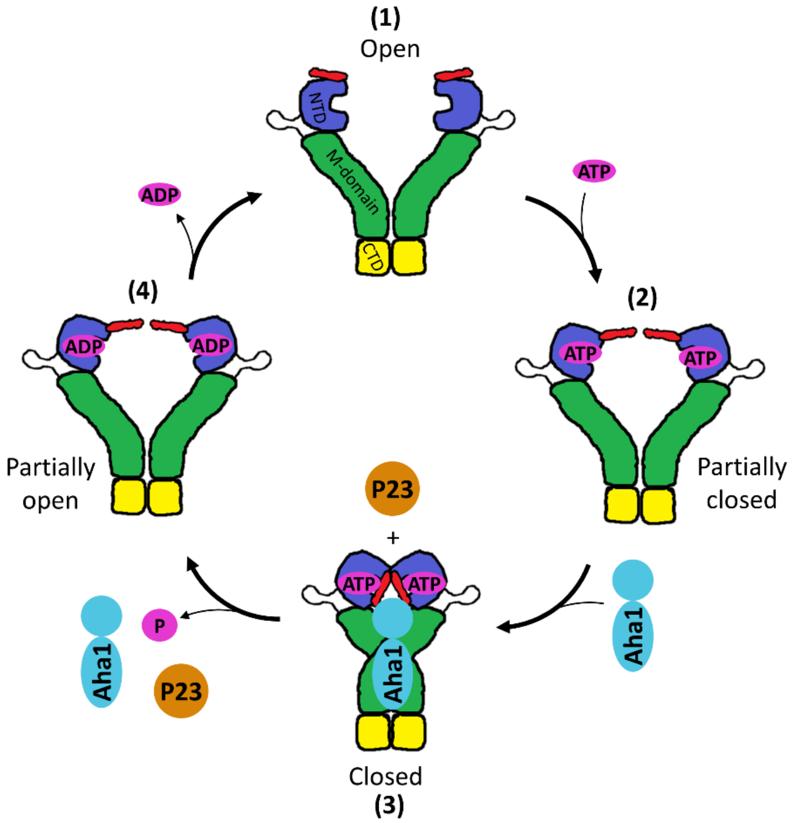


Figure 1.8: Schematic of Hsp90's nucleotide dependent conformational cycle: (1) Nucleotide-free open conformation; (2) ATP binding induces NTD restructuring and the slow closure of the ATP-lid; (3) Aha1 accelerates ATP-lid closure which promotes NTD dimerization and a conformational transition to the closed state; (4) ATP hydrolysis and the subsequent binding of ADP triggers protomer uncoupling and a conformational transition back to the open state. Reproduced with permission from Penkler et. al 2018 [191]

(3) The catalytically active closed conformation of Hsp90 has been described through crystal structures of ATP bound yeast Hsp82 in complex with p23/Sba1 [163] and human Hsp90ß in complex with Cdc37 and Cdk4 [192]. Both structures demonstrate asymmetric protomer arrangements, displaying inter-protomer twisting around the central axis, with fully closed ATP-lids and projection of the MiD catalytic loop towards the respective ATP binding pockets. In the case of yeast Hsp82, the interaction of p23/Sba1 at the NTD suggests stabilization functionality for the co-chaperone [163].

(4) ATP hydrolysis and the subsequent binding of ADP at the NTD, triggers further conformational restructuring and a shift back to a partially open intermediate state. The co-chaperones Aha2 and p23/Sba1 are released from the complex permitting subsequent protomer uncoupling at the NTDs (Figure 1.8-4). The ATP-lids are allowed to open enabling the release of ADP from the nucleotide binding pocket completing the conformational switch back to the nucleotide free open conformation.

– *Regulation of Hsp90 function*

Known mechanisms for the regulation of Hsp90 function include transcriptional regulation and post-translational modifications (PTM). In the case of the former, it is widely known that Hsp90 expression is induced by the stress-related transcription factor HSF, which is also a known Hsp90 client [193]. The current understanding for this regulatory mechanism is that Hsp90 together with Hsp70 bind freely available cytosolic HSF and keep it in an inactive state [194]. In the event chaperone proteins are required for other cellular functions and are thus not available for HSF1 inhibition, the active transcription factor increases transcription of the heat shock genes. Given that HSF is also a client of Hsp90, positions the chaperone as a link between the stress status of the cell and the expression of other heat shock proteins. In keeping with its function as a central hub for integrating diverse cellular signals, Hs90 function is also known to be modulated by numerous PTMs, including phosphorylation, acetylation, S-nitrosylation and sumoylation (for reviews see [160,189]). For the most part, phosphorylation is thought to slow down Hsp90's conformational cycle and affect client maturation and co-chaperone interactions [195–199]. Hyper-acetylation of Hsp90 has been shown to lead to an inability to bind the co-chaperone p23.Sba1 [200], while specific changes in the acetylation state are key regulators of co-chaperone binding [201]. S-nitrosylation of Hsp90 at a CTD Cys residue has been implicated in ATPase inhibition and disrupts the activating effect of Hsp90 on endothelial nitric oxide synthase [202]. In addition, S-nitrosylation has also been shown to affect chaperone activity [203]. Interestingly, a picture is emerging from the analysis of various PTMs at different location on Hsp90, that suggests several of the modified sites may function as allosteric switch points that regulates inter-domain communication across the dimer [195,203], raising the possibility that other PTM sites may define additional allosteric switch points.

– *Hsp90 co-chaperones*

Co-chaperones are arguably the most important regulators of Hsp90. Research in recent years has detailed the interaction and functional mechanisms of the majority of known co-chaperones with the molecular chaperone (Table 1.2). Co-chaperone binding sites on Hs90 have been identified in all three functional domains covering a large part of its surface. On current evidence, it appears that some co-chaperones act synergistically through simultaneous binding, and that others compete for overlapping binding sites, and the relative functions of the co-chaperones present in the same complex are thought to integrated for client protein processing. Indeed, some chaperones have been shown to act as regulators for the chaperone cycle and others as adaptors for client protein recruitment. Although in some cases the same chaperone has been shown to carry out both functions. For the most part, the known chaperones can be divided into two main groups, those with a TPR domain which interact with the MEEVD peptide at the CTD of Hsp90, and those with no TPR domain.

Table 1.2 | Hsp90 co-chaperones and their regulatory functions

Co-chaperone	Hsp90 binding site	Function
Aha1	NTD, MiD	Stimulates ATPase activity
Cdc37	NTD, MiD	Prevents ATP-lid closure; maturation of kinases
Chip	MEEVD	E3 ubiquitin ligase
Cyp40	MEEVD	PPIase; maturation of the glucocorticoid receptor
Fkbp51/52	MEEVD	General co-chaperones
Hop	MEEVD, MiD, CTD	Stabilizes the open Hsp90 conformation; Facilitates client transfer from Hsp70-Hsp40 to Hsp90
P23	NTD, MiD	Inhibits Hsp90 ATPase activity; stabilizes the closed Hsp90 conformation
PP5	MEEVD	Dephosphorylates Hsp90; maturation of kinases
Sgta	NTD	Facilitates NLR maturation in plats and kinetochore in yeast
Tah1	MEEVD	Assists in the formation of the Rvb1-Rbv2-Tah1-Phi1(R2TP) complex
Ttc4	MEEVD	Important for Cdc6 interaction
Unc45	MEEVD	Important for myosin-dependent processes

See references [159,188,202] for an extended list of co-chaperones

One of the most well-known TPR domain-containing co-chaperones is the Hsp70-Hsp90 organizing protein (Hop), which is widely thought to be involved in facilitating the transfer of client proteins from Hsp70 to Hsp90 [204,205]. In addition its dual adaptor function for both heat shock proteins, Hop has also been implicated in the inhibition of ATPase activity by stabilizing the open conformation [206–208]. Another TPR-containing co-chaperone is PP5 which is thought to modulate Hsp90's conformational cycle through dephosphorylation of Hsp90 and the Cdc37 co-chaperone [195,209]. Several other TPR-containing co-chaperones such as Fkb51, Fkb52, and cyclophilin 40 (Cyp40), also possess a peptidyl-prolyl *cis-trans* isomerase (PPIase) domain. These co-chaperones are implicated in the regulation of Hsp90's conformational cycle, and due to their chaperone activity may also be involved in client protein recruitment [210,211].

For the non-TPR-containing co-chaperones, Cdc37 appears to be dedicated to kinase maturation. Its binding interaction at the NTD of Hsp90, leads to partial inhibition of ATPase activity through steric interference of ATP-lid closure and thus inhibition of NTD dimerization [170,212]. Cryo-electron microscopy, revealed additional interactions with the MiD of Hsp90 and human Hsp90ß-Cdc37-Cdk4 complex [192]. In contrast to the Hop and Cdc37, Aha1 has been heavily implicated as a strong activator of Hsp90 ATPase activity by promoting transition towards the closed catalytically active state [190]. Like Cdc37, Aha1 binds both the NTD and MiD in an asymmetric manner, and a single Aha1 molecule is sufficient to stimulate ATPase activity [173,190]. Furthermore, Aha1 competes with

Hop for MiD binding residues and is thus able to drive the Hsp90's conformational cycle by displacing Hop [213]. The p23 co-chaperone is involved in latter stages of Hsp90's conformational cycle, where it is involved in stabilizing the closed catalytic state by binding to a groove formed by NTD dimerization. By stabilizing this conformational state, p23 is thought to reduce Hsp90 ATPase activity and thus regulates the progression of the reaction cycle which is beneficial for steroid hormone receptors (SHR) [214,215].

1.2.3 The Hsp70-Hsp90 chaperone cycle

A progressive chaperone cycle for Hsp90 was established mostly through studies using SHRs as clients [140,216]. In this model, the un-liganded hormone receptor interacts with Hsp70 in complex with the J-domain chaperone in an ATP dependent step [217]. Facilitation of client transfer by Hop drives the formation of an Hsp90-Hsp70-client complex. Binding of p23 and immunophilins displace Hop and Hsp70 to form the mature complex. ATPase activity is regulated by p23 allowing for SHR client activation and competent hormone binding, and eventual hydrolysis leads to client release. In the absence of available hormone, the mature complex decays with a reported half-life of about 5 min [218] in an ATPase dependent manner [219] to release the client, which is then able to reenter the cycle by once again binding Hsp90.

1.2.4 Therapeutic targeting of Hsp90 for the treatment of cancer

The interaction of Hsp90 with several disease related peptides implicate the chaperone with the progression and development of several associated pathologies such as protein folding disorders, cancer, and neurological disease [159]. In recent years it has become increasingly clear that deregulation of Hsp90 may present an attractive treatment strategy for these diseases, elevating interest in human Hsp90 as a viable drug target particularly for the treatment of cancer [220–222].

Cancer and its progression is traditionally referred to as a genetically acquired disease that transitions from a benign to a malignant state through the accumulation of genomic mutations. This genetic instability enables cancer cells to acquire the eight hallmarks of cancer as laid out by Hanahan and Weinberg [223]: (i) self-sufficiency in growth signals; (ii) resistance to anti-growth signals; (iii) evasion of apoptosis; (iv) prolonged angiogenesis; (v) tissue invasion and metastasis; (vi) reprogrammable metabolism; and (viii) the ability to evade immune responses. These eight hallmarks reflect genetic alterations in several protective genes responsible for the coordination and regulation of diverse processes such as proliferation, differentiation, growth, and motility. This genetic plasticity demonstrated by cancer cells, enable them to escape many traditional therapeutics that invariably target single signalling nodes or pathways [224]. To date numerous anti-cancer chemotherapeutics have been successfully developed to target proteins implicated with multiple cancer hallmarks, however none have been able to simultaneously affect all eight hallmarks. Given that several of Hsp90's clients are known *bona fide* oncogenic proteins that are linked to all eight hallmarks, such as the HER-2/ErbB2, EGDR, AR, AKt, Raf-1, and BCR-ABL kinases, as well as the HIF-1α and p53 transcription factors [225], implicates Hsp90 in the progression of several cancers. Furthermore, it is well known that cancer cells exploit Hsp90 to facilitate the acquisition and maintenance of their malignant phenotype and often rely on Hsp90's homeostasis functions to overcome hostile conditions brought on due to their relative microenvironments such as acidosis, nutrient depletion, and hypoxia. It is not surprising therefore to note that Hsp90 is upregulated in tumour cells related to endometrial and ovarian carcinomas [226,227] as well as breast [228–230], lung [231], gastrointestinal [232], and prostate [233] cancers.

At a molecular level, Hsp90 functions as an essential buffer for numerous genetic lesions associated with tumour development and thus enable mutant proteins to gain or retain their cellular functions, while simultaneously allowing the cancer cells to tolerate imbalanced signalling of these oncogenic clients and ultimately avoid apoptotic death [220]. It is thus clear that as molecular chaperone Hsp90 is uniquely positioned as a key player for the growth and survival of cancer cells, and its inhibition would ultimately lead to the degradation of the underlying oncogenic clients via the ubiquitin-proteosome pathway. Thus the therapeutic targeting of Hsp90 presents as viable treatment option for cancer therapy in which a possible outcome could be the simultaneous disruption of all the cancer hallmarks upon which cancer cells depend for their survival and growth [234,235].

Details on the current status of Hsp90 inhibition is covered in Chapter 4 of this thesis in addition to the emergence of allosteric drug targeting as a viable discovery approach for the treatment of cancer.

1.3 Aims

The primary focus of this thesis was to contribute to the growing body of evidence for the allosteric modulation of functional conformational dynamics in the 70 kDa and 90 kDa heat shock proteins using dynamics based computational techniques. Enzymatic activity in both molecular chaperones is underpinned by large conformational restructuring that involves complex rotational and translational domain motions. Bound nucleotides and intricately choreographed co-factor interactions are thought to play a crucial role in the allosteric modulation of these conformational rearrangements and thus the functional activity of the molecular chaperones. The overriding aim of this thesis was thus to demonstrate how PRS coupled with MD simulations and DRN analysis can be used as a viable strategy to efficiently and accurately screen proteins for allosteric sites capable of modulating the functional dynamics of the protein with allosteric drug discovery applications in mind.

To achieve this the following objectives were detailed:

i. The technique of PRS coupled with all-atom MD simulations should be assessed using the model system of bacterial Hsp70 DnaK to investigate the regulatory effect of bound nucleotide and/or peptide substrate on the allosteric potential of the chaperone to interconvert between the catalytically active and inactive states. The results should be cross-validated with experimental findings (see Chapter 2).

ii. Homology modelling techniques should be employed to calculate accurate structures of the cytosolic isoform of human Hsp90 in fully-closed and partially-open conformational states to address the apparent lack of experimental structural data.

iii. Long-range all-atom MD simulations and trajectory analysis techniques should be employed to assess the relative effect of bound nucleotide on the conformational dynamics of the chaperone.

iv. PRS coupled with DRN analysis should be used to probe long-range MD simulations of homology models of human Hsp90α in fully-closed (active) and fully-open (inactive) conformational states to investigate and quantify allosteric modulation of conformational dynamics in the chaperone. Allosteric sites capable of selecting coordinate shifts towards either the active of inactive states should be investigated and correlated with available experimental and computational works (ii-iv see Chapter 3).

v. High throughput molecular docking techniques should be used to screen a small compound library to assess the druggability of identified allosteric target sites in human Hsp90α.

vi. Putative allosteric hit compounds with affinity for the proposed allosteric target sites should be used as probes to assess the allosteric potential of the identified target sites to modulate the conformational dynamics of the chaperone and thus proof-of-principle" for the proposed PRS based allosteric screening strategy. Long-range all-atom MD simulations should be conducted for all hit compounds together with a suitable control and an advanced trajectory based analysis techniques used to assess the effect of ligand interactions on the conformational dynamics of the protein (v-vi see Chapter 4).

Chapter 2:

Perturbation response scanning reveals key residues for allosteric control in Hsp70

This chapter describes the Perturbation response scanning technique and assesses its suitability to predict the allosteric potential of individual residues to select functionally relevant conformational changes

Chapter 2 is reproduced in part with permission from the following publications:

Perturbation–Response Scanning Reveals Key Residues for Allosteric Control in Hsp70. David L. Penkler, Özge Sensoy, Canan Atilgan, and Özlem Tastan Bishop. *Journal of Chemical Information and Modeling.* **2017**. 57 (6), 1359-1374 DOI: 10.1021/acs.jcim.6b00775. 2017 American Chemical Society

Contribution: Designed the experimental plan, implemented PRS in the Python programming environment, carried out all experimental procedures, analysed the data, and wrote the first draft of the manuscript.

MD-TASK: a software suite for analyzing molecular dynamics trajectories David K. Brown, David L. Penkler, Olivier Sheik Amamuddy, Caroline Ross, Ali Rana Atilgan, Canan Atilgan, and Özlem Tastan Bishop. *Bioinformatics*. **2017**. 17, 2768-2771. DOI: 10.1093/bioinformatics/btx349. 2017 Oxford University Press.

Contribution: Wrote the Perturbation response scanning script and implemented it as a Python module in the MD-TASK suite. Assisted with test driven development. Wrote the PRS section of the manuscript.

2.1 Introduction

Hsp70 consists of two domains, a nucleotide binding domain (NBD) and a substrate binding domain (SBD). From a functional stand point Hsp70 is required to bind and release client substrates in a tightly orchestrated and controlled manner. This binding and release of substrate requires major global conformational rearrangements, in which the SBD must first be suitably exposed for client binding, and must then be able to react and entrap bound substrate by assuming a conformation that stabilizes the complex. Finally, the chaperone is required to release its bound client at the right moment for downstream processing. The mechanism for client capture and release is still a topic of debate within the field, however the current understanding is that the protein must interconvert between two well described experimentally determined functional states: i) an ATP-bound SBD open conformation, and ii) an ADP-bound SBD closed conformation (Figure 1.5). The exact mechanism for inter-state conversion is not as yet fully understood, however, it is evident that the nucleotide/peptide bound state of the NBD and SBD as well as interactions with co-chaperones such as DnaJ and nucleotide exchange factor (NEF) play crucial regulatory roles. Indeed, there is a growing body of evidence that suggests fine-tuned allosteric inter-domain communication may be the underlying mechanism for inter-state conversion, in which binding events at the NBD [90,91,93,94,101,102,104,236–239] and SBD [240–243] affect the global conformational dynamics in favour of either an open or closed end-point state.

2.1.1 Computational efforts to elucidate Hsp70’s allosteric mechanism of action

To date, aspects of Hsp70’s complex allosteric control mechanism have been thoroughly investigated through the efforts of several research groups (see ref. [129] for a detailed overview), and while questions still remain, it is widely thought that a core set of residues are important for substrate-triggered ATPase activity and ATP-stimulated peptide release. This hypothesis however is difficult to rigorously test using *in vitro* techniques, and a number of research groups have turned to computational approaches for further insight.

Coarse-grained molecular dynamics studies have demonstrated the collective motions involved in allosteric communication between the SBD and NBD [244,245], while all-atom simulations in conjunction with several computational analysis techniques have been utilised to better understand the conformational dynamics of the two end-point states [238,239,246–248]. Given the computational and hardware limitations of MD simulations, the complete transition from the open to the closed state and *vice versa* cannot be observed in all-atom studies, with complete protein conformational transitions typically occurring on the micro-millisecond time scale [23]. Despite this, several studies have demonstrated the elegant use of alternative computational techniques to describe key residue networks essential for inter-domain allostery in Hsp70.

Smock and co-workers defined a structurally contiguous group of residues that form a physical linkage network between NBD and SBD of which the linker is centrally positioned [249]. General *et al.* (2014) combined computational techniques together with experimental validation to report several highly conserved residues in subdomain IA to be key allosteric communicators between the NBD and SBD [250]. Nicolaï *et* al. demonstrated the use of free-energy landscapes to analyse the structural changes induced by ATP binding, reporting 27 internal coordinates that correspond to a 91 residue network thought to be relevant for allosteric communication [239]. Molecular simulations combined with protein stability analysis and network modelling of residue interactions, suggested allostery to be primarily determined by nucleotide induced conformational redistributions in the NBD and SBD, and that several mediating residues with high network centrality are crucial for conformational stability and allosteric communication [251].

In an alternative approach, allosteric inhibition of DnaK was recently investigated by Stetz *et al.* (2016), in which molecular simulations and binding free energy analysis were coupled with extensive network-based modelling of residue interactions, to compare and characterise the molecular signatures of the apo, substrate bound, and allosteric inhibitor bound closed complex conformations. This study postulated a mechanism by which the allosteric inhibitor PET-16 can stabilise the ADP bound closed conformation and inhibit inter-domain allosteric control [252].

In this study, we demonstrate the use of extensive all-atom molecular dynamics simulations of *E.* coli Hsp70/DnaK, coupled with PRS analysis, to investigate the relative effect different combinations of bound ligand at the NBD (ATP/ADP/apo) and SBD (peptide/apo) have on inter-domain allostery, reporting the allosteric potential of each configuration as well as each residues contribution for allosteric conformational modulation.

2.1.2 Theory of Perturbation response scanning

It is well known that protein function is driven by tertiary structure and the ability to assume a functional state in response to environmental factors such as pH, ionic concentration, chemical modification of specific residues such as mutation or post translation modifications, or through interactions with binding partners [253,254]. PRS is a computational technique that informs on global conformational change in response to artificial external force perturbations.

The underlying theory of PRS has been well described in previous studies [255,256]. Briefly, the technique utilizes linear response theory (LRT) to approximate the coordinate shift ($\Delta\boldsymbol{R}_1$) of a given protein conformation ($\boldsymbol{R}_0$) in response to an external force perturbation of the Hamiltonian [255,257,258]:

$$\Delta \boldsymbol{R}_1 = \langle \boldsymbol{R} \rangle_1 - \langle \boldsymbol{R} \rangle_0 \cong \frac{1}{k_B T} \langle \Delta \mathbf{R} \Delta \mathbf{R}^{\mathrm{T}} \rangle_0 \Delta \boldsymbol{F} = \frac{1}{k_B T} \mathbf{C} \Delta \mathbf{F}$$

Equation 2.1

In this equation, the subscripts 0 and 1 denote the unperturbed and perturbed protein structures. The kernel $\mathbf{C}$, is the variance-covariance matrix which can be obtained by either imposing the approximation of harmonic springs between pairs of interacting residue atoms, or directly from the coordinate data of MD trajectories of suitable length [259]. The force vector $\Delta\mathbf{F}$ denotes the coordinate shift corresponding to the externally inserted force on residue *i*, $(\Delta\mathbf{F})^{\mathrm{T}} = \{000 \ldots \Delta\mathbf{F}_x^i \Delta\mathbf{F}_y^i \Delta\mathbf{F}_z^i \ldots 000\}_{1\times 3\mathrm{N}}$.

The PRS technique involves the repetition of Equation 2.1, to scan each residue in the protein one-by-one with a focus on force perturbations that cause a coordinate overlap with an expected conformational change:

$$\Delta \mathbf{R}_1 = \langle \mathbf{R} \rangle_1 - \langle \mathbf{R} \rangle_0$$

Equation 2.2

PRS thus requires two structurally distinct conformations as input, which together describe an expected conformational transition *i.e. open* (initial state) to *closed* (final state). Using these structures, an expected (experimental) conformational change ($\Delta\mathbf{S}$) can be determined by performing a structural alignment and recording the shift in atomic coordinates between the initial and final states. Each residue in the initial state (state 0) is sequentially perturbed according to Equation 2.1 and the linear response of all other residues recorded $\Delta\mathbf{R}$. Predicted coordinate displacements are then compared to the experimental/expected displacements ($\Delta\boldsymbol{S}$) and the overlap assessed to identify residue perturbations that correspond to the expected conformational change.

The PRS methodology has been successfully utilized in several previous studies. Analysis of calmodulin [260] and ferric binding protein [255] revealed a subset of residues implicated in conformational modulation of either protein. These results were later confirmed when *in silico* mutation and MD simulation analysis revealed disrupted conformational dynamics [261,262]. In another study PRS was able to highlight residues involved in the catalytic mechanism and stability of subtilisin in complex with an inhibitor [263]. More recently, PRS was also used to predict residue involved with conformational transitions between the *apo* and *holo* forms of human transferrin [264].

2.1.3 Chapter objectives

In this study, we apply Perturbation response scanning (PRS) in combination with molecular dynamics (MD) simulations as a computational method to assess the global and residue specific allosteric potential of a large multi-domain protein such as Hsp70 that experiences complex conformational rearrangements that are both translational and rotational. Given the regulatory potential Hsp70's binding partners impart, we propose PRS to be an advantageous and highly suitable approach for identifying allosteric sites due to the force perturbation nature of the technique which can be seen to mimic naturally occurring binding interaction forces. In this manner, primary objectives of this chapter are three-fold: i) to assess the relative allosteric effect of bound nucleotide and/peptide substrate on the conformational dynamics of the protein; ii) to identify site specific allosteric hot residues capable of selecting conformational displacements towards known functional states; and iii) to assess the accuracy of PRS analysis for use on large multi-domain proteins by cross-validating allosteric hot spots with experimentally determined function.

2.2 Methodology

2.2.1 Preparation of ligand-bound conformations

A total of 12 unique configurations of DnaK were prepared, six open and six closed conformations in various combinations of bound/unbound ADP/ATP or peptide substrate. Experimental structures of DnaK in open and closed conformation were obtained from the Protein Data Bank; PDB codes 4BQ9 [101] and 2KHO [91], respectively. PDB 4B9Q (residues 2-602) represents the open conformation of the protein, where the SBD is intimately docked to the NBD and SBDα is undocked from the SBDß exposing the substrate binding site. PBD 2KHO (residues 4-603) is representative of the ADP bound closed conformation, in which each domain is independently positioned on either side of the inter-domain linker, and the SBDα is bound to SBDß.

Neither 4B9Q nor 2KHO contained a peptide substrate and as such the NBD co-crystallized with peptide NRLLLTG (PDB ID 1DKZ) was superimposed over the NBD of 4B9Q and 2KHO respectively and the coordinate data for the peptide incorporated into these structures. In the same manner, ADP coordinates were obtained from PDB ID 3ATV, the NBD of human Hsp70 crystallized with ADP. ATP coordinates were obtained from 4B9Q structure.

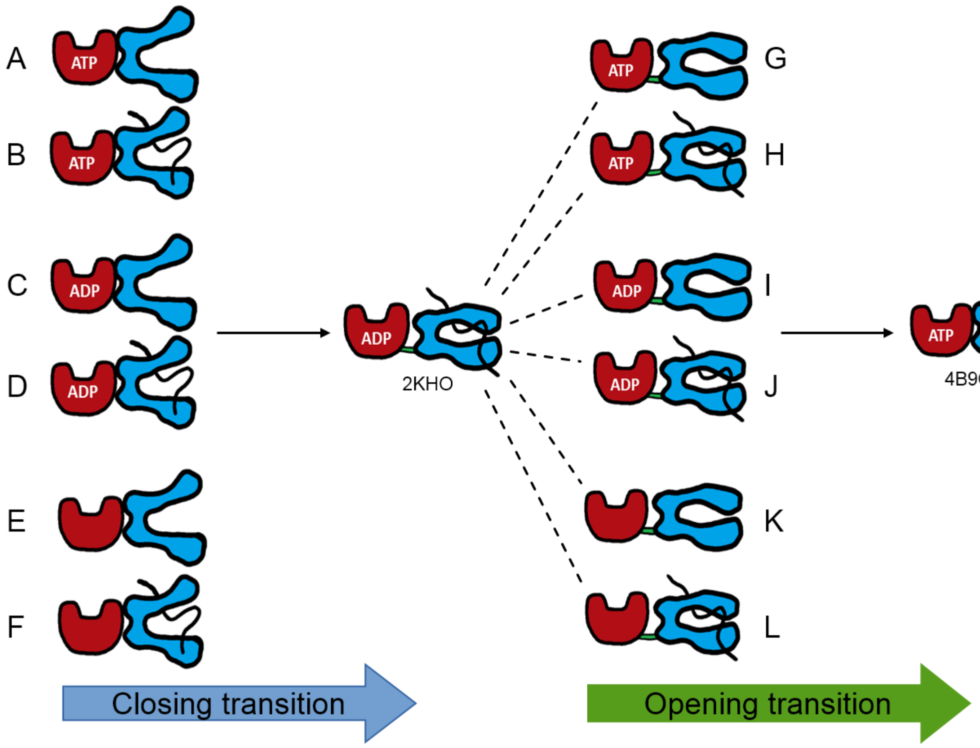


Figure 2.1: Schematic overview of PRS experiments. Showing all 12 unique configurations of DnaK used to investigate the allosteric potential for interstate conversion. The open to closed transition experiments are: A-F while the closed to open transition experiments are: G-L. Reproduced with permission from Penkler et. al 2017 [105]

2.2.2 Molecular dynamics simulations

All MD simulations were performed with the GROMACS 5.1.2 [265–267] employing the CHARMM 22 force-field (with CMAP corrections) [268–270]. Systems were subjected to energy minimization using a conjugate gradient algorithm prior to MD runs. All systems were energy relaxed with 1000 steps of steepest-descent energy minimization. MD simulation on the minimized systems were then carried out in the NVT ensemble at 310 K using the Berendsen temperature coupling algorithm. The final structures were then run in the NPT ensemble at 1 atm and 310 K until volumetric fluctuations stabilized and the desired average pressure maintained, allowing for adequate data points to be collected for further analysis. MD simulations were run for a minimum of 100 ns and maximum of 200 ns until the backbone RMSD of the protein equilibrated with a fluctuation of no more than 3 Å for at least 20 ns. All coordinates were saved at 2 ps intervals for analysis. All MD simulations were carried out using the LINCS algorithm to constrain bond lengths, while the Long-range electrostatic forces were handled using the Fast Particle-Mesh Ewald method (PME), and Van der Waals forces treated with a 0.9 nm cut-off. Periodic boundary conditions were applied in all directions and the simulation time-step was set to 2 fs. Water molecules were modelled using the TIP3P model. NH3+ and COO- groups were used to achieve the zwitterionic form of the protein whereas the N- and the C- termini of the peptide substrate were capped using methyl-carbonyl and methyl-amine, respectively, to mimic as if the peptide were part of a protein.

2.2.3 Perturbation response scanning

The PRS method relies on the assumption that the kernel used in connecting the inserted forces to the displacements originates from a single energy minimum that may be approximated by harmonic springs. While network models ensure this is the case, they are limited in that (i) they require a known PDB structure for network construction, and (ii) the selection of the necessary spacial cut-off distance is ambiguous. MD trajectory data provides an alternative source for the construction of the variance-covariance matrix, provided short enough trajectories are used to construct the Hessian to ensure the matrix represents information obtained from a single potential well [259,271]. Previous studies have explored optimal MD simulation lengths that fit this requirement and using relaxation times of the backbone atom fluctuations, have shown that 20 – 40 ns MD trajectories will carry this information [259,271]. Using MD simulations for the construction of $\mathbf{C}$ is advantageous in that it frees one from the limitation of requiring alternative PDB structures, as one may generate alternative conformations through MD simulation, provided that equilibrium or quasi-equilibrium has been reached for the portion of the MD trajectory that will be used to construct the variance-covariance matrix.

- *Constructing the covariance matrix from MD trajectory data*

Given the extensive MD simulations for each of the 12 Hsp70 configurations, we employ the latter MD based approach for constructing the covariance matrix. First the backbone RMSD was calculated for each MD trajectory and an in-house Python script used to locate the first available 20 ns equilibrated segment defined as a minimum RMSD fluctuation maintained around 3.0 Å, scanning the trajectory in reverse order from the final frame to the initial frame. The equilibrated portion of the trajectory was then extracted and coarse-grained by tracing the coordinates of the C_α atom for each residue in the protein (C_β for glycine). This coarse-grained trajectory was then used to calculate the deviation of residue *j* from the average structure obtained over a time window *w* of observations, which were discretely counted by *w* steps to yield the difference matrix:

$$\Delta\mathbf{R}_j(\boldsymbol{t}) = \mathbf{R}_j(\boldsymbol{t}) - \langle\mathbf{R}_j(\boldsymbol{t})\rangle$$

Equation 2.3

where $\langle\mathbf{R}_j(\boldsymbol{t})\rangle$ is the average structure over the MD trajectory and $\Delta\mathbf{R}$ is the $3N$ x *w* difference matrix of the deviation of all atomic coordinates $\mathbf{R}_j(t)$ from the average structure over a time window *w*. The programmatic method used for determining the average structure and subsequent difference matrix is summarised in Figure 2.2.

Finally, the covariance matrix ***C*** was calculated by:

$$\boldsymbol{C} = \Delta\mathbf{R}\Delta\mathbf{R}^{\boldsymbol{T}}$$

Equation 2.4

In this manner a total of 12 separate ***C*** matrices were constructed, each pertaining to one of the 12 conformational states (Figure 2.1).

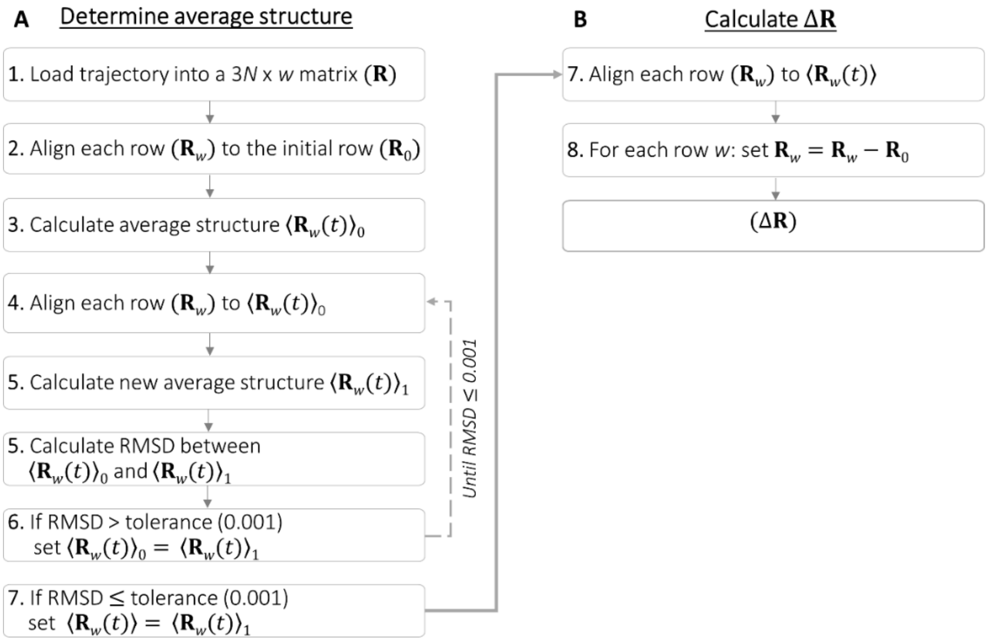


Figure 2.2: Pseudo-algorithm describing the programmatic scheme for calculating the difference matrix. (A) Method used for iteratively determining the average structure according to an RMSD tolerance value of 0.001. (B) Method for calculating $\Delta\mathbf{R}$

- *Defining the perturbations force vectors*

It is important to note that in PRS there is no *a priori* assumption on the direction or magnitude of the perturbing force. Rather the direction of the force vectors are chosen at random to ensure that no bias is attributed to the specific contact topology or solvent exposed nature of the particular residue being perturbed. The random force vectors are thus uniformly distributed within a sphere enveloping the residue of interest in an isotropic manner by applying a total 250 force perturbations. Since the response vectors are calculated by LRT, the resultant displacements will be directly proportional to the imposed force. Since we are only interested in the relative conformational displacements, a force of unit magnitude is applied throughout.

– *Defining the expected experimental conformational change*

For each PRS experiment, the coordinates of the initial state (states 0 in Equation 2.1) were set based on initial frame of the respective 20 ns equilibrated MD trajectory segments used to prepare the covariance matrix. The final end-point state in each experiment was set to the known experimental structures of the closed ADP bound state (PDB ID: 2KHO) for the closing transition experiments (Figure 2.1 A-F) and the open ATP bound state (PDB ID: 4B9Q) for the opening transition experiments (Figure 2.1 G-L). The experimental difference () is calculated by superimposing the final state on the initial state using the Kabsch algorithm [272], and computing the targeted (experimental) residue displacement vectors by subtracting the initial state coordinates from the final state coordinates.

– *Assessing the quality of the response vector*

Each force perturbation yields a corresponding $\Delta \boldsymbol{R}$ response vector that describes the coordinate shift of all other C_α atoms in the protein, thus yielding a total 250 responses. To determine which force perturbation yields the best corresponding overlap with the expected conformational change ($\Delta \boldsymbol{S}$), the quality if each $\Delta \boldsymbol{R}$ response is assessed and the force with the best coordinate fit chosen. Previous work has demonstrated the use of direct overlaps to compare the coordinate shifts between $\Delta \boldsymbol{R}$ and $\Delta \boldsymbol{S}$ [256]. However, given the large scale transitions expected for Hsp70, the overlap, which is calculated as the dot product:

$$(\Delta \boldsymbol{R}_1 \cdot \Delta \boldsymbol{S})/(|\Delta \boldsymbol{R}_1||\Delta \boldsymbol{S}|)$$

Equation 2.5

is not expected to be close to 1 because the direction of the initial motion may be entirely different from the overall displacements that characterise the end points of the conformational change. This problem has been previously discussed and alternative approaches offered [273,274]. In this study, we rather utilize the Pearson’s correlation coefficient between $\Delta \boldsymbol{R}$ and $\Delta \boldsymbol{S}$ as a suitable comparative metric and calculate C_i by correlating the response (ΔR_k) with the experimental displacements averaged over all affected residues *k*:

$$C_i = \frac{\sum_{k=1}^{N}[(\Delta R_k)^i - (\overline{\Delta \boldsymbol{R}})^{\boldsymbol{i}}](\Delta S_k - \overline{\Delta \boldsymbol{S}})}{(\boldsymbol{N}-\boldsymbol{1})\boldsymbol{\sigma_R \sigma_S}}$$

Equation 2.6

The over-bar represents the average, ΔS_k are the experimental displacements, and σ_S and σ_R are the corresponding root mean squared values. Thus ΔR_k is compared with ΔS_k and the goodness of fit assessed with the Pearson correlation coefficient to calculate C_i for each residue i for each isotropic force perturbation. For each residue, only the maximum C_i is retained – where a C_i value close to 1 implies good agreement with the experimental changes, and values close to zero, no agreement or lack of correlation with the experimental findings. A schematic representation of high versus low correlation coefficients is shown in Figure 2.3, where high correlations for the perturbed residue V440 of the open NBD$_{(ATP)}$-BD$_{(apo)}$ complex selects conformational displacements in favour of the expected closing transition. Conversely perturbation of a residue in the SBDα yields a low correlation coefficient value and engenders SBD movement towards the NBD in a stabilised open conformation.

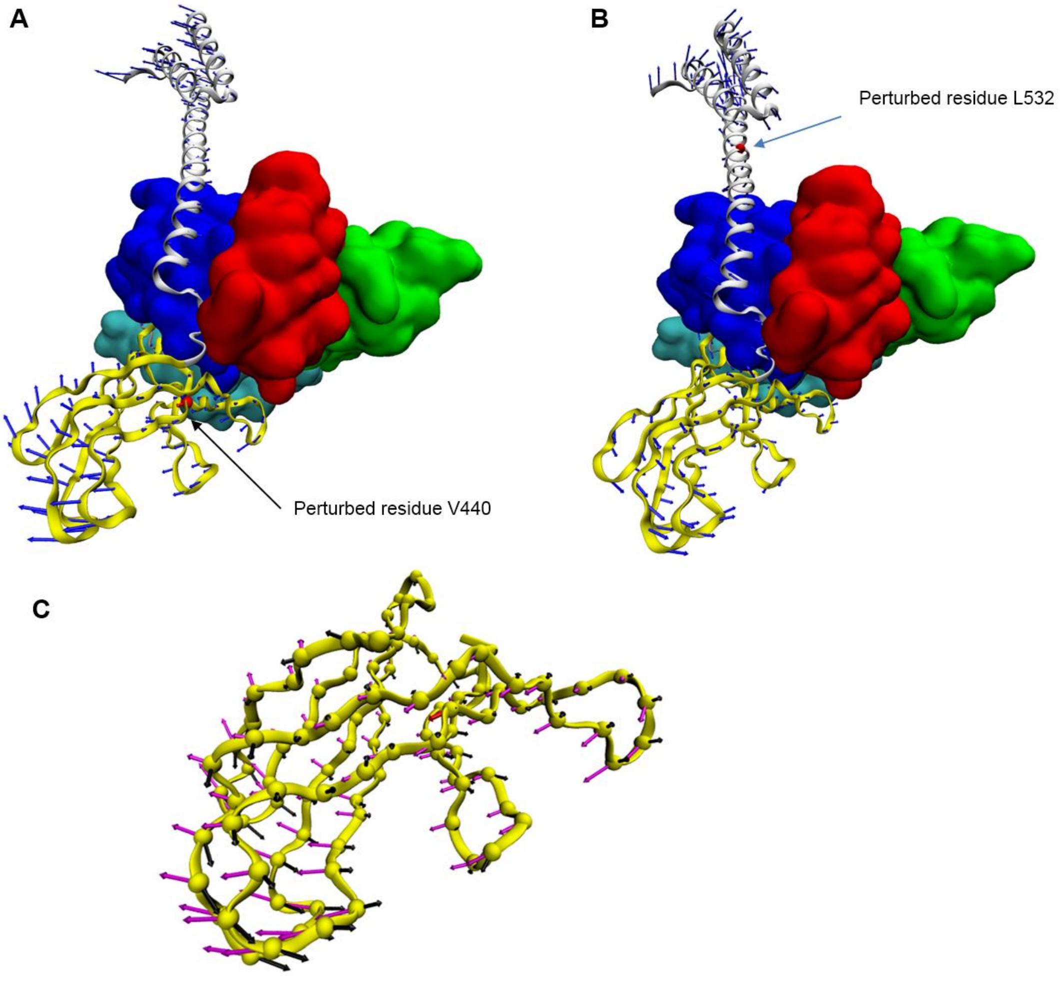


Figure 2.3: Schematic representation of high vs low C_i for the closing transition. Showing the NBD$_{(ATP)}$-BD$_{(apo)}$ complex, **(A)** Perturbation of the functionally important trigger residue V440 (C_i = 0.83) in the SBDß (yellow) results in residue displacements that when mapped as vectors, show that the direction of the SBDß and SBDα are mostly in opposite directions towards each other, in a motion expected for the closing transition. **(B)** Perturbation of the SBDα residue L532 (C_i =0.49), results in residue displacements back towards the NBD (surface representation). **(C)** Zoomed in representation of the SBDß showing in clearer detail the opposing displacement vectors from the two scenarios: A (magenta) and B (black). Reproduced with permission from Penkler et. al 2017 [105]

The residue-based displacements analysed in PRS are critically dependent on the alignment between the two conformational structures. Given Hsp70's large conformational changes that involve both rotational and translational transitions, it is important to note that a whole protein alignment approach was followed in this study, using the Kabsch algorithm [272]. The rationale behind this choice of alignment strategy is best explained using schematics (Figure 2.4). Hsp70 has two domains that are essentially joined by a hinge (the linker) (A I - blue). Let us consider it to have two possible conformations as a result of hinge motions (A II - red). By aligning the whole system of conformation II on conformation I (B), the required conformational change from I → II can be described by the green arrows, and the best single residue perturbation to select the required conformational change demonstrated by the black arrow. Similarly after superimposing the predicted structure (C - blue) onto conformation II (C - red), the force perturbation response vectors can be described by the green arrows (C). In PRS we are correlating the green arrow vectors sets from B and C. By restricting the alignment procedure to just a single domain as demonstrate in D, there exists only one perturbation position to obtain the conformation transition, giving rise to a single response vector (E - green arrow). If we were to analyse the displacements in C and E using overlaps, C would be much higher due to the greater average angular change (magnitude is lost but directionality is retained). By analysing C and E using Pearson correlations, C would again be higher because the magnitude of the three points of motion are better predicted than by the single large point in E.

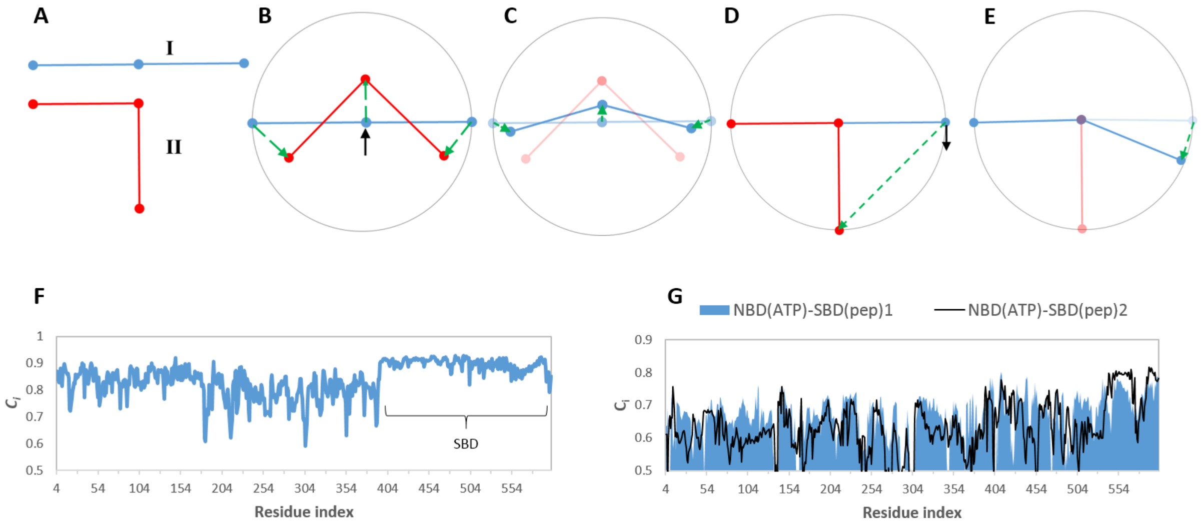


Figure 2.4: Schematic describing the rationale behind a whole protein alignment strategy. (A-E) Illustration of force and displacement vectors (See text for details). **(F)** Example Pearson's correlation coefficients after of domain specific alignments. **(G)** Graphical demonstration of the reproducibility of the whole alignment approach for replicate simulations of the ATP and peptide bound open configurations. Reproduced with permission from Penkler et. al 2017 [105]

We tested this theory by guiding the alignment procedure for the closing transition of $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration, superimposing only the NBD of the two end point structures.

Following this approach, the C_i values become drastically elevated with a distinct bias for the SBD, indicating that the conformational change is linear and that biasing the alignment to one domain may in fact be destroying this linearity by removing the rotational scale (Figure 2.4-F).

Finally the reproducibility of the aforementioned PRS analysis approach was tested for replicate independent MD trajectories for the closed $NBD_{(ATP)}$-$SBD_{(pep)}$ complex (Figure 2.4-G) , in which we note similar C_i profiles that when compared by Pearson's correlation yield a correlation coefficient of 0.81.

2.2.4 Inter-subdomain contact analysis

Inter-subdomain contact residues were determined using the Protein-Interaction Calculator (PIC) [275]. The equilibrated initial structure used for PRS analysis for the open and closed $NBD_{(ATP)}$-$SBD_{(pep)}$ configurations were submitted to the PIC web-server using the default parameters. In house Python scripts were used to parse the interaction data to filter out all relative inter-subdomain interactions.

2.3 Results and Discussion

2.3.1 MD simulations reveal configuration specific inter-domain rearrangements

A total of 12 Hsp70 configurations, six in the closed conformation (NBD-SBD undocked) and six in the open conformation (NBD-SBD docked), were prepared using the experimental NMR and X-ray crystal structures 2KHO [91] and 4B9Q [101] respectively (Figure 2.1). Each conformation having a unique combination of bound nucleotide and/or peptide substrate (see methodology Section 2.2.1). We refer to each configuration with the following nomenclature $NBD_{(ATP/ADP/apo)}$-$SBD_{(pep/apo)}$ to describe the un/bound state of the NBD and SBD respectively. Each configuration was submitted to MD simulation for a minimum of 100 ns and a maximum of 200 ns, to ensure a 20 ns equilibrated trajectory segment could be obtained in requirement for the construction of the covariance matrix (see methodology Section 2.2.3). The initial frame for each equilibrated segment was used as either the initial (perturbed) or final (expected) state structure in the respective PRS calculations and it is thus important to consider the structural arrangement of each configuration post long range all-atom MD simulation, to investigate the effect if any of un/bound nucleotide and/or substrate on the global conformation and inter-domain rearrangement of each configuration. Table 2.1 summarizes each of these simulations in terms of simulation length, equilibrated region and minimum RMSD fluctuation of the equilibrated region.

Table 2.1 | Summary of MD trajectory segments for PRS analysis

	Configuration		Replicate	Simulation length (ns)	Equilibrated region (ns)	ΔRMSD (Å)
	Nucleotide	Substrate				
Closed	ATP	✕	1	100	80-100	2.0
	ATP	✓	1	200	180-200	2.0
	ATP	✓	2	150	130-150	2.0
	ADP	✕	1	100	80-100	2.0
	ADP	✓	1	100	61-81	2.0
	ADP	✓	2	100	61-81	2.0
	apo	✕	1	150	120-140	2.3
	apo	✓	1	150	68-88	3.0
Open	ATP	✕	1	100	56-76	2.3
	ATP	✓	1	200	116-136	2.3
	ADP	✕	1	100	53-73	2.3
	ADP	✓	1	100	80-100	2.0
	apo	✕	1	100	62-82	2.3
	apo	✓	1	100	66-86	2.3

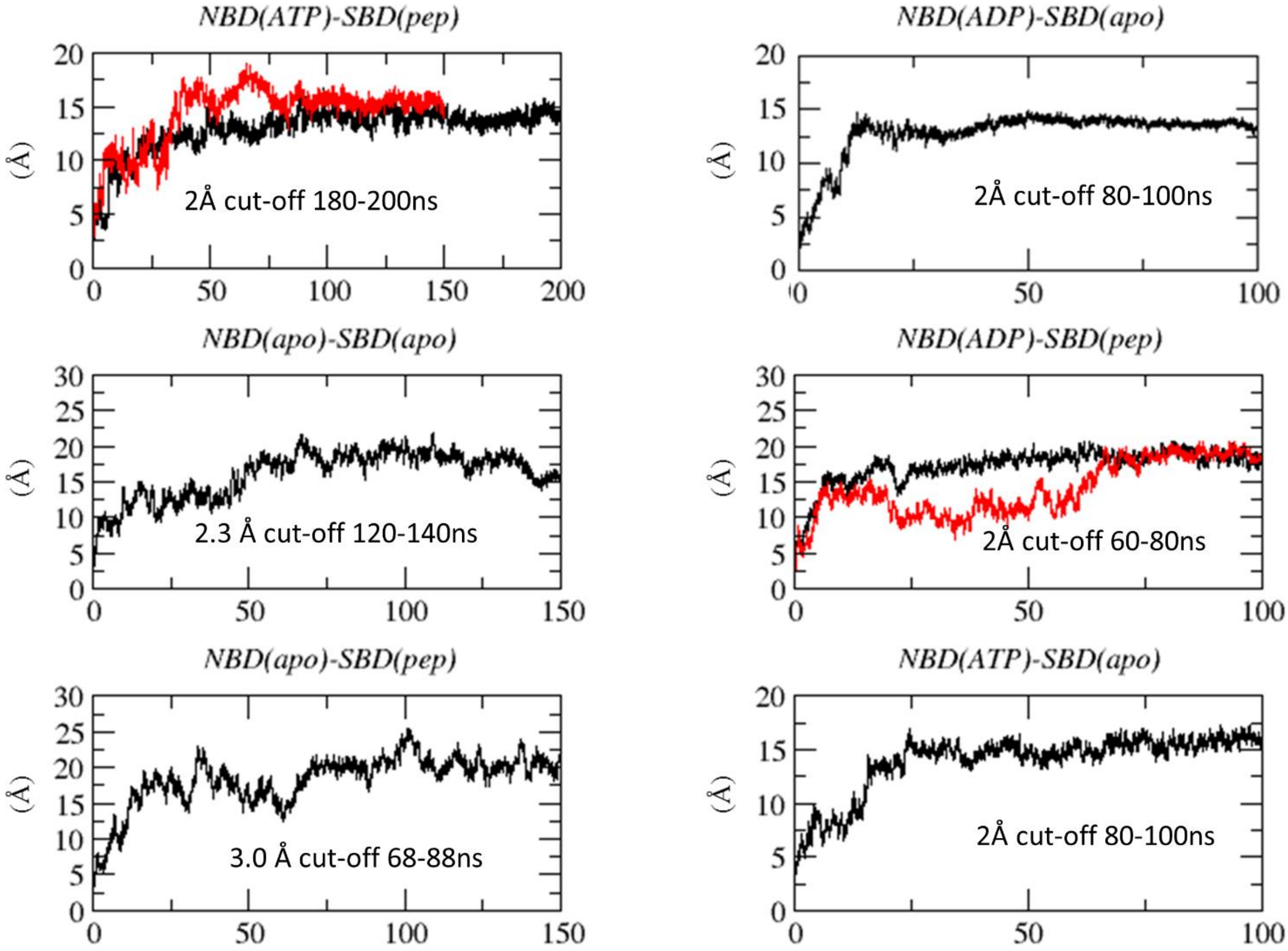


Figure 2.5: Backbone RMSD plots for the closed conformation complexes. Replicate MD simulations are coloured red. Each plot is annotated with the minimum RMSD cut-off and time points for the 20 ns equilibrated regions. Reproduced with permission from Penkler et. al 2017 [105]

– *Conformational dynamics of the closed conformation complexes*

The closed complex configurations were submitted for a minimum 100 ns MD simulation length and looking at the RMSD plots in Figure 2.5 it is evident that variation in the backbone flexibility of all six complexes stabilized towards the end of the simulation run. However it is evident that the degree of stabilization for this metric differed with respect to the un/bound configuration of the NBD and SBD respectively. In each case a stable 20 ns equilibrated segment was located at a maximum RMSD fluctuation of 2.3 Å. Visual inspection of the equilibrated 3D structures reveal nucleotide/peptide dependent conformational repositioning of the NBD and SBD with the inter-domain linker acting as a hinge between the two domains (Figure 2.6).

First we compare the relative effect of bound ATP and ADP on the conformational dynamics of the closed state complexes in the absence of peptide substrate. Over the course of the MD simulation addition of bound ADP at the NBD, causes the SBD of the $\mathrm{NBD_{(ATP)}}$-$\mathrm{SBD_{(apo)}}$ complex to rotate 180° relative to the NBD, forming stabilising contacts between the SBDβ loops L_{23}, L_{67} and the linker and between the linker and subdomain IA (Figure 2.6). This stabilisation between both functional domains is evident in a tightly equilibrating RMSD plot after 50 ns (Figure 2.5). In comparison, the addition of ATP caused the $\mathrm{NBD_{(ATP)}}$-$\mathrm{SBD_{(apo)}}$ complex to only stabilised after 70 ns, courtesy of interactions between the SBDβ and loop 210 (L_{210}) of subdomain IIA, as well as additional interactions between loops L_{23}, L_{67}, and the inter-domain linker (Figure 2.6). In the absence of bound nucleotide and peptide substrate, the $\mathrm{NBD_{(apo)}}$-$\mathrm{SBD_{(apo)}}$ complex experienced a greater degree of backbone flexibility requiring in increase simulation time of 150 ns to yield a suitable equilibrated trajectory segment between 120-140 ns, with a maximum backbone RMSD fluctuation of 2.3Å (Figure 2.5). Visual inspection of this structure revealed several stabilizing domain-linker contacts particularly with the SBDβ loops L_{23}, L_{67} and the NBD loop Asp20-Thr23, as well as additional contacts between Ser505 in the SBD and loop Gly132-Glu137 in the NBD (Figure 2.6).

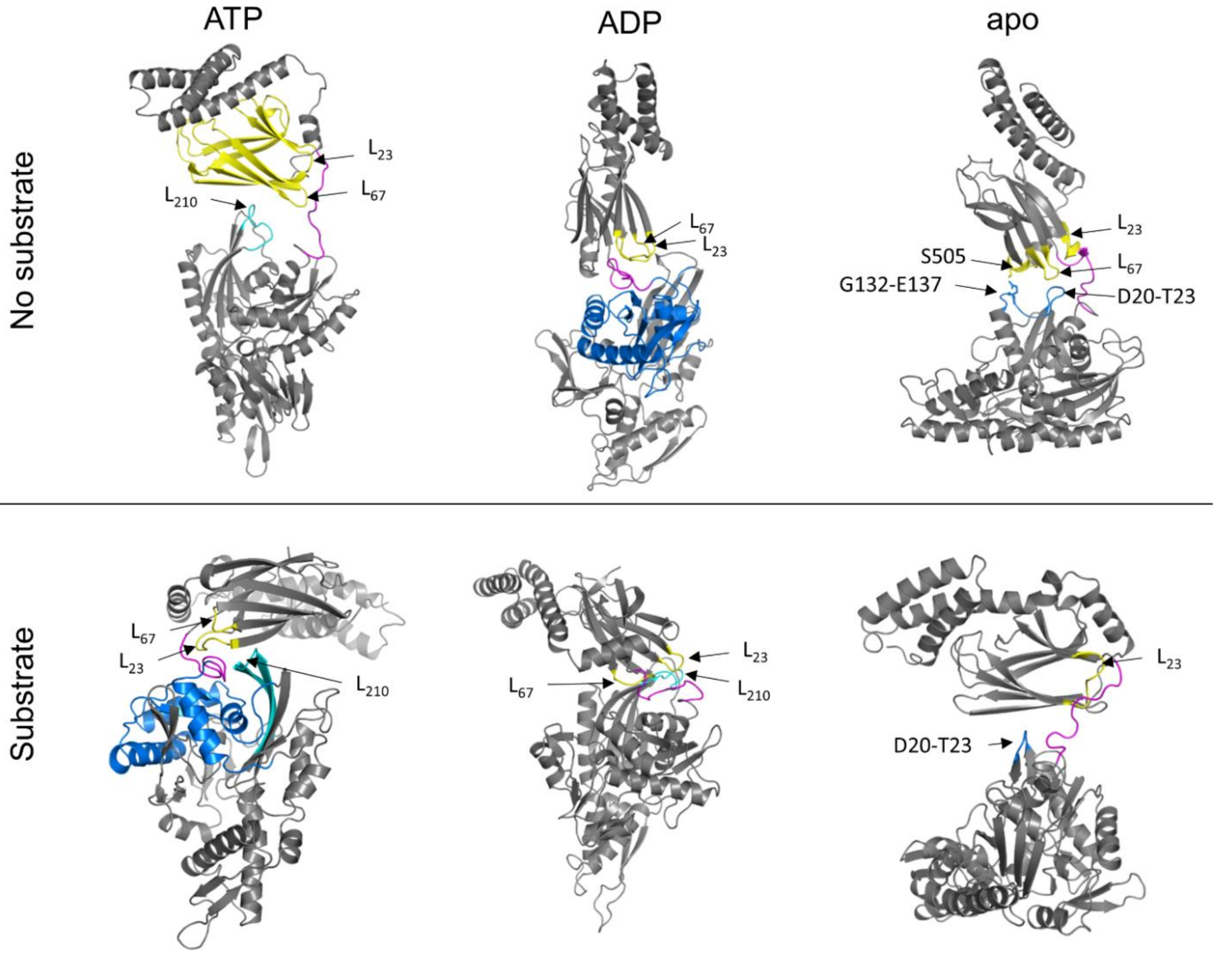


Figure 2.6: Structural arrangements of the six closed conformation configurations after MD simulation. Stabilising interactions between the domains and the inter-domain linker are indicated. Interacting residues are coloured according to the subdomain location: IA – blue, IIA – cyan, SBDß –yellow, linker – magenta. Reproduced with permission from Penkler et. al 2017 [105]

Next we look at the effect bound peptide at the SBD has on the conformational dynamics of the closed conformation complexes. In the ADP bound state the backbone RMSD of the NBD$_{(ADP)}$-SBD$_{(pep)}$ complex equilibrated after 50 ns (Figure 2.5), with stabilising contacts once again forming between the linker and loops from both domains: L$_{23}$, L$_{67}$ in the SBDβ, and L$_{67}$ and L$_{210}$ of subdomain IIA (Figure 2.6). Increased backbone flexibility in the presence of ATP required a longer MD simulation length of 150 ns to achieve an RMSD fluctuation of 2 Å for the NBD$_{(ATP)}$-SBD$_{(pep)}$ complex (Figure 2.5). In this equilibrated structure, the linker appears to interact extensively with several NBD residues that form a cleft between subdomains IA and IIA, as well as additional contacts with loops L$_{23}$ and L$_{67}$ in the SBD. Bound peptide at the SBD in the absence of nucleotide, demonstrated the most backbone flexibility of all six configurations. Despite simulating the NBD$_{(apo)}$-SBD$_{(pep)}$ complex for an extended 150 ns an equilibrated segment with a minimum RMSD fluctuation of 3.0 Å could be obtained between 68-88 ns (Table 2.1, Figure 2.5). Stabilising contacts for the equilibrated structure were observed between the linker and L$_{23}$ in the SBD and loop Asp20-Thr23 in the NBD (Figure 2.6).

Our observations of conformational flexibility and domain rearrangement of NBD and SBD for the closed configurations are largely in line with previous findings. Solution NMR studies have demonstrated that in the presence ADP the SBD and linker collide randomly with the IA and IIA subdomains, but that the IA/IIA binding cleft remains closed due to ADP driven restructuring of the NBD [91]. Here we report a similar behaviour for our MD simulations in which the 20 ns stable segments invariably involve some form of stabilising contact between the linker and subdomain IIA. The only evidence of direct contact between the SBDβ and NBD was observed for the NDB(ATP)-SBD(pep) complex via contacts between L_{23} and subdomain IA – contacts that have been previously suggested to be a direct result of ATP binding [130]. Furthermore, the linker in this complex is positioned within the IA/IIA cleft in an orientation that is in agreement with the experimental crystal structures of the open conformation of DnaK [101,102]. This structural rearrangement is also in agreement with similar MD studies by Chiappori and colleagues [238], in which the authors use extensive trajectory analysis techniques to conclude that the orientation of the linker within the IA/IIA binding cleft is in line with the proposed allosteric intermediate state previously proposed by Zhuravleva and colleagues [104].

Given the significance of the positioning of the inter-domain linker in response ATP binding, replicate simulation runs of the $NBD_{(ADP)}$-$SBD_{(pep)}$ and $NBD_{(ATP)}$-$SBD_{(pep)}$ complexes were produced to validate the findings of an ATP driven intermediate state in which the linker docks to the NBD IA/IIA cleft. The $NBD_{(ADP)}$-$SBD_{(pep)}$ complex once again reached a stable equilibrium within 100 ns (Figure 2.5 - red plot), while the $NBD_{(ATP)}$-$SBD_{(pep)}$ complex stabilised after 150 ns (Figure 2.5 - red plot). Comparing the replicate equilibrated 3D structures from the respective MD runs, the $NBD_{(ATP)}$-$SBD_{(pep)}$ complex assumes slightly different SBD orientations with respect to the NBDs, but notably the linker orientates with in the IA/IIA binding cleft in an almost identically manner (Figure 2.7 B - spheres). For the $NBD_{(ADP)}$-$SBD_{(pep)}$ complexes, the linker appears to assume variable orientations and is unable to bind the IA/IIA cleft in either replica (Figure 2.7 C - spheres). These data and the findings of previous studies support the hypothesis that an intermediate conformational structure may exist between the closed ADP bound configuration and the ATP bound open configuration but only in the presence of bound ATP at the NBD [104,238,239].

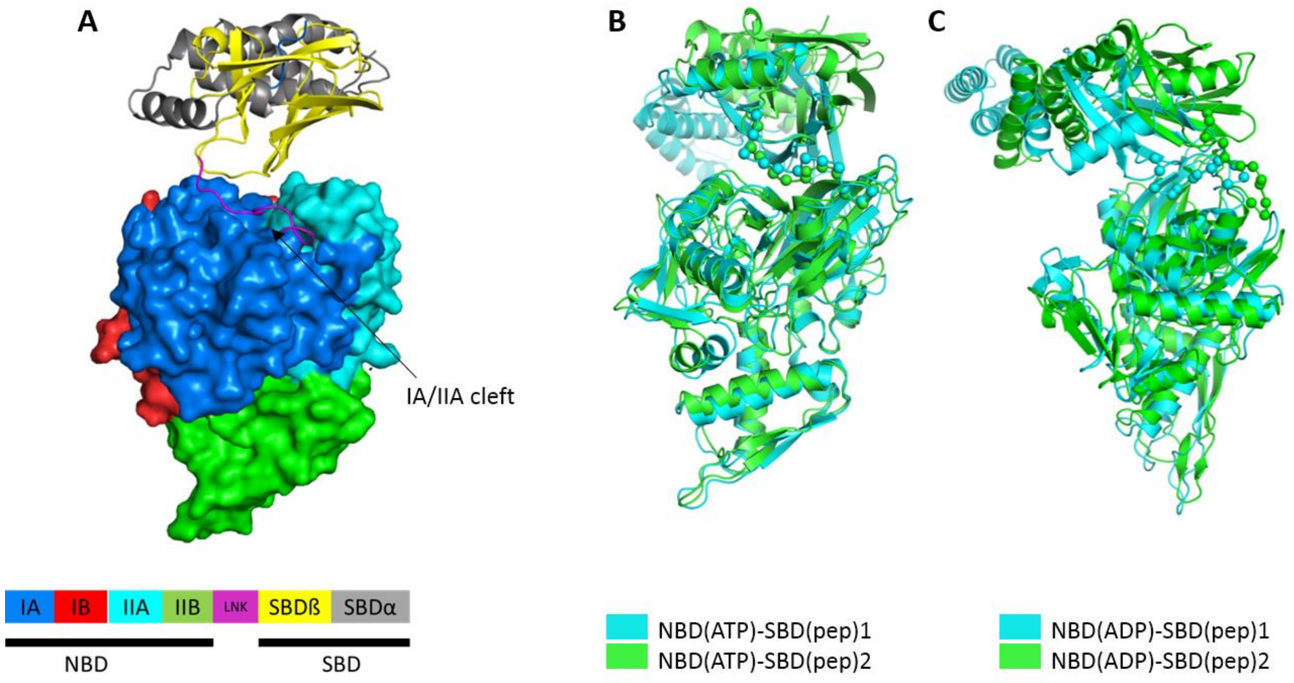


Figure 2.7: Detailed illustration of the linker docking to the IA/IIA binding cleft on the NBD. (A) The NBD is depicted in surface representation while the SBD in cartoon representation. The subdomains and linker are coloured according to the key. The docked linker (magenta) is oriented with in the subdomain IA/IIA binding cleft. **(B-C)** Structural representation of the superimposition of duplicate ATP (B) and ADP (C) peptide bound complexes, demonstrating the reproducibility of the MD simulations with respective RMSD values of 4.21 Å and 5.40 Å and showing near identical linker positioning in the presence of ATP (C) and variable linker positioning for ADP (B). Linker residues are shown as spheres. Reproduced with permission from Penkler et. al 2017 [105]

- *Conformational dynamics of the open conformation complexes*

For the open conformation complexes neither bound nucleotide, nor bound peptide substrate had observable effects on the global conformation dynamics and inter-domain structural arrangement of the SBD in relation to the NBD (Figure 2.9). Despite moderate flexibility of the SBDα, the backbone RMSD in all six configurations stabilised within 100 ns except for the NBD$_{(ATP)}$-SBD$_{(pep)}$ complex, which stabilised after 130 ns (Figure 2.8). The only notable difference between the configurations was elevated maximum RMSD displacements (>10 Å) for the ATP bound and apo complexes compared to the ADP bound complexes. Details of the stable segments used for PRS analysis are summarised in Table 2.1, and the maximum RMSD fluctuation of 2.3 Å for these segments likely account for the flexibility observed in the SBDα.

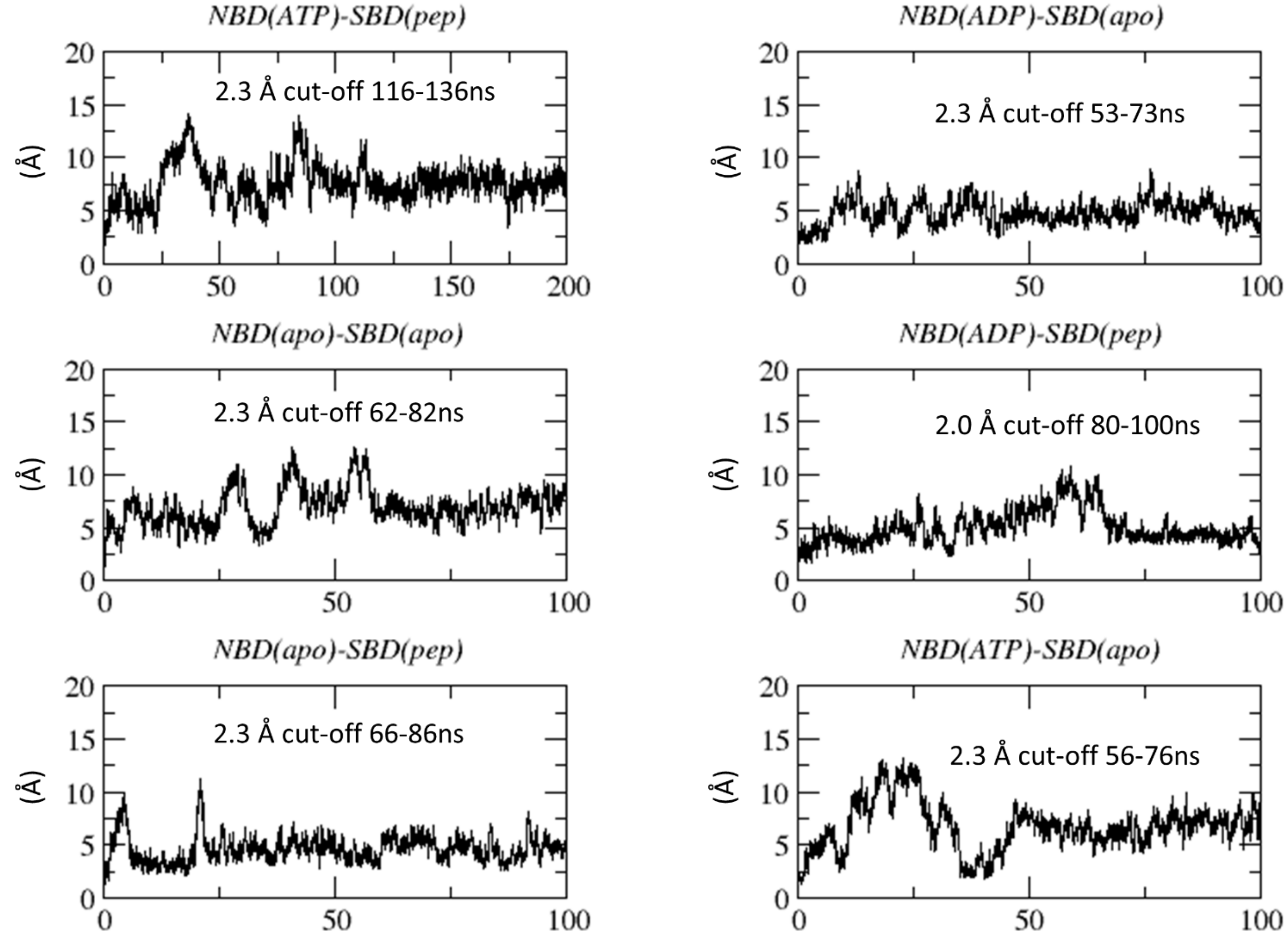


Figure 2.8: Backbone RMSD plots for the open conformation complexes. Each plot is annotated with the minimum RMSD cut-off and time points for the 20 ns equilibrated regions. Reproduced with permission from Penkler et. al 2017 [105]

Previous studies have reported several key contacts (Lys$_{414}$-Asp$_{326}$; Asn$_{415}$-Thr$_{221}$; Gln$_{442}$-Asp$_{148}$; Asp$_{481}$-Arg$_{151}$/Arg$_{167}$/Ile$_{168}$; Leu$_{484}$-Asp$_{148}$) between the SBD and NBD in the open ATP bound conformation, that form part of an important stabilising hydrogen bond network [101,238], that has been implicated in inter-domain allostery [103,120,276,277]. We examine these contact pairs in the equilibrated structures of each open conformation complexes to determine whether bound nucleotide and/or peptide influence the formation these contacts, and thus indirectly affect the stability of the conformation (Figure 2.9). It is evident that residue Asp$_{481}$ appears to be central to this network forming a triad of paired contacts with Arg$_{151}$, Arg$_{167}$, and Ile$_{168}$ respectively. In the NBD$_{(ATP)}$-SBD$_{(pep)}$ and ADP bound complexes with/without bound peptide, Asp$_{481}$ interaction with Arg$_{151}$ and Ile$_{168}$ are absent in the equilibrated structures (Table 2.2, Figure 2.9).

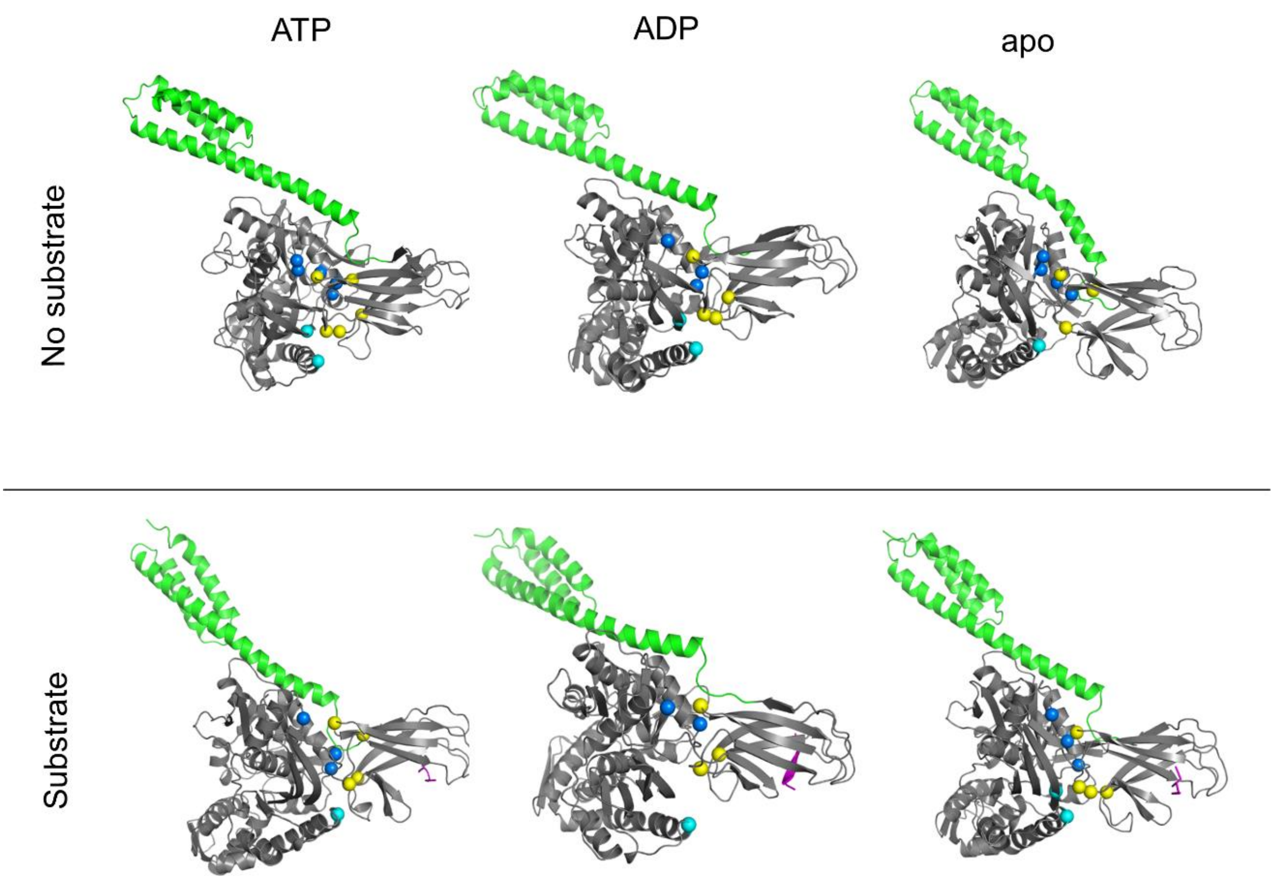


Figure 2.9: Structural arrangements of the six open conformation configurations after MD simulation. Stabilising interactions between domains are represented as spheres and coloured according to their subdomain location: IA – blue, IIA – cyan, SBDß – yellow. Reproduced with permission from Penkler et. al 2017 [105]

Furthermore, there is a complete loss of contact between Leu$_{484}$ and Asp$_{148}$ in these complexes – an interaction that has previously been reported to play a crucial role in sensing bound peptide at the SBD [103,120,237,240,249,276–278] (Table 2.2). Indeed, this interaction is only observed in the absence of bound peptide for the NBD$_{(ATP)}$-SBD$_{(apo)}$ and NBD$_{(apo)}$-SBD$_{(apo)}$ complexes, suggesting that peptide binding trigger localized restructuring that disrupts this contact pair. In fact the NBD$_{(ATP)}$-SBD$_{(apo)}$ complex was the only configuration that maintained all of the aforementioned hydrogen bond contacts with the exception of Asp148 and Gln442 (Table 2.2).

Table 2.2 | Summary of inter-domain hydrogen bond contacts for the equilibrated open conformation complexes. Showing how peptide binding reduces the number of contact pairs.

Contact pair	414	415	442		481			484
	326	221	147	148	151	167	168	148
NBD$_{(ATP)}$-SBD$_{(apo)}$	✓	✓	✓		✓	✓	✓	✓
NBD$_{(ATP)}$-SBD$_{(pep)}$	✓		✓	✓		✓		
NBD$_{(ADP)}$-SBD$_{(apo)}$	✓	✓	✓	✓		✓		
NBD$_{(ADP)}$-SBD$_{(pep)}$	✓			✓		✓		
NBD$_{(apo)}$-SBD$_{(apo)}$	✓				✓	✓	✓	✓
NBD$_{(apo)}$-SBD$_{(pep)}$	✓	✓	✓		✓	✓		

In summary, analysis of the all-atom MD simulations for the closed and open conformations provides structural evidence for the regulation of conformational dynamics by bound nucleotide at the NBD and/or bound peptide at the SBD. For the closed conformation Bound ATP and peptide causes a binding cleft to for between the IA and IIA subdomains on the NBD allowing the inter-domain linker to dock to the NBD in a stable manner. For the open conformation, peptide binding appears to disrupt an inter-domain hydrogen bond network that is crucial for maintaining the open state. Evidence that this network breaks down in the presence of bound peptide supports the hypothesis that these residues are involved in the allosteric recognition of bound peptide at the SBD and thereby act as a trigger for a conformational shift to the closed conformation [103,120,237,240,249,276–278].

2.3.2 Measuring the allosteric potential of Hsp70 by PRS analysis

In this study PRS was carried out following previously published methodology to quantify the allosteric potential of individual residues to select conformational displacements in response to random force perturbations [263,264,279] (see sections 2.1.2 and 2.2.3 for more details). By assessing each of the 12 Hsp70 configurations, 12 unique PRS experiments could be conducted and the results compared and contrasted with respect to the bound configuration of the respective NBD and SBD (Figure 2.1). The 12 Hsp70 configurations were divided into two groups based on their initial state conformations (open or closed), and PRS used to assess the allosteric potential for single residue perturbations to select a conformational shift towards the opposite conformational state i.e. open state → closed state and *vice versa*. We refer to these experiments as being either closing (Figure 2.1 A-F) or opening (Figure 2.1 G-L) transitions respectively. PRS requires structural information for the start and end-point states to determine the expected conformational displacement when transitioning from one state to another. For the closing transition experiments the final end-point state was set to the know NMR structure PDB ID 2KHO [91] and for the opening transition experiments, the x-ray structure PDB ID 4BQ9 [101]. In each case the initial frame from the equilibrated MD trajectory was set as the initial state structure.

For each set of PRS experiments (closing/opening), we identify all residues that are accentuated by PRS analysis as being residues whose perturbation selects a coordinate change towards the target conformation. The selection of accentuated residues was conducted depending on the distribution of the observed correlation coefficient values (C_i) using the following approach: (i) First the residues are ranked in descending order of C_i and plotted on the same set of axes to allow for the identification of the maximum C_i value (C_i^{max}) for the given transition, an exercise that also affords the opportunity for a quantitative comparison between the different configurations. (ii) Next, two correlation threshold values are set – a minimum threshold of $C_i \geq 0.60$ [264] and an upper threshold that is equal to the maximum C_i value observed over all experiments less 0.10 i.e. $C_i^{max} - 0.10$ for a given transition. (iii) If the majority of residues have C_i greater than the upper threshold, we do not report any residues to be accentuated by PRS and interpret this behaviour as a natural propensity for the protein to interconvert to the target conformation. (iv) Conversely, if the majority of residues have $C_i < 0.60$, this behaviour is interpreted as resistance of the protein to interconvert to the target conformation. (v) Finally, the C_i profiles of those experiments with select residues that are accentuated by PRS are analysed in detail, and the results critically compared to previous experimental and computational findings.

– *The closing transition experiments*

In the following section we present the PRS analysis data for the closing transition experiments (Figure 2.1 A-F) by plotting the C_i values for all six experiments as descending curves on the same set of axes (Figure 2.10). Looking at the descending correlation curves for the closing transition, a maximum $C_i = 0.89$ is observed for the NBD$_{(ADP)}$-SBD$_{(pep)}$ complex and as such the upper threshold is set to 0.79 (Figure 2.10 - black line). In the presence of ADP, 86% of residues have $C_i > 0.79$, for the NBD$_{(ADP)}$-SBD$_{(apo)}$ and 43% for the NBD$_{(ADP)}$-SBD$_{(pep)}$ complex (Figure 2.10 yellow & green respectively). The elevated number of accentuated residues for these complexes suggests that when bound by ADP, Hsp70 has a natural tendency to inter-convert from the open state conformation to the closed state conformation. Conversely, the presence of ATP alone the NBD$_{(ATP)}$-SBD$_{(apo)}$ complex records the lowest number of $C_i > 0.79$ (6%) (Figure 2.10 dark blue), while the peptide bound equivalent, NBD$_{(ATP)}$-SBD$_{(pep)}$ records 31% $C_i > 0.79$ (Figure 2.10 red). Interestingly, the latter complex shares a similar descending correlation curve to the nucleotide free complexes recording 32% $C_i > 0.79$ for NBD$_{(apo)}$-SBD$_{(apo)}$ complex and 30% for the NBD$_{(apo)}$-SBD$_{(pep)}$ (Figure 2.10 brown & light blue respectively).

We note that all six configurations share similar shaped descending C_i curves, in which the first ranked residues experience a sharp drop in C_i, followed by a quasi-linear behaviour before experiencing another sharp drop in C_i for the final ranked residues. This data trend is not necessarily universal however, for a protein that displays simple hinge bending motions, it is possible that there will be a multitude of residues whose perturbation would accomplish such a conformational change and one would expect the C_i curves to gradually decrease having many residues with high C_i. Conversely, for proteins with complex motions e.g. a mix of shearing and hinge motions or rotational and translational motions typical of the multi-domain proteins such as Hsp70, there are likely only a few residues capable of selecting such a complex conformational change. This would explain the sudden drop in the initial part of the curve. In the low correlation extreme, there will always be residues that resist forcing due to high connectivity in the structure such that their motion does not accomplish any productive change e.g. residues located in the core of the protein, and thus a sudden decrease in the C_i values at this end is expected for all proteins.

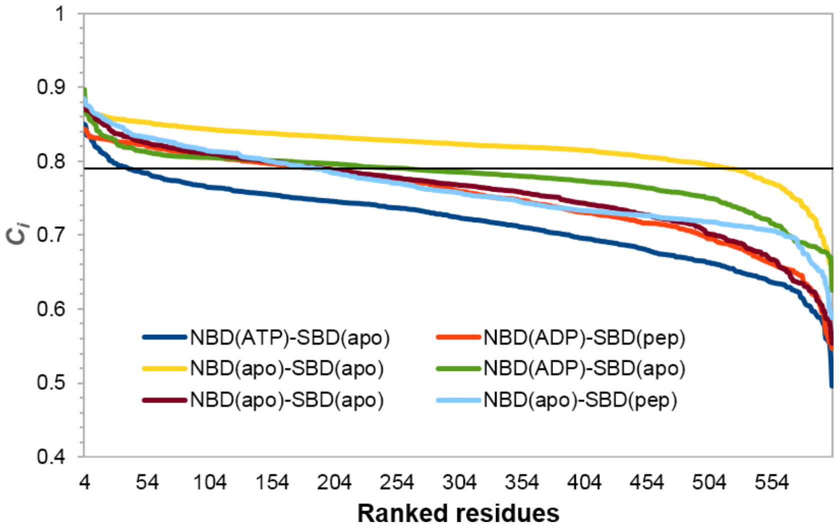


Figure 2.10: Descending C_i profiles for the closing transition experiments. Graphically indicating the propensity for allosteric inter-conversion to the closed state for each configuration. The upper correlation threshold value of 0.79 is shown as a black line. Reproduced with permission from Penkler et. al 2017 [105]

- *ADP drives the closing transition*

NMR studies have shown that the nucleotide free and ADP bound states of Hsp70 appear to be more dynamic than the ATP bound state, particularly with respect to subdomain tilting and shearing movements in the NBD coupled with increased SBDα-SBDβ and NBD-SBDβ distances [90,280]. MD studies based on DnaK homology models have demonstrated that bound ADP promotes spontaneous closing in to a semi-closed intermediate state [248] in which the SBD does not interact with the NBD [239]. Our PRS data is in agreement with these findings in which the ADP bound configurations recorded elevated accentuated residue counts: 86% in the absence of bound peptide and 43% in the presence of bound peptide. These results suggest that under ADP bound conditions, Hsp70 has a natural propensity to interconvert to the closed conformation.

- *Peptide binding allosterically activates the open ATP bound complex*

Previous studies on mitochondrial Hsp70 conformational dynamics have indicated that in the presence of ATP the open conformation is more stable compared to the ADP or nucleotide free state, although undocking of the otherwise bound NBD and SBD is transiently observed on the time-scale of seconds [281]. MD simulations of an ATP bound DnaK homology model indicated the NBD to have decreased mobility, with the SBD remaining stably docked to the NBD [239,248]. Hydrogen exchange mass spectrometry as well as NMR experiments have demonstrated that addition of a peptide substrate at the SBD in conjunction with bound ATP at the NBD causes the SBDβ to undock from the NBD leading to a conformational transition to a semi-closed intermediate state which is thought to be allosterically active [104,282].

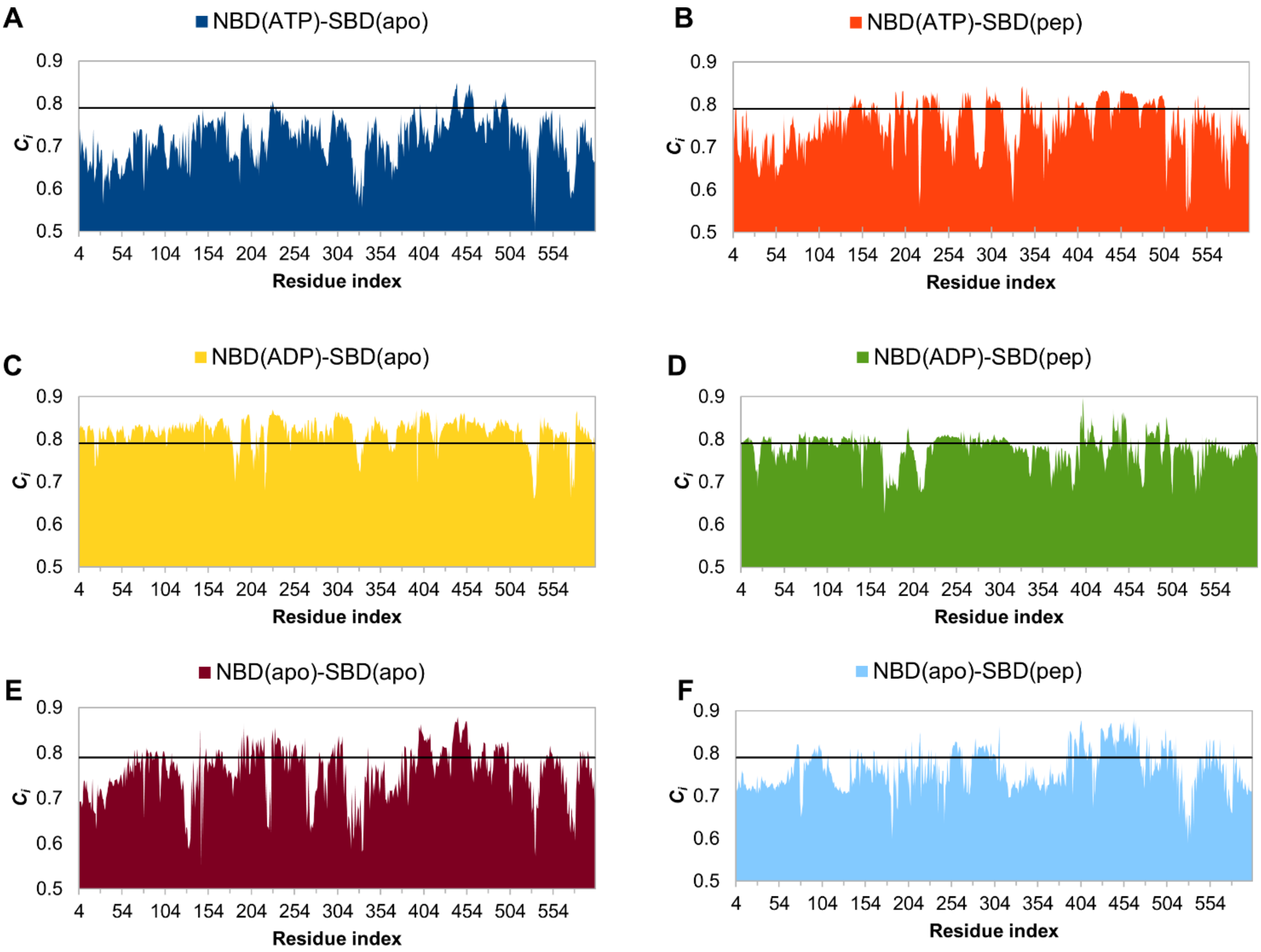


Figure 2.11: Correlation profiles for the closing transition experiments. Showing the distribution of residues with $C_i > 0.79$: A – 6%, B – 31%, C – 86%, D – 43%, E – 32%, F – 30%. Each configuration is colour according to the respective descending curves in Figure 2.10. The upper threshold value of 0.79 is shown as a black line. Reproduced with permission from Penkler et. al 2017 [105]

PRS analysis of the open ATP bound complexes indicates that in the absence of bound peptide only 6% of are capable of selecting a transition towards the closed conformation, while 31% of residues are accentuated in the peptide bound configuration (Figure 2.11 A & B). The increased number of allosterically active residues in the latter peptide bound configuration suggests this complex to have a higher propensity to interconvert to the closed state. Taken together these data are in line with the abovementioned experimental work and support the hypothesis that in the absence of bound peptide the open $NBD_{(ATP)}$-$SBD_{(apo)}$ complex is relatively stable and that peptide binding at the SBD may allosterically activate the protein and thereby enable transition towards an allosterically active semi-closed intermediate state [104,280]. Indeed, in a previous study statistical coupling analysis was used to define a contiguous group of 115 co-evolving residues implicated in mediating allosteric coupling in Hsp70 [249]. Of these 115 residues, 64 are also accentuated in our PRS analysis of the $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration, while only 12 residues are observed for the $NBD_{(ATP)}$-$SBD_{(apo)}$ complex – again these data point to an allosterically active peptide bound complex.

Visual comparison of the C_i profiles for the ATP complexes (Figure 2.11 A & B) indicates that residues accentuated by PRS tend to be grouped by residue number, forming distinct peaks of contiguous residues. Comparing the location of the C_i peaks for these complexes, we find that the distribution of C_i for the peptide bound configuration spans both the NBD and SBD with the exception of subdomain IB in the NBD (Table 2.3). For the peptide free configuration C_i are predominantly constrained to the SBDβ with only four observations in the NBD (subdomains IIA and IIB, Table 2.3). Bellow we analyse the residue composition of each peak in the $NBD_{(ATP)}$-$SBD_{(apo/pep)}$ complexes for functional and allosteric importance by drawing comparisons with previous experimental and computational studies.

- *Functional significance of PRS accentuated residues in the NBD*

Starting with the NBD of the $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration, residues P_{143}-F_{146}, R_{151}, and E_{171}, have been previously shown to play a role in the a functionally relevant proline switch mechanism [237]. Of these residues Y_{145} and F_{146} play a crucial role in stabilizing the open conformation through aromatic interactions with P_{143} [130,237]. Residues E_{171}, G_{198}, T_{199}, K_{270}, S_{274}, T_{344}-R_{355} are reported to be either implicated in direct interactions with bound nucleotide or the coordination of ATP hydrolysis, their mutation leading to a loss of allosteric control [94,237]. Also in the NBD are the hinge residues L_{227}, G_{228}, and G_{229} which are intricately involved in coordinating shearing motions between subdomain IA and IIA. Alanine point mutations of these residues resulted in reduced *in vitro* and *in vivo* chaperone activities in DnaK [236]. All of these NBD residues were found to be accentuated by PRS in the ATP and peptide bound complex.

ATP hydrolysis has been previously shown to be synergistically stimulated >1000-fold by the simultaneous binding of peptide substrate and the JDP co-chaperone [117,120,130,237], highlighting the functional importance of JDP interactions in Hsp70's ATPase cycle. The interaction of JDP with Hsp70 is not yet fully understood and to date only residues Y_{145}, N_{147}, D_{148}, R_{167}, T_{170}, E_{217}, V_{218}, G_{400}, D_{526}, and G_{539} have been implicated in JDP binding [120,130,283,284]. Of these residues, only N_{147} and D_{526} are not accentuated by PRS in the NBD-$SBD_{(pep)}$ configuration, however none were observed to be accentuated in the $NBD_{(ATP)}$-$SBD_{(apo)}$ complex suggesting JDP binding to be a putative allosteric activator for the closing transition but only in conjunction with simultaneously bound substrate. A further study by Gao and co-workers demonstrated that the SBDα is essential for JDP binding and ATPase stimulation [285]. In line with these results, we found several PRS accentuated residues in the SBDα of the NBD$_{(ATP)}$-SBD$_{(pep)}$ complex but not in the peptide free complex. Indeed, residues M$_{515}$, E$_{520}$, G$_{539}$, L$_{543}$, V$_{550}$, E$_{551}$ G$_{554}$, and A$_{582}$ could be naturally stimulated by JDP binding at the C-terminal of Hsp70.

Table 2.3 | Summary of all accentuated residues for the ATP bound complexes.

Location	NBD$_{(ATP)}$-SBD$_{(apo)}$	NBD$_{(ATP)}$-SBD$_{(pep)}$	References
IA	-	**7**, 113, 120, **127**, **139**, 140, **141**, **143**, **144*-146**, **148**, 149-150, **151**, **152** 153, 156, 164, 166, **167**, **170**, **171**, 381	2–4,14,15,24
IB	-	-	-
IIA	226	**192**, **193**, **195**, **198-200***, **201**, **205**, 207*, **216***, **217***, **218**, 224, **225**, 226, **227***, 310-314, **316**, 338, **339**, **340**, **344-347**, **349-351**, **354**, **361**, **362**	14,15,20, 23,49,50
IIB	**228**-**229**, 232	**228**, **229**, 230, **231**, 232-234, **235***, **236**, **237***, 238-240, 243, 267, **270**, 271, 273, **274**, 275-281, **297***, **298**, **299***, 303*-304*-308*-309*	14,20,24
linker	-	**392-393**	2,14,15
SBDβ	**399, 418-419,438-440-442**, 444, 446, 450, 451-453-461 , 485, 487, 489, **494**, 495-498, **499**, 500	**398***, **400*-409*, 410***, 413, **414***, **419***, **425*-431*, 432**, **433***, **434-436*-438*-440*, 446***, **447***, **449***, **450** 451*, 453*, **454**, 455-456, **457***, **458***, **459**, 460-461, **462-467***, **468***, **469-471***, **472**, 473, **474**, 475-**479-482**, 483, **484**, 485*, **486**, 487, **488 489-490***, **493**, **494** 497*, 498*, **499***, 500*-502*	9,10,15,23, 37,49,57
SBDα		**503**, **504**, **505**, **515**, **520**, **539***, 540, 543, 550*-551, 554, 582	4,15,23,71

Residues shown in **bold** are previously reported functional or allosterically relevant residues.
Residues marked with * were also observed in the nucleotide free complexes

Binding interactions with ATP has been shown to trigger intra-domain conformational restructuring of NBD, resulting in the formation of a cleft between the IA and IIA subdomains. This structural rearrangement provides the inter-domain liker with a suitable binding groove to drive inter-domain docking between the NBD and SBD. The positioning and orientation of the linker in this binding cleft has been shown to be essential for both ATP-stimulated peptide release as well as polypeptide-stimulated ATP hydrolysis [103,120,276–278]. We find that the allosteric mediator linker residues L$_{392}$ and D$_{393}$ to be accentuated by PRS in the NBD$_{(ATP)}$-SBD$_{(pep)}$ complex. *In vitro* mutation of these residues was shown to impair inter-domain allostery [120,276]. In addition to the inter-domain linker, PRS also accentuates several residues lining the IA/IIA cleft (Y$_{145}$, D$_{148}$, K$_{155}$, E$_{217}$ and V$_{218}$) as well as the linker coordinating residue P$_{419}$ in the SBDß. Mutational studies have also implicated the former set of residues in inter-domain allosteric communication [276,283], while the latter P$_{419}$ has been shown to be crucial for maintaining allosteric inter-domain communication [101,102,249]. Another study suggests that peptide binding coupled with JDP interactions may be necessary to reposition the linker within the IA/IIA binding cleft and in doing so may provide structural stimulation of the inter-domain contact residues R_{151}, K_{155}, and R_{167}, which in turn act as a trigger for the proline switch and eventually the initiation of ATP hydrolysis [237,276]. It is evident from our PRS data that peptide binding is necessary in order to accentuate the aforementioned linker coordinating residues as well as the proline switch and JDP binding residues as none are accentuated in the peptide deficient $NBD_{(ATP)}$-$SBD_{(apo)}$ complex.

In addition to the formation of the IA/IIA cleft in response to ATP binding, a further effect of bound ATP is the rotation of the NBD lobe I relative to lobe II. This structural rearrangement enables the formation of inter-domain contacts between the NBD and SBD which in turn primes the ATPase domain for ATP hydrolysis [286,287]. Peptide binding and the subsequent restructuring of the SBD is thought to destabilise the inter-domain contacts allowing back rotation of lobe I, which in turn provides the necessary switch required for ATPase activation. Premature back rotation of lobe I and thus ATP hydrolysis is prevented through several crucial stabilising contacts between the NBD and SBDß including residues D_{148}, K_{414}, Q_{442}, D_{481}, L_{484}, S_{505}, and G_{506} [101,102,130], all of which are also accentuated by PRS in allosterically active $NBD_{(ATP)}$-$SBD_{(pep)}$ complex.

- *Functional significance of PRS accentuated residues in the SBD*

Next the SBDβ is investigated and in particular residues that are known to be involved in the recognition and binding of peptide substrates. A recent study characterised the residue composition of the peptide binding cavity of DnaK to include residues I_{401}-T_{409}, Q_{424}-H_{439}, N_{458}, R_{467}-M_{469}, I_{472}, and V_{474} [288]. PRS analysis of the NBD(ATP)-SBD(pep) complex yields several distinct peaks in the SBDβ that overlap significantly with these binding cavity residues including G_{400}-T_{410}, V_{425}-V_{440}, and S_{453}-E_{473}. Also in the NBD(ATP)-SBD(pep) complex, PRS accentuates residues S_{398}, T_{403}, G_{405}, T_{428}, D_{431}, I_{438}, F_{457}, L_{459}, and G_{468} of the SBDß all of which experience large NMR chemical shift changes in the presence substrate is bound [103], further illustrating the relative importance of peptide binding for allosteric activation of the open complex. Despite there being several PRS peaks in the SBDβ of the $NBD_{(ATP)}$-$SBD_{(apo)}$ complex (Figure 2.11 A) there is a complete absence of these particular substrate biding cavity residues (with the exception of N_{458}) in this peptide free complex.

An *in silico* alanine scanning mutation study on the full-length ATP bound open structure of DnaK, found the highly conserved SBDß residues S_{427}, T_{428}, M_{469}-I_{472} to be particularly intolerant to mutation [289]. Here, we find these residues to also be accentuated by PRS in the $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration, but not the $NBD_{(ATP)}$-$SBD_{(apo)}$ complex (Table 2.3). Since these are also sites directly involved in substrate recognition, their natural perturbation in the open state by substrate binding may provide the necessary trigger for inter-state conversion towards the closed state. In the same study, mutations of the SBDß residues G_{405} and V_{407} severely disrupted the intra-domain communication pathways between the SBDα and SBDß [289]. The SBDα residues V_{533} and R_{536} are bridged over the G_{405} and V_{407} and positioned around the peptide binding site via a hydrophobic patch and G405S and M408I mutants have been experimentally shown to diminish peptide binding [242]. Given their location and contribution to peptide recognition, it is not surprising that these residues are also accentuated by PRS for the $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration and not the $NBD_{(ATP)}$-$SBD_{(apo)}$ complex. Collectively our PRS results suggest that perturbations arriving at these residues may be central in controlling the grip of DnaK over the substrate and that this role may be perfectly fulfilled by JDP in the initial substrate binding step.

PRS was previously used by General *et al*. in an earlier experimental study to validate inter-domain allostery in DnaK using a homology model of the ATP bound open conformation [250]. In this study, several groups of effector/sensor residues were reported in the IA, IIB, and SBDß subdomains. We find that the C_i profile of our $NBD_{(ATP)}$-$SBD_{(apo)}$ configuration closely resembles that of General *et al*., but note that only the SBDß residues reported therein (F_{426}, A_{435}, I_{462}, Q_{471}-I_{472}, and K_{491}) overlap with our PRS hot residues (Table 2.3). In the aforementioned study, ANM was used to construct the Hessian matrix and the quality of predicted effector residues is evaluated by the magnitude of the resulting displacements rather than by correlation with a known experimental displacement. While this generates a similar profile to the one seen in our study the definition of allosteric hot residues is not the same. However we note, that all effector residues reported by General and co-workers overlap with accentuated residues observed for the allosterically active $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration in our study.

- *Peptide binding residues may be allosteric mediators of the closing transition*

Thus far we have reported a number of residues accentuated by PRS for the six open conformation configurations and find evidence that suggests the bound state of the NBD and SBD affects the number and distribution of accentuated residues. This finding prompts the question, are there any residues that are accentuated regardless of bound nucleotide or peptide substrate? The C_i profiles illustrated in Figure 2.11 indicate the SBDß to be consistently accentuated in all six configurations, and analysis of these peaks reveals 10 residues (I_{438}-V_{440}, $K4_{46}$, D_{450}, N_{451}, Q_{456}, H_{485}, Q_{497}, and T_{500}) located in the peptide binding cavity overlap all six configurations. Of these residues, special mention must be made of I_{438}-V_{440} and H_{485} which have been implicated in peptide recognition and allosteric signal propagation [241,288,290], while a recent study has shown that alanine mutations to V_{440} and the nearby L_{484} abrogates substrate simulated ATPase activity [130]. Indeed the authors of that study suggest a plausible mechanism for ATPase activation, whereby peptide binding exerts a force through V$_{440}$ and L$_{484}$ onto the NBD residue D$_{148}$, which acts as a release trigger for lobe I back rotation and activation of ATP hydrolysis [130]. While it is as yet unclear what the functional relevance is of the remaining seven residues their conserved accentuation by PRS together with the key substrate recognition residues I$_{438}$-V$_{440}$, and H$_{485}$ in all six configurations, suggests these residues may act as key allosteric hotspots for initiating the closing transition in response to peptide binding.

The nucleotide free complexes favour the closed conformation

Previous work has demonstrated that in the absence of bound nucleotide the SBD and NBD are mostly separated and experience transient docking on a 5 ms timescale [106,239]. Furthermore, a recent MD study demonstrated that the SBDα occasionally explores partially closed and open conformations in the absence of nucleotide and substrate [248]. For our NBD$_{(apo)}$-SBD$_{(apo/pep)}$ configurations, approximately 30% of residues are accentuated in each and of these residues 53% are observed in both configurations. Furthermore, 64 of these residues overlap with accentuated residues in the NBD$_{(ATP)}$-SBD$_{(pep)}$ (Table 2.3*), suggesting these nucleotide free configurations have a similar tendency to interconvert to the closed conformation. Indeed the functional residues E$_{217}$, G$_{400}$, G$_{539}$ (JDP interactions); D$_{398}$ (inter-domain linker); G$_{400}$-T$_{410}$, Q$_{424}$-V$_{440}$, N$_{458}$, R$_{467}$-M$_{469}$, and I$_{472}$ (substrate binding cavity) are conserved in all three complexes. From these data we conclude that the presence of bound peptide has little impact on the allosteric nature of the nucleotide free complexes and that the number of accentuated residues is conserved and distributed similarly to that of the NBD$_{(ATP)}$-SBD$_{(pep)}$. This suggests the nucleotide free complexes to have a natural tendency to interconvert toward the closed conformation.

The opening transition experiments

In this section the PRS analysis data obtained for the six opening transition experiments (Figure 2.1 2G-L) is discussed. Starting with the descending C_i curves depicted in Figure 2.12, the C_i^{max} for the six closed conformation configurations is found to be 0.80 for NBD$_{(ATP)}$-SBD$_{(pep)}$ complex and the upper C_i threshold accordingly set to 0.70. Visual inspection and comparison of the descending curves indicates a distinct pairing of experiments based on the identity and configuration of un/bound nucleotide at NBD. Interestingly the ATP bound complexes (Figure 2.12 blue & red) record the highest C_i values. A total 44% of residues record $C_i > 0.7$ for the NBD$_{(ATP)}$-SBD$_{(apo)}$, and 34% of residues in the NBD$_{(ATP)}$-SBD$_{(pep)}$ complex. IN the presence of ADP, only 3% of residues have $C_i > 0.7$ for the NBD$_{(ADP)}$-SBD$_{(apo)}$ complex and a mere 1% in the NBD$_{(ADP)}$-SBD$_{(pep)}$ complex (Figure 2.12 yellow & green). In contrast, no residues in the nucleotide free configurations have $C_i > 0.7$ (Figure 2.12

brown & cyan), a behaviour that can be interpreted as a natural resistance of the complex to interconvert towards the open ATP bound conformational state.

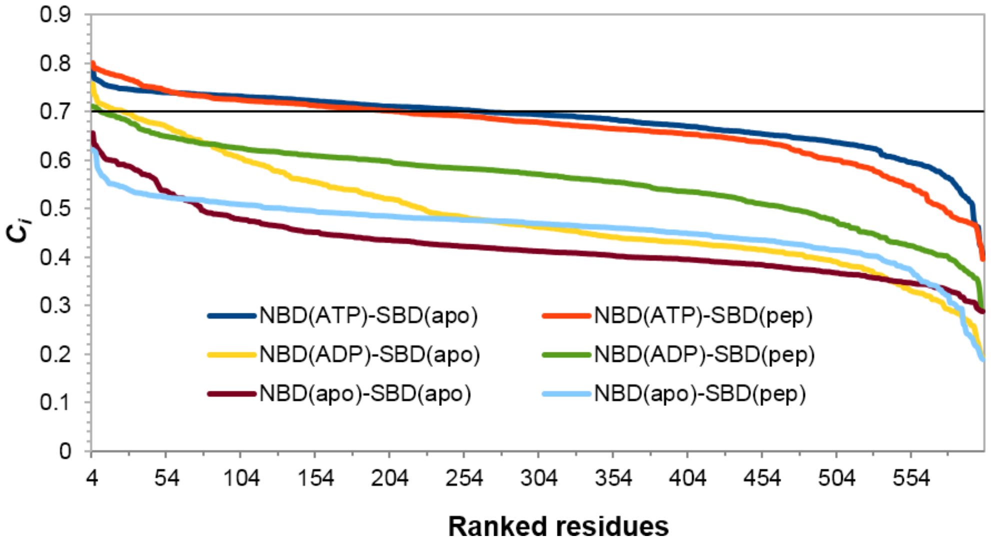


Figure 2.12: Descending C_i profiles for the open transition experiments. Graphically illustrating the propensity for allosteric inter-conversion for each configuration. The upper correlation threshold value of 0.70 is shown as a black line. Reproduced with permission from Penkler et. al 2017 [105]

- *ATP binding may promote an allosterically active intermediate state*

The elevated number of residues with $C_i > 0.7$ for the ATP complexes indicates that ATP binding may promote an allosterically "sensitive" or active state. This observation is in agreement with previous work which posits the notion of there being an allosterically active intermediate state between the closed $\text{NBD}_{(\text{ADP})}\text{-SBD}_{(\text{pep})}$ complex and the open $\text{NBD}_{(\text{ATP})}\text{-SBD}_{(\text{apo})}$ complex when ADP is exchanged for ATP [91,104,238,239]. This hypothesis is indeed supported by our MD simulations of the closed $\text{NBD}_{(\text{ATP})}\text{-SBD}_{(\text{pep})}$ complex which revealed an equilibrated structure that closely resembles such an intermediate state (Figure 2.6& Figure 2.7). This transient structure has also been reported in other MD simulation studies [238,239]. It is interesting to note that the $\text{NBD}_{(\text{ATP})}\text{-SBD}_{(\text{pep})}$ and $\text{NBD}_{(\text{ADP})}\text{-SBD}_{(\text{pep})}$ complexes share similar C_i distribution profiles, but that in the presence of ADP the C_i values are ±0.1 lower than when ATP is present (Figure 2.13 B), an observation that provides further evidence in support of the hypothesis that ATP binding favours allosteric signal propagation.

- *Bound peptide decreases the allosteric potential of the closed conformation*

In the absence of substrate, 44 % of residues in the $\text{NBD}_{(\text{ATP})}\text{-SBD}_{(\text{apo})}$ complex record $C_i > 0.70$, putatively suggesting that under these conditions Hsp70 may have a natural propensity to interconvert towards the open conformational state (Figure 2.13 B). On the other hand addition of peptide substrate at the SBD appears to decrease the allosteric potential of the protein particularly at the SBD. This observation coupled to previous studies which have shown that peptide interactions influence the conformational dynamics of the SBD dramatically reducing the frequency of spontaneous α-lid opening [101,281], suggest peptide binding interactions maintained at the SBD my serve a roll in stabilising the complex in the closed state by decreasing the allosteric potential of the protein.

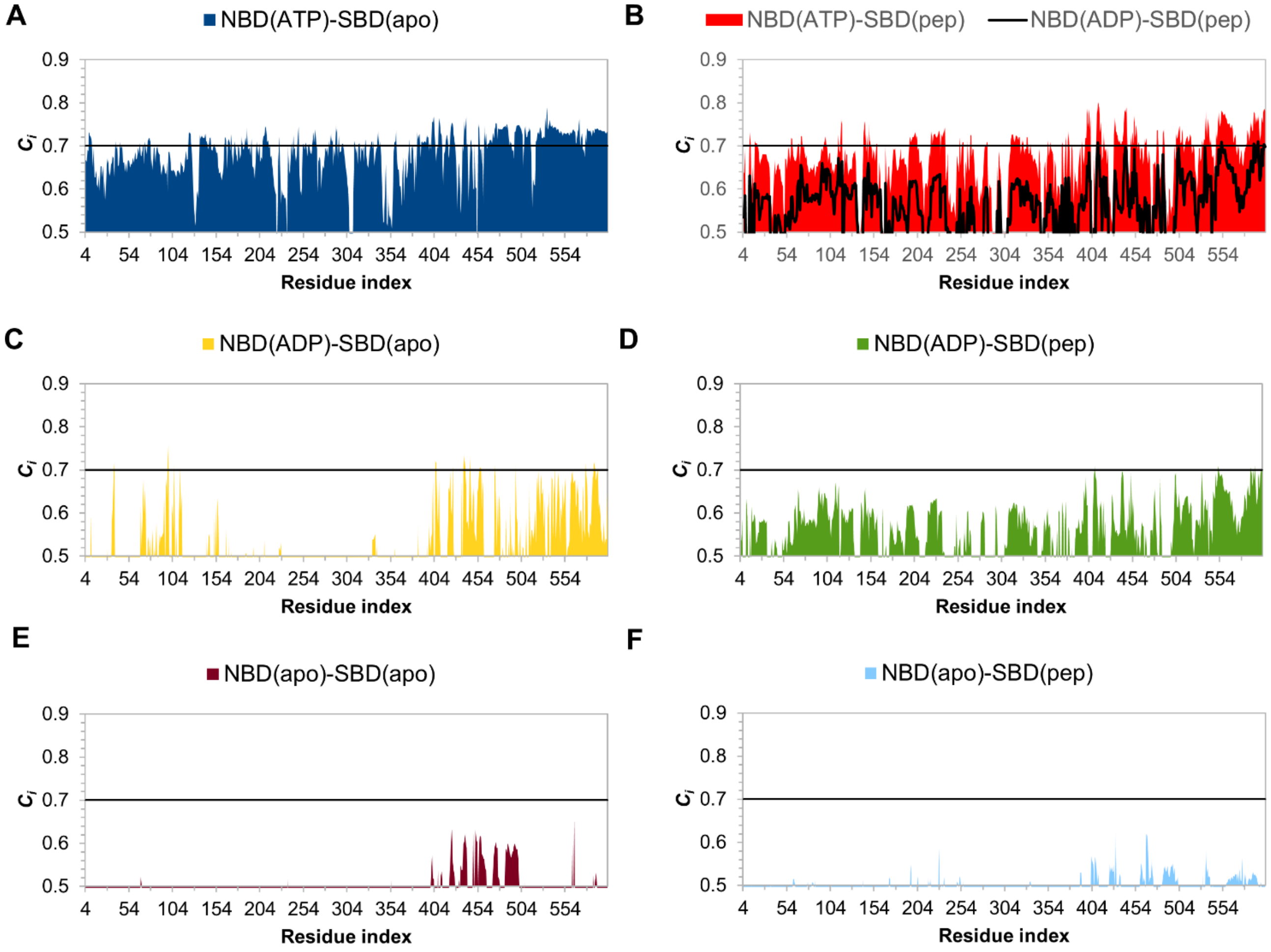


Figure 2.13: C_i Profiles for the opening transition experiments. Showing the distribution of residues with $C_i > 0.70$. A – 44%, B – 34%, C – 3%, D – 1%, E – 0%, F – 0%. Each configuration is colour according to the respective descending C_i curves in Figure 2.12. The upper threshold value of $C_i > 0.70$ is shown as a black line. In B, the data for D is superposed (black curve) to illustrate the profile similarities when ATP or ADP are present. Reproduced with permission from Penkler et. al 2017 [105]

It is also interesting to note that the highest C_i values for NBD$_{(ATP)}$-SBD$_{(pep)}$ complex were recorded in the SBDß, an observation that may be significant in that peptide disassociation must occur before the protein can fully resume the open conformation [101,130,281,291]. Previous biochemical studies have shown that residues R$_{536}$-Q$_{538}$ constitute an important hinge site for substrate release and that "latch" residues D$_{431}$, R$_{467}$, D5$_{40}$, H$_{544}$, and K$_{548}$ form crucial hydrogen bonds between the

SBDα and SBDß, together these residues may be important for maintaining the stability of substrate bound closed conformation [292–294]. This hypothesis has been further supported by a computational study which found allosteric inhibition of DnaK by PET-16 at the SBD, in which ligand interactions disrupt the proposed stabilising network, ultimately leading to the loss of substrate affinity [252]. With the exception of H_{544}, these residues collectively record the highest C_i values in the $NBD_{(ATP)}$-$SBD_{(pep)}$ complex and indeed much higher than those recorded for the $NBD_{(ATP)}$-$SBD_{(apo)}$ complex in which there is no client peptide present.

PRS also accentuate residues S_{398}, G_{400}, K_{414}, G_{443}, and E_{444} in the $NBD_{(ATP)}$-$SBD_{(pep)}$ complex, all of which have been implicated in modulating substrate specificity [242]. Similarly, residues F_{146} and D_{148} have been implicated in an intricate mechanism for the modulation of substrate release [130], both of which are also accentuated in this configuration. Taken together, these findings indicate that disassociation of substrate together with ADP/ATP exchange is necessary for the successful transition towards the open state conformation [101,281].

Functional relevance of accentuated residues in the $NBD_{(ATP)}$-$SBD_{(pep)}$ complex

Given the hypothesis that the $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration may be an allosterically active intermediate, the functional relevance of some key resides within the accentuated peaks of this complex are discussed in the context of previous experimental studies (Table 2.4 bold). Starting with the NBD: residues R_{71}, P_{143}, A_{144}, and F_{146} have been previously implicated in the mechanistic functioning of the proline switch [237]; G_{196}-T_{199} are known to be involved in direct interactions with bound nucleotide [94]; while residues G_{223}, L_{227}-G_{229} form important hinge sites in the NBD [236]. In the SBD, mutation of K_{414} has been shown to lead to a loss of inter-domain allosteric communication [240]. Residues N_{98}-G_{99} [239], F_{457} [249], and T_{437}-H_{439} [249,288] have been implicated in allosteric signal propagation. Finally, several accentuated residues in this particular complex correspond with residues that have also been suggested to be involved in allosteric signal propagation in two previous computational works [239,249].

Table 2.4 | Summary of accentuated residues in the $\text{NBD}_{\text{(ATP)}}$-$\text{SBD}_{\text{(pep)}}$ complex

Location	$\text{NBD}_{\text{(ATP)}}$-$\text{SBD}_{\text{(pep)}}$	References
IA	**11**, **12**, **17-19**, 113, **116**, 117, 142, **143**, **144**, **146**, **148**, 149-150, 156-157, **167**, 370, 373-374, 376-377, 380	3,4,15,23,50
IB	59-60, **66**, **68**, **70**, **71**, 73, **96-99**	3,20,23,24
IIA	**195-201**, 202-203, 219, 220, **221**, 222, **223**, **225**, 226, **227**, 311-315, **316**, 317-321, 324-325, 329-330, 333, **353**, 356,	14,15,20,23,49,50
IIB	**228-229**, **231**, 232-233-235, **252**, **255**, 256, 261, 264-265, 281, 283-284,	14,20,24
linker	**392**	2,14,15
SBDβ	**396-410**, 411, **412**, 413, **414**, **417**, **429-440**, 441, **442-444**, **448**, **450**, 451-453, **454**, 455, **457**, **464-467**, 478, **479**, 485, **486**, 498, 500-502	9,10,15,23,37,49,57,59
SBDα	**503-508**, 511, **512**, 514, **515**, 516-517, **520**, 530, 532-533-534-**536-538**, **540**, 546-**548**-568, 571, 581-602	4,15,23,60,71,72

Residues shown in **bold** are previously reported functional or allosterically relevant residues.

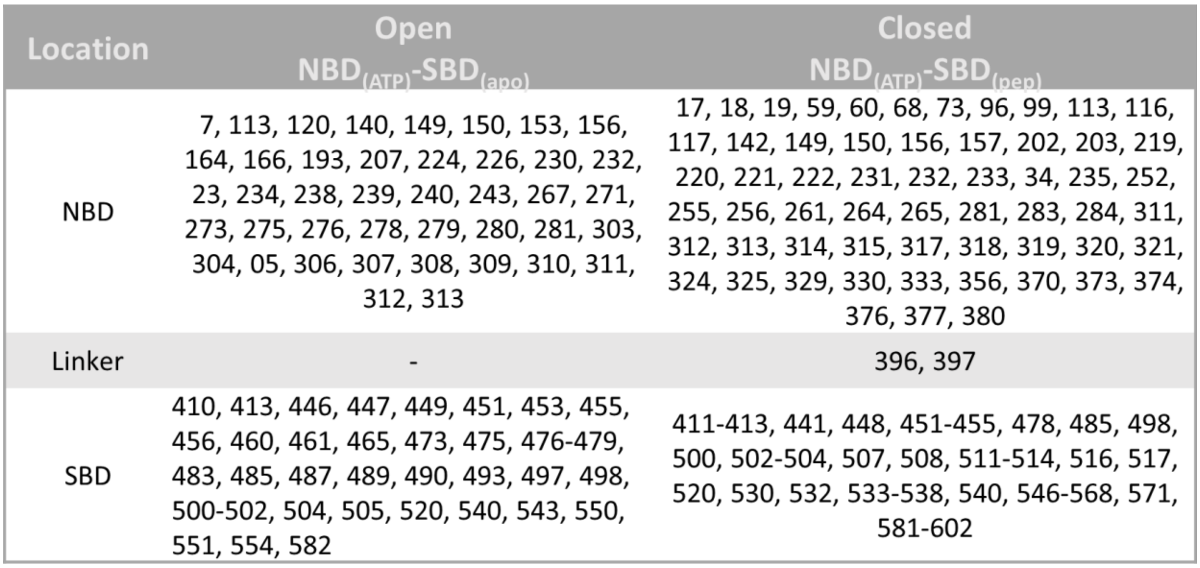
- *Overlapping closing and opening PRS hot residues point to key allosteric mediator residues*

Taken collectively, the data present thus far supports the assumption that Hsp70 may adopt an allosterically active intermediate state during conformational switching between the open and closed conformational states, but only when it is bound by both ATP and peptide substrate, irrespective transition direction. Comparison of the location of hot residues for these complexes reveals 78 overlapping accentuated residues in both configurations which include the previously validated functional residues belonging to: the proline switch, the NBD hinge residues, multiple residues from the substrate binding cavity including the highly conserved I_{438}, the substrate trigger residues D_{148}, V_{440} and H_{485}, and the allosteric modulator K_{414}. The fact that these critical functional residues overlap both the opening and closing transitions may highlight allosteric mediatory roles for these particular residues implicating them in allosteric signal propagation between the NBD and SBD in response to bound nucleotide and/or peptide substrate.

- *PRS hot residues unique to this study point to inter-subdomain interactions*

By filtering out all accentuated residues with known functional or allosteric implications in the $\text{NBD}_{\text{(ATP)}}$-$\text{SBD}_{\text{(pep)}}$ configurations for the closing and opening transitions, only those residues unique to this study are selected (Table 2.5) and mapped to their respective perturbed initial structures (Figure 2.14) and their significance analysed from a structural standpoint. For the closing transition, hot residues appear to be clustered around the inter- and intra-subdomain interfaces, particularly between the SBDβ-IA; IIA-IIB; and the linker-IA (Figure 2.14 A). No hot residues were observed for the IB in subdomain, while several residues appeared to be accentuated at the extreme C-terminal end of the SBDα. For the opening transition, PRS extensively accentuated residues located in SBDα, including several residues that interface with the loop regions of the SBDβ (Figure 2.14 B). For this transition PRS also accentuated several residues that interface between the IA-IB and IB-IIB subdomains, as well as several isolated residues in subdomain IIA.

Table 2.5 | Accentuated residues for the open and closed $NBD_{(ATP)}$-$SBD_{(pep)}$ configurations unique to the present study



Location	Open $NBD_{(ATP)}$-$SBD_{(apo)}$	Closed $NBD_{(ATP)}$-$SBD_{(pep)}$
NBD	7, 113, 120, 140, 149, 150, 153, 156, 164, 166, 193, 207, 224, 226, 230, 232, 23, 234, 238, 239, 240, 243, 267, 271, 273, 275, 276, 278, 279, 280, 281, 303, 304, 05, 306, 307, 308, 309, 310, 311, 312, 313	17, 18, 19, 59, 60, 68, 73, 96, 99, 113, 116, 117, 142, 149, 150, 156, 157, 202, 203, 219, 220, 221, 222, 231, 232, 233, 34, 235, 252, 255, 256, 261, 264, 265, 281, 283, 284, 311, 312, 313, 314, 315, 317, 318, 319, 320, 321, 324, 325, 329, 330, 333, 356, 370, 373, 374, 376, 377, 380
Linker	-	396, 397
SBD	410, 413, 446, 447, 449, 451, 453, 455, 456, 460, 461, 465, 473, 475, 476-479, 483, 485, 487, 489, 490, 493, 497, 498, 500-502, 504, 505, 520, 540, 543, 550, 551, 554, 582	411-413, 441, 448, 451-455, 478, 485, 498, 500, 502-504, 507, 508, 511-514, 516, 517, 520, 530, 532, 533-538, 540, 546-568, 571, 581-602

Previous studies have shown that allosteric signal propagation from the NBD to SBD and *vice versa* involves the intricate structural rearrangement of its respective subdomains highlighting the relative importance of key inter-domain contact residues [94,130,241]. PRS accentuation of several experimentally validated key inter-domain contact residues (D_{148}, K_{414}, Q_{442}, D_{481}, L_{484}, S_{505}, and G_{506}) [101,102,130] have already been discussed in context of allosteric signal propagation. To assess the whether PRS accentuates other meaningful inter-subdomain interactions and thus mediating residues for allosteric signal propagation, the Protein Interaction Calculator (PIC) [275] was used to identify inter-subdomain interacting residues in the initial perturbed $NBD_{(ATP)}$-$SBD_{(pep)}$ configurations for the closing and opening transition experiments and the results correlated with the PRS accentuated residues unique to this study (Table 2.6).

Table 2.6 | PRS hot residues involved in inter-subdomain interactions for the $\text{NBD}_{(\text{ATP})}$-$\text{SBD}_{(\text{pep})}$ complex

Closing transition						Opening transition					
Hot residue	Location	Interacting pair		Location	Interaction type	Hot residue	Location	Interacting pair		Location	Interaction type
		A	B					A	B		
113	IB	73	113	IA	HI	73	IB	73	112	IA	HI
113	IB	111	113	IA	HI	73	IB	73	146	IA	HI
113	IA	113	507	SBDα	HI	73	IB	73	153	IA	HI
149	IA	149	454	SBDß	HI	142	IA	142	216	IIA	HI
150	IA	150	73	IB	MC-SC HB	150	IB	74	150	IA	MC-SC HB
150	IA	150	73	IB	MC-SC HB	150	IB	74	150	IA	MC-SC HB
153	IB	73	153	IA	HI	150	IA	150	72	IB	MC-SC HB
207	IA	176	207	IIA	HI	150	IA	150	72	IB	MC-SC HB
207	IA	179	207	IIA	HI	150	IA	150	75	IB	MC-SC HB
226	IIA	226	71	IB	MC-SC HB	150	IA	150	75	IB	MC-SC HB
226	IIB	231	226	IIA	MC-SC HB	232	IIB	232	312	IIA	HI
226	IIB	231	226	IIA	MC-SC HB	232	IIB	232	346	IIA	HI
226	IIA	226	85	IB	SC-SC HB	312	IIB	232	312	IIA	HI
267	IB	55	267	IIB	SC-SC HB	396	SBDß	396	512	SBDα	HI
267	IB	55	267	IIB	SC-SC HB	397	SBDß	397	503	SBDα	HI
305	IIB	305	348	IIA	HI	448	SBDß	448	529	SBDα	HI
306	IIA	351	306	IIB	SC-SC HB	452	SBDß	452	506	SBDα	MC-SC HB
306	IIA	351	306	IIB	SC-SC HB	454	SBDß	454	503	SBDα	HI
309	IIB	309	346	IIA	HI	478	SBDß	478	503	SBDα	HI
309	IIB	309	348	IIA	HI	502	SBDß	502	388	LNK	MC-SC HB
311	IIB	235	311	IIA	SC-SC HB	503	SBDß	397	503	SBDα	HI
311	IIB	235	311	IIA	SC-SC HB	503	SBDß	399	503	SBDα	HI
311	IIB	235	311	IIA	SC-SC HB	503	SBDß	454	503	SBDα	HI
311	IIB	235	311	IIA	SC-SC HB	503	SBDß	478	503	SBDα	HI
312	IIB	232	312	IIA	HI	503	SBDß	484	503	SBDα	HI
313	IIB	232	313	IIA	HI	504	SBDα	504	388	LNK	SC-SC HB
338	IA	179	338	IIA	HI	512	SBDß	396	512	SBDα	HI
338	IIA	338	365	IA	HI	530	SBDß	446	530	SBDα	SC-SC HB
381	IA	381	386	LNK	HI	533	SBDß	407	533	SBDα	HI
381	IA	381	389	LNK	HI	534	SBDß	409	534	SBDα	SC-SC HB
413	LNK	395	413	SBDß	MC-SC HB	534	SBDα	534	409	SBDß	SC-SC HB
501	SBDß	501	503	SBDα	HI	534	SBDα	534	409	SBDß	SC-SC HB
504	IB	76	504	SBDα	MC-SC HB	549	SBDα	549	467	SBDß	MC-SC HB
						549	SBDα	549	467	SBDß	MC-SC HB
						585	SBDß	424	585	SBDα	SC-SC HB
						585	SBDß	424	585	SBDα	SC-SC HB
						585	SBDα	585	424	SBDß	SC-SC HB
						585	SBDα	585	424	SBDß	SC-SC HB
						592	SBDß	471	592	SBDα	MC-SC HB
						592	SBDß	471	592	SBDα	MC-SC HB
						593	SBDα	593	471	SBDß	MC-SC HB
						595	SBDß	469	595	SBDα	MC-SC HB
						596	SBDα	596	491	SBDß	MC-SC HB
						596	SBDα	596	491	SBDß	MC-SC HB
						598	SBDß	469	598	SBDα	HI
						601	SBDß	469	601	SBDα	HI

HB = Hydrogen bond; HI = Hydrophobic interaction; MC = Main chain; SC = Side chain

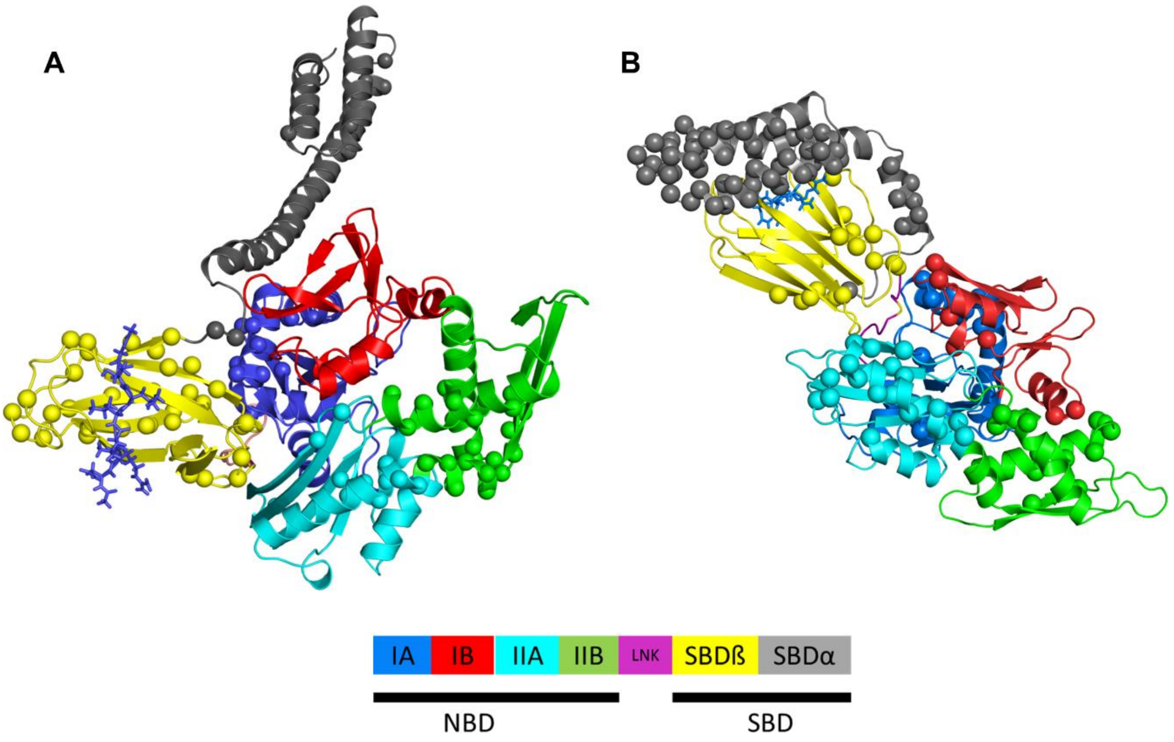


Figure 2.14: Structural representation of unique PRS hot residues. For the (A) open and (B) closed NBD$_{(ATP)}$-SBD$_{(pep)}$ configurations. PRS hot residues are indicated as spheres and coloured according to subdomain location. Reproduced with permission from Penkler et. al 2017 [105]

For the closing transition PRS accentuates several inter-subdomain interacting residues (P$_{113}$, A$_{149}$, Q$_{150}$, A$_{153}$, I$_{207}$, H$_{226}$, E$_{267}$, L$_{305}$, E$_{306}$, V$_{309}$, D$_{311}$, L$_{312}$, V$_{313}$, I$_{338}$, V$_{381}$, A$_{413}$, I$_{501}$, and S$_{504}$) that span all seven subdomains (Table 2.6). Several of these residues (A$_{149}$, Q$_{150}$, A$_{413}$, I$_{501}$, and S$_{504}$) are in close proximity to the abovementioned experimentally validated inter-domain contact residues. For the opening transition, PRS accentuates residues I$_{73}$, V$_{142}$, Q$_{150}$, F$_{232}$, L$_{312}$, P$_{396}$, L$_{397}$, A$_{448}$, K$_{452}$, L$_{454}$, I$_{478}$, K$_{502}$-S$_{504}$, I$_{512}$, D$_{530}$, V$_{533}$, Q$_{534}$, Q$_{549}$, E$_{585}$, A$_{592}$, Q$_{593}$, S$_{595}$, Q$_{596}$, and I$_{601}$, several of which are located in the SBDα and appear to interact with the SBDß binding loops through hydrogen bond and hydrophobic interactions (Table 2.6). Furthermore, for several interacting pairs (F$_{232}$-L$_{312}$, P$_{396}$-I$_{512}$, L$_{397}$-A$_{503}$, L$_{454}$-A$_{503}$, and I$_{478}$-A$_{503}$) both interacting residues are accentuated by PRS. Only residues Q150, L312 and S504 are accentuated in both configurations.

2.3.3 Mechanistic interpretation of identified PRS allosteric hot residues

Thus far results generated by PRS analysis of the molecular chaperone Hsp70 have been largely correlated with experimentally validated functional sites on the protein several of which have been implicated in allosteric modulation of the chaperone cycle. Here the results garnered from this study are summarised and interpreted in light of Hsp70's chaperone cycle (see Figure 1.6 for a schematic representation) and the relative role of the PRS accentuated residues discussed in terms of their potential for allosteric modulation of Hsp70.

- *Phase 1-2: Substrate binding and ATP hydrolysis*

Our PRS data suggests that bound ATP alone is insufficient to trigger the closing transition, this configuration recording the lowest allosteric potential out of all the open state complexes. Conversely, the addition of a client peptide at the SBD appears to increase the allosteric potential across both domains of the protein in a manner suggestive of a putative allosteric intermediate state. PRS accentuates many key functional residues in this configuration (Figure 2.15 A). Most notable are residues D_{148}, R_{151}, K_{414}, V_{440}, D_{481}, and L_{484} (Figure 2.15 A-black), which have recently been heavily implicated in a complex mechanism for peptide stimulated ATPase activation [130]. The accentuation of several known JDP interaction residues supports the theory that JDP interactions may be necessary to overcome the energetic barrier present between the open and closed states. PRS also accentuates several groups of residues that have be implicated in the coordination of the bound nucleotide (Figure 2.15 A- red); as well as the known proline switch [237] (Figure 2.15 A-yellow) and NBD hinge residues [236] (Figure 2.15 A-cyan). The accentuation of several residues known to be implicated in substrate binding provides evidence that the binding forces involved in substrate loading may be sufficient to initiate an allosteric signal at the SBD, which is propagated to the NBD via other hot residues such as those placed at inter-subdomain interfaces, culminating in ATP hydrolysis.

- *Phase 3: The closed ADP bound functional state*

PRS analysis indicates that irrespective of bound peptide substrate the presence of ADP at the NBD allosterically favours a conformational transition towards the closed state, supporting previous findings that ATPase activity and the subsequent biding of ADP drives the closing transition [239,248]. Our data also shows that ADP provides the closed form of the protein with a natural resistance to interconvert back to open state suggesting it stabilises the chaperone in a functionally closed state.

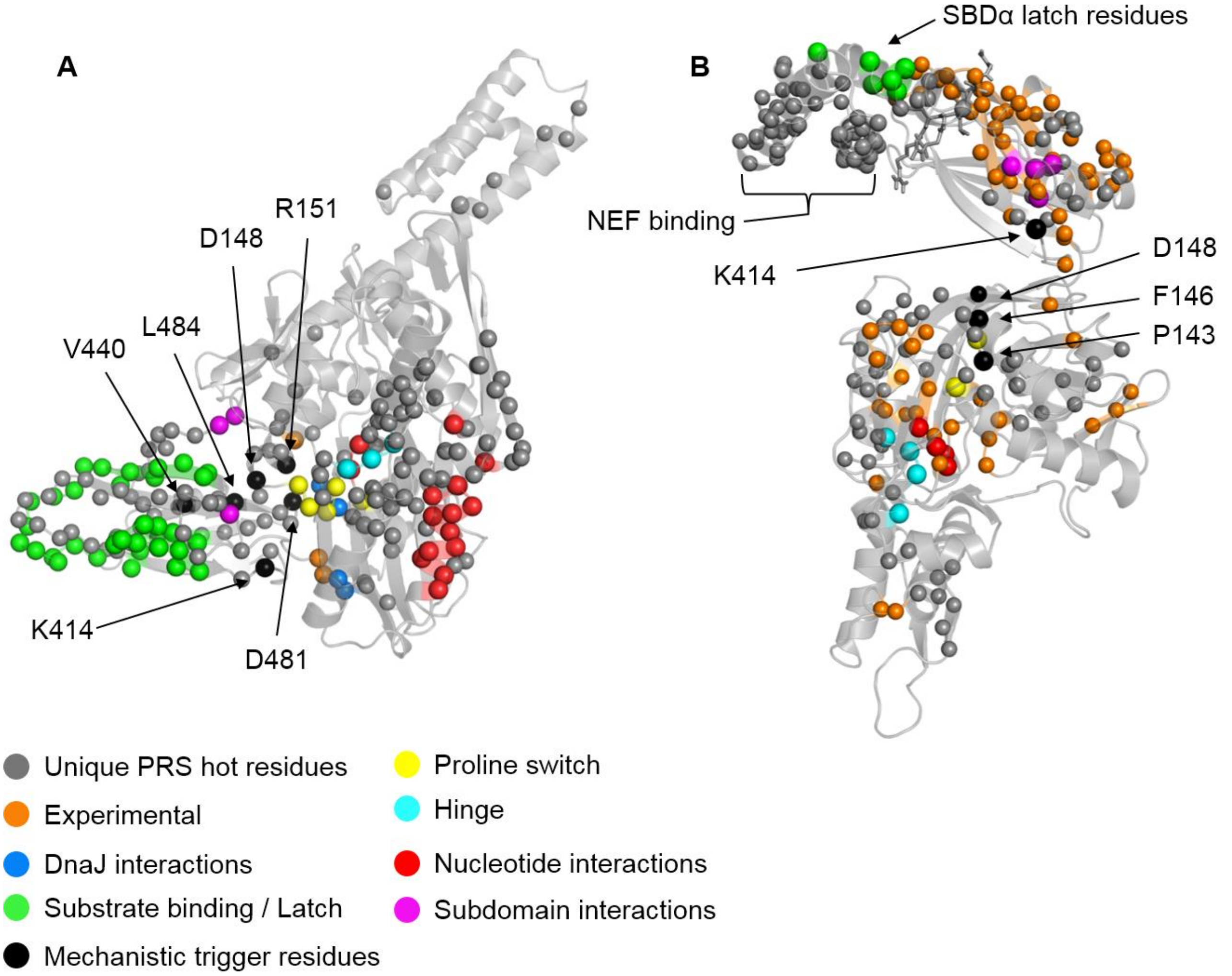


Figure 2.15: Structural representation of residues accentuated by PRS. Showing the **(A)** Closing and **(B)** Opening $NBD_{(ATP)}$-$SBD_{(pep)}$ configuration transitions. Residues unique to this study are shown as grey spheres, residues that overlap with experimentally determined functional residues are coloured according to known function (inset legend), or coloured orange. Reproduced with permission from Penkler et. al 2017 [105]

- *Phase 4: Nucleotide exchange and conformational rearrangement*

Replacing ADP with ATP at the NBD appears to allosterically activate the closed conformation complexes, drastically increasing the number of accentuated residues. Our MD simulations of the ATP and substrate bound complex indicate rearrangement of SBD with respect the NBD in line with a previously reported intermediate conformational state. PRS analysis of this complex once again accentuates several key functional residues, including the nucleotide coordinating residues, the Proline switch, and the NBD hinge residues (Figure 2.15 B). Hsp70's nucleotide exchange is heavily dependent on NEF facilitation. PRS accentuates several residues of unknown function in SBDα that have recently been implicated in NEF interactions by Melero and co-workers [127]. This observation may suggest that external force perturbations by other binding partners such as NEF may be required to drive the opening transition.

- *Phase 5: Client peptide release stabilises a conformational return to the open state*

The SBD residues were among the highest correlating residues for the ATP and peptide bound closed complex. PRS accentuates multiple residues involved in peptide binding interactions, including several SBDα residues in direct contact with loops of the SBDß, including the known latch (D_{431}, R_{467}, D_{540}, and K_{548}) and hinge residues (R_{536}-Q_{538}) [292] (Figure 2.15 B). These data suggest that conformational dynamics of the SBD together with peptide release may be important determinants for the opening transition and supports the theory that a stable open conformation can only be achieved after successful dissociation of the client substrate [101,241].

2.3.4 Implementation of the PRS algorithm in the MD-TASK software suite

Traditional MD trajectory analysis techniques are often incorporated in MD packages such as GROMACS and AMBER, however there are scenarios that depending on the research questions being asked where researchers are required to write standalone analysis scripts. This task may be challenging for some, especially if the analysis requires mathematical and complex programming knowledge. To address this particular need for the analysis of MD trajectories the Rhodes University Bioinformatics Institute (RUBi) collaboratively developed an MD trajectory analysis tools suite called MD-TASK [295] which is comprised of a set of trajectory analysis techniques involving dynamic residue networks analysis, Perturbation response scanning, and dynamic cross-correlation analyses.

The PRS methodology presented in this chapter was implemented in the Python programming environment as a stand-alone tool in the MD-TASK tools suite that can be executed from the command line interface. The tool accepts an MD trajectory in several formats (.xtc, .trr, .dcd) as input together with an expected/known end-point structure and the number of force perturbations to be tested per residue (default value set to 250). The tool begins by reducing the MD trajectory to $C_ß$ atoms ($C_α$ for glycine residues) which it then uses to calculate the variance-covariance matrix ($\boldsymbol{C}$). The expected displacement vector ($\Delta\boldsymbol{S}$, Equation 2.6) is calculated by aligning the first frame in the trajectory to the supplied end-point structure and the difference in *x, y,* and *z* coordinates for each $C_α$ calculated. Then in a sequential manner the algorithm performs the required number of force perturbations on each residue in the protein, calculating a Pearson's correlation coefficient between the overlap of the expected and predicated conformational displacement vectors. Finally the best correlation coefficient value for each residue is stored in a data frame and outputted to the user in a comma separated file (.csv) together with a line plot generated using the matplotlib module. MD-TASK has been open-sourced and is available for download from https://github.com/RUBi-ZA/MD-TASK, implemented in Python and supported on Linux/Unix.

2.4 Conclusions

In this chapter, the usefulness of PRS as a technique for the prediction of a protein allostery in large multi-domain proteins was thoroughly investigated for the 70 kDa heat shock protein using atomistic MD simulation trajectories as a kernel for predicting coordinate shifts in response to single residue perturbations. The results correlate well with experimental data for known functional sites implicated in Hsp70's nucleotide driven allosteric chaperone cycle. Based on these findings PRS coupled with all-atom MD simulations presents as a suitable and computationally efficient approach for investigating the allosteric protein of large multi-domain proteins and in is sensitive to alternate binding partner configurations such as nucleotides and polypeptide substrate.

Overall, the data presented here demonstrate that the simultaneous binding of ATP and client peptide in the Hsp70 system may engender allosterically active intermediate states for both the closing and the opening transitions. These intermediate state are highly sensitive to external force perturbations at select residue sites. In nature, these force perturbations could be linked to functional events, such as peptide binding/release, nucleotide exchange, co-factor binding as well as inter-subdomain allosteric communication. Given the significant overlap of the PRS hot residues with numerous experimental works, it stands to reason that the unique hot residues reported herein with as yet unknown function, may in fact point to functionally important sites. Those that appear on the external extremities of the protein could be involved in co-factor binding interactions such as those with JDP and NEF, while those that are buried within the core of the protein may be implicated in allosteric signal propagation across inter-subdomain interfaces. Collectively data may provide further insight into Hsp70's allosteric mechanism of action and thus present as a suitable platform for future mutational studies.

Chapter 3:

Allosteric Modulation of Human Hsp90α Conformational Dynamics

This chapter describes a dynamics based computational study in which Perturbation response scanning is coupled with dynamic residue interaction network analysis and long-range molecular dynamics simulations for the quantification and identification of allosteric control elements in the 90 kDa heat shock protein.

Chapter 3 is reproduced in part with permission from the following publication:

Allosteric Modulation of Human Hsp90α Conformational Dynamics. David L. Penkler, Canan Atilgan, and Özlem Tastan Bishop. *Journal of Chemical Information and Modeling.* 2018. 58 (2), 383-404 DOI: 10.1021/acs.jcim.7b00630. 2018. American Chemical Society.

Contribution: Conceived the study together with Özlem Tastan Bishop, experimental design, carried out all experimental procedures, and wrote the first draft of the manuscript.

3.1 Introduction

3.1.1 Computational efforts to understanding allosteric regulation in Hsp90

To date, a number of computational studies aimed at elucidating the mechanistic regulation of Hsp90's conformational dynamics have been carried out based largely on the available experimental structural data. Atomic resolution models of allosteric regulation in Hsp90 have been proposed, in which reported cross-talk between the NTD and CTD implicate ATP binding and hydrolysis at the NTD to have a strong influence on conformational dynamics at the CTD [296,297]. These studies lead to the prediction of allosteric inhibitor ligands that bind a CTD allosteric target site [298–300]. Coarse grained and atomistic MD simulations coupled with energy landscape analysis were used to demonstrate the global motions and functional dynamics of Hsp90, reporting a network of conserved regions within the protein that may be important for regulating intra-protomer communications in response to ATP hydrolysis [301]. A comprehensive all atom simulation study of several Hsp90 crystal structures from different organisms demonstrated how the functional dynamics of Hsp90 are modulated by the motion of quasi-rigid domains involving two sets of conserved residues that form inter-domain hinges between the NTD-MiD and MiD-CTD [302]. Atomistic simulations coupled with force distribution analysis of the HtpG structure revealed allosteric communication pathways connecting a similar MiD hinge and putative client binding site to the NTD nucleotide binding site [303]. Thermodynamic analyses based on MD simulations were used to describe structural rearrangements involved in Hsp90's conformational cycle, leading to the identification of several key interactions that may govern conformational change [304]. Structural stability analysis and protein network modelling based on atomistic simulations of Hsp90 were used to characterize the evolution of interaction networks and allosteric communication during interstate transitions, reporting a small number of conserved functional residues that may be central modulators of several functions including allosteric communication, structural stability, and co-chaperone binding [305]. More recently a comprehensive structural and dynamic network analysis of posttranslational modification (PTM) sites across Hsp90 proteins was conducted, in which the specific role of regulatory PTMs hotspots in the allosteric mechanism of Hsp90's conformational cycle were identified [306].

3.1.2 Allosteric regulation of Hsp90 function through co-factor interactions

With the elucidation and identification of Hsp90's numerous co-chaperones and binding partners, it has become evident that Hsp90's interaction with a broad range of co-chaperones tailor the flexibility of the dimer to suite its functional needs. In this manner, evidence has been shown directly implicating co-chaperones with the modulation of Hsp90's ATPase activity by controlling the conformational transitions involved in client loading and release. Functional dynamics investigations of Hsp90’s interactions with the co-chaperones Aha1 and p23 demonstrated that binding of these cofactors allosericallymodulate Hsp90’s conformational dynamics [307]. Furthermore, analysis of the interaction networks involving Hsp90 complexes with the cofactors Cdc37, Sgt1, Rar1 [308], and P53 [309] demonstrated how small-world networks involving highly connected residues at the inter-protein interfaces correspond to known functional sites.

3.1.3 Allosteric regulation of Hsp90 function via small molecule ligands

Traditionally, therapeutic efforts for the inhibition of Hsp90 have largely focused on competitively targeting the nucleotide binding site. However, with the implication of co-chaperone interactions for client based function, coupled with a growing body of evidence of allosteric modulation of conformational dynamics, drug discovery efforts are moving towards the targeting of specific co-chaperone binding sites, or allosteric pockets, as an alternative avenue for the inhibition of Hsp90 by disrupting its conformational dynamics and thus the loading and release of selective clients. Interestingly, the associated allosteric regulation of Hsp90 by co-chaperone interactions is believed to only occur for the cytosolic isoform of the protein [159,310]. This compartmentalization would further maximize the selectivity for potential inhibitors. Progress along these avenues for the treatment of human related diseases has been limited however, due to a lack of structural data for human Hsp90α which in turn has limited the discovery of druggable allosteric sites. Indeed, Hsp90’s dynamic nature raises the possibility that allostery in the chaperone may be conformation dependent. The structure of yeast Hsp82 has been well characterized for the closed ATPase active conformation, and as a result most structure based allosteric investigations of Hsp90 have only been applied to this functional state.

3.1.4 Chapter objectives

In this study, the lack of structural data for Hsp90 in conformations other than the closed catalytically active state is addressed, by employing a multi-template homology modelling technique to calculate accurate full-length structures of human Hsp90α in the three functionally relevant conformations: the fully-closed ATPase active state, the partially-open/closed intermediate state, and the fully-open “v-like” inactive conformation. By subjecting these structures to long-range all-atom MD simulations in the presence/absence of nucleotide at the NTD, we look to quantify the effect of nucleotide on protein dynamics using several MD based analysis techniques. Finally we apply PRS to every possible configuration to assess the allosteric potential of the protein under the different nucleotide conditions and identify select sites capable of allosterically modulating the dynamic conformational ensemble of the protein in favour of the opposite conformational state. i.e. inactive (open) → active (closed) and *vice versa*.

3.2 Methodology

3.2.1 Homology modelling

The theory of homology modelling is built on the concept that protein function is conserved through tertiary structure (Illergård et al. 2009), thus allowing for the structural determination of an unresolved protein sequence form a closely related homolog. Homology modelling is thus four-step process including template identification, target-template alignment, model building, and validation. Given that protein structure is inferred on sequence homology alone, the identification of the template structures and the alignment to the target sequence is of utmost imporance as the down stream model build is entirely dependent on the accuracy thereof. Another important aspect to consider is the quality of the template structure, as any incongruencies in the resolution of the template i.e. missing residues or atoms, will be extraoplated into the target sequence model. The model building process can be largely automated using various online web-serves, however a manual approach to model build with software such as the MODELLER package developed by Sali and Blundell affords the researcher the ability to fine tune the alignement file and refinement levels during model building and this increase the overall accuracy of the final model.

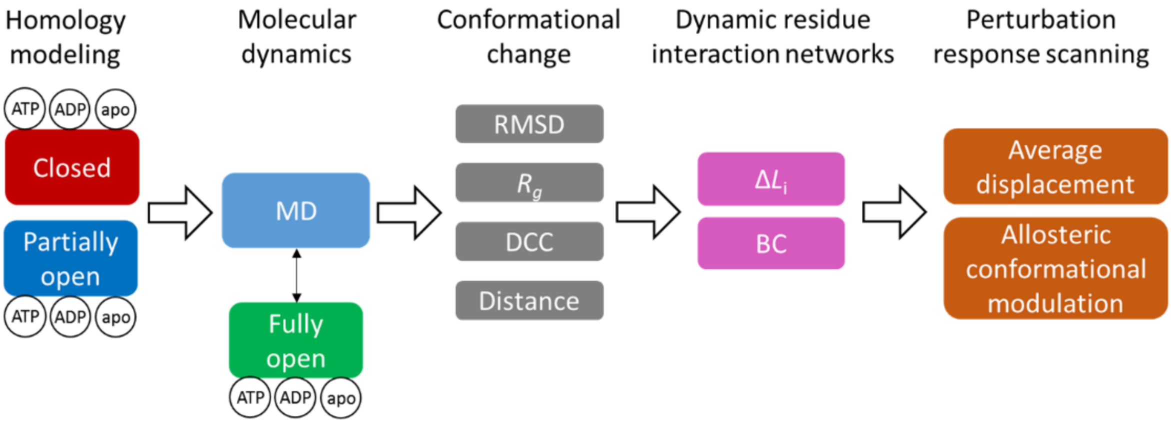
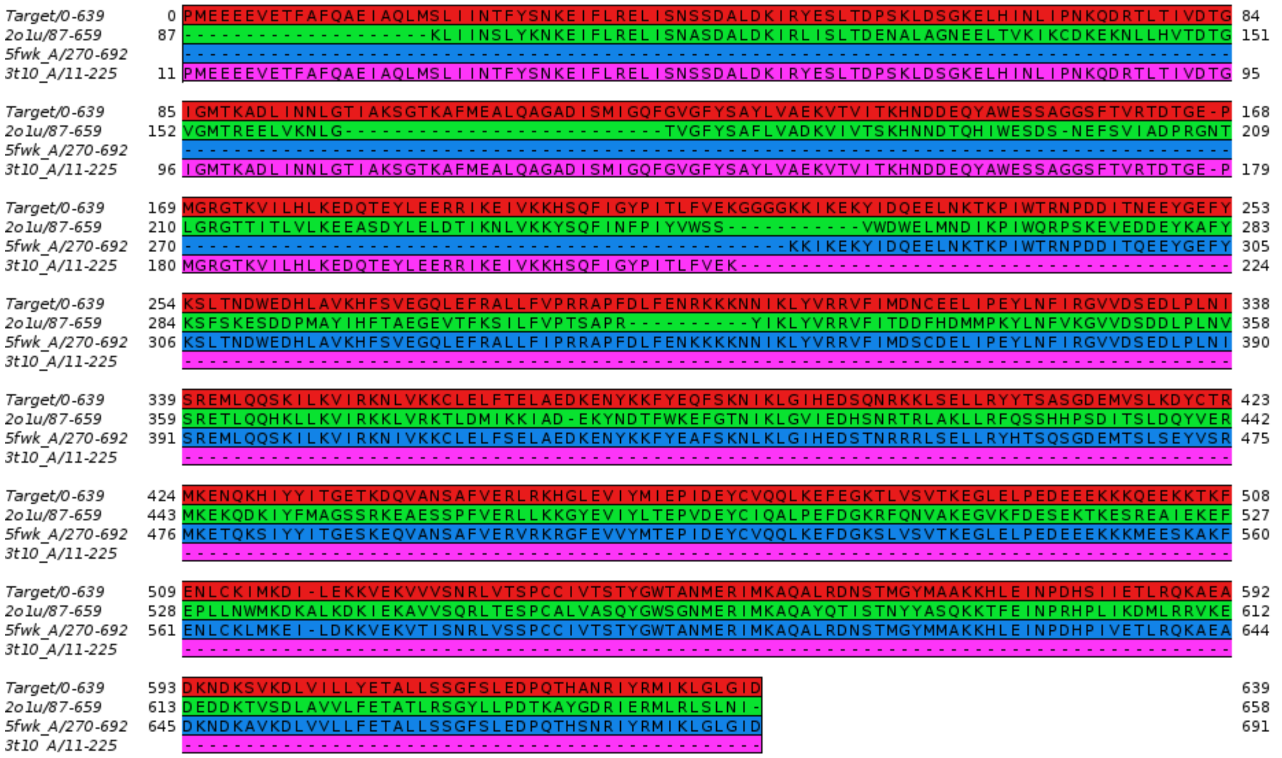


Figure 3.1: Schematic overview of experimental approach: Homology modelling was employed to obtain full-length structures of human Hsp90α in the closed and partially-open conformations. Both structures were submitted to long range MD simulations. The fully-open structure which was obtained from the trajectory of the partially-open structure in complex with ADP was also submitted MD simulation to total nine unique trajectories with different nucleotide configurations. Each trajectory was submitted to downstream analysis to assess the effect of bound nucleotide on conformational change and allosteric signal propagation. Reproduced with permission from Penkler et. al 2018 [191].

MODELLER is implemented within the Python programming environment. The underlying modelling algorithm calculates 3D protein models using the satisfaction of spatial restraints, a technique well known in NMR. Spatial restraints are expressed as a probability density function for each protein sub-structure. Models are built to satisfy these spatial restraints based on the alignment of target to template sequences [311]. The required input for MODELLER, includes the alignment file containing the target-template alignment, as well as the original PDB file of the template. To ensure that model building returns an optimum protein model, MODELLER utilizes an Lamarckian iterative process whereby several models are calculated, and models that are not statistically significant discarded and the remaining models sampled [312]. Given that loop regions are highly variable and thus very difficult to predict and model, MODELLER includes some elementary *ab initio* structure prediction, which greatly improves the modelling of these regions.

In this study the 3D structure of cytosolic human Hsp90 (Hsp90α) was modeled in a partially-open and fully-closed conformation using the full protein sequence NP_005339.3, obtained from the NCBI database, as the target sequence. Alignments were carried out using PROMALS3D [313] and modelling calculations performed using MODELLER [314]. For the closed conformation, the full-length cryo-EM structure of human Hsp90β in complex with ATP (PDB ID: 5FWK) [192] was used as a structural template. 5FWK represents the full protein structure except for the linker between the NTD and MD (res 228-277), which due to its inherent flexibility is yet to be crystallized. For this reason, residues 230-275 were deleted from the target sequence, and replaced with four glycine residues (G-G-G-G) as a structural placeholder.

Modelling of the open conformation required the use of a multi-template homology modelling approach to ensure a “v-like” open dimer, with an extended orientation for each protomer. The canine endoplasmic reticulum paralog of Hsp90, GRP94, demonstrates a partially-open ‘v-like’ conformation with dimerization at the CTD (PDB ID: 2O1U) [188]. This structure was used as the primary template for MODELLER to ensure model building followed this structural arrangement. Figure 3.2 illustrates the final alignment used to construct the partially-open structure showing 2O1U in green. Given that 2O1U and the target sequence only share 53% sequence identity, 5FWK (blue) was used as a secondary template, to provide cover for the missing region L290-R299. Furthermore, 2O1U lacks structural data for the ATP-lid domain (res 98-123), while 5FWK is representative of the structure with a closed ATP-lid. Thus, to ensure that the ATP-lid was modeled in its open conformation, the NTD (res 15-278) of 5FWK were excluded from the alignment file and the open ATP-lid NTD crystal structure of human Hsp90 (PDB 3T10) was used as a second template for this domain (magenta).



```
Target/0-639      0 PMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNSSDALDKIRYESLTDPSKLDSGKELHINLIPNKQDRTLTIVDTG  84
2o1u/87-659      87 --------------------KLIINSLYKNKEIFLRELISNASDALDKIRLISLTDENALAGNEELTVKIKCDKEKNLLHVTDTG 151
5fwk_A/270-692      -------------------------------------------------------------------------------------
3t10_A/11-225    11 PMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNSSDALDKIRYESLTDPSKLDSGKELHINLIPNKQDRTLTIVDTG  95

Target/0-639     85 IGMTKADLINNLGTIAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHNDDEQYAWESSAGGSFTVRTDTGE-P 168
2o1u/87-659     152 VGMTREELVKNLG--------------------------TVGFYSAFLVADKVIVTSKHNNDTQHIWESDS-NEFSVIADPRGNT 209
5fwk_A/270-692      -------------------------------------------------------------------------------------
3t10_A/11-225    96 IGMTKADLINNLGTIAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHNDDEQYAWESSAGGSFTVRTDTGE-P 179

Target/0-639    169 MGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQFIGYPITLFVEKGGGGKKIKEKYIDQEELNKTKPIWTRNPDDITNEEYGEFY 253
2o1u/87-659     210 LGRGTTITLVLKEEASDYLELDTIKNLVKKYSQFINFPIYVWSS-----------VWDWELMNDIKPIWQRPSKEVEDDEYKAFY 283
5fwk_A/270-692  270 -------------------------------------------------KKIKEKYIDQEELNKTKPIWTRNPDDITQEEYGEFY 305
3t10_A/11-225   180 MGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQFIGYPITLFVEK---------------------------------------- 224

Target/0-639    254 KSLTNDWEDHLAVKHFSVEGQLEFRALLFVPRRAPFDLFENRKKKNNIKLYVRRVFIMDNCEELIPEYLNFIRGVVDSEDLPLNI 338
2o1u/87-659     284 KSFSKESDDPMAYIHFTAEGEVTFKSILFVPTSAPR----------YIKLYVRRVFITDDFHDMMPKYLNFVKGVVDSDDLPLNV 358
5fwk_A/270-692  306 KSLTNDWEDHLAVKHFSVEGQLEFRALLFIPRRAPFDLFENKKKKNNIKLYVRRVFIMDSCDELIPEYLNFIRGVVDSEDLPLNI 390
3t10_A/11-225       -------------------------------------------------------------------------------------

Target/0-639    339 SREMLQQSKILKVIRKNLVKKCLELFTELAEDKENYKKFYEQFSKNIKLGIHEDSQNRKKLSELLRYYTSASGDEMVSLKDYCTR 423
2o1u/87-659     359 SRETLQQHKLLKVIRKKLVRKTLDMIKKIAD-EKYNDTFWKEFGTNIKLGVIEDHSNRTRLAKLLRFQSSHHPSDITSLDQYVER 442
5fwk_A/270-692  391 SREMLQQSKILKVIRKNIVKKCLELFSELAEDKENYKKFYEAFSKNLKLGIHEDSTNRRRLSELLRYHTSQSGDEMTSLSEYVSR 475
3t10_A/11-225       -------------------------------------------------------------------------------------

Target/0-639    424 MKENQKHIYYITGETKDQVANSAFVERLRKHGLEVIYMIEPIDEYCVQQLKEFEGKTLVSVTKEGLELPEDEEEKKKQEEKKTKF 508
2o1u/87-659     443 MKEKQDKIYFMAGSSRKEAESSPFVERLLKKGYEVIYLTEPVDEYCIQALPEFDGKRFQNVAKEGVKFDESEKTKESREAIEKEF 527
5fwk_A/270-692  476 MKETQKSIYYITGESKEQVANSAFVERVRKRGFEVVYMTEPIDEYCVQQLKEFDGKSLVSVTKEGLELPEDEEEKKKMEESKAKF 560
3t10_A/11-225       -------------------------------------------------------------------------------------

Target/0-639    509 ENLCKIMKDI-LEKKVEKVVVSNRLVTSPCCIVTSTYGWTANMERIMKAQALRDNSTMGYMAAKKHLEINPDHSIIETLRQKAEA 592
2o1u/87-659     528 EPLLNWMKDKALKDKIEKAVVSQRLTESPCALVASQYGWSGNMERIMKAQAYQTISTNYYASQKKTFEINPRHPLIKDMLRRVKE 612
5fwk_A/270-692  561 ENLCKLMKEI-LDKKVEKVTISNRLVSSPCCIVTSTYGWTANMERIMKAQALRDNSTMGYMMAKKHLEINPDHPIVETLRQKAEA 644
3t10_A/11-225       -------------------------------------------------------------------------------------

Target/0-639    593 DKNDKSVKDLVILLYETALLSSGFSLEDPQTHANRIYRMIKLGLGID                                       639
2o1u/87-659     613 DEDDKTVSDLAVVLFETATLRSGYLLPDTKAYGDRIERMLRLSLNI-                                       658
5fwk_A/270-692  645 DKNDKAVKDLVVLLFETALLSSGFSLEDPQTHSNRIYRMIKLGLGID                                       691
3t10_A/11-225       -----------------------------------------------
```

Figure 3.2: Graphical illustration of the final homology modelling alignment file for the partially-open conformation. The target sequence is coloured red and the template structure sequences coloured: 2O1U green, 5FWK blue, and 3T10 magenta.

A total of 100 models were calculated for each dimerized conformation (open and closed), and the models with the lowest normalized DOPE score (DOPE-Z score) selected for further evaluation by VERIFY3D [315], Errat [316] and ProSA [317]. To obtain each of these conformations in complex with ATP and ADP, Pymol was used to extrapolated the respective nucleotide's structural coordinates from representative crystal structures by superimposing PDBs 3T0Z [318] and 1BYQ [319] for ATP and ADP respectively. The apo NTD was obtained by omitting nucleotide coordinate data. In this manner, a total of six human Hsp90α configurations were prepared, three in a partially-open conformation and three is a fully-closed conformation.

3.2.2 Molecular dynamics simulations

MD simulations were carried out on all nucleotide bound configurations of the homology models. All MD simulations were performed using GROMACS 5.1.2 [265] within the CHARMM 36 force field [268–270], using a orthorhombic periodic box with a clearance space of 1.5 nm. Water molecules were added as solvent, and modeled with the TIP3P water model. The system was neutralized using a 0.15 M NaCl concentration. Prior to production runs, all molecular systems were first energy minimized using a conjugate-gradient being energy relaxed with up to 50 000 steps of steepest-descent energy minimization, and terminated when the maximum force < 1000.0 kJ/mol/nm. Energy minimization was followed by equilibration, first in the NVT ensemble at 310 K using Berendsen temperature coupling, and then in the NPT ensemble at 1 atm and 310 K until the desired average pressure (1 atm) was maintained and volumetric fluctuations stabilized. All production simulations were run for a minimum of 100 ns and a maximum of 200 ns, to ensure that the backbone root-mean-square deviation (RMSD) of the protein equilibrated with a fluctuation of no more than 3 Å for at least 20 ns. Data pertaining to the protein and nucleotide were saved at 2 ps intervals for further analysis. All simulations utilized the LINCS algorithm for bond length constraints. The fast particle mesh Ewald method was used for long-range electrostatic forces, the switching function for van der Waals interactions was set to 1.0 nm and the cut-off set to 1.2 nm. $NH3^+$ and COO^- groups were used to achieve the zwitterionic form of the protein. Periodic boundary conditions were applied in all directions and the simulation time step was set to 2 fs. Duplicate simulations were carried out for all closed and FO complexes.

3.2.3 Dynamic cross-correlations

Dynamic cross-correlations (DCC) describe the correlation of motions between each C_α atom in a protein over the course of an MD trajectory. DCC is calculated based on the reduction and normalization of the 3N x 3N covariance matrix (**C**). The MD-TASK [295] software suite was used to calculate DCC for each MD trajectory. The kernel **C** was constructed from the MD trajectory with a sampling step of 100 ps intervals, and the average deviation of each C_α atom from a mean structure representative of the trajectory length calculated. The essential directions of correlated motions were determined by diagonalizing **C** to get the N x N correlation matrix (Corr):

$$\mathbf{Corr}_{ij} = \frac{\langle \Delta \boldsymbol{r}_i \cdot \Delta \boldsymbol{r}_j \rangle}{\sqrt{\langle \Delta \boldsymbol{r}_i^2 \rangle} \cdot \sqrt{\langle \Delta \boldsymbol{r}_i^2 \rangle}}$$

3.2.4 Dynamic residue network analysis

To analyse inter- and intra-domain communication, the protein is represented as a residue interaction network (RIN), where the C_β atoms of each residue (C_α for glycine) are treated as nodes within the network, and edges between nodes defined within a distance cut off of 6.7 Å [46]. In this manner, the RIN was constructed as a symmetric N × N matrix, where the ijth element is assigned as 1 if residue i is connected to residue j and a zero if no connection exists.

In this study, MD-TASK [295] was used to construct dynamic residue interaction networks (DRN) for each MD trajectory, in which RINs are constructed for every nth frame of the trajectory using a 200 ps time interval, to build a DRN matrix. By iterating over the DRN, each RIN is analysed in terms of the average of shortest path length (L_i) and betweenness centrality (BC) of each residue.

- *Shortest path length (L_i)*

The shortest path length (L_{ij}) between two residues *i* and j is defined as the minimum number of nodes that need to be crossed to reach *j* from *i*. The average L_i is then calculated as the average number of steps that the node/residue may be reached from all other residues in the in the network:

$$\boldsymbol{L_i = \frac{1}{N-1}\sum_{j=1}^{N} L_{ij}}$$

Here, we analyse the change in reachability (ΔL_i) of each residue by monitoring how L_i shifts over the course of the MD trajectory and thus over time:

$$\boldsymbol{\Delta L_i = \frac{1}{N}\sum_{m=1}^{N} \left(L_i^0 - L_i^n\right)}$$

where L_i^0 denotes the average shortest path length for residue i at time zero, and L_i^n the average shortest path length at frame n.

- *Betweenness Centrality (BC)*

BC is defined as the number of shortest paths running through a node/residue for a given RIN, and provides a measure of usage frequency each node during navigation of the network. Here, BC was calculated using MD-TASK based on Dijkstra's algorithm [320] and the data rescaled in the range of 0.0 – 1.0 to be comparable between conformers. Finally, the average BC and ΔL_i are calculated over the DRN, as this measure provides an indication of residues that experience permanent changes in ΔL_i and BC as opposed to minor fluctuations over the course of the MD trajectory [295].

3.2.5 Perturbation response scanning

The theory of PRS has been previously discussed in sections 2.1.2 2.2.3 and we follow the same methodology in this study and again construct the Hessian matrix from the MD trajectory data by selecting 20 ns stable segments based on the RMSD fluctuation profiles. Each residue in the protein was perturbed 250 times using random force perturbations of a single unit measure.

- *Effector and sensor analysis*

Effector residues are define as those residues whose perturbation causes large displacements elsewhere in the protein. Sensor residues are residues that are sensitive to perturbations elsewhere in the protein. Both data sets are constructed from the response vectors obtained after 250 force perturbations of each residue in the protein. Each response vector for each perturbation for each residue was stored in a 3*N* x 3*N* x 250 matrix where *N* is the total number of residues in the protein. An average displacement matrix is obtained by averaging each matrix coordinate over the 250 perturbations resulting in a 3*N* x 3*N* response matrix, where effector residues are detected on the vertical axis and sensor residues on the horizontal axis. Thus for any given residue *i,* the average effect 250 perturbations at *i* can be assessed by examining the matrix row corresponding to *i.* Likewise sensor residues can be assessed by considering the column corresponding to residue *i* on the horizontal axis. Python scripts using the numpy module were used to perform all matrix operations.

- *Response correlation analysis*

For each PRS experiment, the initial and final structures (states 0 and 1 in Equation 2.1) were defined as follows: The coordinates of state 0 were set to the initial frame of the equilibrated MD segment, while the coordinates for state 1 were set to the initial equilibrated frame from the opposite conformation. Thus for the open to closed transitions, state 1 was set to the equilibrated structure of the ATP bound closed configuration. Conversely, for the closed to open transitions, the final structure was set to the equilibrated ATP bound open conformation. The coordinates of state 1 were superimposed on state 0 using Kabsch algorithm [272] in a whole protein alignment approach and the experimental difference vector calculated (ΔS). Each residue in the initial state was sequentially perturbed with 250 fictitious forces of random direction and a single unit magnitude. The quality of the response vectors were assessed using Pearson's correlation coefficient with (ΔS) see section 2.2.3 for more details.

3.3 Results and Discussion

3.3.1 Homology modelling

Structures of human Hsp90α in closed and partially open 'v-like' conformations were obtained by homology modelling (Figure 3.3). In each case 100 unique homology models were calculated and ranked by normalized DOPE-Z score. The normalized DOPE-Z score is an atomic distance-displacement statistical potential used to evaluate the fold of a model and is dependent on a sample structure of a native protein [321,322]. A value less than -0.5 is deemed an accurate representation of the native structure, and the lowest scoring models were selected for further model evaluation and verification. Table 3.1 summarizes the results of the best scoring models after separate model evaluation with the Verify3D, Errat, and ProSA techniques.

Table 3.1 | Summary of model evaluation before and after energy minimization

Conformational state		DOPE Z-score	Verify3D	Errat	ProSA Z-score
Un-minimized	Closed	-1.24	93.55 %	92.38 %	-10
	Open	-1.13	94.10 %	95.28 %	-11
Minimized	Closed	-1.82	95.22 %	94.71 %	-11
	Open	-1.74	96.13 %	96.86 %	-11

Verify3D is an online web-server that determines the relative compatibility of a 3D atomic coordinate structure with its own 1-Dimensional amino acid sequence, assigning structural classes based on secondary structure, location, and local environment, comparing the results with known good structures [323,324]. The percentage of residues within the protein that score more than 0.2 is reported, and a minimum of pass rate of 80% is required to pass model evaluation. Errat analyses the statistics of all non-bonded interactions between carbon, nitrogen, and oxygen atoms reporting the percentage of residues that fall below a rejection confidence level of 95% [316]. The ProSA web-server calculates the overall quality of a protein structure based on a Z-score which indicates the goodness of fit between the modeled structure and native structures of a similar size [317].

Both, the selected open and closed homology models pass each of these evaluation tests, recording scores in the good to very good range (Table 3.1). Appendix I-I shows a schematic illustration of the secondary structure qualities. It is important to note that model evaluation was carried out before and after energy minimization, to assess the impact of possible bad contacts within the structure. Erroneous regions highlighted in the pre-minimized structures were mostly confined to loop regions, particularly the missing linker region between the NTD and the MiD, all of which improved post energy minimization (Table 3.1). Given that the final models were to be used as starting structures for long range MD simulations and that they were first energy minimized, it was not necessary to carry out a loop refinement strategy as this would be handled through all-atom simulations.

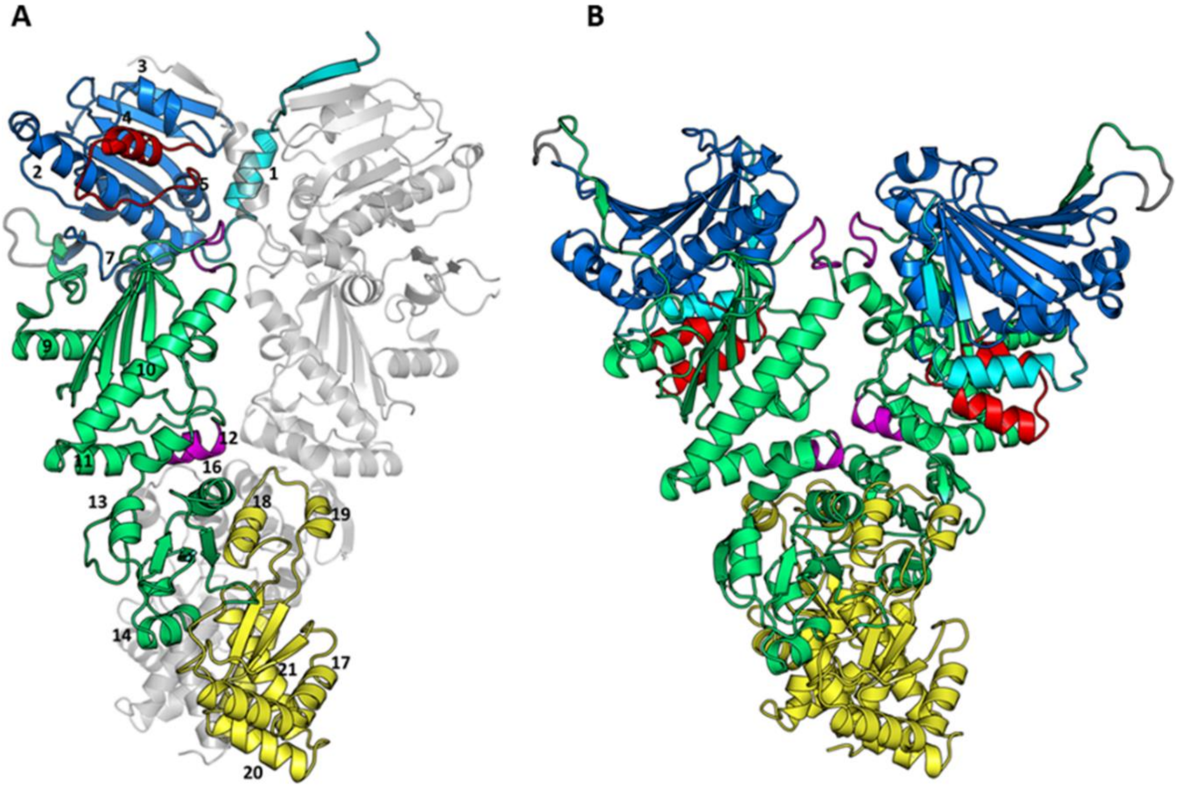


Figure 3.3: Structural representation of the human Hsp90α homology models. (A) Fully closed conformation, **(B)** partially open conformation. Indicating the locations for the NTD (blue); charged linker (grey); M-domain (green); CTD (yellow); ß-strand$_1$-helix$_1$ interface (cyan); ATP-lid (red); and M-domain/CTD hinge (magenta). Helices are labelled according to the numbering convention used throughout the text. Reproduced with permission from Penkler et. al 2018 [191]

3.3.2 Effect of nucleotide binding on conformational dynamics

The NTD nucleotide bound configuration (ATP/ADP/apo) of Hsp90 has been implicated with the modulation of conformational dynamics that underpin the chaperone's functional cycle (Figure 1.8). To determine the relative effect of bound nucleotide on the conformational dynamics (opening and closing transitions) of human Hsp90, all-atom MD simulations in explicit water were carried out on nucleotide bound (ATP/ADP) and apo fully-closed (FC), partially-open (PO), and fully-open (FO) complex configurations, in a total of nine unique simulation runs (Table 3.2). Duplicate MD simulations were carried out for the closed and FO complexes to confirm MD convergence totalling 1.8 µs of simulation data for 12 separate MD runs (Table 3.2). The FC and PO simulations were initiated based on the energy minimized homology models, while the FO simulations were based on a structure extracted from the PO MD trajectory in complex with ADP (PO-ADP). The resultant trajectories were analysed in terms of their backbone RMSD, residue root-mean-square-fluctuations (RMSF), radius of gyration (R_g), inter-protomer distance, dynamic cross-correlation (DCC) of atomic motions, and conformational restructuring.

Table 3.2 | Summary of MD simulations and equilibrated PRS segments.

	Nucleotide	Label	Simulation length (ns)	Equilibrated region (ns)	ΔRMSD (Å)	Replicate
FC	ADP	FC-ADP	100	80-100	2.0	✓
	ATP	FC-ATP	100	80-100	2.0	✓
	apo	FC-apo	100	80-100	2.0	✓
PO	ADP	PO-ADP	200	-	-	✕
	ATP	PO-ATP	200	-	-	✕
	apo	PO-apo	200	-	-	✕
FO	ADP	FO-ADP	200	65-85	2.0	✓
	ATP	FO-ATP	200	180-200	2.3	✓
	apo	FO-apo	200	80-100	3.0	✓

FC = Fully-closed; PO = Partially-open; FO = Fully-open

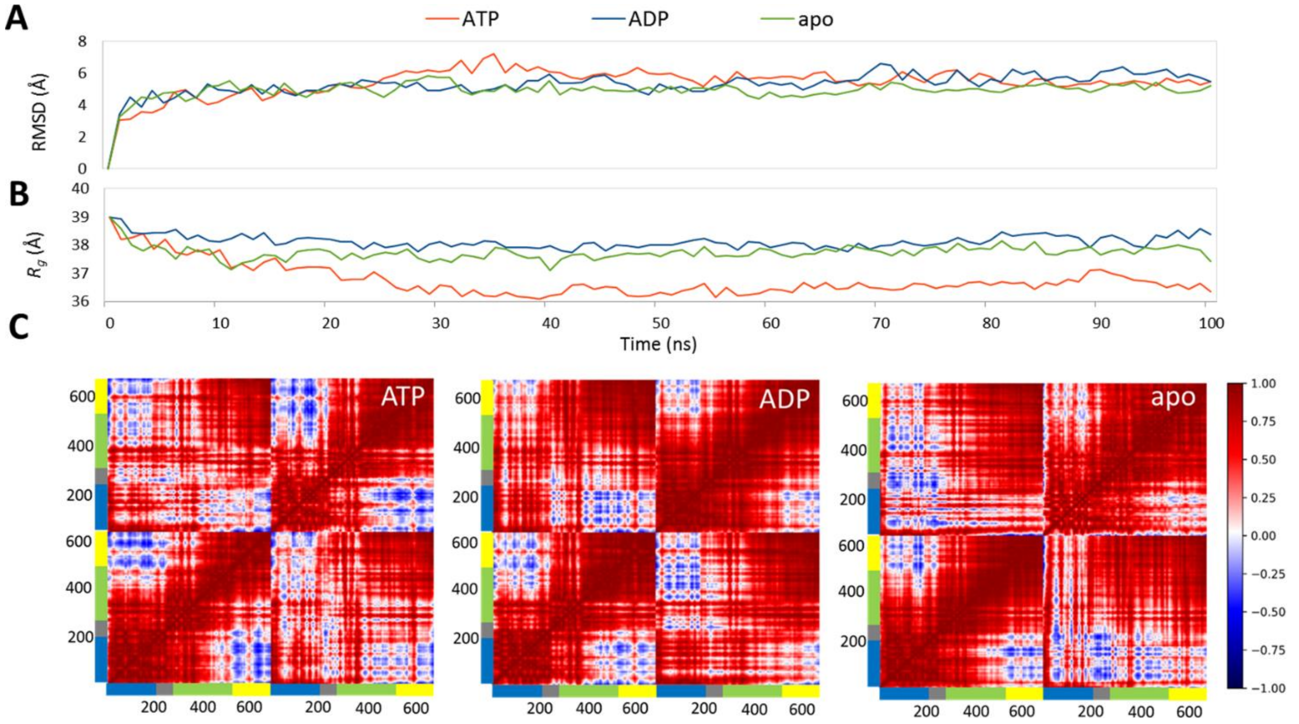


Figure 3.4: Global conformational dynamics metrics for the closed conformation complexes. (A) backbone RMSD; **(B)** radius of gyration; **(C)** DCC heat maps showing correlated motions (positive) in red and anti-correlated motions (negative) in blue. The colour bar accompanying each heat map represents each functional domain: NTD (blue), linker (grey), M-domain (green), and CTD (yellow). Reproduced with permission from Penkler et. al 2018 [191]

- *Fully-closed complexes*

Backbone RMSD analysis of the closed conformation complexes indicates fairly stable structures over 100 ns, with RMSD values converging around 5-6 Å (Figure 3.4-A). RMSF analysis for these trajectories (Appendix I-II) revealed prominent NTD fluctuations surrounding residues belonging to the loop R_{60}-E_{75}, the ATP-lid (res 112-136), and the charged linker (res 220-282), with subtle differences between the two protomers. The only notable difference between the different nucleotide configurations are slightly larger values for the ATP complex. This observation is in agreement with a previous MD study on yeast Hsp90 [297].

Despite RMSD convergence and similar RMSF values between the three configurations, visual inspection of the final structures (100 ns) reveals subtle differences in the NTDs of each complex when compared to their initial structures (0 ns). The NTD of the ATP complex becomes more compact compared to the NTDs of ADP and apo complexes (Figure 3.5). Quantifying the radius of gyration (R_g) over the MD trajectory provides a metric for evaluating the degree of compactness of a protein of interest [325]. R_g analysis for the closed complexes shows an initial decrease for the ATP complex which eventually stabilizes around 36.5 Å (Figure 3.4-B), suggesting a tightening of the dimers around the central axis. The ADP and apo complexes show a smaller initial decrease compared to the ATP complex, both stabilizing around 38.5 Å, suggesting these complexes to be more 'relaxed' around the central axis.

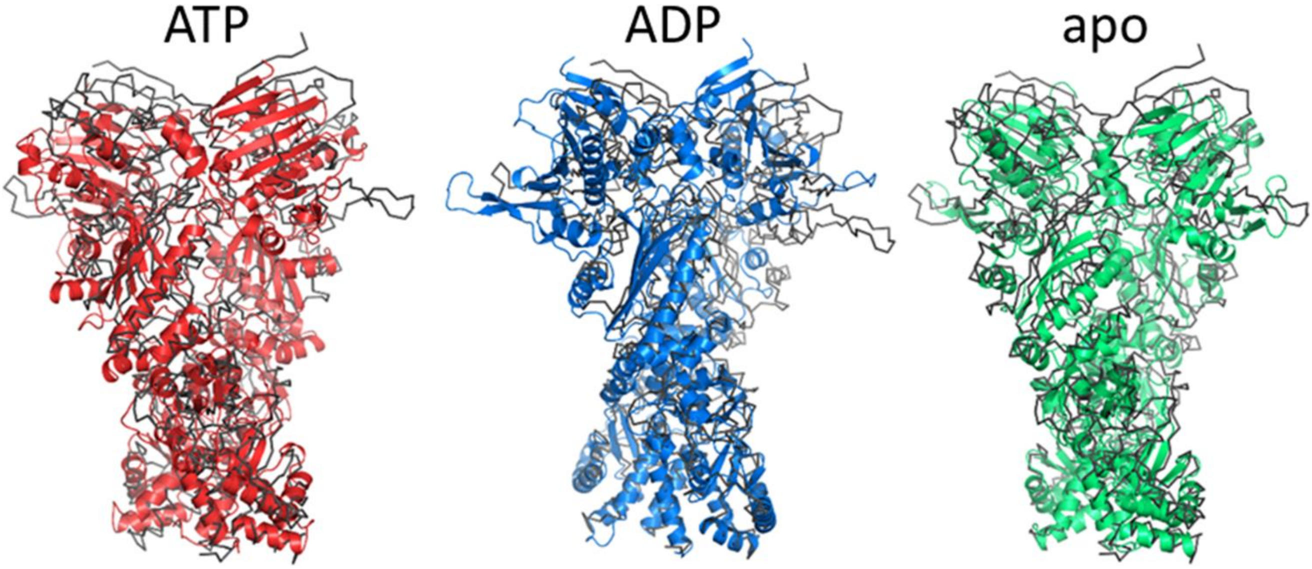


Figure 3.5: Structural representation of the final closed conformation structures after 100 ns MD simulations. Showing the ATP (red); ADP (blue); and apo (green) complexes as cartoons superimposed on the original start structure in grey ribbons. Reproduced with permission from Penkler et. al 2018 [191]

The apparent compactness of the ATP complex is also confirmed when analysing the MD trajectories in terms of correlated $C_ß$ atom motions. In this analysis, each correlation matrix (Figure 3.5-C) represents the correlation between any two $C_ß$ atoms over the course of an MD trajectory. Positive values approaching 1 describe correlated motions (in the same direction) and negative values close to -1 describe anti-correlated motions (in opposite directions). In the ATP complex, strong anti-correlation is observed between the NTD and CTD within and between individual protomers, suggesting these domains to have anti-correlated motions either towards ($\overset{\downarrow}{\uparrow}$) or away ($\overset{\uparrow}{\downarrow}$) from each other in the presence of ATP. The degree of anti-correlation between these domains is reduced in the ADP and apo complexes, especially in protomer 2, supporting the hypothesis that these complexes resemble more relaxed structures. Looking at the motions between the NTDs of each protomer, positive correlated motions are observed between the NTDs of the nucleotide bound complexes, suggesting that the NTDs move in a concerted fashion relative to one another. Strikingly, the converse is observed for the apo structure, in which the NTDs show anti-correlated motions relative to one another suggesting movement in opposite directions. This evidence may be indicative of protomer uncoupling at the NTD which precedes transition towards the open state. We note that these observations are unlikely to be attributed to a lack of MD convergence, as replicate DCC heat maps computed for separate trajectories demonstrated highly reproducible results (Appendix I-III). Furthermore, we suggest that the differential behaviour observed for the two protomers may be evidence of previously reported asymmetric conformational dynamics of Hsp90 [160,326], which is a trademark characteristic that is also observed in the DRN and PRS analyses (see sections 3.3.3 & 3.3.5).

– *Partially-open complexes*

The PO complexes of human Hsp90α were calculated based primarily on the canine Grp94 template (PDB ID: 2O1U), a structure that is representative of a partially open intermediate structure that may represent the native conformation of the protein in the absence of bound client proteins or co-factors [188]. The atomic coordinates for the ATP-lid are missing from 2O1U and the ATP-lid was modelled in the open configuration (Figure 3.6), to yield a putative intermediate structure between the fully-closed catalytically active state and the native inactive fully-open state.

The backbone RMSD plots of the PO complexes (Figure 3.7-A) indicate minor conformational restructuring within the first 35 ns, after which the ATP and apo complexes stabilize converging to a value around 5 Å, while the ADP complex undergoes extensive conformational fluctuations and is unable to equilibrate over the 200 ns simulation. Superimposing the CTDs of the final structures with the initial PO structure shows that each complex configuration yields a unique final structure after the 200 ns MD simulation (Appendix I-IV). The ATP complex undergoes minor conformational change in which the NTDs are reoriented towards the M-domain of the opposite protomer (Appendix I-IV-red). This reorientation appears to tighten the two protomers and supports previous observations that these conformational rearrangements are necessary for catalytic activation [162,183]. The ADP complex appears to undergo extensive conformational restructuring, whereby the protomers disengage from one another in a distinct opening of the molecular clamp that likely explains the jump in RMSD at 30 ns. The final PO-ADP structure closely resembles an open 'v-like' structure (Figure 3.7-D) that has been previously reported for HtpG (PDB ID: 2IOQ) [180]. Indeed, superimposition of the 2IOQ on the final PO-ADP structure revealed near identical tertiary structures (Appendix I-IV), recording a backbone RMSD deviation only 8.23 Å. Like the PO-ATP complex, the PO-apo configuration also shows minor deviation from the initial structure. The only notable conformational change being the NTD of the second protomer, which appears to relax and extend in an upward flexing motion away from the MiD of the opposite protomer (Appendix I-IV-green).

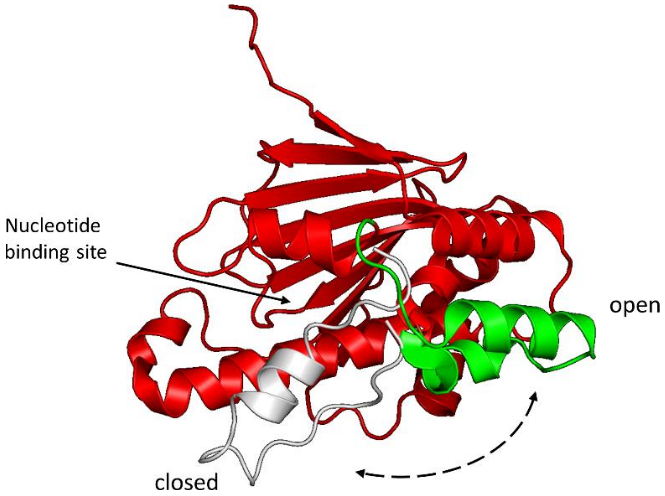


Figure 3.6: Illustration the NTD ATP-lid positons. The NTD is coloured red and the ATP-lid coloured according to the two possible positions open (green) and closed (grey).

Next we measure the time evolution of NTD inter-protomer distance, defined by the distance between the center of mass of each the NTDs, over the course of each MD simulation to quantify the conformational sampling of the individual protomer with respect to each other. This data would inform on the stability/flexibility of the protomers over the course of the MD simulation. For the PO-ADP complex it is evident that protomer uncoupling likely occurs after 20 ns, the inter-protomer distance increasing by ±10 Å to fluctuate around 30 Å for the remainder of the MD simulation (Figure 3.7-B). In contrast, both the ATP and apo complexes maintain their NTD inter-protomer distance, fluctuating around 21 Å.

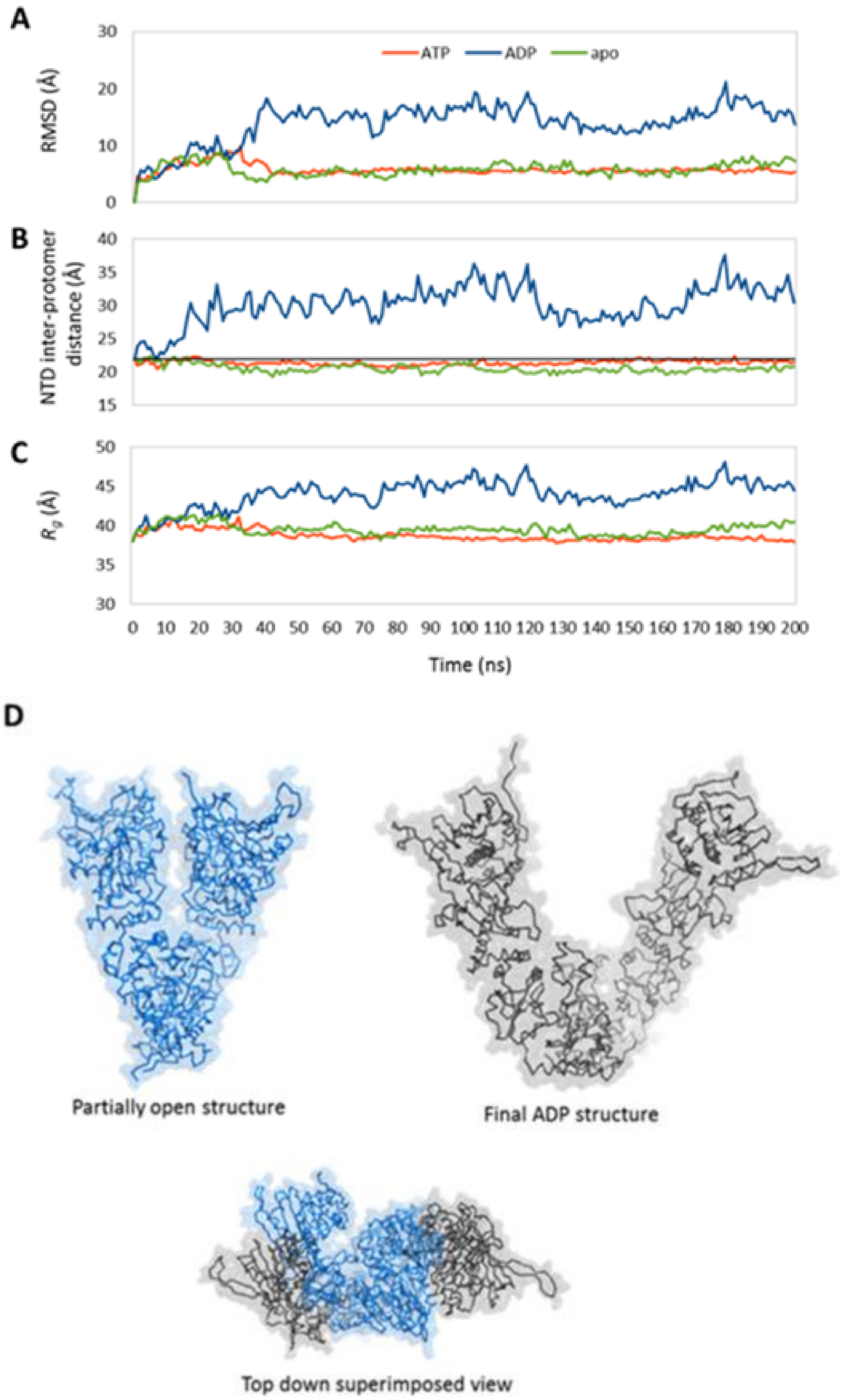


Figure 3.7: Global conformational dynamics metrics for the partially-open conformation complexes. (A) Backbone RMSD, **(B)** NTD inter-protomer distance (black line represents the initial distance at 0 ns), **(C)** *Rg* plots, **(D)** 3D illustration of dimer opening in the PO-ADP complex, showing the initial partially open structure (blue) and the final structure after MD simulation (grey). Reproduced with permission from Penkler et. al 2018 [191]

RMSF analysis of the PO complexes (Appendix I-V) shows an elevated degree of residue fluctuation in the NTDs of all three complexes, with slightly higher fluctuations for the ADP complex. From a functional stand-point, notable NTD fluctuations include: the N-terminal β-strand$_1$ (res 15-26), β-sheets 3-5 (res 65-75, 160-80), the ATP-lid (res 112-136), and the charged linker (res 210-290).

An interesting difference between the different PO configurations is observed in the MiD (res 288-549), where fluctuations are significantly higher for the ADP complex, particularly the catalytic loop region (res 391-406). All three complexes show similar residue fluctuations at the CTD; however, the ADP complex once again experiences higher fluctuations, this time at helix$_{18}$ and helix$_{19}$ (res 601-632). Collectively this data suggests that the ADP complex is more flexible than the ATP and apo complexes, and its increased fluctuations in all three subdomains may be evidence of a 'relaxed' ADP state. This apparent increase in flexibility for the ADP complex may be an important structural characteristic for the destabilization of anchoring inter-protomer interactions at the MiD and CTDs and thus facilitation of the opening transition. The converse may be true for the PO-ATP and PO-apo complexes, in which protomer rigidity may help maintain inter-protomer interactions, which stabilise the closed state.

R_g analysis for these complexes supports this observation (Figure 3.7-C), in which the PO-ADP complex becomes considerably less compact over time fluctuating around 45.0 Å. Indeed, the initial increase in R_g occurs soon after the protomers first disengage (Figure 3.7-B). In contrast, the PO-ATP and PO-apo complexes maintain their original structures, fluctuating around 38.0 Å, the former becoming slightly more compact than the latter in agreement with the previous visual observations made (Appendix I-IV-A).

Looking at the inter-protomer interactions present in the final MD simulation structures, the only interactions conserved in all three complexes are those involved in CTD dimerization, the notable difference being helix$_{18}$ (res 601-632). In the initial closed and PO structures, helix$_{18}$ is seen to form intimate interactions with the identical residues in the opposite protomer, forming an important interface between the protomers. Despite minor fluctuations of these residues over the course of the MD simulations (Appendix I-V) both the apo and ATP complexes appear to retain this inter-protomer interaction, while in the ADP complex they are completely abolished, allowing helix$_{18}$ and helix$_{19}$ to fold back towards their respective protomers. The CTD has been previously implicated in the allosteric modulation of dimerization in yeast Hsp90 [195,203,297,299]. This data could describe how ADP may allosterically modulate conformational dynamics, in which it increases protomer flexibility, allowing the NTDs to become extended triggering disengagement of helix$_{18\text{-}19}$ interface, which in turn causes further destabilization of other inter-protomer interactions in the MiDs, resulting in a transition to the open state.

– *Fully open 'v-like' complexes*

All-atom MD simulations of the PO complexes revealed drastic conformational changes in the ADP configuration that closely resembled the fully-open 'v-like' conformation (Figure 3.7-D & Appendix I-IV). This structure provided a suitable starting point for further MD investigations regarding the effect of bound nucleotide on the conformational dynamics of the fully-open (FO) conformation. As such, the final structure of the PO-ADP complex was selected from the 200 ns trajectory, and FO-ATP, FO-ADP, and FO-apo complexes prepared in the same manner as the previous two conformations, and each complex submitted to 200 ns MD simulations.

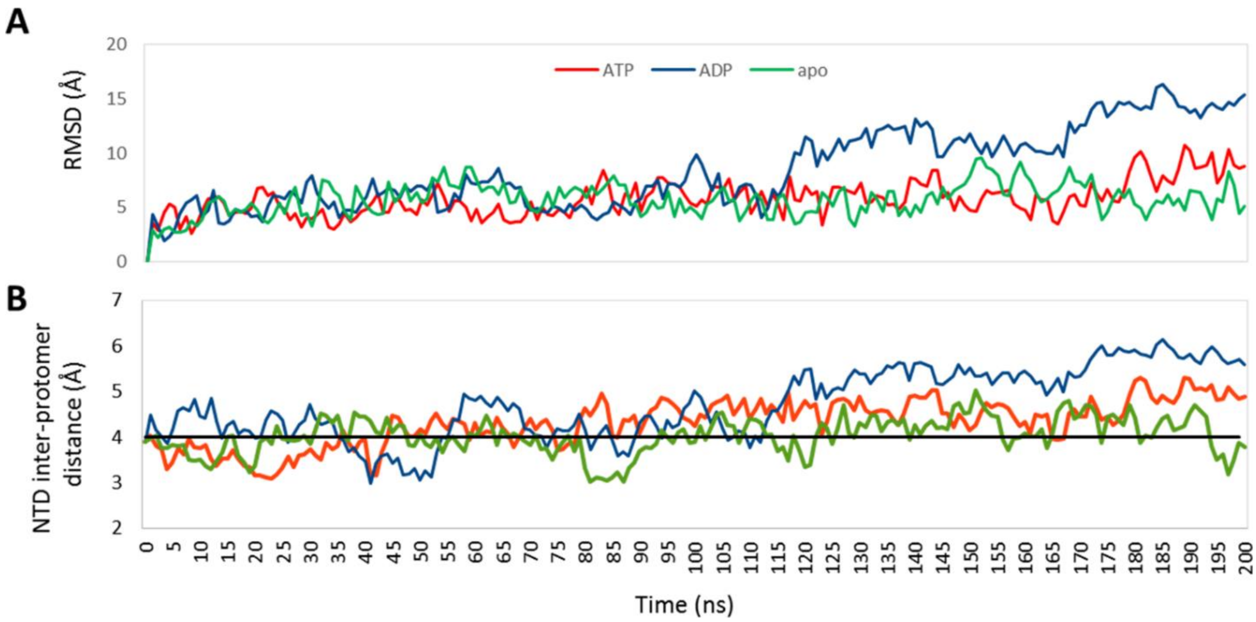


Figure 3.8: Conformational dynamics metrics of the FO complexes. (A) Backbone RMSD, **(B)** NTD inter-protomer distance. The black line represents the initial distance at 0 ns. Reproduced with permission from Penkler et. al 2018 [191]

The RMSD plots of the fully open complexes (Figure 3.8-A) indicate minor backbone fluctuations for the ATP and apo complexes, with RMSD values converging around 6 Å. In contrast, the FO-ADP complex appears to undergo further conformational changes at 100 ns, which correspond to a slight increase in NTD inter-protomer distance (~2 Å) (Figure 3.8-B). RMSF analysis of the FO complexes (Appendix I-VI) reveals similar residue fluctuation profiles to the PO complexes (Appendix I-V), where the NTD and M-domains experience higher fluctuations compared to the CTD. Comparing the final structures of each complex, the only noticeable difference is observed for the ATP complex in which the NTDs appear to re-orient inwards, towards their respective MiDs, a movement that may be indicative of an early closing motion.

Overall, the data presented in here demonstrate that major conformational dynamics are experienced for the PO complexes in response to the ligand bound state of the NTD, while more subtle structural rearrangements are experienced by the FC and FO complexes. In each case, ATP appears to confer structural rigidity and 'tensing' of the dimer, while ADP enables a more 'relaxed' flexible complex. The apo state engenders opposing correlated motions in the FC complexes; however the absence of nucleotide has little or no effect on the PO and FO complexes which appear largely stable. This contrasting behaviour may be in agreement with previous reports of stochastic dynamics observed for the apo protein [181,327,328]. These findings are in agreement with several previous studies [159,326,328], and the lack of global restructuring reticent of either the opening or closing transitions is likely in part due the limited sampling time scales available for this study. Thus, to probe our data for further insights regarding nucleotide modulation of inter-state transition, the MD trajectories of the closed and FO complexes were subjected to DRN analysis to investigate the effect of bound nucleotide on intra-protein communication, as well as PRS analysis to identify single residues involved in modulating inter-state conversion.

3.3.3 Dynamic residue interaction network analysis

Dynamic residue networks (DRNs) [295] were utilized to analyse the effect of ligand binding on residue connectivity over time. DRNs were constructed for each MD trajectory by treating $C_{ß}$ atoms (C_{α} for glycine) as nodes in the network and connections between nodes established based on a distance cut-off of 6.7 Å (see Methods for details), and the resultant DRNs analysed in terms of average long-range residue reachability (L) and average betweenness centrality (BC). In graph theory, the reachability of a residue is defined as the number of connections required to reach residue i from j using the shortest possible path.

The average reachability of a residue (L_i) is thus defined as the average number of steps required to reach residue i from any other residue in the network. Here we consider the average change in path length and monitor how L_i shifts over time with respect to initial structure. By using this approach, changes in the global conformational structure that occur during the MD simulation that are large enough to invoke a change in intra-protein communication should be capture by monitoring ΔL_i. ΔL_i is widely thought to be a suitable metric for identify regions of a protein that are actively involved allosteric signal propagation [289].

The metric BC is related to L in that it is a measure of how often on average a residue is utilized in shortest path navigation and has been previously shown to be an effective measure for the identification of functional residues implicated in intra-protein communication [191,329], protein-ligand [330] and protein-protein binding sites [331,332], as well as non-synonymous SNP analysis [333,334].

– *Nucleotide configuration modulates long-range residue reachability*

In this section we compare and contrast the relative effect bound nucleotide at the NTD has on the change in average path length of the FC and FO complexes. Regions of particular interest are highlighted and the results discussed in context of previously validated experimental findings.

· *Fully-closed complexes*

The ΔL_i profiles for the FC complexes (Figure 3.9) show that the ATP and apo complexes experience mostly negative ΔL_i values over the entire length of the protein, while the ADP complex experiences both positive and negative change. Negative ΔL_i can be described as a shortening of the average path length residue *i* must transverse to get to any other residue in the network, while positive ΔL_i values indicate increased path lengths. The global negative ΔL_i observed in the ATP and apo complexes may suggest that these NTD conditions favour more compact structures, in which the physical compression of the protein along the central axis established numerous alternate routes between residues and thus decreases the average path length. Conversely, the increased path lengths observed for the ADP complex may suggest a more relaxed elongated structure, shaping the flow of network communication along select routes. Indeed, this explanation is in agreement with the conformational dynamics results (section 3.3.2), and supports the hypothesis of structural tensing and relaxing under ATP and ADP conditions, respectively.

Analysis of select regions of the protein that display elevated ΔL_i points to structural elements that are important for inducing intra-protein communication. Looking at the respective ΔL_i profiles in more detail, it is interesting to note a differential degree of change in some regions depending on the ligand bound configuration of the NTD. Residues residing in ß-strand$_1$ (res 15-26) and the ATP-lid (res 112-136), residues D$_{302}$, S$_{330}$-L$_{335}$, helix$_{10}$ (res 407-427), and several residues neighbouring Y$_{604}$ and Y$_{627}$ all record sharp positive ΔL_i in the presence of ADP, but negative changes when ATP is present. Putting these findings into context of protein function; previous studies have shown that ß-strand$_1$ swapping between protomers contributes to the stabilization of NTD dimerization in the FC conformation [163,335,336], and its deletion has been shown to adversely affect ATPase activity in mitochondrial and yeast Hsp90 homologs [337,338]. The ATP-lid has been intricately implicated in the progression of conformational change, acting as a trigger for NTD dimerization in response to bound ATP [163,179,187,336]. Helix$_{10}$ has been previously implicated in nucleotide driven allosteric signal propagation between the NTD and M-domain in *E. coli* HtpG [303]. Residues Y$_{604}$ and D$_{621}$-Y$_{627}$ reside in helix$_{18}$ and helix$_{19}$ respectively, these sites undergo large negative ΔL_i in the ATP and apo complexes, but display positive ΔL_i in the ADP complex. This differential behaviour may suggest a tighter binding of these interface elements in the ATP and apo complexes and a looser dimerized state in the presence of ADP. This observation is in agreement with the conformational dynamics analysis of the PO structures (see section 3.3.2) in which interactions between these residues from either protomer are lost in the presence of ADP and retained in the ATP and apo complexes.

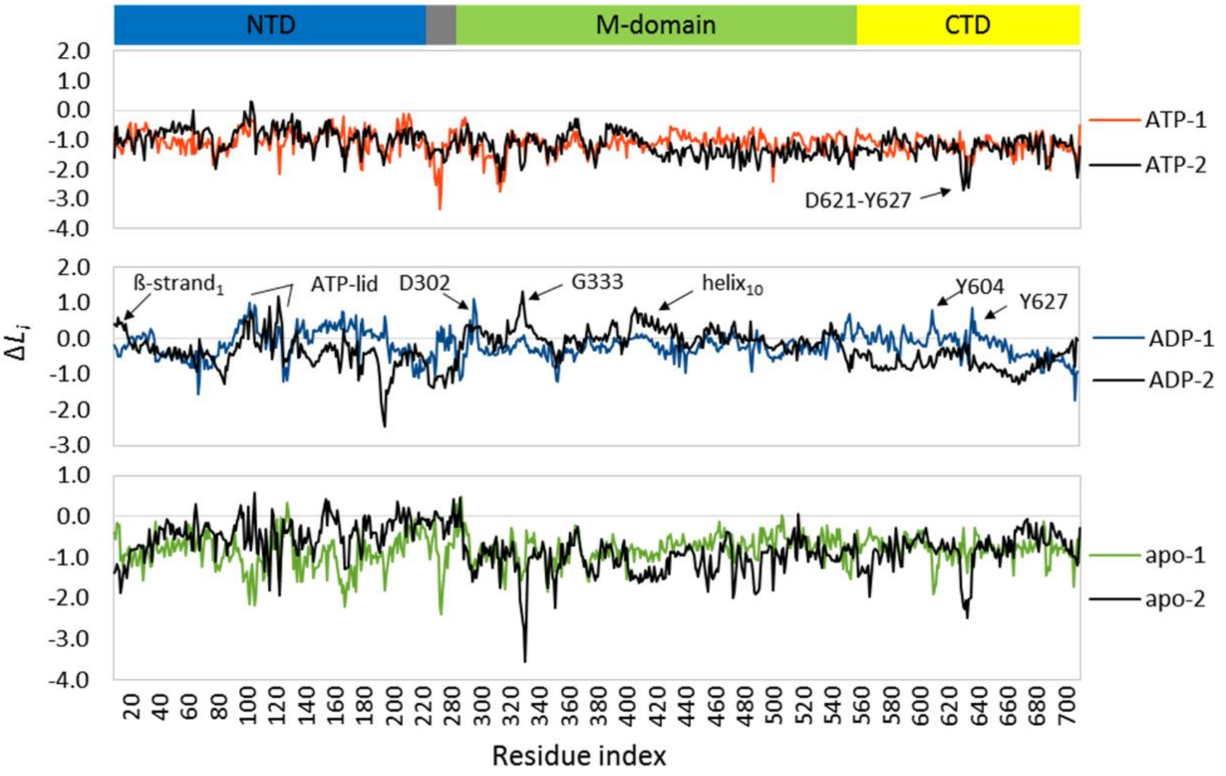


Figure 3.9: Change in reachability (ΔL_i) for the closed conformation complexes. Protomer 1 is coloured by NTD configuration (ATP–red, ADP–blue, apo–green), and protomer 2 is in black. Reproduced with permission from Penkler et. al 2018 [191]

- *Fully-open complexes*

Next we focus our attention on the FO conformation complexes, and note ΔL_i to be largely negative for FO-ATP complex, but that it fluctuates for the in the present of ADP or in the nucleotide-free state (Figure 3.10). As previously seen in the closed complexes, certain regions of the protein respond differentially in response to the bound configuration of the NTD and can be linked to experimentally determined protein functions. Starting with the NTD, residues E_{200}-K_{224}, forming ß-strand$_8$ located at the interface between the NTD and the charged linker experience negative ΔL_i in the ATP complex, and positive ΔL_i for the ADP and apo configurations, suggesting reduced flexibility in this region when ATP is present. These residues have been directly implicated in the regulation of the human Hsp90α ATPase cycle, and mutations to select residues in this region have been shown to lead to increased NTD flexibility and severely altered chaperone activity causing a shift in the conformational equilibrium in favour of the open state [339]. In the MiD, residues R_{367}-E_{380} and L_{396}-Q_{404} are in close spatial proximity to one another and all experience sharp positive ΔL_i in the presence of ADP. Positioned at the NTD-MiD interface, these residues have been reported to form a nucleotide sensitive hinge, that is implicated in allosteric signal propagation from the NTD to the MiD [302,305]. In the CTD, the inter-protomer dimerization interface (helix$_{18}$ and helix$_{19}$) experience positive ΔL_i in all three complexes, however in the presence of ADP these regions record a much higher values, once again suggesting looser dimerization interactions under these conditions.

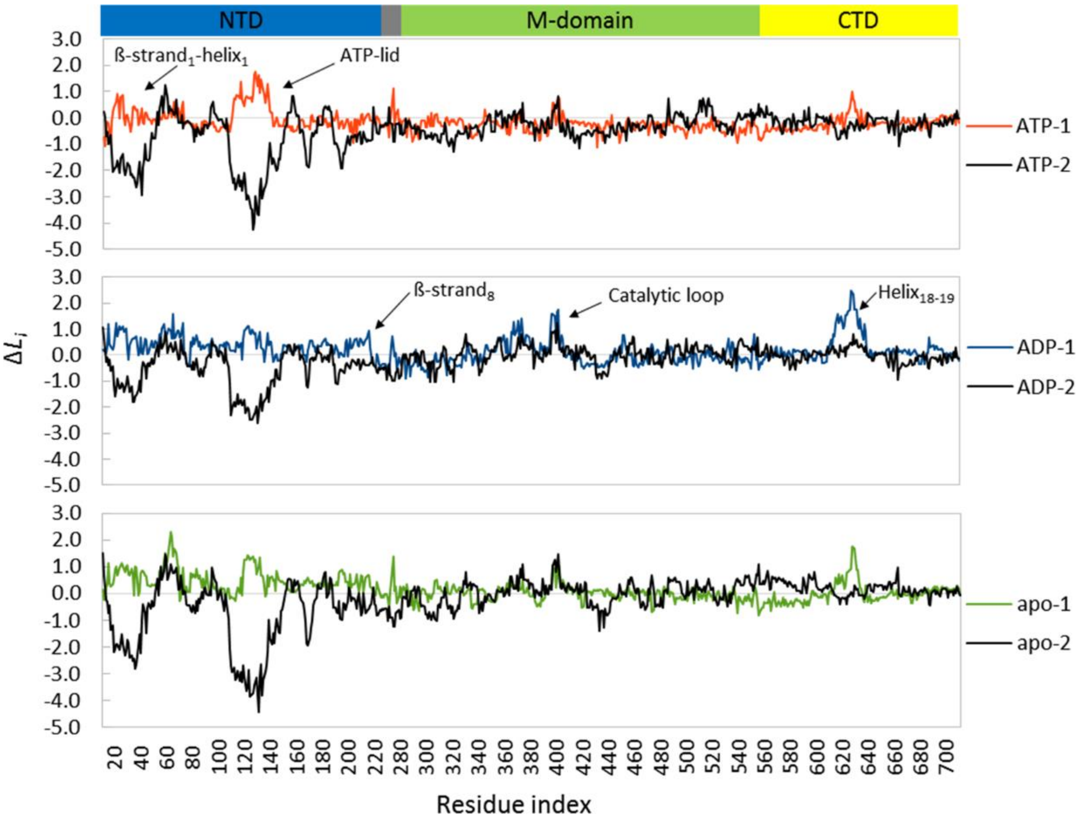


Figure 3.10: Change in reachability (ΔLi) for the fully open complexes. Protomer 1 is coloured by NTD configuration (ATP–red, ADP–blue, apo–green), and protomer 2 is in black. Reproduced with permission from Penkler et. al 2018 [191]

- *Differential residue reachability between protomers*

We note that a largely differential behaviour of ΔL_i between the protomers of both the FC and FO complexes, and tentatively suggest that this may arise due to asymmetrical modulation of protein dynamics. In the closed complexes, notable differences in ΔL_i values between protomers are observed for the ADP and apo complexes, while in the FO complexes large contrasting ΔL_i values are recorded in the ß-stand$_1$, helix$_1$, and the ATP-lid regions regardless of bound nucleotide. This differential modulation is unlikely to arise due to asymmetry in the starting structures, with backbone superimposition of the protomers yielding RMSD values of 0.8 Å and 3.1 Å for the closed and FO conformations respectively, indicating a high degree of symmetry between protomer 1 and protomer 2 for both conformations. Furthermore, ΔL_i analysis on replicate trajectories demonstrated highly reproducible results (Appendix I-VII and Appendix I-VIII), eliminating MD convergence as a possible cause of asymmetry. As such, we suggest that this differential behavior may be further evidence of previously reported asymmetrical protomer dynamics of the protein in response to the nucleotide bound configuration [160,326,338].

Overall, ΔL_i provides a unique metric for monitoring the effect of nucleotide modulation on local and global restructuring and its implication on long range communication in Hsp90, highlighting several regions of the protein that may be involved in steering conformational change. Here, the analysis confirms the tensing effect of ATP, with the protomers of both conformations experiencing negative ΔL_i values. Conversely, protomer relaxation afforded by ADP binding is also confirmed, with the average path length increasing in both ADP bound conformations. Regions that experience large ΔL_i are most notable in the open complexes and include residues belonging to the functionally important ß-strand$_1$, helix$_1$, ATP-lid, and catalytic loop [163,179,336], as well as the helix$_{18}$/helix$_{19}$ CTD interface.

- *Betweenness centrality points to putative communication hubs*

BC has been shown to be an effective measure for identifying functional residues implicated in the control of intra-protein communication [289,333], as well as protein-ligand [330] and protein-protein binding sites [331,332] (see review [49] for other examples). Here, BC is calculated for the closed and FO complexes to identify individual/groups of residues important for controlling intra- and inter-protomer communication.

The respective BC profiles of both conformations (Figure 3.11), reveal that bound nucleotide has little impact on modulating communication hubs in either conformation, with similar trends observed for all three configurations. Thus, to simplify data presentation, the average BC over all three configurations for each conformation is shown (Figure 3.11-black curves). Peak locations in each conformation describe residues with high BC and thus high frequencies of usage in intra-protein communication. As previously mentioned, BC for a given residue *i* is a measure of the number of shortest paths that visit *i*, and points to locations on the protein that may be important communication hubs given their high frequency of usage.

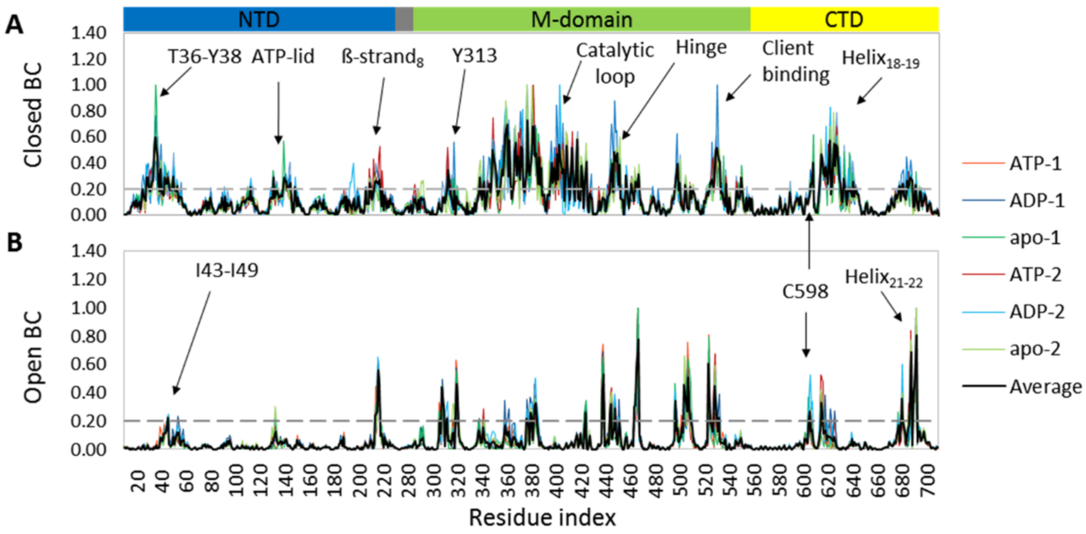


Figure 3.11: Betweenness centrality (BC) profiles. Showing peak regions in the **(A)** closed and **(B)** open complexes that may constitute putative communication hubs. Dashed grey lines indicates the peak residue threshold. Reproduced with permission from Penkler et. al 2018 [191]

- *Comparison of the closed and open complexes*

The BC profiles for the fully-closed and fully-open conformation complexes are shown side-by-side in Figure 3.11 for comparative purpose. A minimum BC threshold of 0.20 was arbitrarily set to distinguish peak BC residues (grey line). At a glance it is evident that more residues are selected by BC in the closed complexes compared to the open complexes. Furthermore, it is apparent that BC selects contiguous residue in select regions of the closed complex, and individual more localised hotspots in the open conformation complexes. The increased number of BC residues in the closed conformation complexes may be attributed to the inherent compactness and rigidity of the structure, as well as the increased number of intra- and inter-protomer contacts. Taken together these attributes are expected to influence BC by increasing the number of possible shortest paths in the network. A scenario that would likely result in an increased number of possible communication hubs. Indeed the converse would be true for the FO complexes, which likely experience localised BC hot residues due to the dynamic nature of its protomers that only interact via the CTD.

Despite the difference in the total number of BC residues, both conformations share several overlapping peak residues that correlate well with known functional sites, particularly those involved in regulating ATPase activity and conformational dynamics. The peak residues listed by BC for the closed and open complexes are summarized in Table 3.3 and are mapped to their respective 3D structures in Figure 3.12, and their respective functional importance is discussed below.

Table 3.3 | Summary of residues with high BC for the fully-closed and fully-open complexes

	NTD	M-domain	CTD
Closed	I34-D54; M130-A145; I206-L220	F312-N318; L351-V388; L396-F428; F441-E451; Y493-E497; Y520-V530; L541-K546	C598-T601; W606-K632; L666-N686
Open	I43-I49; M130-F134; S211-I214	F312-Y313; H323-L324; L340-L341; P344-R345; I361-Y364; R367-F369; I378-G387; L423-F424; F437-Y438; F441-E451; L462-Y465; I494; Q501-F507; Y520-Y528	C598-T601; T607-L619; L671; L678; D680-Q682

- *Functional relevance of BC hotspots*

Starting with the NTD, phosphorylation and mutation studies have linked T_{36} and Y_{38} to ATPase regulation and altered binding affinity of co-chaperones Aha1 and Cdc37 [179,197,340]. Computational studies have implicated both residues in inter-domain signal propagation [203,297,341]. Both residues are only selected in the closed complexes where they are likely involved in stabilising NTD dimerization by contributing to crucial interactions with residues in the catalytic loop (S_{391}-S_{406}) and residues R_{46}-I_{49} [162,163]. The latter set of residue are included in the region I_{43}-I_{49} which is listed by BC in both conformations. Mutation studies have shown that $E_{47}A$ abolishes ATP hydrolysis suggesting a catalytic function *in vivo* [319]. In the closed conformation, $E_{47}A$ would have a negative impact on catalytic loop interactions and thus decrease ATPase activity. In the open state, I_{43}-I_{49} are positioned in closed proximity to the open ATP-lid, presumably helping to stabilise it in the open configuration (see Figure 3.6). Furthermore, BC also lists the ATP-lid residues M_{130}-F_{134} in both conformations. In the open complex, these residues are uniquely positioned over I_{43}-I_{49}, while in the closed complex they are positioned over residues S_{53} and D_{54}, which together form part of a hydrophobic patch known to be involved in ATP-lid stabilization [179]. The physical repositioning of the ATP-lid over the nucleotide binding site is thought to provide the trigger for ATPase activity [342]. In sum, these findings point to I_{43}-I_{49} and M_{130}-F_{134} as crucial communication hubs in both conformational states.

Next, residues I_{206}-L_{220} of ß-strand$_8$ and helix$_7$ are located at the interface between the NTD and the charged inter-domain linker and have been implicated in chaperone secretion and modulation of the ATPase activity, where $I_{218}A$ and $L_{220}A$ mutations are thought to alter the topology of the charged linker, causing a reduction in co-chaperone associations and ultimately leading to *in vivo* cell death [339]. Of these residues, only S_{211}-I_{214} are listed in both conformations, and we note that for the closed complexes, the presence of ATP results in higher BC values for these residues. The reduced

BC of these residues in the ADP and apo complexes is significant in that the mutation I_{218}A has also been shown to impact conformational dynamics, shifting the conformational equilibrium in favour of the open state [343]. It is possible that I_{206}-L_{220} may act as a crucial communication hub capable of sensing bound nucleotide at the NTD. Under rigid ATP conditions I_{206}-I_{220} may help stabilize the complex in a closed conformation. Eventual ATP-hydrolysis and subsequent protomer relaxation in response to ADP binding may destabilize this region, and ultimately trigger a conformational shift to the open sate.

Looking at the listed residues in the MiDs of both conformations; phosphorylation of Y_{313} has been shown to promote Aha1 binding affinity by inducing structural rearrangements that favour co-chaperone binding [344]. Residues V_{368}-I_{370} have been previously identified as key inter-domain contacts in a comprehensive network analysis of yeast Hsp90 (V_{348}-I_{350}) [305]. Of these residues, the mutation F_{369}A in yeast has been found to drastically decrease ATPase activity by disrupting the positioning R_{400} in the catalytic loop which in turn leads to destabilization of the ATP-lid [162,187,345]. Residues L_{396}-L_{409} are listed in the closed but not the open complexes, represent the MiD catalytic loop where they are thought to form part of an important inter-domain hinge between the NTD and M-domain [302,346]. In addition to the this residues S_{399} [195], R_{400} [162], E_{401} [190], and Q_{404} [162] have been implicated in modulating ATPase activity and allosteric signal propagation [297,305].

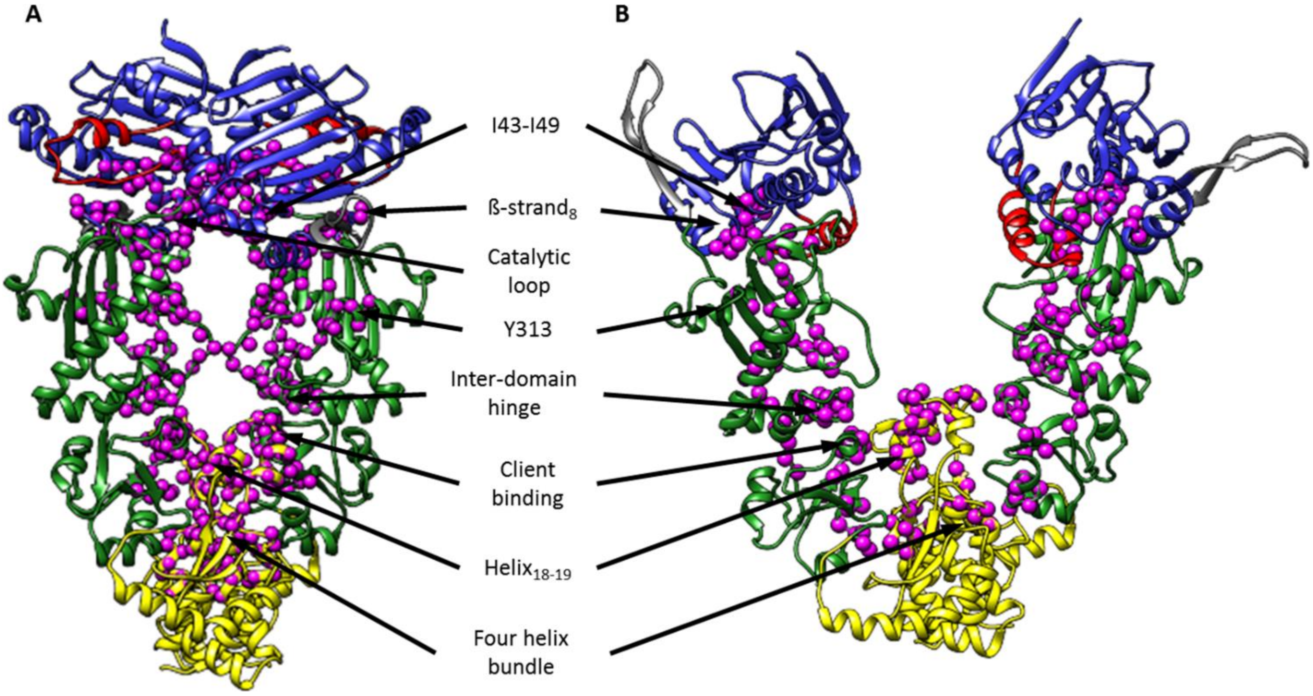


Figure 3.12: Structural mapping of high BC residues. (A) Closed complex and (B) fully-open complex. Showing peak residues as magenta spheres. Reproduced with permission from Penkler et. al 2018 [191]

The aromatic cluster formed by F_{384} and F_{441} are listed by BC in both conformations, and are thought to be important allosteric control points for nucleotide dependent conformational change, acting as a mechanical hinge between the MiD and CTD [302,346]. Reside E_{451} is also selected in both conformations and mutation studies have shown that $E_{451}A$ decreases inter-protomer interactions leading to decreased ATPase activity [346], while $E_{451}K$ is thought to perturb the MiD structure impacting gluticoid receptor binding [347]. Phosphorylation mimicking mutations to S_{505} is fatal for yeast viability (S_{485}) and is thought to impart a resistance to ATP induced conformational transition to the closed state, and stabilise the open state [195]. Here, S_{505} is only selected in the open complexes suggesting an important role in this conformation, possibly as a post translational modification site for regulating the closing transition. Residues E_{527}, Y_{528}, and Q_{531} have been shown to form part of an important region for client binding in both yeast and *E. coli* Hsp90s [348] and have also been implicated in allosteric regulation [308], while $T_{545}I$ mutation has been shown to decrease client binding in *E. coli* [347].

In the CTD, BC lists several prominent overlapping peaks in both conformations: The C_{598}-T_{601} peak includes C_{598} which has been heavily implicated as a crucial allosteric switch point, capable of regulating the ATPase cycle [203,297]. Peak T_{607}-L_{619} in the open conformation corresponds to helix_{18} at the CTD interface. In the closed conformation this peak is extended to include residues W_{606}-K_{632} representing a coiled loop between $\text{helix}_{18\text{-}19}$. Of these residues: W_{606} is thought to be involved in client binding and chaperone activity [348]; phosphorylation mimicking mutations to S_{623} and M_{625} affect inter-subunit communication leading to decreased ATPase rates in yeast [195]; while phosphorylation of Y_{627} is thought to induce structural rearrangements that reduce the binding affinity of co-chaperone Aha1 [344]. Finally, several residues at the four helix bundle ($\text{helix}_{21\text{-}22}$) located at the extreme C-terminal dimerization site are selected in both conformations. Included in this residue peak is D_{680} which has been shown to play an important role in maintaining inter-protomer contacts [346]. Several other residues from this site in yeast Hsp90 have been reported to participate in fast allosteric communication with residues in the NTD [297], suggesting a nucleotide sensitive regulatory role.

In summary, BC selects several groups of residues in both the closed and FO structures regardless of bound nucleotide suggesting key intra- and inter-protomer communication points. These residues correspond largely to stable regions of the protein and overlap with known functional sites several of which are implicated in conformational dynamics and allosteric mediation. Furthermore, we note that BC also lists several additional residues in regions of as yet unknown function.

3.3.4 Relationship between the BC, *L* network descriptors and residue fluctuations

In this study BC and L_i were calculated based on data extracted from MD trajectories to make allowance for the potential effect protein dynamics may have on the respective network measures. Thus far, we have discussed how global domain restructuring in response to bound nucleotide impacts the change in reachability in both the closed and FO complexes. In this section, we focus on the relative impact local dynamics may have on BC and L_i. Residue fluctuation (RMSF) is a primary example of localised protein dynamics, in which select sites on the protein experience conformational variance relative to the global structure. Structural repositioning of flexible residues is expected to influence the local network by breaking previously established connections between neighbours and/or forming new ones altogether. This means that RMSF has the potential to cause shifts in the local networks that are likely to impact the BC and L_i metrics. To assess these assumptions, we take RMSF as an appropriate measure of residue fluctuation and compare its relationship to BC and L_i respectively.

- ***L_i is proportional to RMSF:***

When considering average path length, residues that experience minor fluctuations over the MD trajectory are likely to experience increased connectivity and thus shorter average path lengths. Highly mobile sites on the other hand are expected to experience lower, more intermittent connectivity and thus longer average path lengths. In this manner, RMSF and L_i could be seen as being proportional to each other, and indeed this observation that has been previously reported for static protein x-ray crystal structures without molecular dynamics data [46]. Inspection of the comparative plots for the fully-closed and fully-open complexes (Figure 3.13) reveals a similar trend between L_i and RMSF for both conformations, particularly for the peak regions.

Interestingly, comparison of the L_i and RMSF curves using the Pearson's correlation coefficient indicates much higher correlations for the FO complexes compared to the closed complexes (Table 3.4). This observation can be explained by the relative plasticity of the two structures at equilibrium. As previously discussed and shown in this study, the closed conformation is the most stable form of the protein compared to the FO conformation which is significantly more flexible. Comparison of their respective RMSF plots reveals on average higher fluctuations for the FO complexes compared to the closed complexes (Appendix I-II and Appendix I-VI respectively). This difference in distance fluctuation is significant in that the underlying local network graphs are likely to be more severely impacted by large distance fluctuations which may cause the distance between connected residues to exceed the 6.7 Å distance cut-off used to construct the network. In the event that this does occur degree of connectivity within the local networks when averaged over the whole trajectory would be significantly reduced.

For the FO complexes, the NTD and MiD of both protomers experience increased RMSF compared to the CTD which is more stable due to its involvement in inter-protomer dimerization. The aforementioned hypothesis is echoed in the comparison of the correlation coefficients for these domains which reveal higher correlations for NTDs and MiDs compared to the more stable CTD.

Conversely, in more stable structures such as the closed complexes, the likelihood of residue fluctuations affecting path length is decreased due to the inherent rigidity of the structure, where the magnitude of the fluctuations may be insufficient to exceed the 6.7 Å distance cut-off. In this scenario, previously established connectivity between neighbouring residues would remain intact despite minor local residue fluctuations, and thus cause the trend between L_i and RMSF to deviate and negatively impact the correlation coefficient. An example of this is observed when comparing the correlation coefficients for the closed conformations which show subtle differences between the different nucleotide configurations. Therein, the apo form records slightly higher correlation coefficients compared to the ATP and ADP configurations (Table 3.4). This observation supports the aforementioned proportional relationship between L_i and RMSF, in that the apo complex which experiences a higher degree of flexibility compared to the nucleotide bound complexes (see previous sections), records slightly larger residue fluctuations (Appendix I-II), and thus a stronger correlation with L_i.

In addition to local fluctuations, we note that the degree of compactness of the network graph could also be a contributing factor to minor deviations between RMSF and L_i. For example, the closed conformation complexes (and the CTDs of the FO complexes) have more compact local networks due to their 3D structural arrangements, and are thus afforded a greater number of possible paths to navigate the network by. This increased selection of possible routes may enable L_i to utilize alternate routes to avoid local fluctuations, and ultimately L_i to remain fairly stable compared to RMSF for a given location.

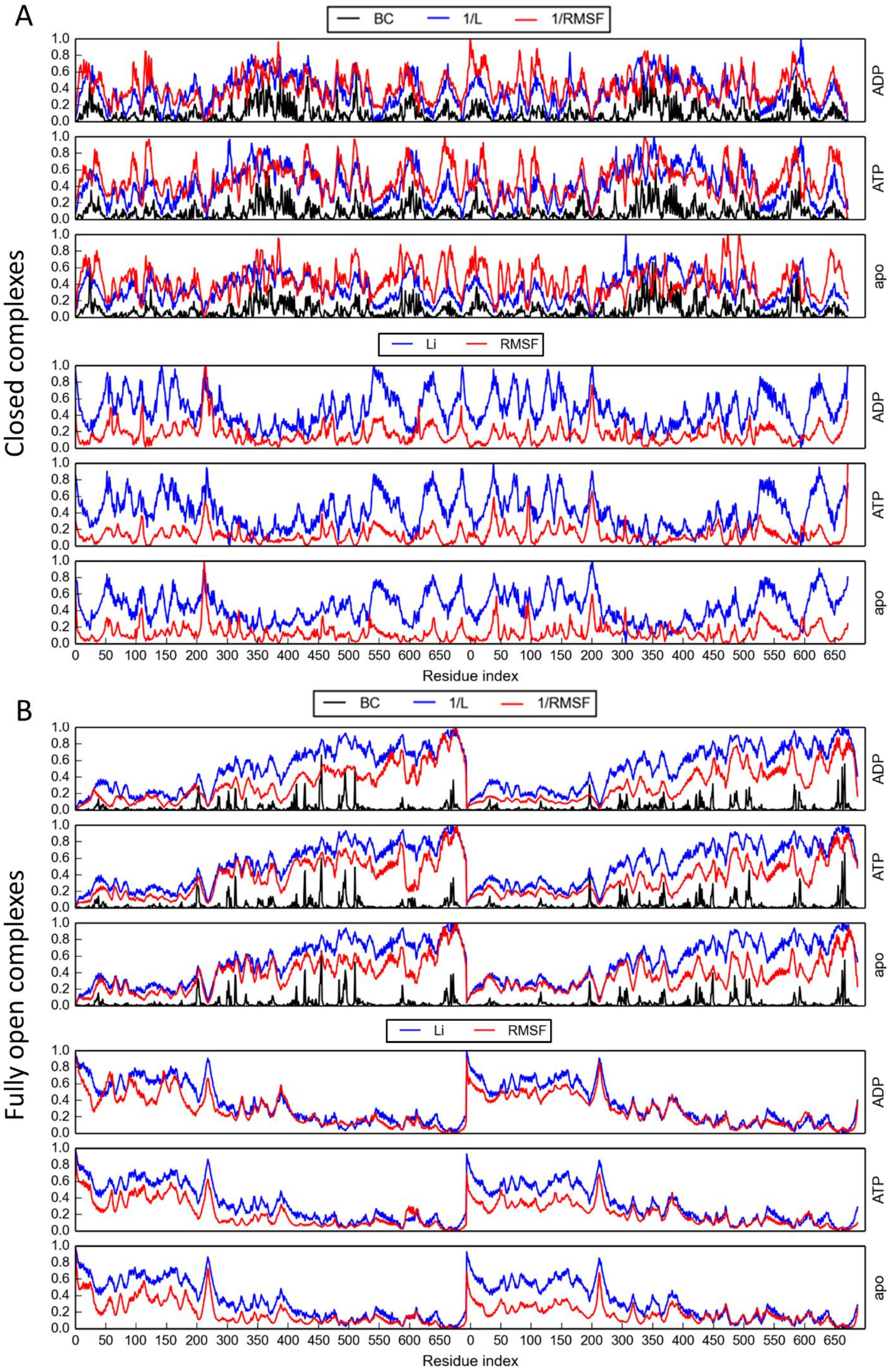


Figure 3.13: Comparative plots for BC, L_i and RMSF. For the ATP bound (A) Fully-closed conformation and (B) Fully-open complexes. Reproduced with permission from Penkler et. al 2018 [191]

We test these hypotheses on the FO ATP complex by comparing RMSF with different L_i that are calculated based on incremental distance cut-offs (6.7 Å, 9.0 Å, and 12.0 Å) (Figure 3.14). In this analysis, the correlation between L_i and RMSF is seen to hold for the flexible NTDs and M-domains regardless of cut-off, but decreases severely (up to 3-fold) for the more compact rigid CTDs as the cut-off increases, demonstrating how the potential effect of RMSF on local network arrangements decreases when the magnitude of the fluctuations is close to or less than the specified distance cut-off. These observations are in agreement with a previous study that demonstrates how increased cut-offs (>8.5 Å) cause L_i to gradually converge towards the theoretical limit [46].

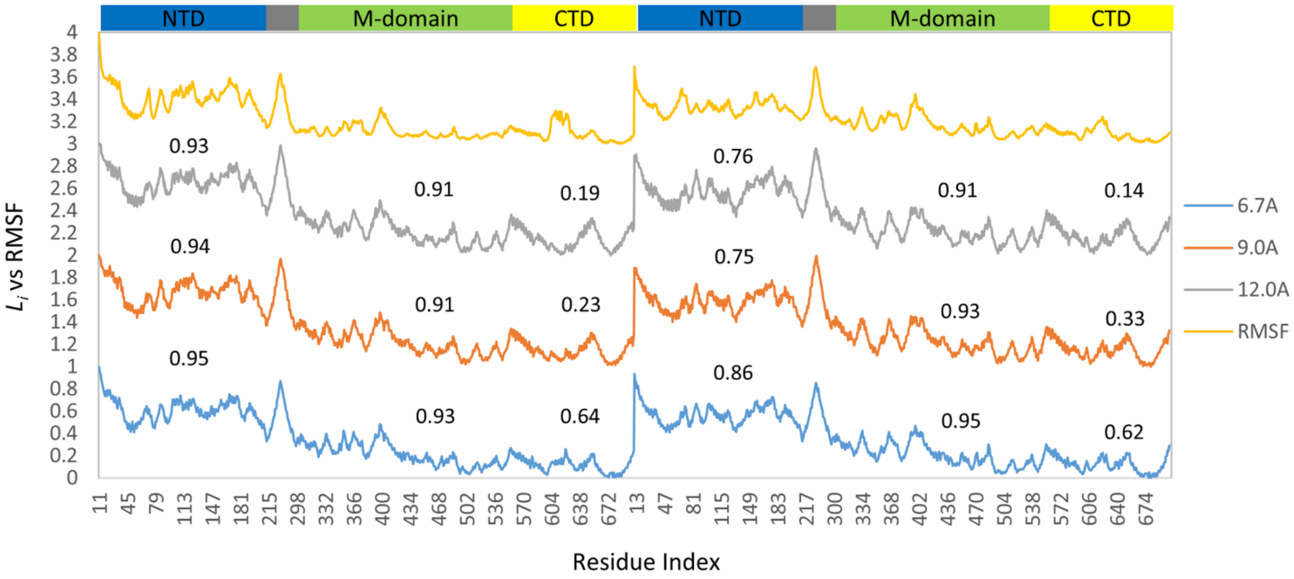


Figure 3.14: Correlation between RMSF and L_i with incremental distance cut-offs. Several network graphs were constructed based on 6.7 Å, 9.0 Å and 12.0 Å distance cut-offs for the FO ATP bound complex. Whole protein correlation were 6.7 Å – 0.95, 9.0 Å – 0.93 and 12.0 Å – 0.91. Reproduced with permission from Penkler et. al 2018 [191]

In summary, we suggest that L_i and RMSF share a proportional relationship, and that deviation from this trend may arise when residue fluctuations are of an insufficient magnitude to affect the connectivity of the local network, or when the structure is suitably compact such that alternative routes may be taken to avoid increased L_i in response to fluctuation of the local network. Furthermore, we note that care should be taken when setting the distance cut-off for the construction of the network graph, and note that if the cut-off is too large and the system is predominantly rigid/compact, residue fluctuations will have a reduced effect on the observed local network.

Table 3.4 | Pearson's correlation coefficient between BC, ΔL_i^{-1} and RMSF^{-1}. Correlations were calculated for each protomer as a whole and for each separate domain (NTD, MiD, CTD).

			Protomer 1				Protomer 2			
			NTD	MiD	CTD	Whole	NTD	MiD	CTD	Whole
Fully-closed	BC vs L^{-1}	ADP	0.80	0.70	0.83	0.74	0.76	0.74	0.73	0.73
		ATP	0.79	0.66	0.81	0.70	0.71	0.64	0.75	0.66
		apo	0.74	0.69	0.70	0.70	0.76	0.61	0.66	0.64
	BC vs RMSF^{-1}	ADP	0.53	0.55	0.66	0.57	0.6	0.51	0.65	0.51
		ATP	0.57	0.55	0.57	0.55	0.55	0.44	0.53	A.48
		apo	0.60	0.37	0.53	0.47	0.58	0.41	0.55	0.45
	RMSF vs L	ADP	0.53	0.71	0.69	0.62	0.63	0.70	0.74	0.62
		ATP	0.63	0.75	0.73	0.66	0.69	0.70	0.36	A.57
		apo	0.80	0.65	0.67	0.71	0.71	0.65	0.67	0.63
Fully-open	BC vs L^{-1}	ADP	0.61	0.32	0.6	0.31	0.58	0.33	0.54	0.37
		ATP	0.59	0.32	0.54	0.31	0.64	0.33	0.52	0.35
		apo	0.58	0.32	0.62	0.33	0.61	0.37	0.56	0.37
	BC vs RMSF^{-1}	ADP	0.22	0.25	0.46	0.20	0.58	0.29	0.25	0.31
		ATP	0.54	0.33	0.31	0.32	0.59	0.38	0.29	0.33
		apo	0.52	0.40	0.45	0.35	0.45	0.56	0.35	0.44
	RMSF vs L	ADP	0.53	0.89	0.79	0.91	0.91	0.96	0.67	0.97
		ATP	0.95	0.93	0.64	0.95	0.85	0.96	0.61	0.95
		apo	0.89	0.86	0.76	0.93	0.88	0.80	0.77	0.86

- *BC is inversely proportional to L_i and RMSF*

As previously mentioned, BC is intricately coupled to L_i, in that it is a measure of the number of shortest paths that pass through a given node and thus informs on a residues frequency of usage. Regions of the protein that record high are likely to be structurally stable and experience minor fluctuations, increasing their availability to participate in network communication. Based on this assumption, RMSF and L_i could be expected to correlate inversely with respect to BC. In this study, both conformations demonstrate this displaying an inverse relationship between the peak BC regions both L_i and RMSF, where high BC regions correlate with regions of the protein that experience minor residue fluctuations and shorter L_i (Figure 3.13). We note that in both conformations, stronger correlations are observed between L_i and BC than between RMSF and BC (Table 3.4). This observation can be explained by the aforementioned inferences regarding the relationship between L_i and RMSF, whereby L_i may provide a more accurate description of the local network surrounding high BC regions as it is potentially a more stable descriptor compared to RMSF. Furthermore, we note that the Pearson's correlations between BC-RMSF and BC-L_i to be higher for the closed conformation complexes than the open complexes and suggest this to be a limitation of the Pearson's correlation coefficient which is more suitable to linear data sets. Nevertheless, good agreement is seen for peak BC regions, which correlate to some degree with local peaks in the RMSF and L_i curves but not necessarily for non-BC regions (Figure 3.13.)

Finally, as with L_i, we examine whether the distance cut-off used to construct the network graph has an effect on the inverse relationship between BC and RMSF for FO-ATP complex (Figure 3.15). In this analysis, increased distance cut-offs appear to have little effect on the selection of high BC regions, with peak residues remaining consistent over all three distance cut-offs examined (6.7 Å, 9.0 Å, and 12.0 Å). We note however that the frequency of usage for these peak residues is altered by changing the cut-off, but with no observable trend. For example, residues 501 and 523 have a higher degree of usage in protomer 1 for the 9.0 Å cut-off, compared to the 6.7 Å and 12.0 Å cut-offs. From a structural perspective, the larger cut-off distances appear to increase BC for flexible regions of the protein such as the flexible helix$_{17}$ and helix$_{18}$ regions (res 600-630). This observation is likely to arise when the magnitude of the residue fluctuations is insufficient to exceed the cut-off for connected residues and thus enable a higher probability for shortest path navigation via these flexible regions over the course of the MD trajectory. Comparison of the Pearson's correlation coefficient between RMSF^{-1} and BC for each distance cut-off confirms this, showing similar correlation coefficients over the three domains regardless of the cut-off used, except the CTDs which record stronger correlations for the smaller the distance cut-offs. This is likely due to the reduced impact of the flexible helix$_{17}$ and helix$_{18}$ for the larger distance cut-offs.

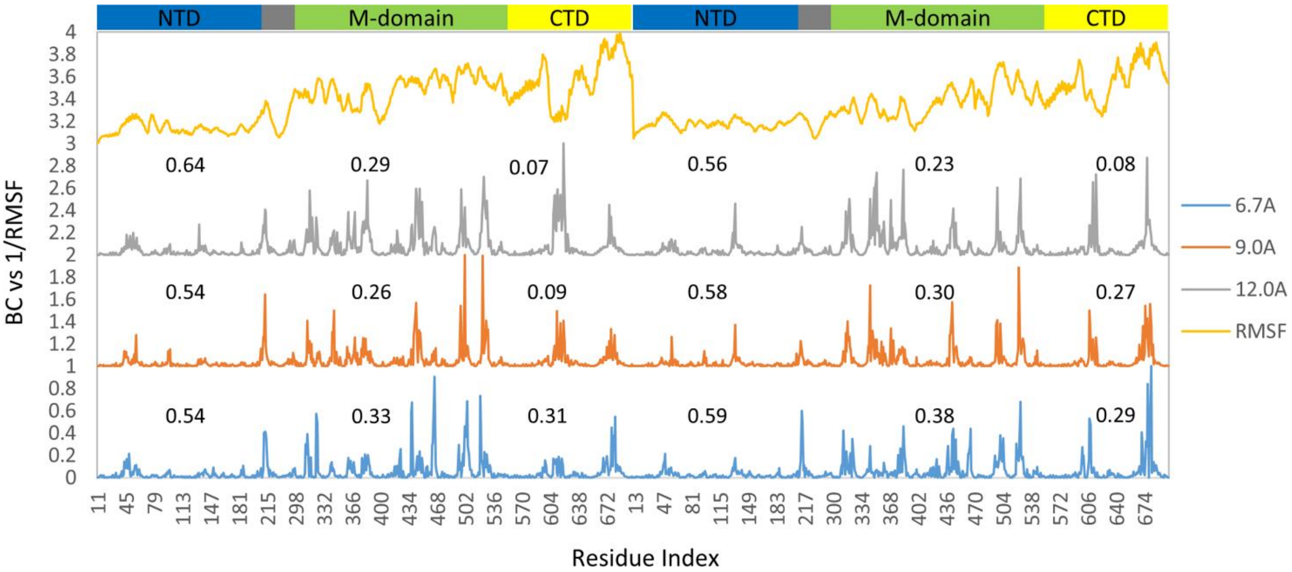


Figure 3.15: Inverse correlation between BC and RMSF, L_i with incremental distance cut-offs. Several network graphs were constructed based on 6.7 Å, 9.0 Å and 12.0 Å distance cut-offs for the FO-ATP complex. Whole protein correlation were 6.7 Å – 0.32, 9.0 Å – 0.25 and 12.0 Å – 0.20. Reproduced with permission from Penkler et. al 2018 [191]

To conclude this analysis, we find RMSF and L_i to be largely proportional to one another, but note that this relationship is intricately linked to (i) the magnitude of the distance fluctuations, (ii) the distance cut-off used to construct the network graphs, as well as (iii) the relative compactness of the local networks. Furthermore, we show for the first time that there exists an inverse relationship between these network measures and BC that holds true even for increased distance cut-offs, and note that the L_i descriptor may be better suited to inform on the functional relevance of the local network compared to RMSF. These observations were all confirmed for separate replicate trajectories in which similar trends were observed (Appendix I-IX). Finally, we propose that since not all residues displaying small fluctuations act as communication bridges, we suggest BC and L_i to be alternative and potentially more sensitive quantities to follow for the identification of novel functional sites on proteins.

3.3.5 Perturbation response scanning analysis

In this section PRS analysis was used to assess the allosteric potential of individual residues sited in the FC-ATP and FO-ATP complexes to selection conformational displacement towards the FO and FC conformations respectively.

PRS relies on linear response theory to predict conformational displacements in response to external force perturbations. The underlying theory and methodological approach is discussed at length in Chapter 2 of this thesis. In this study, we once again sequentially perturb each residue 250 times with random force perturbations. The resultant displacement of the whole protein in response to each perturbation was recorded in a 3D matrix and the collective data analysed in terms of: (1) average relative displacement - to gain insight on protein sensitivity to external force perturbation (new to this study), and (2) conformational overlap with a known target state - to assess the potential of each residue to select displacements that are representative of an expected conformational change (see Section 2.2.3). Together these analyses provide insight regarding sites on the protein that are potentially involved in modulating conformational dynamics.

- *Average residue displacements identify putative allosteric effector and sensor residues*

Analysis of the average response of the whole protein to 250 random force perturbations of each residue allows for the identification of allosteric sensor residues, which are defined as being distant sites on the protein that are allosterically sensitive to external force perturbations. Conversely, residues whose perturbations result in large displacements elsewhere in the protein can be seen as allosteric effectors, representing sites on the protein likely involved in external binding events such as specific co-chaperone and client binding in the case of Hsp90. Figure 3.16 shows the average response matrices for the FC-ATP and FO-ATP complexes as respective heat maps, where the ik^{th} element in a matrix refers to the average response (displacement) of residue k to external perturbation at residue i. Thus, bright spots indicate peak residues that experience large displacements, while darker areas denote residues that experience little or no displacement.

Accompanying each map, are the most influential 'effector' residues (rows) in the vertical right hand bar plots, and the most sensitive 'sensor' residues (columns) in the lower horizontal column plots. Each plot corresponds to the average displacement across the row or column.

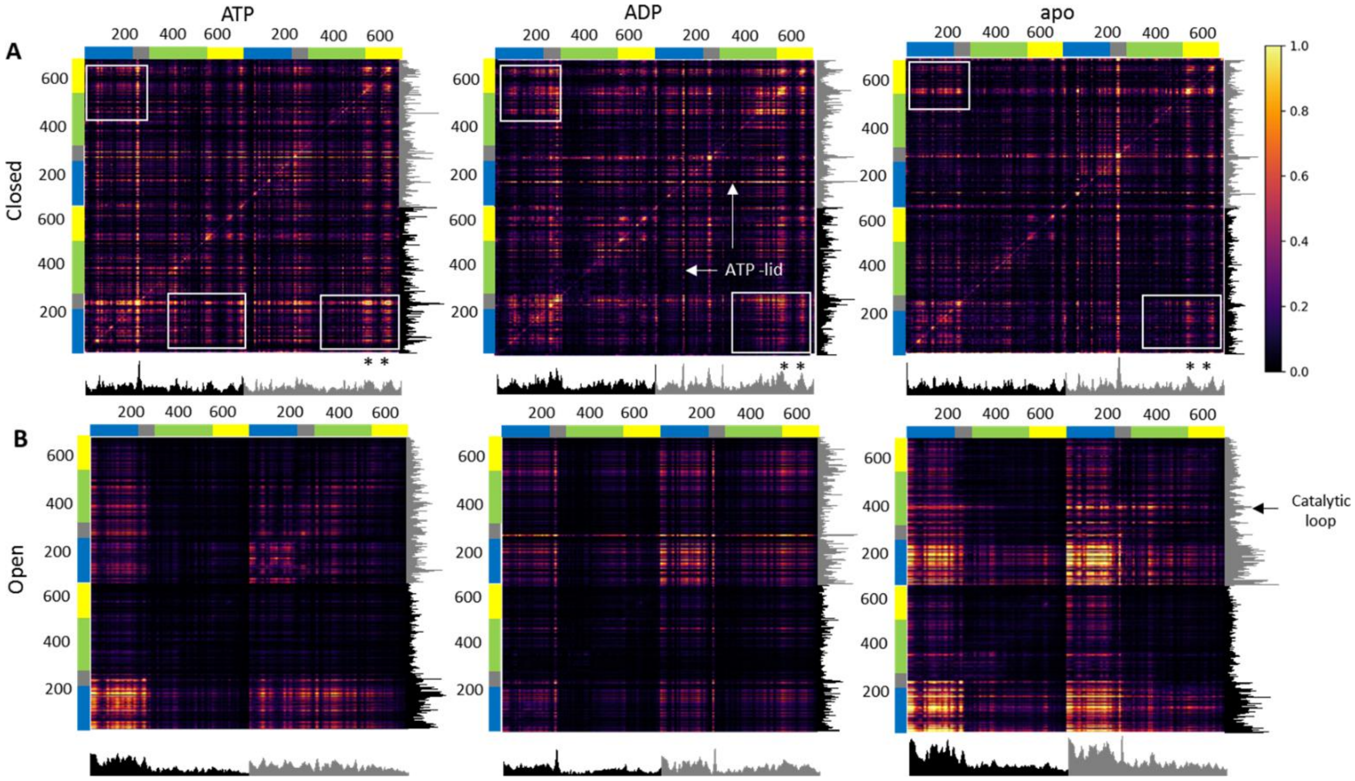


Figure 3.16: **PRS response maps showing effector and sensor regions of the protein.** (A) Closed and (B) fully open complexes. Columns denote sensor residues, while rows represent effector (perturbed) residues. Bright spots indicate highly sensitive residues that experience large displacements. Accompanying each map are 'effector' bar plots on the right-hand-side and at the bottom 'sensor' column plots (Protomer 1-black, Protomer 2-grey) which represent the sum of the matrix elements. (*) Denotes helix17 and helix20. Reproduced with permission from Penkler et. al 2018 [191]

Starting with the closed conformation complexes (Figure 3.16-A), comparing the displacement intensities of the different NTD configurations reveals the ADP complex to be the most sensitive to external force perturbations, with the apo complex being the least sensitive. This observation may suggest the ADP closed complex to be more amenable to conformational change compared to the other configurations, an observation that is in agreement with the hypothesis that ADP imparts protomer flexibility. Strong sensor and effector signals are recorded at the linker regions of all three complexes. These residues correspond to the G-G-G-G insert introduced during the homology modelling process to cover the missing linker region (see section 3.2.1). The high degree of displacement for this region can be expected as glycine residues are known to be highly mobile elements.

Several overlapping sensor peaks are observed in all three closed configuration complexes including: ß-strand$_1$ (res 15-20) ; helix$_1$ (res 21-36); the first turn of the NTD ß-sheet (res 81-88); the ATP-lid (res 120-129); ß-strand$_4$ (res 170-180); the loop connecting helix$_{12}$ and helix$_{13}$ (res 469-472); as well as helix$_{17}$ (res 550-590) and helix$_{20}$ (res 638-660) at the CTD interface. Nucleotide dependent differences are also observed for residues that correspond to Aha1 binding (res 300-308) [173], the catalytic loop and inter-domain hinge [302] (res 391-406), the phosphorylation site Y$_{313}$ [344], helix$_{13}$ (res 478-490), and the putative client binding site at helix$_{16}$ (res 523-530) experience increased sensitivity in the presence of ADP, but not the ATP and apo complexes.

Studying the corresponding effector sites for the FC complexes, inter-protomer coupling is observed between the NTD and CTD in all three complexes, where perturbations at the CTD (helix$_{17}$ and helix$_{20}$) of protomer 2 are strongly detected at the NTD of protomer 1 and *vice versa* (Figure 3.16-A, white boxes). For the ATP complex, NTD perturbations in protomer 1 appear to illicit a stronger CTD response in both protomers compared to the ADP and apo complexes, while CTD perturbations in protomer 2 of the ADP and apo complexes appear to illicit a stronger response at the NTD of protomer 1 compared to the ATP complex. This observed allosteric coupling between the two terminals is in agreement with a previous computational study demonstrating efficient allosteric communication between the NTD and CTD in response to bound nucleotide [297]. Other notable effector residues include the ATP-lid (res 121-126), particularly in the ADP bound complex, helix$_{10}$ (res 406-425), and residues surrounding A$_{469}$ and G$_{515}$ in the M-domain. These sites are significant in that ATP-lid stabilization is necessary to maintain the closed 'active' state, while helix$_{10}$ has been implicated in allosteric signal propagation between the NTD and MiD in HtpG [303]. Furthermore, all of the MiD sites are located on the surface of the protein, in positions easily accessible to co-factor binding.

The FO conformation complexes (Figure 3.16-B) show a clear differentiation between the different nucleotide configurations. However, all three complexes show strong allosteric coupling between the NTDs , where perturbations at the NTD of one protomer lead to strong displacement signals at NTD of the second protomer. Interestingly, the apo complex appears to be the most sensitive to NTD perturbations while the ADP complex appears to be the least sensitive. This observation suggests the apo complex to be more amenable to conformational change, which is an observation that is in line with previous reports describing stochastic conformational dynamics for apo Hsp90 [181,327,328]. Looking in more detail, it is also apparent that perturbations arriving at the NTD result in displacements at the MiD and CTD depending on the nucleotide configuration. MiD displacements at ß-turn E$_{332}$-E$_{336}$ and the catalytic loop (res 395-407) are observed within the perturbed protomer of the ATP and apo complexes, while CTD sensitivity at helix$_{17}$ (552-570) is observed in the presence of ADP. Most of these nucleotide specific sensor sites are also listed as allosteric effectors for the respective complexes, implying allosteric coupling between these sites.

- *Perturbation of key residues reveals sites induces functional conformational change*

In this section, PRS is used to identify the allosteric potential for select residue sites to select functional conformation in response to force perturbations. Each residue is perturbed at random a total 250 times and the resultant displacement vector compared to the expected conformational displacement required to reach the opposite functional state known experimental displacement changes i.e. fully-closed → fully-open and *vice versa*, and goodness of fit for this comparison is measured using the Pearson's correlation coefficient. In this manner, 250 correlation coefficients were recorded for each residue, and the highest correlation (C_i) selected to represent the maximum potential for that residue to invoke the expected conformational change towards the opposite state.

Table 3.5 | Summary list of PRS peak residues for FC and FO complexes.

	ATP	ADP	apo
Closed	23-41; 74; 80-83; 87; 105; 109; 110-119; 123; 124; 128-139; 146; 167; 194; 195; 197; 198-211; 221-223; 285; 286; 287; 288; 289; 295; 358; 366-368; 395-401; 404; 411	16; 19-50; 54; 56; 57; 60; 83; 88; 89; 91-93; 107-113; 116; 119; 130-132; 134; 135; 137-139; 142-145; 167; 174; 182; 184; 194; 196; 197; 203; 205; 206; 209-211; 214; 219; 227; 280; 282; 283; 291; 292; 338; 397-400	19-45; 82; 83; 85-89; 105; 106; 108-119; 123; 124; 127; 129; 130-136; 139-149; 165-167; 187-192; 195; 196; 203; 204; 207; 395-399; 407
Open	20-50; 64; 65; 67; 74; 78; 80-92; 98; 105-111; 113-150; 163-170; 173; 174; 186-214; 219-228; 278-283; 287-296; 298; 312-316; 319; 320; 326; 330-338; 341; 343; 347; 360; 361; 363; 376-393; 402-414; 416-418; 420; 443-449; 451-454; 458; 459; 469; 471; 472; 474; 476; 479; 489; 495; 497; 523-525; 527; 529; 530; 540; 541; 544-551; 562; 675; 699	24; 26; 29; 45-53; 78-93; 111-115; 135-137; 185; 186; 195-215; 218-224; 285-287; 296-330; 338; 388; 389; 398; 400; 410; 413	15; 29; 31; 34-42; 68; 69; 73; 74; 84; 85; 116-129; 143; 144; 191-200; 203; 206; 207; 210; 222; 223; 227; 228; 278-284; 286-291; 307; 332; 410; 412-414; 416; 525-530; 532; 533; 597; 607; 608; 654

Thus, the fully-closed complexes are assessed in terms of the opening transition and the final conformational state set to an equilibrated FO-ATP structure. Likewise an equilibrated FC-ATP structure set as the final state for the fully-open complex experiments to assess the closing transition. The results of these analyses shown in Figure 3.17 and Figure 3.18 respectively. To ensure the reproducibility of this technique for large flexible systems such as Hsp90, duplicate PRS calculations for each complex were calculated from separate MD trajectories and report good overlap for peak residues (Appendix I-X and Appendix I-XI). The highest C_i residues are summarized in Table 3.5 and their relative functional importance is discussed in more detail bellow.

- *Opening transition*

For the fully-closed conformation complexes (Figure 3.17-A), PRS indicates that regardless of bound nucleotide, no single residue is capable of selecting the open conformation in response to external force perturbation, with peak residues recording maximum C_i values of ± 0.60. Despite this, several C_i residue peaks ($C_i > 0.50$) appear to correspond to several functional sites previously identified and discussed in the BC analysis. Accentuated residues belonging to ß-strand$_1$ (res 16, 19, 20); helix$_1$ (res 23-40); the ATP-lid (res 107-113, 116, 119, 130-132, 134, 135, 137-139, 142-145); and the catalytic loop (res 397-400) form a central dimerization hub at the NTD and have been previously described as forming part of an NTD-MiD mechanical hinge [302] (Figure 3.17-B). Destabilization of these inter-protomer interfaces is crucial before protomer uncoupling and a conformational shift toward the open state can occur [340].

PRS also accentuates residues residing in helix$_2$ (res 41-65); ß-strand$_2$ (res 88-94); helix$_5$ (res 203, 205, 209-211); and ß-strand$_8$ (214, 219). Of these, helix$_2$ has been previously reported to play a crucial role in propagating allosteric signals from the NTD to the MiD in *E. coli* Hsp90 [303], while residues residing in ß-stand$_2$ have been associated with long range allosteric communication with the CTD [297]. The NTD/linker interface formed by helix$_5$ and ß-strand$_8$ has been previously implicated in modulating chaperone affinity [339] as well as modulation of conformational dynamics in response to bound nucleotide [343]. Comparing the highest C_i between complexes, residues belonging to the ATP-lid recorded the highest C_i values in the ATP and apo complexes (0.62 and 0.67 respectively), while residues belonging to ß-strand$_1$, helix$_1$, and the catalytic loop recorded the highest C_i values for the ADP complex.

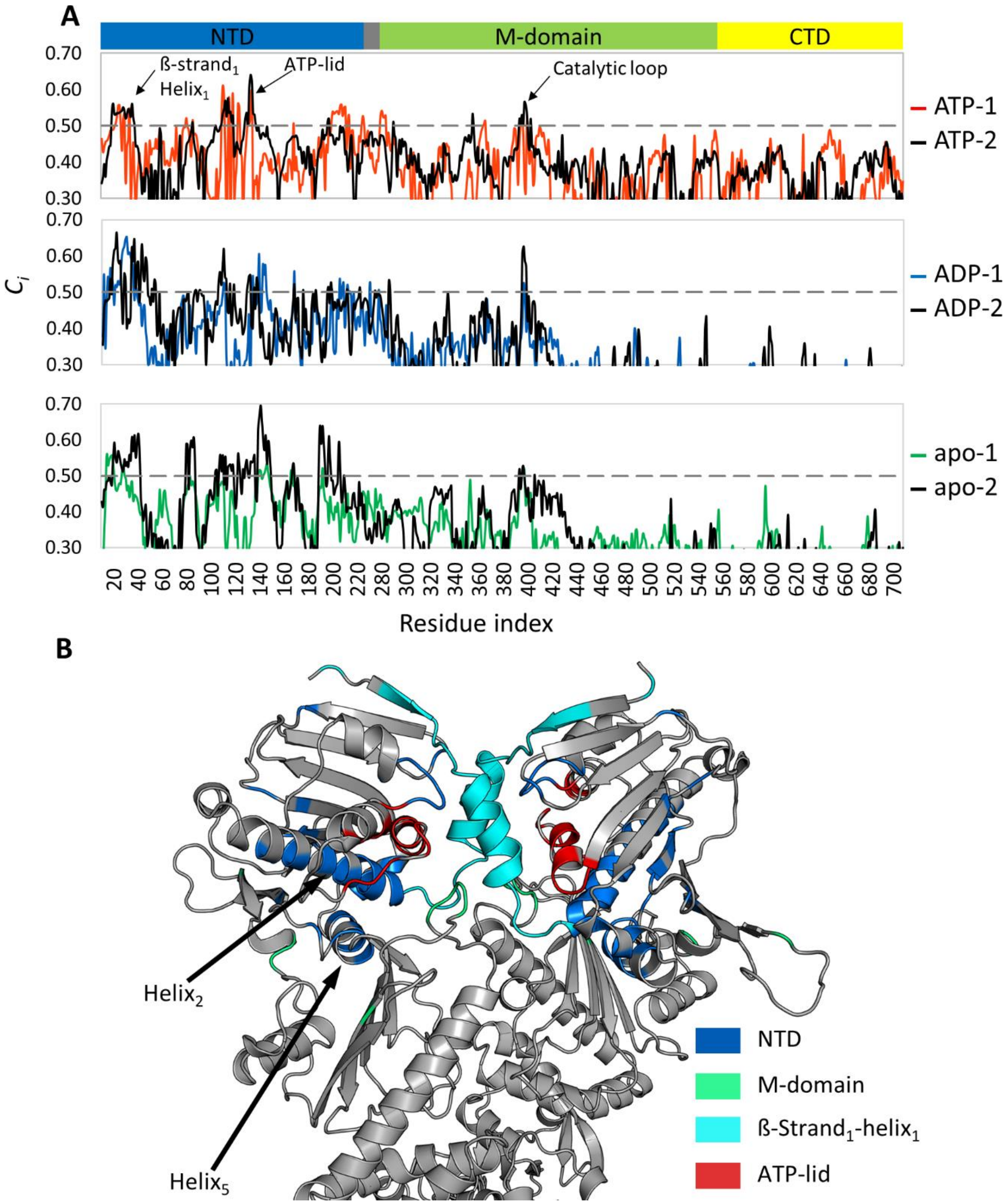


Figure 3.17: Residues accentuated by PRS in the closed conformation complexes. (A) PRS profiles showing accentuated residues which select the open conformation. Protomer 1 is coloured by NTD configuration (ATP–red, ADP–blue, apo–green), and protomer 2 in black. **(B)** Zoomed in view of the NTD of the closed conformation, showing the accentuation of the NTD dimerization interface of the ADP bound complex. Residues accentuated by PRS are coloured according to domain location. Penkler et. al 2018 [191]

- *Closing transition*

For the fully-open conformation complexes, the nucleotide configuration appears to have a marked differential effect on the potential of each complex to interconvert to the closed state (Figure 3.18-A). Comparing the percentage of residues with $C_i > 0.75$, the ATP (33%) and ADP (11%) complexes record far fewer than the apo complex (85%). The elevated number of $C_i > 0.75$ in the apo complex is interpreted to be indicative of a natural propensity for this complex to interconvert to the closed state, an observation in keeping with the hypothesis that the apo protein behaves stochastically [181,327,328]. In the case of the nucleotide bound complexes, PRS accentuates select residues capable of eliciting a conformational transition towards the closed state.

Looking at the C_i profile for the ATP bound complex (Figure 3.18-A), PRS accentuates several residue clusters in the NTD and MiD, which once again overlap with functional elements. In the NTD, these residues include: ß-strand$_1$ (res 20-24); helix$_1$ (res 25-35); helix$_2$ (res 43-53); the ATP-lid (res 113-136); ß-strand$_5$ (res 163-170, 173-174); ß-strand$_6$ (res 186-190); helix$_{6-7}$ (res 191-214); ß-strand$_8$ (res 219-224). The functional importance of ß-strand$_1$, helix$_{1-2}$, ß-strand$_8$, the ATP-lid, and the catalytic loop have already been discussed in context of the opening transition, however accentuation of these elements by PRS for the FO-ATP complex, and thus the closing transition also bears functional significance. It is widely believed that ATP binding initiates the closing transition by triggering the uncoupling of ß-strand$_1$ and helix$_1$ from the ATP-lid, allowing the latter close over the nucleotide binding pocket (see Figure 3.6). This conformational repositioning of the ATP-lid not only entraps bound nucleotide, but also facilitates N-terminal dimerization through exposure of a hydrophobic surface that is essential for successful ß-strand$_1$ swapping and stabilization of the closed active complex [340]. In the MiD of the FO-ATP complexes PRS also accentuates: helix$_9$ (res 312-316); several residues belonging to the ß-sheet (res 319, 320, 326, 330-338, 341, 343, 347, 360, 361, 363, 376-393); the catalytic loop (res 402-410); helix$_{10}$ (res 411, 412, 416-418, 420); helix$_{12}$ (res 443-449, 451-454); helix$_{16}$ (res 523-525, 527, 529, 530); and lastly loop residues T540-L551. It is well known that ATPase activation of the closed complex is only fully achieved when the MiDs dock onto their respective NTDs, a conformational rearrangement that enables the MiD catalytic loop to be repositioned such that R4$_{00}$ projects into the nucleotide binding pocket to coordinate with the γ-phosphate of ATP and thus facilitate the hydrolysis reaction [162,183]. It is also interesting to note that PRS accentuates ß-strand$_1$, helix$_1$, and the ATP-lid in the FO-ATP complex, but not in the FO-ADP complex (Figure 3.18-A), a finding that supports the hypothesis that ATP binding allosterically activates the closing transition, and that bound ADP stabilizes the 'v-like' open conformation.

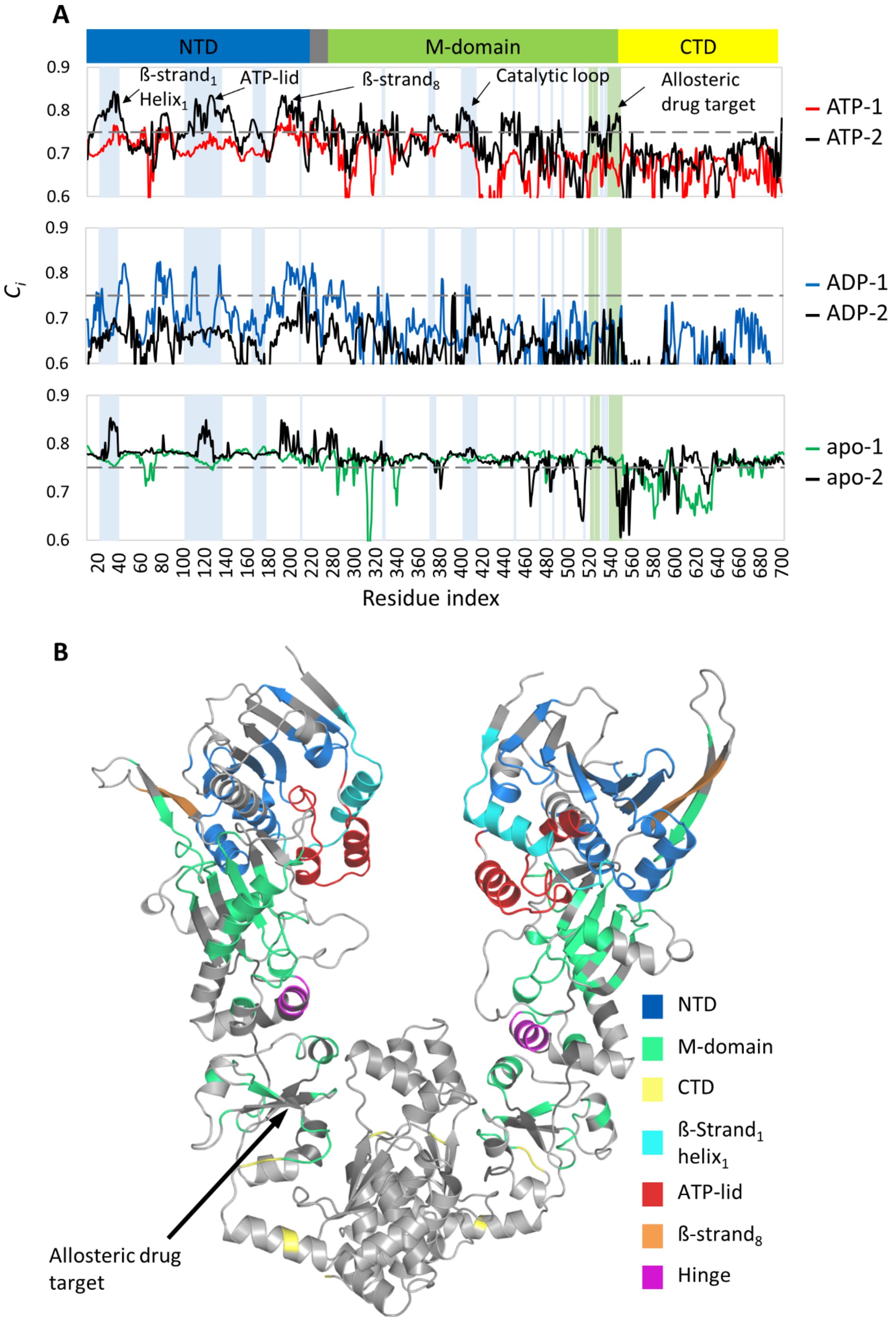


Figure 3.18: Residues accentuated by PRS in the open conformation complexes. (A) PRS profiles showing accentuated residues capable of selecting the closed conformation. Protomer 1 is coloured by NTD configuration (ATP–red, ADP–blue, apo–green), and protomer 2 in black. Blue shading represents overlap with Aha1 binding residues and the green shading client binding residues. **(B)** Structural mapping of residues accentuated by PRS for the FO-ATP complex coloured by location. Penkler et. al 2018 [191]

The conformational dynamics required for ATPase activation in Hsp90 are thought to be the rate limiting step and an inherently slow process, recording rate constants in the order of minutes [161,184,185,349], and a recent study has linked the rate constant of ATP hydrolysis with the rate constants of ATP-lid closure, ß-strand swapping, and intra-protomer association of the NTD and M-domains [187]. Furthermore, the authors of this study propose a two-step mechanism for ATP-lid closure and find evidence of a previously unknown mode of action for the co-chaperone Aha1, where they suggest its involvement at this stage of the conformational cycle facilitates early mobilization of the ATP-lid and thus enhanced ATPase activity [187]. It is thus interesting to note the residues accentuated by PRS in the ATP complex are clustered around NTD and M-domain sites that have been previously implicated in Aha1 binding [162,350] (Figure 3.18-A, blue shading). Given this correlation, we propose that the accentuating perturbations arriving at ß-strand$_1$, helix$_1$, and the ATP-lid may occur naturally through binding interactions with Aha1, providing possible evidence in support of Aha1 assisted ATP-lid closure and accelerated inter-state conversion to the closed state.

Also of functional importance and accentuated by PRS in the ATP complex are residues belonging to helix$_{12}$ located at the MiD-CTD hinge [302,305], as well as helix$_{16}$ and loop T$_{540}$-L$_{551}$ which are thought to be important regulatory sites for client binding [305,347,348] (Figure 3.18-A, green shading). Interestingly, this CTD region has been previously reported to be of interest as an allosteric drug target site [299], and subsequent drug discovery studies have demonstrated how bound ligands at this site can allosterically enhance ATPase activity through the asymmetric modulation of protomer conformational dynamics [300,351].

On the basis of the correlation between our PRS data and Aha1 binding we propose PRS to be a suitable technique for identifying *de facto* co-factor binding sites that may be specifically involved in allosteric modulation of conformational dynamics. To tests this hypothesis, we analysed the PRS data for the FO-ATP complex in terms of sites known to be involved in binding interactions with the heat shock organizing protein (HOP) (res: E$_{307}$, E$_{311}$, K$_{314}$, N$_{318}$, W$_{320}$, D3$_{22}$, K$_{431}$, E$_{432}$, E$_{473}$, A$_{469}$, T$_{482}$, and E$_{486}$) [352,353]. HOP is thought to modulate Hsp90's conformational dynamics by interrupting NTD dimerization through steric interference rather than allosteric mechanisms [159], and thus presents as a suitable negative control. With the exception of K$_{314}$ and W$_{320}$, PRS does not accentuate the HOP binding residues in the FO-ATP, confirming that co-factor binding residues identified by PRS are likely allosteric modulators of conformational dynamics.

3.4 Conclusions

To date, numerous biochemical and computational studies have made considerable advances towards understanding the modulation of Hsp90's complex conformational cycle. In this study, we have assessed the current opinions regarding Hsp90's conformational dynamics with respect to the human Hsp90α isoform in a comprehensive analysis of full-length homology models of the human Hsp90α using all-atom MD simulations coupled with DRN analysis and PRS techniques.

Within the limitations of all-atom MD simulations, our results are in agreement with previous studies describing a differential effect of bound nucleotide on conformational dynamics in which ATP drives NTD dimerization and the closing transition, while the ADP/apo complexes favour the open Hsp90 dimer [163,180,188,328,354]. Regardless of the conformation, bound ATP was found to 'tense' the dimer complex in favour of the closed conformation, while ADP appeared to 'relax' the protomers affording a greater degree of flexibility. Bound ADP in the partially-open complex increased protomer flexibility, resulting in global conformational rearrangements representing the opening transition yielding the fully-open client loading conformation [180,182,188]. Analysis of the change in residue reachability over time confirmed the 'tensing' and 'relaxing' effect of ATP and ADP respectively, with notable reduction in average path length in the former, and variable increased path lengths in the latter. Locations in the protein recording large changes in reachability corresponded to key functional elements such as ß-strand$_1$, helix$_1$, ATP-lid, and catalytic loop, implicating these regions in steering conformational dynamics. Betweenness centrality analysis provided a measure for distinguishing sites in the protein responsible for controlling intra-protein communication. These sites were found to correlate closely with known functional sites, demonstrating how BC analysis may be used to identify functional sites on a protein.

Finally PRS was used to probe both the fully-closed and fully-open conformation complexes, to assess each residue's allosteric potential to affect conformational change functionally related or otherwise. Analysis of the average response of the whole protein to external force perturbation revealed an inter-protomer allosteric coupling between the NTD and CTD of the closed complexes, in which perturbations arriving at the terminal end of one protomer are felt or 'sensed' at the opposite terminal of the other protomer. This observation is in agreement with a recent study demonstrating efficient communication between the two domains, leading to the discovery of a CTD allosteric drug target site [203,297,299,355]. In a similar manner, analysis of the open complexes demonstrated allosteric coupling between the NTDs of each protomer, the apo complex showing particular sensitivity to perturbations at the NTDs, and an increased sensitivity of the ATP complex over the ADP complex, suggesting ATP driven allosteric activation of this complex.

It is well documented that regulation of Hsp90's conformational cycle is impacted by various co-chaperone interactions [159,326,328]. The perturbations utilized in PRS provide a novel way of introducing external influences on select residues that could be likened to the natural forces involved in co-factor and/or client binding. PRS analysis of the closed complexes revealed no single residue perturbations capable of selecting a coordinate change towards the open conformational state and it is thus likely that the opening transition occurs spontaneously without the aid of external binding partners. Interestingly, the highest C_i residues selected by PRS for the closed complexes include residues that make up the primary NTD dimerization site, including ß-strand$_1$, helix$_1$, ATP-lid, and M-domain catalytic loop, elements that have been previously reported to form an important NTD/M-domain hinge implicated in conformational dynamics [302]. The implication of the co-chaperone p23 in stabilizing the closed conformation may present as an alternative functional explanation for the accentuation of these dimerization regions, in which co-chaperone uncoupling from this region may invoke 'pulling' forces that in turn provide the necessary trigger for the opening transition initiating an allosteric signal that is propagated to the CTD.

PRS analysis of the open conformation complexes revealed several sites in the NTD and MiD capable of selecting a functionally relevant conformational displacements towards the closed form. Most of the residues at these sites also correlate to known functional sites, while several map to regions on the surface of the protein believed to be involved in co-chaperone binding [159]. In particular, PRS accentuates a number of residues in the ß-strand$_1$, helix$_1$, ATP-lid, and catalytic loop regions of the FO-ATP complex. These residues correspond with sites implicated in Aha1 binding interactions, suggesting that perturbations at these sites may arrive naturally through co-chaperone binding interactions. This hypothesis is in keeping with recent reports implicating Aha1 in early ATP-lid mobilization [187], and acceleration of ATPase activation [190], and allosteric modulation of Hsp90's conformational dynamics in favour of the closed conformational state [307]. The accentuation of these residues in the FO-ATP but not the FO-ADP complex, reinforces the hypothesis that ATP allosterically activates or primes the open complex for conformational transition towards the closed state.

Finally it should be noted that the perturbation nature of allostery has been widely examined in several experimental and computational studies [356]. One such study presented an alternative structure-based statistical mechanical model of allostery (SBSMMA) [357] which evaluates the free energy of allosteric signalling caused by the ligand binding and mutations as perturbations introducing a quantitative measure of per-residue allosteric potential defined as *microscopic potential*. This approach is in agreement with the methodology described in this thesis and similarities can be made regarding the stabilizing (UP) and destabilizing (DOWN) perturbations, which are defined by the authors as either increasing (stiffening) or decreasing (loosening) interactions with the residue’s neighbours respectively. Indeed, this precise aspect of dynamic allostery can be investigated and quantified from a granular and global perspective by coupling DRN analysis with PRS.

In conclusion, this study presents PRS coupled with MD simulations and DRN analysis as a novel package for detecting allosteric control elements in a large multi-domain proteins that are capable of modulating the conformational dynamics. From a drug discovery perspective, the methodology presented here could be applied to any allosteric system to rapidly identify potential allosteric sites that are capable of fine-tuning the conformational dynamics. These sites could provide the basis for small molecule allosteric drug targeting studies.

Chapter 4:

Modulation of Human Hsp90α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain

This chapter describes a follow up study to the findings presented in Chapter 3 of this thesis in which previously identified allosteric control elements located that the C-terminal domain of Hsp90α are probed with small drug like molecules to assess the allosteric potential of ligand binding to modulate the conformational dynamics of the chaperone.

Chapter 4 is reproduced in part with permission from the following publication:

Modulation of Human Hsp90α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain. David L. Penkler and Özlem Tastan Bishop. *Scientific Reports.* **2018**. (*Accepted. In press).* 2018. Nature

Contribution: Conceived the study with assistance from Özlem Tastan Bishop, designed and carried out all experimental procedures, and wrote the first draft of the manuscript.

4.1 Introduction

4.1.1 Hsp90 inhibition

The first notion of the relative importance of ATP hydrolysis in Hsp90 became evident when Whitsell and colleagues first discovered Hsp90 inhibition by the ansamycin antibiotic Geldanamycin in 1994 (GA) [358]. This was soon followed by the alternative inhibition by the macrolide radicicol by Schulte et al. 1998 [359]. Later crystallography efforts revealed that both GA and radicicol preferentially bind to the ATP binding pocket and thereby act as competitive inhibitors of nucleotide binding and thereby inhibit ATPase activity in Hsp90 [172] and ultimately the formation of the active mature client complex. Inhibition of Hsp90 by GA and radicicol meant that client proteins that would have otherwise been rescued by Hsp90 are targeted for ubiquitination and proteasomal degradation. Armed with the knowledge and prospect of inhibiting Hsp90 by small molecule interaction at the NTD provided a novel research tool and lead to an explosion of Hsp90 related studies regarding Hsp90's chaperone activity, enabling the identification of numerous client protein and co-chaperones. To date, numerous Hsp90 inhibitors that target the ATPase domain have been identified, most of which demonstrate potent anti-proliferative effects (see ref [341,360,361] for review). The progression of these compounds through clinical trials has been severely limited however due to high levels of toxicity associated with an induced heat shock response [220]. For this reason, recent research efforts have shifted focus away from the NTD ATPase binding site to targeting distant functional sites on the protein as alternative drug treatment strategies.

One approach saw the design and development of inhibitors that target the formation of specific Hsp90-co-chaperone complexes, which led to the discovery of inhibitors such as Derrubone, Withaferin A, and Celastrol which block the formation of the Hsp90-Cdc37 complex [362] and 1,6-dimethyl-3-propylpyrimido[5,4-e][1,2,4]triazine-5,7-dione that interrupted complex formation with HOP [363]. However the discovery of a secondary nucleotide binding site at the CTD led to the design and development of several CTD allosteric inhibitors including the coumarin based antibiotic Novobiocin that has been shown to allosterically destabilise the Hsp90 dimer by interfering with CTD dimerization leading to the dissociation of bound clients [178,364,365]. These CTD inhibitors of Hsp90 displayed a distinct advantage over the NTD ATPase inhibitors, in that they do not appear to induce secondary heat shock responses and thus incur lower levels of toxicity [362].

The discovery of Novobiocin and the coumarin based inhibitors was a fundamental development for the advent of allosteric modulators of Hsp90. Several studies reported evidence of allosteric coupling between the NTD and CTD [191,297,302,346,366] raising the possibility that nucleotide driven conformational restructuring may be influenced and fine-tuned by allosteric ligand binding events at the CTD. This line of enquiry soon lead to the design and development of several alternative CTD modulators such as the biphenyl scaffold and noviose based compounds as well as Epigallocatechin-3-gallate and coumermycin A1, all of which resulted in specific client protein degradation most likely due the inhibition of NTD dimerization and thus ATPase activity (see [367] for a comprehensive review). A more recent biochemical study provided evidence of Bisphenol A based CTD inhibitors of human Hsp90α that bind a site separate from the nucleotide binding site [368] and display anti-proliferative activity in several tumour cell lines [369]

It soon became evident that allosteric modulation by ligand interaction was not exclusive to the CTD with the emergence of several alternative NTD and MiD targeted allosteric inhibitors. The natural compound Gambogic acid was shown to bind the NTD but at a distinct site from the ATPase domain [370]. This molecule was shown to inhibit cell proliferation, determine degradation of stringent client proteins, disrupt co-chaperone interactions with Hsp70 and Cdc37 and increase Hs70 expression levels. In another study the McAlpine group identified allosteric modulators derived from the natural compound Sansalvamide A, which were shown to interact with at the NTD:MiD interface of Hsp90, where it interfered with the several binding partners including HOP, inositol hexakisphosphate kinase 2, and FKBP52 [371–373].

Inasmuch as allosteric inhibitors offer alternative Hsp90 disease related treatment strategies, allosteric modulators that enhance protein function offer a novel tool for probing mechanistic information on biochemical processes [374]. In this context, Yokoyama and colleagues have reported goniothalamin to target the NTD of *E. coli* HtpG where it acts as an ATPase activator [375]. The Buchner laboratory identifies allosteric activators of chaperone closure kinetics using FRET-based screening analysis, that also bind the NTD at a pocket close to helix α2 [376]. Computational studies based on the closed conformation of yeast Hsp90 identified a putative allosteric binding site located at the MiD:CTD inter-protomer interface that is implicated in allosteric modulation of conformational dynamics that favour ATPase activation [297,299,366]. Subsequent to this discovery, follow up studies led to the development of several allosteric CTD ligands that that target this allosteric pocket and appear to enhance ATPase activity up to six-fold by promoting conformational dynamics in favour of the ATPase active closed conformation [300,351]. Interestingly over enhancing ATPase activity by allosteric intervention has been shown to have a detrimental effect on cell viability through mutational and biophysical analysis [346,377]. More recently, Prodromou and colleagues demonstrate that Dihydropyridines derivatives that target the MiD:CTD interface of Hsp90 are capable of inducing ATPase activation and upregulate the heat shock response that induces a neuroprotecive effect in mouse models of Alzheimer Disease [378].

As is evidenced by the aforementioned studies, allosteric modulation of Hsp90 is gaining interest and showing great promise as an alternative and attractive drug treatment strategy. Allosteric modulators can be used as probes to antagonize and fine tune chaperone function by inducing function specific conformational changes.

4.1.2 Chapter objectives

This chapter describes a follow up study to the findings presented in Chapter 3 of this thesis. In the previous chapter, PRS coupled with DRN analysis was presented as an effective method for the prediction of functional sites on human Hsp90α that may be involved in allosterically modulating the chaperones conformational dynamics. Of particular interest, the previous study demonstrated allosteric coupling between the NTD and CTD in response to force perturbations at novel allosteric hot residues at either end, giving rise to the hypothesis that ATPase activity at the NTD may be allosterically modulated by binding events at the CTD. Herein, we set out to test this hypothesis by conducting a series of molecular docking experiments targeted at two novel allosteric binding sites predicted by PRS analysis that are located at the CTD of the fully-open ‘v-like’ conformation of human Hsp90α. We show for the first time *in silico* evidence for the allosteric modulation of not only the human protein, but also for the fully-open structure for which experimental data does not yet exist.

The main objectives of this study were thus two-fold: (i) to assess the drugability of PRS identified allosteric sites by high throughput molecular screening, and (ii) to evaluate the allosteric effect bound ligands at these sites have on the conformational dynamics of the fully-open form of the protein using MD based analysis techniques. It is important to note that drug like lead compound identification was not a primary aim of the study. Overall, the methodologies presented here were designed in such a way as to provide ‘proof of concept’ for the use of PRS, DRN, and molecular docking as a suitable collective approach for identifying and testing putative allosteric binding sites on a protein that are capable of modulating the functional dynamics of the protein.

4.2 Methodology

4.2.1 Molecular docking

The 3D coordinate data for human Hsp90α in an open “v-like” conformation was obtained from previous MD simulations of a homology model of human Hsp90 in the presence of various nucleotide conditions [191]. Clustering analysis was carried out on a 400 ns all-atom MD trajectory of Hsp90α in complex with ATP according to the methodology described by Daura and co-workers [379], and the frame with the smallest RMSD to the average of the largest cluster selected the representative protein receptor. Natural compounds indigenous to Southern Africa were obtained from SANCDB [380], while the known Hsp90 inhibitor Novobiocin was retrieved for the ZINC database. Both the compounds and the protein were prepared for docking using the AutoDockTools software suite [381]. The previously identified allosteric binding sites were mapped onto two separate grids with AutoGrid4 [381]. Site-1 was centred over the four-helix bundle located at the CTD interface (residues G668-A685), and Site-2 on the adjacent pocket located on protomer B (residues H491-V508, V542-Q550, P596-N609). All docking calculations were performed with AutoDock4 [381] using the Lamarckian Genetic algorithm with a population size of 150, and the number of evaluations and generations set to 10 000 000. A total of 100 docking runs were performed for each compound, and each run evaluated with the semi-empirical scoring function supplied by AutoDock4. Clustering of the docked conformations was carried out using AutoDockTools with a cut-off of 1.5 Å. Docking of a compound was deemed reproducible if the largest cluster exceeded 50 % of the total runs, with an average energy score of less than -8.00 kcal/mol. The compounds were ranked by average energy score, and the lowest scoring conformation of the best candidates selected as the representative conformation for long range all-atom MD simulations.

4.2.2 Molecular dynamics

All MD simulations were performed using GROMACS 5.1.2 [265,267] with in the CHARMM 36 force field [268–270] using the methodology previously described in section 3.2.2. All trajectory clustering analyses were carried out with the GROMACS 5 gmx_cluster function using the gromos method described by Daura and co-workers [379] using a cut-off of 0.2 nm.

4.2.3 Protein-ligand interactions

A custom Python script was designed to analyse protein-ligand interactions over the course of each MD trajectory. Briefly, the MD trajectory was read frame-by-frame every 100 ps using iterator tool from the MDTraj module. Each frame was converted to PDB format and then passed to the Protein

Ligand Interaction Profiler (PLIP) ^{17} as input to predict all protein-ligand interactions present at that time point. Thus for a sampling step of 100 ps, each 200 ns trajectory was reduced to exactly 2000 frames and thus 2000 protein-ligand interaction reports from PLIP. Each PLIP report was parsed using a custom parsing tool, to extract and separate all interaction types between the ligand and either protomer. In this manner for each time point, those residues involved in either hydrophobic or hydrogen bond interactions with the bound ligand were extracted and stored in a suitable data frame allowing for the time evolution of bond formation to be assessed. Finally the matplotlib module was used to write out interaction heat maps for each protomer in a binary manner where 1 = interaction present and 0 = no interaction for all residues involved in ligand binding over the MD trajectory.

4.2.4 Dynamic residue networks

Inter- and intra-domain communication was analysed using the L_i and BC network descriptors. Both metrics were calculated using MD-TASK [295] (see section 3.2.4 for more details on the theory). A 200 ps time interval and a interaction distance cut-off of 6.7 Å as laid out in [46]. To assess the effect of bound ligand with respect to the ligand-free complex the Δ*L* and ΔBC metrics were considered:

$$\Delta L = L_i^{ligand} - L_i^{ligand-free}$$

$$\Delta \mathrm{BC} = \mathrm{BC}_i^{ligand} - \mathrm{BC}_i^{ligand-free}$$

4.2.5 Inter-domain contacts

Inter-domain contacts between the NTD and MiD were evaluated using the contact_map.py script from MD-TASK [295] with a distance cut-off of 6.7 Å. A custom Python script was designed to run contact_map.py every 200 ps and parse out any contracts between the two domains and the resultant data collated into a single data frame for plotting purposes. The percentage interaction metric was calculated by determining the fraction of frames that contained a particular interaction for the total 1000 frames.

4.2.6 Communication propensity

The pairwise communication propensity (*CP*) describes the efficiency of communication between residues *i* and *j* and is based on the notion that signal transduction events in proteins are directly related to the distance fluctuation of the communicating atoms [297,382]. *CP* is thus defined as a function of the inter-residue distance fluctuations, where residues whose Cα -Cα distance fluctuate with low intensity are seen to communicate more efficiently (faster) compared to residues whose distance fluctuations are large in which the amplitude of the fluctuations results in slower inter-residue communication. *CP* is calculated as the mean-square fluctuation of the inter-residue distance

$$CP = \langle \left(d_{ij} - d_{ij,ave}\right)^2 \rangle$$

where d_{ij} is the time dependent distance between the Cα atoms of residues i and j respectively, and the brackets denote the time-average over the trajectory. In this study, the N x N ΔCP matrix is used to assess the relative impact each compound has on the internal dynamics and overall flexibility of the protein, and is calculated as the difference matrix between each ligand bound complex to the ligand-free (ATP only) complex.

4.2.7 Principal component analysis

Principal completed analysis (PCA) was used for the analysis of global motions present over the course of several MD trajectories [383]. This technique involves two main steps: 1) the construction the covariance matrix, **C**, based on the positional deviation of each Cα atom, and 2) dimensionality reduction of **C** by diagonalization, to obtain the eigenvectors and eigenvalues. Each $3N$ x N covariance matrix was calculated based on an ensemble of protein structures obtained from the respective MD simulation and the elements of **C** defined as

$$\boldsymbol{C} = \langle (x_i - \langle x_i \rangle)\left(x_j - \langle x_j \rangle\right) \rangle$$

where x_i and x_j are atomic coordinates of each Cα atom, and the brackets denote the average. Eigenvectors with the largest eigenvalues are representative of the slowest modes, and generally are associated with large-scale movements in proteins, which are responsible for protein function. All PCA analyses were conducted using the GROMACS 5 software suite. The covar function was used for the construction and diagonalization of **C**, not including bound ligands and excluding the first 10 ns of trajectory data to avoid equilibration artefacts. The anaeig function was used to project the MD trajectories onto the main eigenvectors.

For the conformational sampling experiment, PCA analysis was conducted using a concatenated trajectory comprised of 20 ns equilibrated trajectory segments of human Hsp90α in the fully-closed, partially-closed, open, and fully-open conformations obtained from previous MD simulations [191]. The 20 ns trajectory segments were concatenated using the GROMACS 5 trjcat program, and PCA analysis conducted on the resultant trajectory as described above.

4.3 Results and discussion

4.3.1 Identification of ligand binding sites at the CTD

This study set out to identify potential ligand binding sites at the CTD of the FO ‘v-like’ conformation of human Hsp90α that are in close proximity or overlap directly with allosteric control elements identified by PRS analysis. These sites are then probed with a library of small natural compound ligands and the allosteric potential of the bound ligands assessed in terms of conformational dynamics.

The first step in this study was thus to screen the CTD for putative binding sites capable of binding small molecule ligands. This screening process was achieved using the FTMap webserver [384] with an unbiased screen positioned over the MiD:CTD region of the FO 'v-like' structure obtained from our previous study on human Hsp90α [191] (see 3.3.2). FTMap is a computational mapping server that is capable of identifying binding hotspots on static macromolecules by assessing the major contributions to ligand-binding free energy. Briefly, FTMap accepts a PDB coordinate file of a macromolecule and probes its surface with billions of small organic molecules, and scores their pose based on the free energy of biding using a detailed energy expression [384]. Regions of the macromolecule that bind clusters of multiple probe types are interpreted putative ligand binding hotspots in good agreement with experimental data. The authors of FTMap claim that it is more accurate than the classical mapping methods such as GRID and MCSS, and is faster than the more-recent mixed molecular dynamics based protein mapping approaches.

The results from the FTMap analysis indicated two candidate ligand binding sites at the CTD of Hsp90α (Figure 4.1). Closer inspection of these binding sites revealed sufficient overlap with several allosteric hot residues previously identified by PRS. Site-1 is located precisely at the CTD interface in a groove positioned at the four-helix bundle (helix$_{21,22}$) (Figure 4.1 & Figure 4.3-yellow). Site-2 is adjacent to Site-1 and represents a sub-pocket positioned at the MiD:CTD interface of each protomer (Figure 4.1 2A and 2B, Figure 4.3-blue). Each sub-pocket is formed by residues belonging to helix$_{18}$, ß-sheet$_{17-18}$, and helix$_{22}$, as well as several residues located at the MiD:CTD hinge region. For this study we selected Site-2B for targeted molecular screening experiments due to the increased diversity of the clustered organic probes and hereon refer to Site-2B as Site-2.

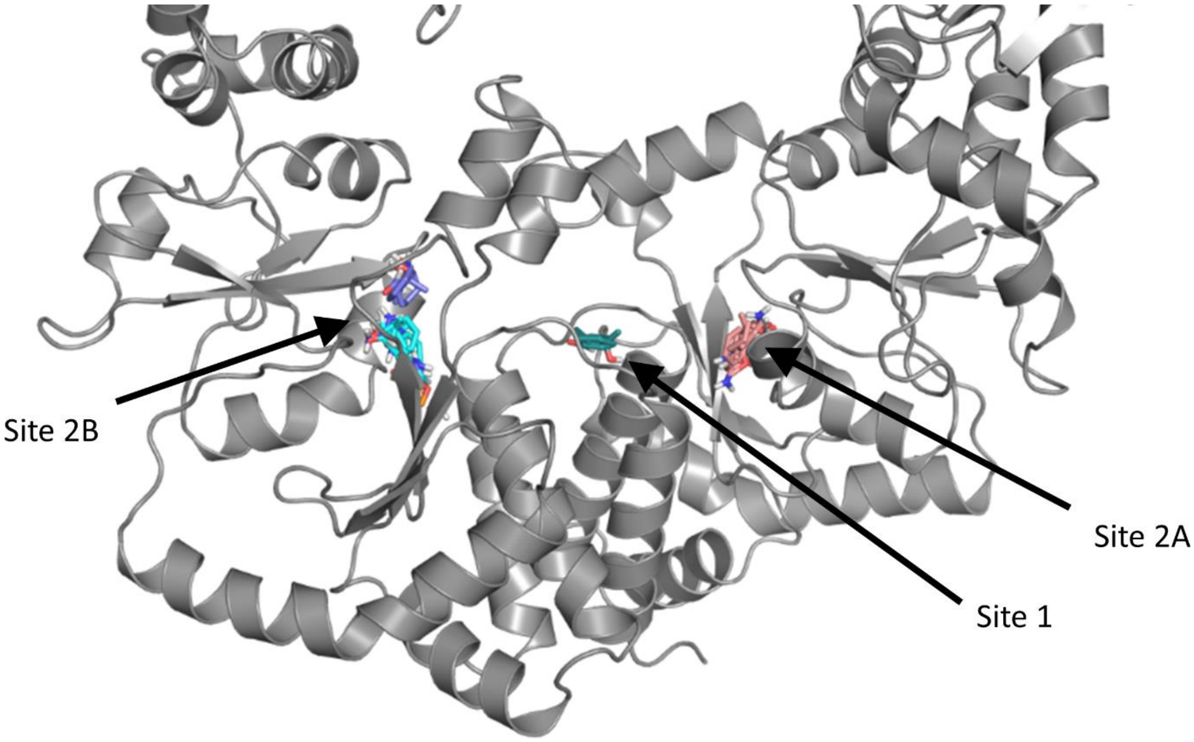


Figure 4.1: FTMap screen of the CTD: Showing two putative ligand binding sites: Site 1 located at the four-helix bundle, and Site 2 located adjacent to Site1 and is mirrored in both protomers (Site 2A and Site 2B respectively). Reproduced with permission from Penkler et. al. 2018 [385]

4.3.2 Identification of novel natural compounds that interact with the CTD

The next step was to perform molecular docking techniques to separately screen a library of 702 South African natural compounds against Site-1 and Site-2. The natural compounds were obtained from the SANCDB in a ready to dock format [380]. The known CTD inhibitor Novobiocin was included in both screens as a control. Due to computational constraints, the protein receptor was kept rigid and the ligands allowed flexibility around available rotatable bonds. Each compound was re-docked 100 times to ensure conformational diversity and the results filtered and analysed in terms of conformation cluster size and average estimated free energy of binding. Table 4.1 provides a summary of the best docked compounds for each site.

Table 4.1 | Summary of the best docked compounds for Site-1 and Site-2

Compound	Name	Site	Cluster size	Average energy (kcal/mol)	Lowest energy (kcal/mol)
Novobiocin	Novobiocin	1	81	-8.75	-8.97
SANC309	3'-Bromorubrolide F	2	79	-9.13	-9.20
SANC491	Cephalostatin 17	1	93	-11.41	-11.44
SANC518	20(29)-Lupene-3β-isoferulate	1	84	-11.03	-11.46

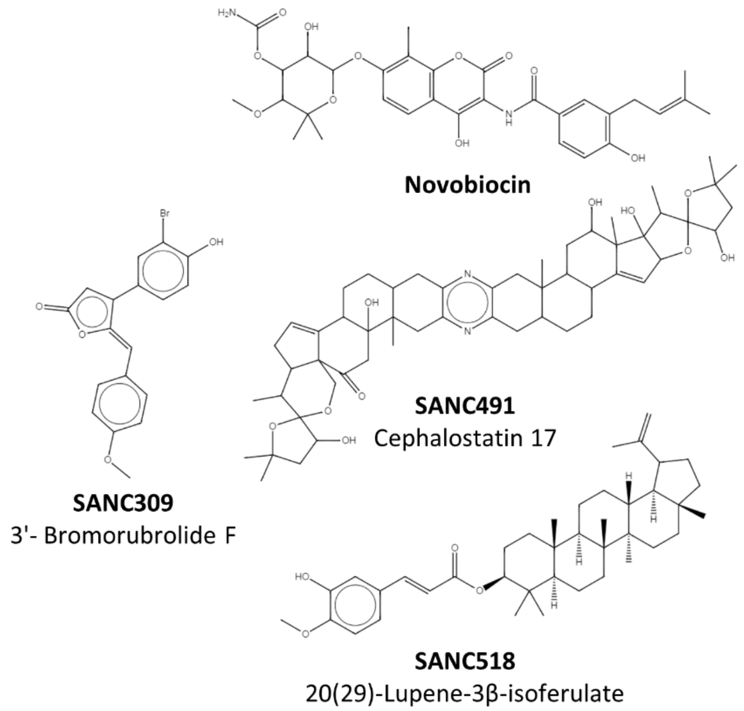


Figure 4.2: 2D representation of Novobiocin and the three natural compounds that bind the CTD. Reproduced with permission from Penkler et. al. 2018 [385]

Of the 702 natural compounds screened, only three putative hits against either CTD binding site were observed (Table 4.1, Figure 4.2). Novobiocin, Cephalostatin 17 (SANC491) and 20(29)-Lupene-3β-isoferulate (SANC518) reproducibly docked to Site-1 (Figure 4.3). All three ligands recorded similar docking orientations, in which each compound lines the length of the binding groove positioned over the four-helix bundle with their terminal moieties extending out towards the Site-2 sub-pocket in protomer B (Figure 4.3 - inset). This observed binding site and orientation is in agreement with previous reports for Novobiocin [386]. From a docking quality perspective both SANC491 and SANC518 recorded lower binding free energy scores (<-11.0 kcal/mol) compared to Novobiocin (-8.97 kcal/mol) suggesting higher binding affinities for the natural compounds.

Looking at Site-2, only 3'-Bromorubrolide F (SANC309) was able to dock to the small sub-pocket in protomer B with a high degree of reproducibility and suitable binding free energy scores (-9.13 kal/mol) (Figure 4.3 – red). To fully examine the stability of ligand binding and to gain further insights as to the allosteric effect of each ligand, all-atom MD simulations were carried out on the protein-ligand complexes for a total 200 ns. In each case, the ligand conformation with the lowest binding free energy score from the largest cluster was selected as a representative start structure.

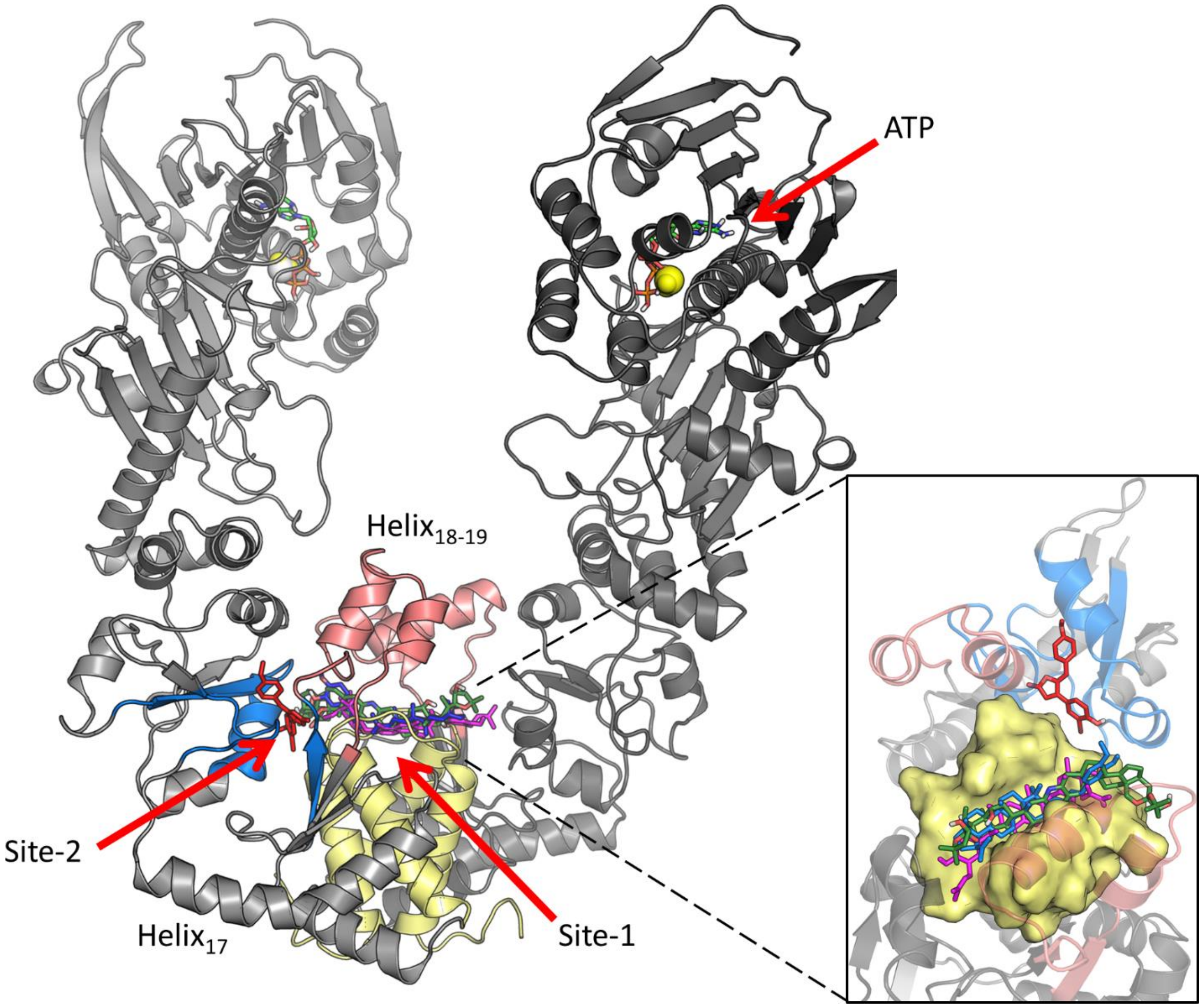


Figure 4.3: Illustration of docked compounds to Hsp90α in the open conformation. The location of the different binding site residues are shaded: Site-1 helix18-19 (red), helix21-22 four-helix bundle (yellow) and Site-2 sub-pocket (blue). The NTD location of ATP and magnesium ions (spheres) are also shown. Inset – zoomed in view of docked compounds SANC309 red, SANC491 green, SANC518 blue, and Novobiocin magenta. Reproduced with permission from Penkler et. al. 2018 [385]

4.3.3 Characterisation of protein-ligand interactions

Ligand binding stability was assessed by monitoring the binding site residue contribution to protein-ligand interactions over time. The PLIP program was used to evaluate hydrophobic and hydrogen bond interactions every nth frame in the trajectory using a 200 ps time interval, and the cumulative data depicted as interaction heat maps for each interaction type for each protomer (Figure 4.4).

Starting with binding Site-1, Novobiocin, SANC491, and SANC518 primarily interact with residues (residues L_{672}-Q_{682}) from both protomers that line a distinct binding groove located at the four-helix bundle of the CTD inter-protomer dimerization interface (Figure 4.4 – yellow). Of these residues, L_{627}, S_{677}, L_{678}, and P_{681} are the most consistent contributors to ligand interactions over the 200 ns MD trajectories. In addition to stable hydrophobic (residues L_{627} and P_{681}) and hydrogen bond interactions (residues S_{677} and L_{678}), the terminal moieties of the Site-1 ligands also interact with residues (T_{495}-F_{507}) that line the entrance of the protomer B sub-pocket binding site (Site-2) (Figure 4.4 – blue).

The only notable difference between the Site-1 ligands was the formation of additional interactions between SANC491 and $helix_{18}$ of protomer A (residues R_{612}-T_{624}) (Figure 4.4 – red). The observed orientation of the binding Site-1 ligands together with their corresponding interacting residues is in close agreement with a previous combined biochemical and *in silico* study of Bisphenol A based allosteric inhibitors of human Hsp90 [369]. Furthermore, interacting residues L_{672}, S_{674}, and P_{681} are closely positioned to and overlap with the CTD allosteric hotspots residues T_{599}-W_{606}, and T_{669}-L_{678} previously implicated in NTD allosteric signalling and control of conformational dynamics [191,297].

Looking at the binding Site-2 residues, SANC309 appears to interact exclusively with residues belonging to protomer B (residues T_{495}-F_{507} and S_{543}-K_{546}, Figure 4.4 – blue) with the exception of hydrogen bond interactions with the four-helix bundle through residue Q_{682} in protomer A (Figure 4.4 – red). In protomer B, residues Q_{501}, T_{545} and K_{546} form stable hydrophobic interactions with SANC309, while interactions with the remaining sub-pocket residues appear to be more transient over the 200 ns MD trajectory (Figure 4.4 – blue). The protein-ligand interaction landscape observed for SANC309 is to the best of our knowledge novel to the current study and notably overlaps with several allosteric hotspot residues (T_{495}, E_{497}, T_{545}, and K_{546}) that have been previously implicated in selecting conformational displacements in favour of the closed conformation when externally perturbed [191].

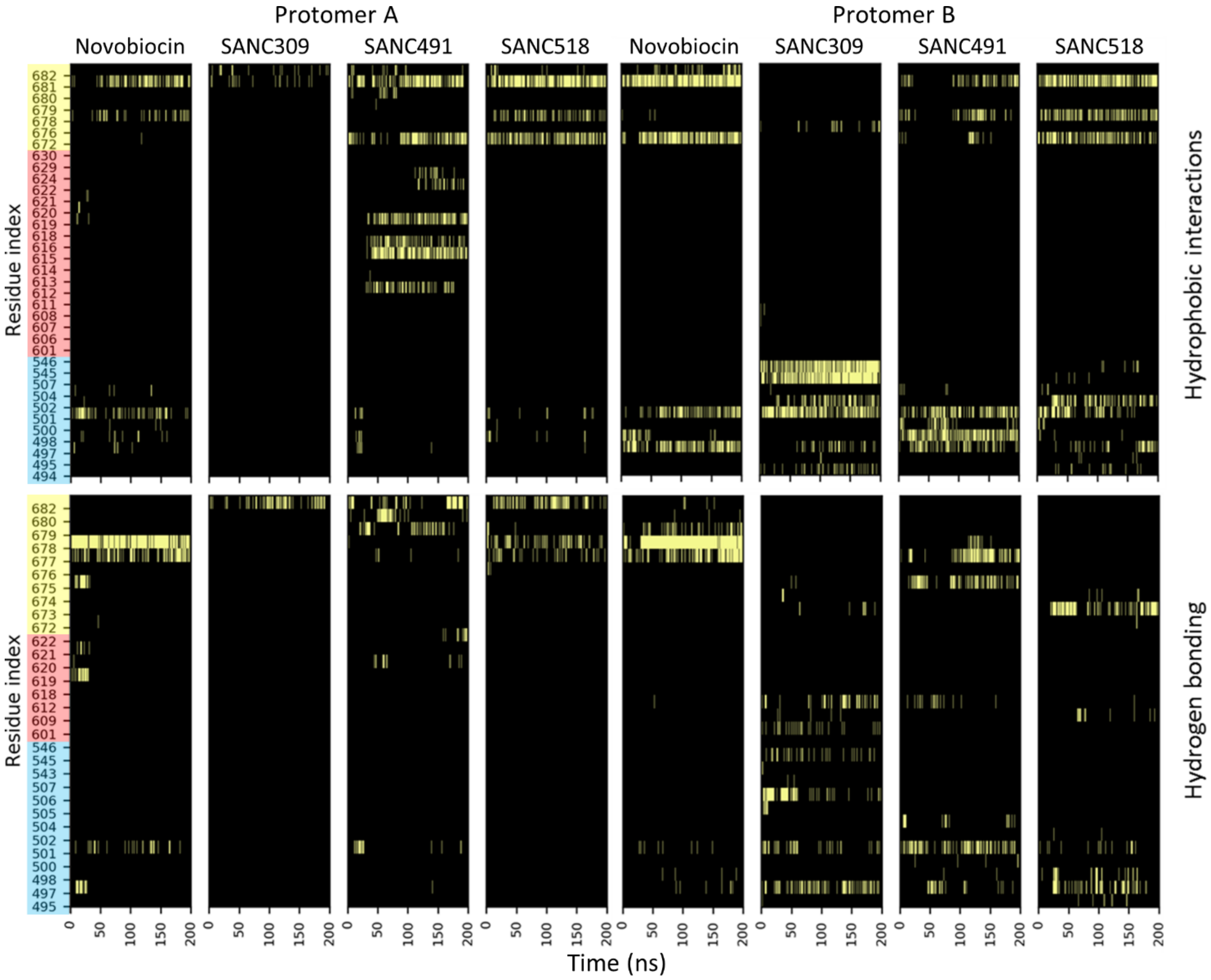


Figure 4.4: Time evolution of binding site residue contribution to protein-ligand interactions. Showing hydrophobic and hydrogen bond interactions separately. Detected interactions are depicted by light bars, the absence of ligand interaction is denoted by a black bar. Y-axis residue shading represents the different binding site residues: blue - sub-pocket; red – helix$_{18}$; yellow - four-helix bundle (see Figure 4.3). Reproduced with permission from Penkler et. al. 2018 [385]

Overall, MD simulations revealed stable protein-ligand complexes over 200 ns, and the interaction profiles for both Site-1 and Site-2 overlap with known allosteric sites opening the possibility for external modulation of Hsp90α conformational dynamics through ligand binding interactions. We investigate this avenue by monitoring the effect each ligand has on the global and internal dynamics of the dimer compared to a ligand-free system and assess the respective allosteric potential of Site-1 and Site-2.

4.3.4 CTD ligands modulate protein flexibility

– *Backbone RMSD analysis*

Backbone root-mean-square-deviation (RMSD) analysis serves as a measure for monitoring conformational variation over an MD trajectory. Whole protein RMSD analysis indicates that in the ligand-free state, Hsp90 experiences variable backbone flexibility recording RMSD values that are normally distributed around a mean of 0.70 nm with a standard deviation of 0.16 nm (Figure 4.5 and Appendix II-I). This data supports the current opinion of inherent plasticity for Hsp90 in the fully-open conformation [291]. Addition of either Novobiocin or SANC518 at binding Site-1 appears to improve the stability of the dimer compared to the ligand-free system, recording normally distributed RMSD values centred on reduced means of 0.56 and 0.57 nm respectively. The fact that these complex record standard deviations of 0.12 and 0.16 nm respectively confirms a degree of variable flexibility for the open conformation.

In contrast, the presence of either SANC309 or SANC491 at binding Site-2 and Site-1 respectively appears to enhance conformational flexibility, each complex recording increased mean RMSD values of 0.76 and 0.81 nm, coupled with increased standard deviations of 0.26 and 0.21 nm respectively (Figure 4.5 and Appendix II-I). Furthermore, the bimodal RMSD distribution for SANC309 is in stark contrast to the normal distributions of the four remaining complexes, suggesting the protein samples diverse conformational states with respect to the starting structure. The statistical significance of these findings is reported in Appendix II-II. In summary, the data presented here tentatively suggests an increased flexibility for SANC309 and SANC491 and a maintained if not reduced flexibility for the Novobiocin and SANC518 complexes.

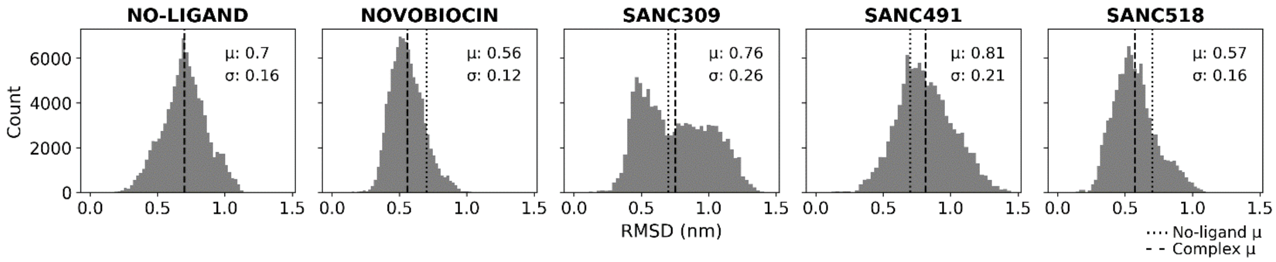


Figure 4.5: Whole protein backbone RMSD distribution plots. Comparison of the histogram distribution plots for each ligand bound complex with the ligand-free state demonstrates variable conformational flexibility. In each case, the shift in the conformational distribution over the 200 ns MD trajectory can be visually assessed by comparing the mean (μ) of each complex (dashed line) with the ligand-free complex (dotted line). Reproduced with permission from Penkler et. al. 2018 [385]

The observed plasticity of the fully-open Hsp90 dimer could be explained by independent conformational restructuring of either protomer. To address this the RMSD profiles for each protomer in isolation are examined (Figure 4.6 and Appendix II-I). With the exception of SANC518, the protomer RMSD distribution for each complex is normally distributed, recording similar average and standard deviation values (μ: ±3.2 nm, σ: ±0.06) compared to the ligand-free system, suggesting the protomers move independently of each other to some degree within the limitations of CTD dimerization. In contrast, the presence of SANC518 appears to enhance individual promoter flexibility, shifting the distribution mean from 0.32 nm and 0.38 nm to 0.43 and 0.51 nm for protomers A and B respectively (Figure 4.6 and Appendix II-II). The marked difference in protomer B of this complex is coupled to a bimodal distribution with peaks centred at ±0.38 and ±0.70 nm respectively, suggesting conformational sampling of two dominant conformations with the former likely corresponding to that of the ligand-free system.

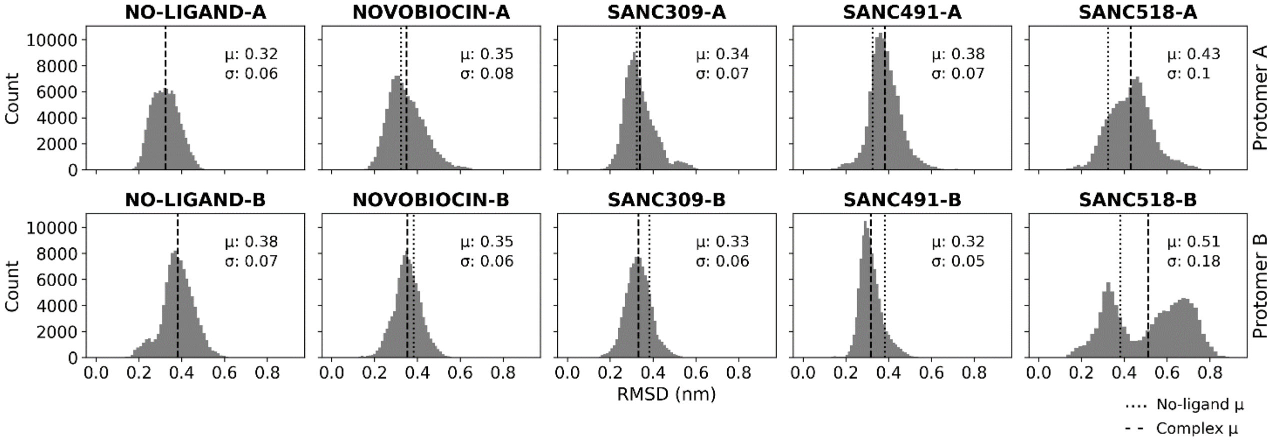


Figure 4.6: Protomer backbone RMSD distribution plots. Comparison of the histogram distribution plots for each ligand bound complex with the ligand-free state. In each case, the shift in the conformational distribution over the 200 ns MD trajectory can be visually assessed by comparing the mean (μ) of each complex (dashed line) with the ligand-free complex (dotted line). Reproduced with permission from Penkler et. al. 2018 [385]

- *Inter- and intra protomer repositioning*

Backbone restructuring of the individual protomers relative to the starting structure informs on secondary structure repositioning however it does not account for whole protomer repositioning over the course of the MD simulation. Indeed, it is likely that the observed backbone flexibility/mobility of the whole protein dimer may result from conformational repositioning of the semi-rigid protomers relative to one another in either a linear opening/closing manner, or through perpendicular rotation about the C-terminal dimerization axis in a hinge bending like motion.

- *Inter-protomer distance analysis*

The distribution of the inter-protomer distance, defined as the measured distance between the center of mass for each NTD, was used to inform on the propensity for Hsp90 to populate conformations distinct from the starting structure in a linear opening/closing motion over the course of the 200 ns MD trajectory (Figure 4.7). In the absence of CTD ligands, the distribution is positioned around the initial NTD-NTD distance of 7.6 nm, recording a mean distance of 7.0 nm which indicates limited conformational variation under these conditions (Figure 4.7 and Appendix II-III). For the ligand bound systems, Novobiocin and SANC518 appear to populate conformations analogous to the ligand-free system with mean values of 7.4 and 7.1 nm respectively (Figure 4.7). Conversely, the addition of SANC309 and SANC491 cause a shift in the distance distribution away from the ligand-free system. SANC491 populates conformations with an increased average inter-protomer distance of 8.1 nm (Figure 4.7), suggesting a conformational preference for a more open "v-like" structure. SANC309 on the other hand, shifts the distribution to the left (µ 6.6 nm) and as seen in whole protein RMSD analysis records a bimodal distribution, suggesting population of two dominant conformations, one analogous to the ligand-free structure (±0.70 nm) and the other in favour of a more closed conformation (±6.0 nm). The statistical significance of these findings is reported in Appendix II-III.

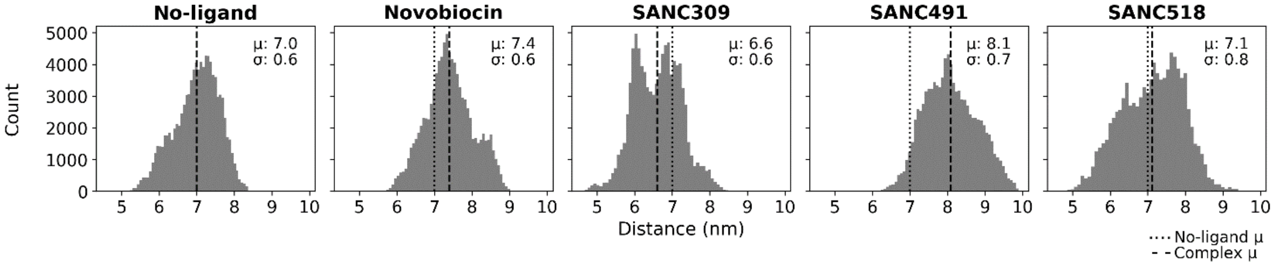


Figure 4.7: Distribution plots of the inter-protomer distance. Inter-protomer distance is defined as the distance between the center of mass of each NTD. Reproduced with permission from Penkler et. al. 2018 [385]

- *Protomer bending and flexing*

Inasmuch as the NTD-NTD distance provides information on the linear opening/closing of the homodimer, the distribution of the NTD-CTD distance for each protomer informs on protomer flexibility around the central axis as seen in hinge bending motions. Shorter NTD-CTD distances could be indicative of protomer bending, while increased NTD-CTD distances could suggest protomer straightening/extension. The NTD-CTD distance is defined as the distance between the centers of mass of the NTD and CTD four-helix bundle, and the time evolution of these measurements is represented as a distribution for each protomer separately (Figure 4.8).

In the absence of bound ligand at the CTD, both protomers form distinct distribution peaks around the mean NTD-CTD distance (µ 8.21 nm: protomer A ; µ 7.63 nm: protomer B) which correspond to the initial NTD-CTD measurements of 8.0 nm and 7.5 nm respectively. This data suggests minimal hinge bending motions and confirms the rigid nature of the protomers under these conditions. Addition of Novobiocin appears to have a marked differential effect on the observed distributions, which are skewed to the right (±9.0 nm – protomer A and ±7.75 nm – protomer B) when compared to the ligand-free system. This positive shift in the inter-domain distance may be indicative of ligand induced sampling of more extended protomer conformations. Indeed, similar albeit exaggerated observations can be made for the other Site-1 ligands, in which both SANC491 and SANC518 induce skewed bimodal distance distributions in protomer A with dominant peaks centred ±9.0 nm, while protomer B experiences tighter skewed distributions at ±8.1 nm. The formation of alternate peaks in the protomer A distributions indicates a wider range of sampling under these conditions providing evidence of ligand induced flexibility around the central axis.

Binding of SANC309 at Site-2 also appears to have a tuneable effect on the flexibility of the protomers when compared to the ligand-free system, particularly protomer B, which forms three distinct distribution peaks (5.0, 8.0, and 9.0 nm), while protomer A forms a skewed distribution peak at ±9.0 nm. This data suggests that ligand interactions at the CTD may allosterically induce protomer flexibility allowing the dimer to sample alternate structural conformations. Comparative statistics between the ligand un/bound complexes are reported in Appendix II-IV.

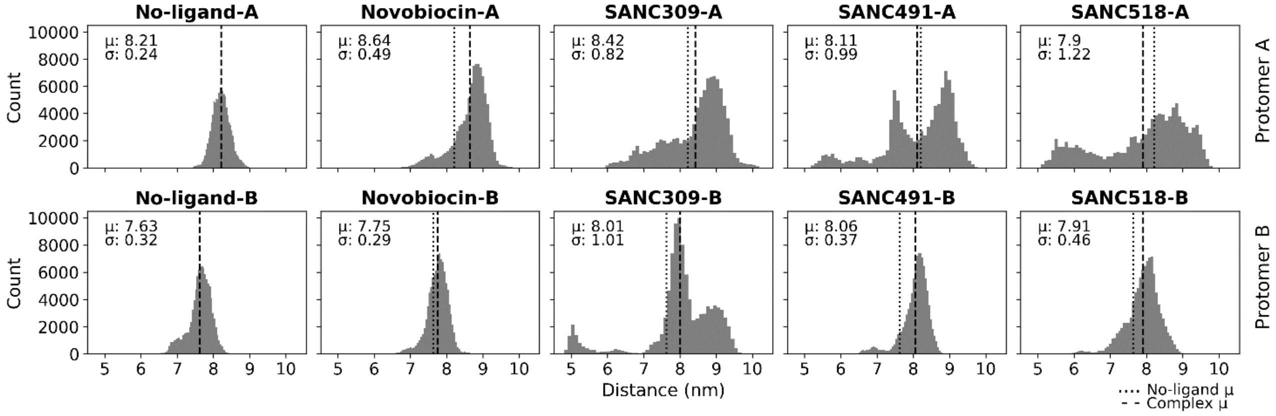


Figure 4.8: Flexing around the central axis of each protomer. Protomer flexing is measured as the distance between the center of mass of each NTD and the CTD interface. Reproduced with permission from Penkler et. al. 2018 [385]

- *Localised flexibility analysis*

Next, we focus on the relative effect of bound ligand on the flexibility of localised regions of the protein. The root-mean-square-fluctuation (RMSF) is a measure of the average positional displacement of each residue over time and indicates flexibility/mobility of individual residue sites. Here, we monitor the relative change in residue fluctuation (ΔRMSF) between the ligand-free and ligand bound complexes to determine the extent to which bound ligands influence intra-protomer flexibility.

The ΔRMSF profiles reveal differential ligand specific modulation of domain flexibility, whereby bound ligands appear to modulate the RMSF of entire domains rather than individual residue sites (Figure 4.9-A). Novobiocin and SANC491 differentially modulate the RMSF of the NTDs of each protomer, causing increased fluctuations in protomer A and a decrease in protomer B (Figure 4.9-A blue shading). Meanwhile, SANC309 and SANC518 increase residue fluctuations at the NTDs of both protomers recording large positive ΔRMSF. Furthermore, SANC309 also experiences increased fluctuations in the MiD of both protomers compared to the Site-1 ligands (Figure 4.9-A green shading).

The collective results presented in this section demonstrate evidence of ligand modulation of both whole protein mobility and domain specific flexibility depending on the identity of the bound ligand. Conformational flexibility and rigid body mobility form the basis of enzymatic catalysis and allosteric modulation [387] and in the case of Hsp90, conformational plasticity is crucial for molecular functionality [159,291]. The increased flexibility experienced by SANC309, especially in the MiD where co-chaperones bind, may enable the protein to overcome energetic limitations allowing it to explore a larger conformational space, and thus aid its conformational search for the closed catalytically active state. Conversely, reduced protomer flexibility in SANC491 may lead to protomer elongation and a more rigid structure that favours the open conformation.

4.3.5 Ligand binding perturbs residue interaction networks

Dynamic residue networks (DRNs) [295] were utilized to analyse the effect of ligand binding on residue connectivity over time. DRNs were constructed for each MD trajectory and the L_i and BC network descriptors used to evaluate the relative effect of CTD bound ligand on the connectivity of the protein, by comparing each protein-ligand complex to the ligand-free ATP-only complex by monitoring the change in L_i (ΔL_i) and BC (ΔBC) over the course of the MD trajectory.

- *Analysis of the change in average residue reachability*

Starting with ΔL_i (Figure 4.9-B), it is evident that irrespective of ligand identity, binding interactions at the CTD increases in L_i, particularly at the NTDs of both protomers (± 1 unit). Residues belonging to the ATP-lid (residues 130-160) experience the largest shift in L_i for this domain, and this observation is notably accentuated in protomer B for the SANC309 complex ($\Delta L_i > 3.0$). This expressed ligand induced modulatory effect on the proteins network confirms NTD:CTD allosteric coupling, whereby ligand binding at the CTD allosterically perturbs the intra-domain interaction network at the NTD.

Given that ΔL_i is a measure of the average change in network communication for the whole protein, it can be influenced in one of two ways: (i) a significant spatial alteration in the local neighbourhood surrounding i can affect the contribution of first and second neighbours to the average number of steps taken [191,388]; (ii) the local network surrounding i remains intact but large conformational changes at distant locations in the protein affect the average path length. As discussed in section 3.3.4, there exists a proportional relationship between RMSF and L_i, when residue fluctuations/displacements exceed the distance threshold used to construct the DRN [191]. This observation is once again evident in the present study, RMSF and L_i record strong Pearson's correlation coefficients (> 0.85) for all five complexes (Appendix II-V A). This relationship does not hold true however when considering ΔRMSF and ΔL_i, which record Pearson's correlations < 0.7 (Appendix II-V B). This observation suggests that the positive ΔL_i experienced in the ligand bound systems is not directly linked to increased residue fluctuations as described in scenario (i), and the scenario presented in (ii) must be considered for an explanation for the increase in L_i.

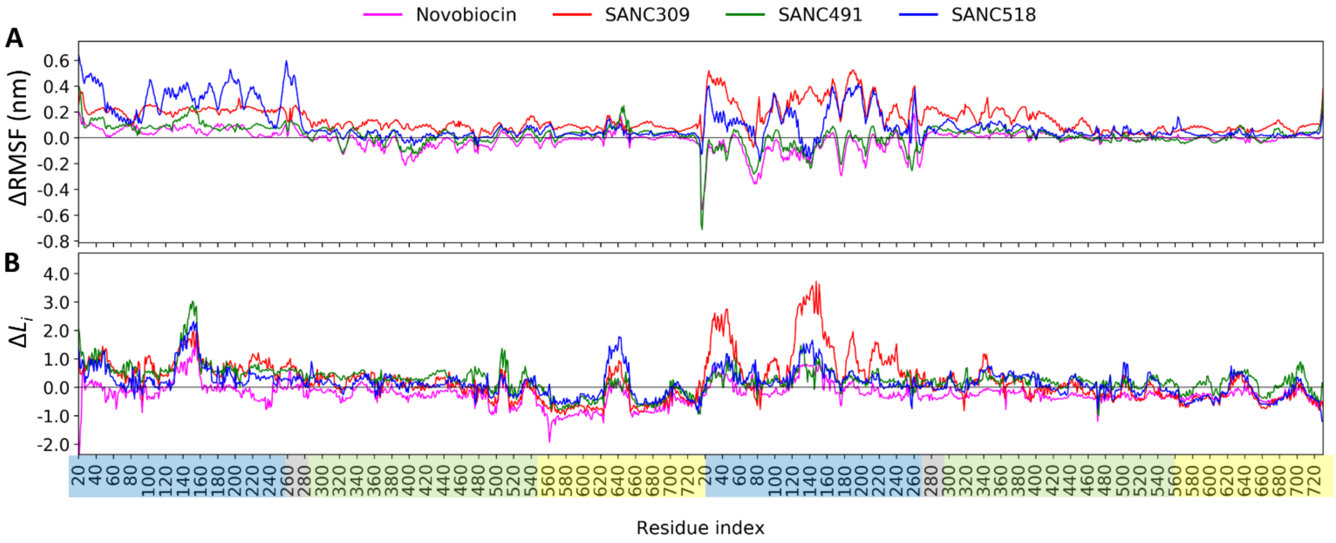


Figure 4.9: Effect of CTD bound ligands on internal dynamics: Showing **(A)** ΔRMSF, and **(B)** ΔL_i plots for each ligand bound complex. Both metrics are calculated as the average difference between the protein-ligand and ligand-free complexes for each residue. Residue indices are coloured by domain: NTD – blue, M-domain – green, CTD – yellow. Reproduced with permission from Penkler et. al. 2018 [385]

The variable flexing/bending of both protomers in response to bound ligand (Figure 4.8), may result in sufficient conformational restructuring so as to disrupt established key inter-domain network contacts between the NTD:MiD which would in turn lead to an increase in the average path length for entire regions of the protein. To investigate this possibility the time progression of all NTD:MiD contacts were quantified over the entire MD trajectory using a 200 ps time interval. All identified inter-domain contacts that maintained a spatial distance ≤ 6.7 Å over the trajectory were captured and quantified as a percentage interaction over time (Figure 4.10-B). Looking at the ligand-free system, it is evident that there are more inter-domain contacts between the NTD and MiD in protomer A compared to protomer B, however two key inter-domain contact regions are present in both protomers (Figure 4.10-A): the NTD residues K_{204}, V_{207}, I_{214}, I_{218}, and L_{220} form stable contacts with the MiD residues L_{290} and N_{291} over 200 ns. In addition to these contacts, residues D_{57}, R_{60}, Y_{61}

and L_{64} in the NTD of protomer A maintain stable inter-domain interactions with R_{366}-V_{368}.

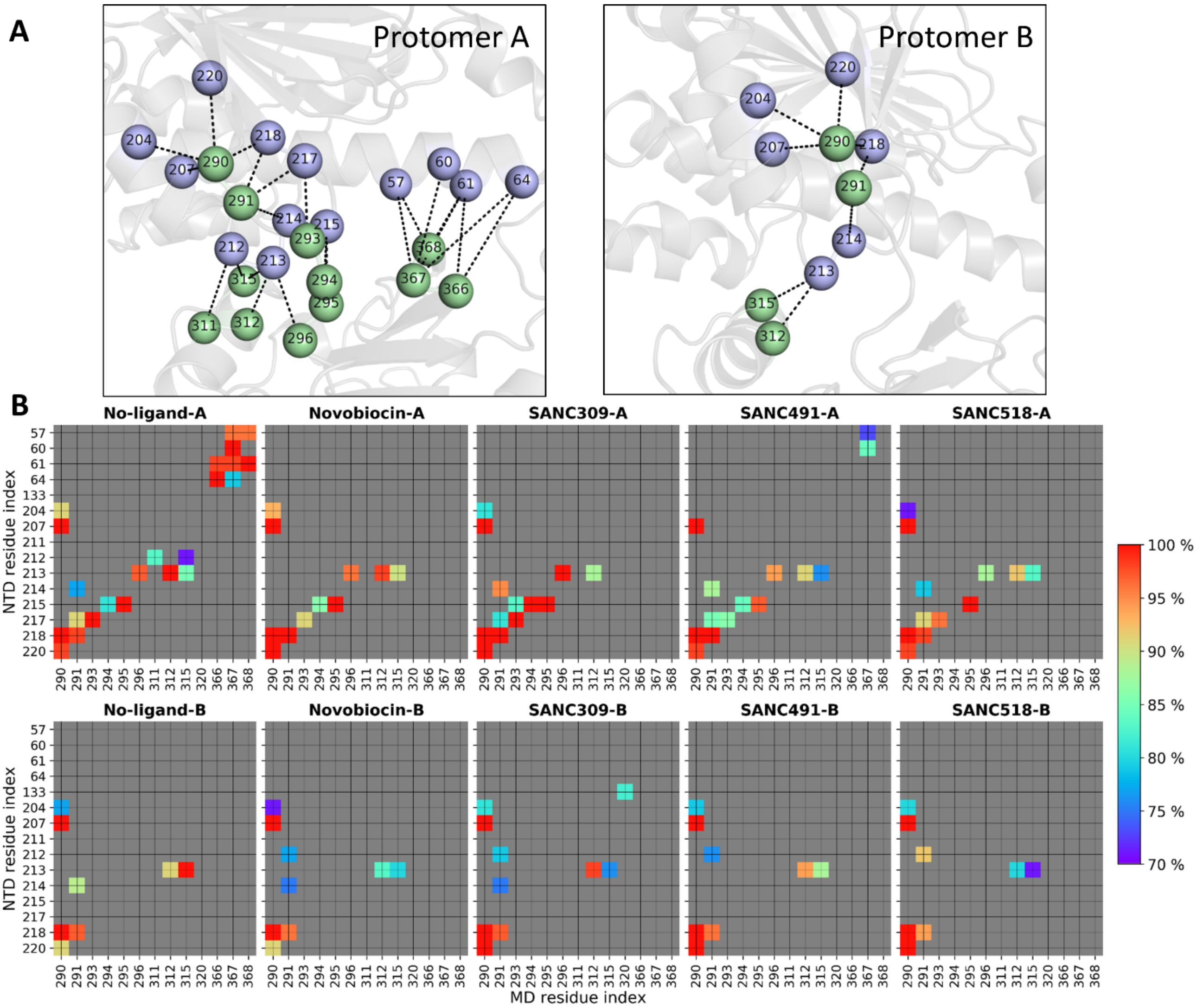


Figure 4.10: Inter-domain contacts between the NTD and M-domain of each protomer. (A) Illustration of the inter-domain contacts between the NTD (violet spheres) and MiD (green spheres) for the ligand-free ATP-only complex. **(B)** Time progression of key contacts between the NTD and MiD over the respective MD trajectory for each complex. Contacts are coloured (blue – red) by percent detected over the trajectory and grey if detected < 70 %. Reproduced with permission from Penkler et. al. 2018 [385]

For the ligand bound systems, it is evident that the inter-domain interactions involving reside L_{290} are maintained in both protomers and those involving N_{291} are weakened in protomer B. Notably, the stable inter-domain interactions between the R_{366}-V_{368} triad in protomer A and the NTD are lost in all ligand bound complexes with the exception of SANC491 which retains partial connectivity between R_{367} and the NTD residues D_{57} and R_{60}. Taken together this suggests that breakdown of stable inter-domain interactions, particularly those involving the R_{366}-V_{368} triad in protomer A and N_{291} in protomer B, may lead to NTD:MiD decoupling which in turn may result in the redirection of inter-domain network communications to pass through the long NTD:MiD linker and thus ultimately increase L_i. This NTD decoupling may explain the enhanced NTD residue fluctuations observed at the

NTD of protomer A (Figure 4.9-A). While the NTD and MiD record elevated ΔL_i, the CTDs experience a ±1 unit decrease in L_i in both protomers. This observation is particularly apparent for residues 550-620 belonging to helix$_{17}$ and helix$_{18}$, and several residue located in the four-helix bundle (residues 660-720). Interestingly, the former set were been previously implicated as sensors to physical perturbations elsewhere in the protein [191] (see section 3.3.5 for details) and have also been implicated in allosteric signal propagation [297]. In this instance, scenario (i) provides a suitable explanation in that ligand binding stabilises the CTD reducing residue fluctuations and thus allowing for stable residue-residue interactions.

– *Betweenness centrality analysis*

As with L_i, ΔBC is also averaged over time and comparison of the different ligand bound complexes reveals a high degree of overlap between the ligand binding sites and regions of the protein that experience large shifts in BC (Figure 4.11), demonstrating the metric's sensitivity to detect putative ligand binding sites. Interestingly, all four ligand complexes experience increased BC at the four-helix bundle in protomer A, but a decrease in protomer B. Furthermore, it is evident that the ligand interactions between SANC491 and helix$_{18}$ (red) in protomer A has a direct impact on BC at this location, possibly due to ligand stabilisation of this inherently flexible region as demonstrated by the differential ΔL_i values (Figure 4.9-A). Indeed, this finding is in agreement with our previous study in which we report an inverse relationship between BC and L_i [191].Overall the communication control hubs identified through BC correlate with ligand binding residues and are in agreement with the residue locations previously identified in section 3.3.3. Ligand binding does not alter or perturb the residue interaction network in such a way that these control points are re-established at different location. Rather, ligand binding appears to alter the frequency of usage of the control point, likely due to "rewiring" if the shortest communication paths.

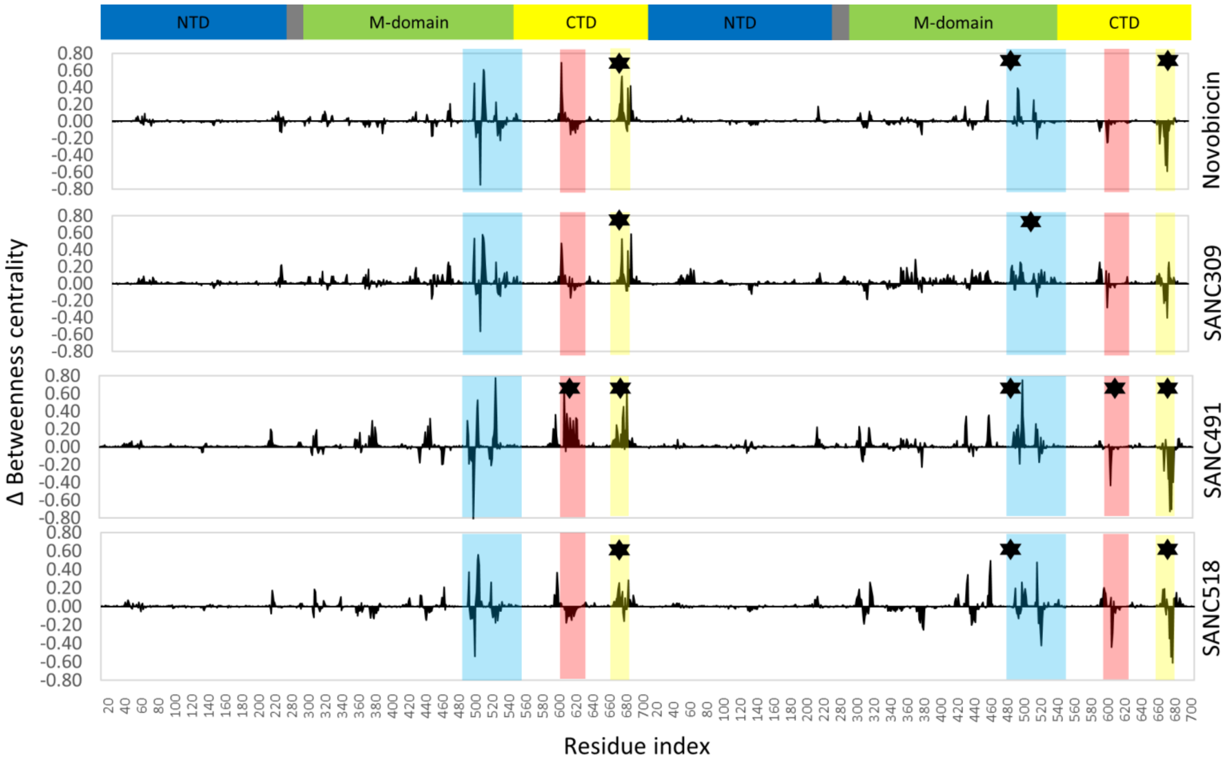


Figure 4.11: Comparative ΔBC profiles. Shaded regions denote ligand binding site residues: blue sub-pocket, red helix$_2$, yellow four-helix bundle. ★ denote residues participating in ligand interactions. Reproduced with permission from Penkler et. al. 2018 [385]

4.3.6 Effect of allosteric ligands on communication propensity

– *Whole protein ΔCP pattern analysis*

Communication propensity (CP) is a measure of the efficiency of communication for a given residue pair *i* and *j* in a protein, and is based on the premise that signal propagation is directly related to distance fluctuation between the communicating residue pair. The concept and methodology for determining CP in proteins was first introduced for elastic network models by Chennubhotla and Bahar [382] and subsequently extended by Morra et. al. for MD trajectories [297]. In the latter methodology CP is defined as the mean-square distance fluctuation between the C_α–C_α atoms of residues *i* and *j* over the course of an MD trajectory. Smaller amplitude inter-residue distance fluctuations are expected to increase communication efficiency compared to that of large amplitude C_α–C_α distance fluctuations [297,382]. It is important to note that CP denotes communication time and thus smaller values represent more efficient commination between any two C_α–C_α atoms. To obtain a compact representation of how bound ligands affect CP, we define the difference matrix between all residue pair CPs determined for the ligand-free and ligand bound systems (Figure 4.12). Thus, positive ΔCP (blue) indicates more efficient (faster) communication with the addition of ligand, while negative ΔCP (red) indicates slower communication (slower) in the presence of bound ligand.

Comparison of the ΔCP matrices in Figure 4.12, reveals a similar block character for regions of the protein that display either efficient (+CP) or slower (-CP) coordination propensities. The intra-protomer coordination patterns provide a visual means by which to compare the inter-domain communication propensities. For the Site-1 complexes, the NTD in protomer A appears to be decoupled from the MiD and CTD as indicated by the red blocks. This decoupling is particularly apparent for SANC518 and resonates with the findings from the previous inter-domain contact analysis (Figure 4.10). In protomer B, the opposite effect is observed as indicated by the large blue blocks in which the NTD experiences more efficient CP between the MiD and CTD, however this observation is less apparent for SANC518 which experiences a degree of inter-domain decoupling in protomer B. Looking at the inter-protomer coordination patterns for these complexes, it is largely evident that ligand binding causes a loss in communication efficiency between protomers. However, in the presence of Novobiocin increased CP is observed between the NTD of protomer B and the MiD:CTD region of protomer A, while SANC491 experiences a moderate increase in CP between at the NTD of protomer A and the CTD of protomer B.

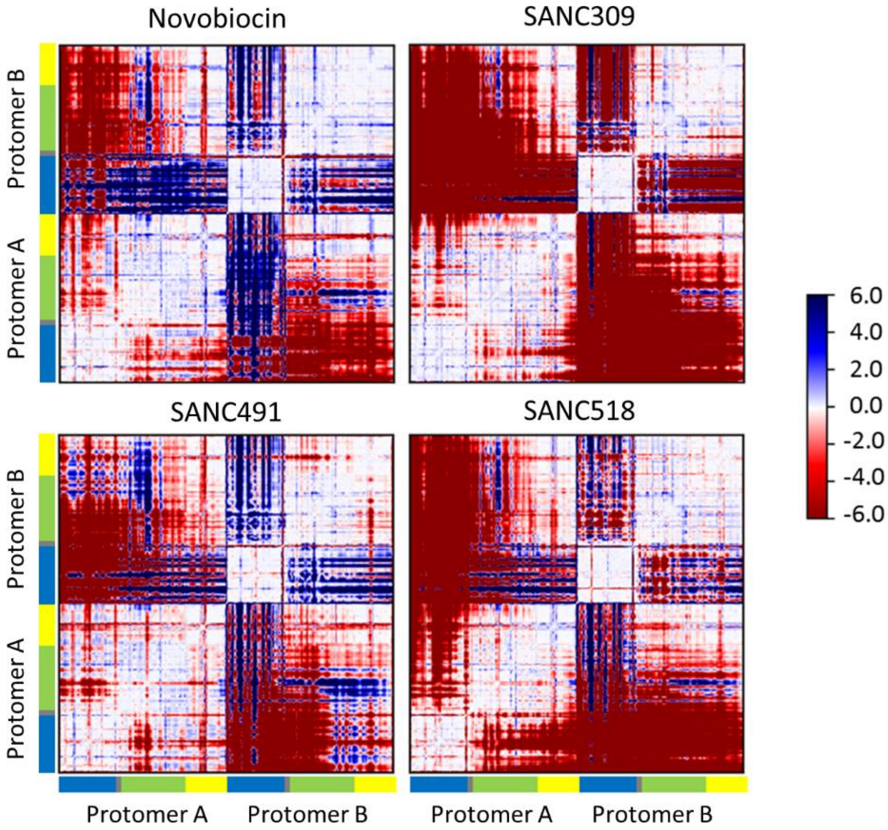


Figure 4.12: Heat map illustration of the change in communication propensity. Matrices calculated as the difference in communication efficiency between the ligand-free complex and the ligand bound complex. Positive values (blue) denote higher communication propensity for the ligand bound complex compared to the ligand-free complex while negative values (red) indicate lower communication propensity for the ligand bound complexes. Reproduced with permission from Penkler et. al. 2018 [385]

The coordination patterns for SANC309, indicate that ligand binding at Site-2 appears to have a drastic effect on CP causing significant intra- and inter-domain decoupling within and between both protomers as evidenced by the large red blocks. It is interesting to note that this observation is particularly apparent in protomer B which provides the sole binding interactions for SANC309. Overall, these observations are consistent with the ΔRMSF results presented in Figure 4.9-A, whereby bound ligands at binding Site-1 appear to increase residue fluctuations at the NTD of protomer A and decrease fluctuations at the NTD of protomer B, while ligand binding at Site-2 increases residue fluctuations throughout both protomers.

– *Fraction of fast communication residues*

Next, the relative contribution of individual residues to long range coordination with distant residues elsewhere in the protein is investigated. For this analysis, the average CP for neighbouring residues (*i* ± 4) for the ligand-free complex was found to be 0.85, and thus CP=0.85 was set as a suitable threshold for discriminating fast communication between residues positioned >80 Å from one another. A similar analysis method was previously described and validated by Morra and co-workers [297]. The protein was sequentially scanned and the fraction of residues in the whole protein capable of communicating with a CP ≤ 0.85 recorded (Figure 4.13).

Looking at the ligand-free complex it is evident that long range communication is established between the NTDs (residues 80-90, 150-160, 170-185) and CTDs (residues 555-580, 640-655, and 690-700) providing further evidence of NTD-CTD allosteric coupling in agreement with previous reports for yeast Hsp90 in the closed conformational state [297]. For the ligand bound complexes, it is evident that none of the binding site residues directly contribute to long range residue communication (Figure 4.13 shaded areas). Rather, the presence and identity of the bound ligand appears to impact the intensity of relative long range coordination of residues in close proximity to the bound ligands. Novobiocin and SANC491 increase the fraction of communicating residues at NTD and CTD of both protomers (residues 80-90, 160-200, 555-580, 640-660 and 690-700). SANC309 appears to significantly reduce the communication efficiency of the CTD, as well as residues 80-90 at the NTD of protomer B, while SANC518 reduces all long range coordination in both protomers.

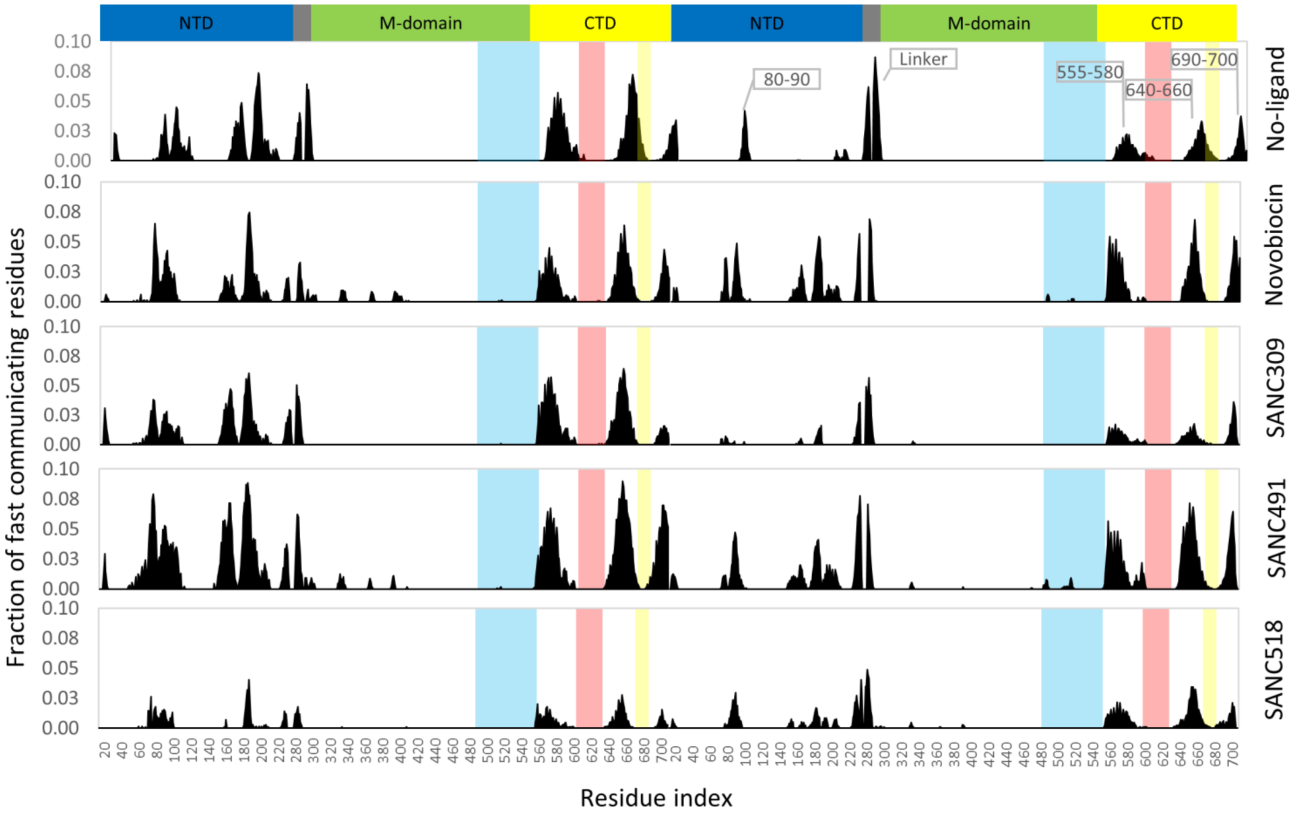


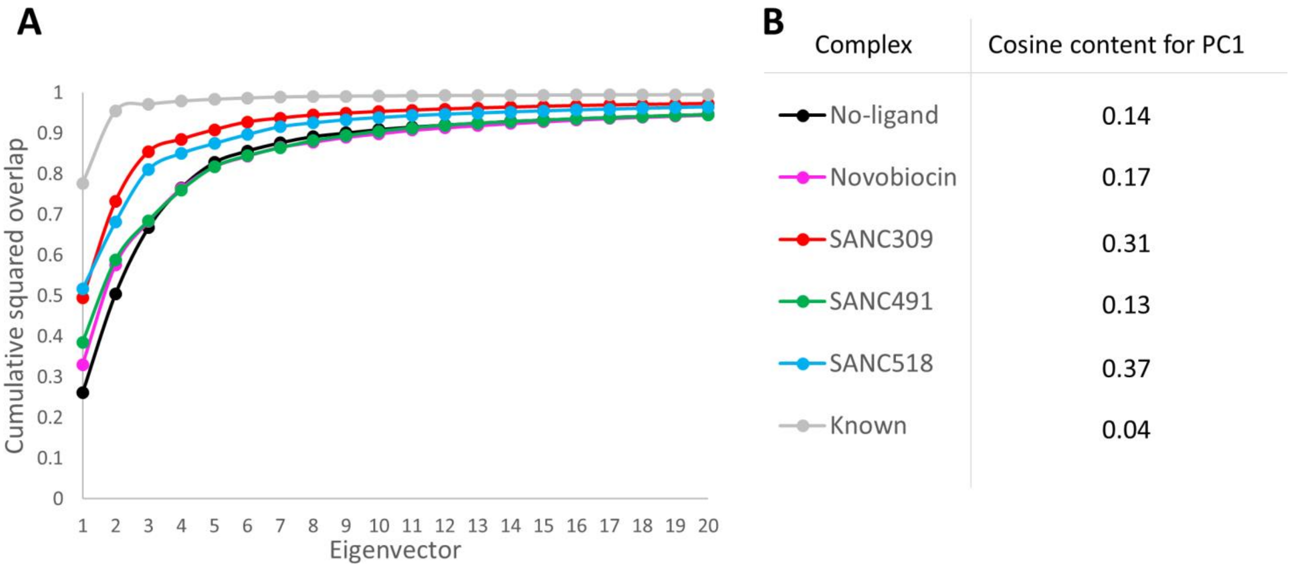
Figure 4.13: Fraction of fast communicating residues over 80 Å apart. Each histogram refers to a single residue and indicates the total fraction of residues that communicate with it. Coloured shading represents the ligand binding residues at Site-1 – yellow and red; and Site-2 – blue. Reproduced with permission from Penkler et. al. 2018 [385]

The data presented here, illustrates that ligand binding at either CTD has a modulatory effect on the communication propensity of the protein and specifically long-range communication between the NTD and CTD. Furthermore, it is evident that the modulatory effect (increased or decreased efficiency of communication) is dependent on the identity of the bound ligand. Over all, SANC309 and SANC518 appear decrease the efficiency of inter-domain and inter-protomer communication, while Novobiocin and SANC491 appear to have an enhancing effect on communication propensity. The inhibitory effect of SANC309 and SANC518 may be due to the fact that both ligand interact with residue 504 in protomer B. Interestingly this residue experience decrease BC in both complexes but not in the Novobiocin and SANC491 complexes.

4.3.7 Essential dynamics analysis

– *Functional relevance of dominant global correlated motions*

Given the tremendous size of Hsp90 (1400 residues), 200 ns MD trajectories are of an insufficient length to access functional global motions such as the full closing motion expected for the ATP bound complex. Essential dynamics (ED) analysis techniques were employed as an alternative method to assess whether ligand binding corresponds to functional global correlated motions. In ED, the MD trajectory is represented as the covariance matrix which is used to calculate the relative eigenvectors and eigenvalues used to descript correlated motion. The former metric represents the correlated displacement of atom groups through essential space, while the latter gives an indication of the magnitude (nm^2). The configurational space represented by eigenvectors can be separated into two subspaces [383]; (i) the essential subspace which represents correlated motions comprising very few degrees of freedom, and likely point to biologically relevant global motions; and (ii) the independent subspace which is constrained to local regions and offers little functional importance. Here, we focus on the former subspace (i) which is often accounted for by the first few low-frequency modes with large corresponding eigenvalues.



Complex	Cosine content for PC1
No-ligand	0.14
Novobiocin	0.17
SANC309	0.31
SANC491	0.13
SANC518	0.37
Known	0.04

Figure 4.14: Cumulative squared overlap for the first 20 eigenvectors. (A) Cumulative squared overlap plots. **(B)** Cosine content for the first eigenvector of each complex. Results for the concatenated known conformational states are labelled “Known” and coloured grey (see text for details). Reproduced with permission from Penkler et. al. 2018 [385]

Analysis of the cumulative squared overlap for the first 20 modes for each complex shows that between 60 % and 80 % of all protein motion is accounted for by the first three eigenvectors (Figure 4.14). In each case the first two eigenvectors account for more than 50 % of all protein fluctuations, suggesting these modes represent functionally relevant protein motions. To gain insight as the direction and magnitude of the dominant modes, the global correlated motions associated with the first eigenvector for each complex are illustrated by projecting the corresponding trajectory onto the eigenvector and interpolating over the two most extreme projections and arrows used to describe the relative atomic displacements (Figure 4.15 and Appendix II-VI – Appendix II-XI).

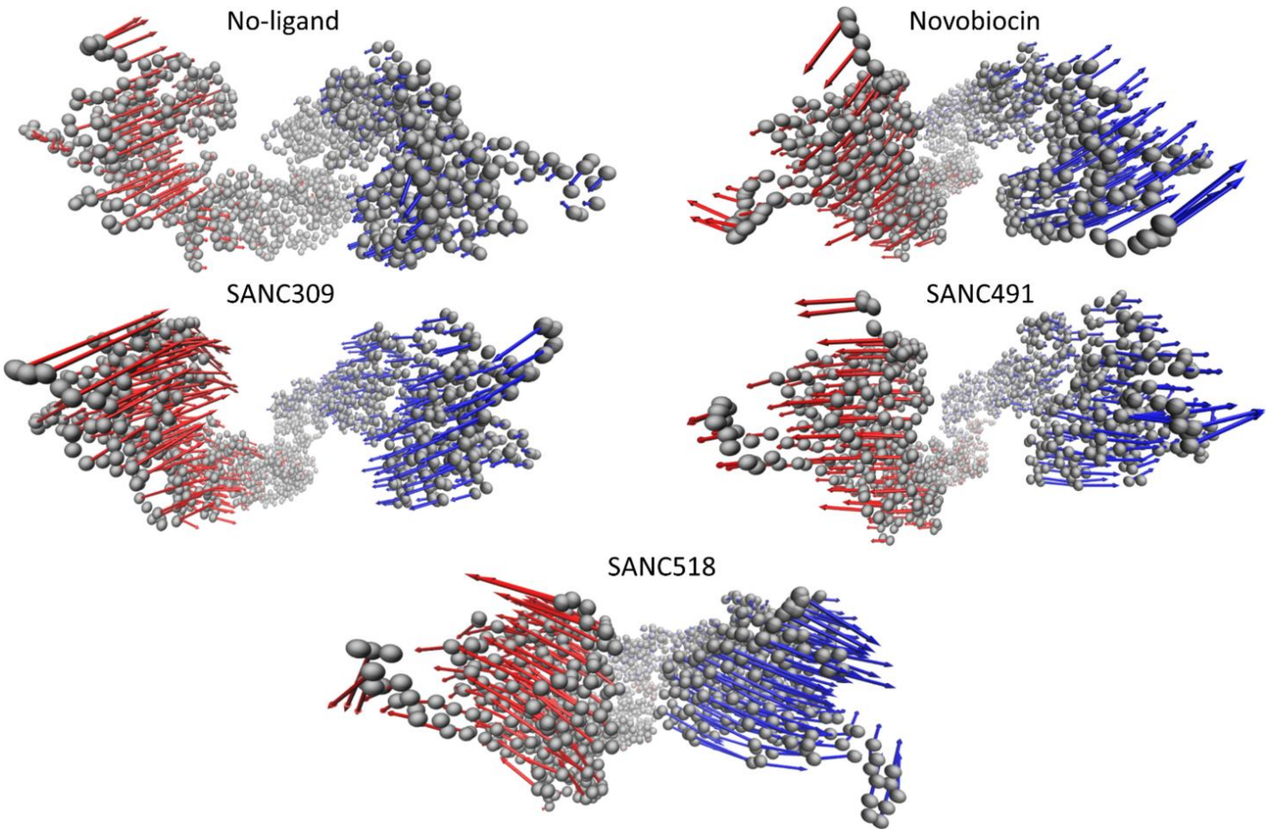


Figure 4.15: Illustration of the global correlated motion for the first eigenvector of each Hsp90 complex. The protein complex is represented by C_α atoms (grey spheres) and the arrows describe the relative direction and magnitude of the correlated motion for protomer A (blue) and protomer B (red). Displacement arrows were drawn for every 2^{nd} C_α atom to simplify presentation. Reproduced with permission from Penkler et. al. 2018 [385]

In the ligand-free state, ED analysis reveals linear global correlated motions for each protomer in the opposite direction towards one other, in a motion that closely resembles the closing transition that is expected under these conditions, in which ATP binding triggers conformational rearrangements towards the catalytically active closed state. Addition of Novobiocin engenders correlated protomer motions that are in opposite directions, in what can be described as an opening like motion. This motion may provide early evidence of protomer uncoupling as Novobiocin is known to disrupt protomer dimerization at the CTD [365]. Interestingly, SANC491 and SANC518 demonstrate similar linear motions to Novobiocin, providing further evidence that these compounds may invoke a similar inhibitory allosteric mechanism. SANC309 on the other hand appears to enhance global correlated protomer motions towards each other in a similar manner to the ligand-free ATP-only complex. This observation coupled with an increased magnitude compared to the ligand-free complex, may suggest SANC309 to be an allosteric activator of human Hsp90α, promoting allosteric closure of the dimer. Indeed, this observation agrees with earlier results in which SANC309 populated conformations with smaller inter-protomer distances (Figure 4.7). Furthermore, we note similar results for the second eigenvectors which describe rotational twisting of the protomers along their central axis in either a clockwise closing or an anti-clockwise opening motions much like the twisting and untwisting of a double stranded helix.

- *PCA analysis to assess conformational change*

To confirm the ED findings presented in the previous section, the change in sampling of the conformational space of the protein was assessed using PCA analysis to compare the aforementioned correlated motions with known structural reference data. An 80 ns concatenated trajectory of four known conformational states was constructed based on 20 ns equilibrated trajectory segments of the ATP-bound trajectories of the fully-closed, partially-open, open, and fully-open conformation complexes (see Chapter 3 section 3.3.2 for trajectory details). PCA analysis was conducted on this trajectory and the MD trajectories of the different complexes projected onto the first three principle components (PC1: linear opening, PC2: twist closing PC3: linear closing) in a pairwise manner to identify ligand induced conformational sampling.

Comparison of the resultant pairwise scatter plots illustrates the conformational sampling of the different complexes with respect to known structural states (Figure 4.16). We note that the initial starting conformation for the different complexes in this study was obtained from the ATP-bound “open state” MD trajectory (Figure 4.16-green). In the absence of CTD ligand, Hsp90 appears to populate conformations in favour of the closed and partially-closed conformations, confirming the dominant global correlated closing motions described in the previous section. In contrast, the addition of Site-1 ligands engenders conformational overlap with the open and fully-open conformational states, confirming the global correlated motions in favour of a more open v-like conformation compared to the initial open structure. Finally, analysis of the SANC309 complex reveals conformational sampling that is analogous to the ligand-free system demonstrating significant overlap with partially-closed conformational state, supporting the observed global correlated motions in favour of the closing transition and thus possibly enhancing the catalytically active conformation.

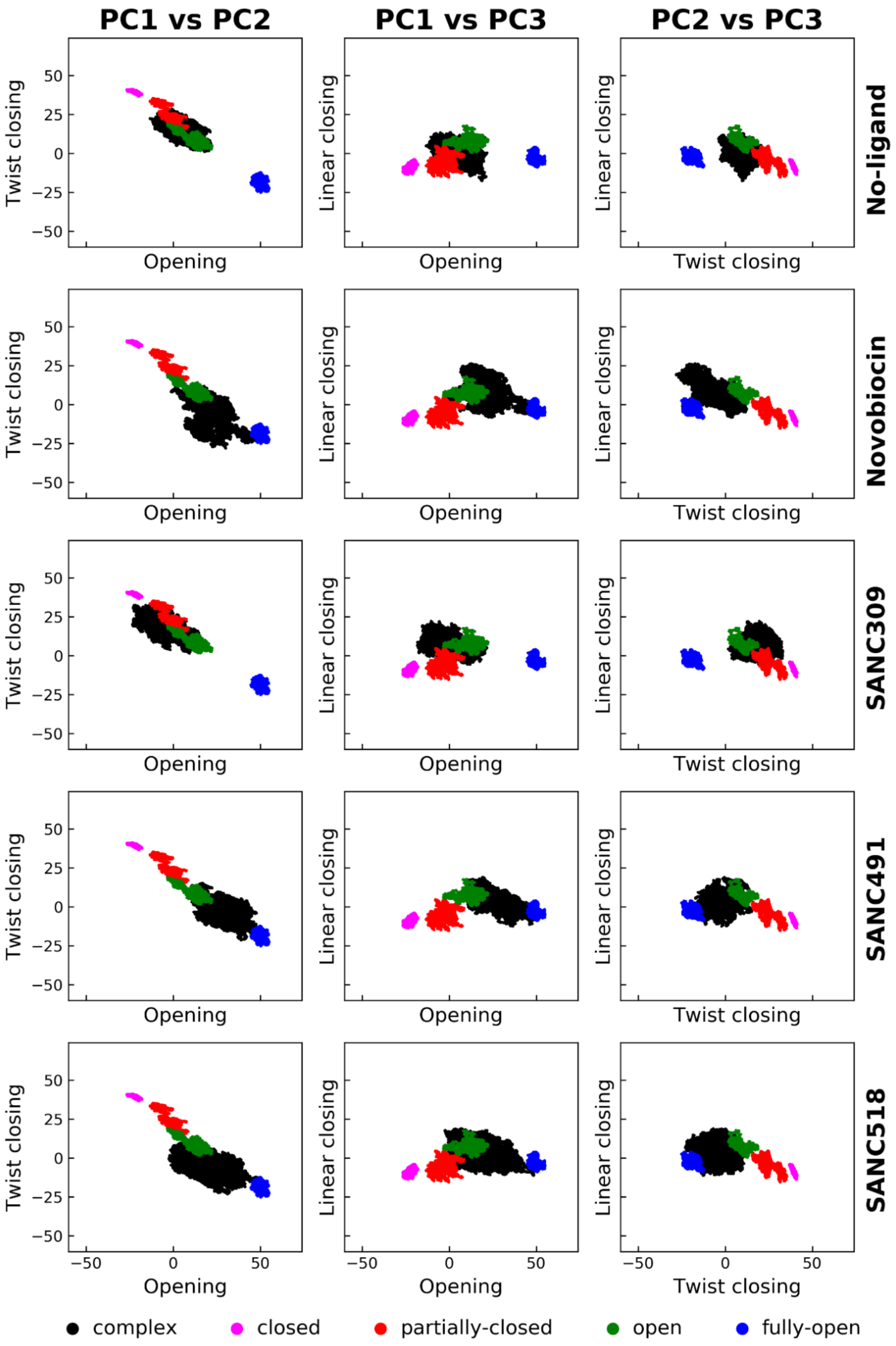


Figure 4.16: Comparative pairwise score plots demonstrating ligand induced conformational sampling. Each row illustrates the conformational sampling of each ligand un/bound complex (black) in relation to the four known conformational states (see legend). Each column describes the projection of each complex onto the top three eigenvectors (PC1: linear opening, PC2: twist closing, PC3: linear closing). Reproduced with permission from Penkler et. al. 2018 [385]

4.4 Conclusions

Overall, the findings presented in this study establish evidence of ligand specific modulation of the conformational dynamics of human Hsp90α in the open "v-like" conformation. We show how natural forces associated with ligand interactions at two putative druggable CTD binding sites direct the conformational sampling of the dimer by fine tuning the internal dynamics of the protein. Of the binding Site-1 ligands, Novobiocin and SANC491 enhanced conformational rigidity through protomer flexing and reduced NTD residue fluctuations resulting in more efficient NTD-CTD allosteric communication. SANC518, on the other hand, behaved differentially, enhancing protomer flexibility leading to reduced NTD-CTD allosteric communication in protomer A.

Given, the overlap in binding site for the Site-1 ligands with known inhibitors such as Novobiocin and several Bisphenol A based inhibitors, coupled with evidence of correlated opening motions, we propose that small molecules targeting the CTD dimerization site located at the four-helix bundle of the open conformation may externally modulate the conformational dynamics in favour of a more open conformation and thus act as allosteric inhibitors of Hsp90α. These compounds may prevent conformational cycling to the closed catalytically active state by either disrupting CTD dimerization as is the case for Novobiocin [365], or by allosterically enhancing the energetic barrier that must be overcome in order to access the ATPase active state [304]. Indeed, it is interesting to note that Cephalostatin 17 (SANC491) is known to be a potent anti-cancer agent [389,390] although its mechanism of action remains unclear. In contrast to the Site-1 ligands, SANC309 at Site-2 greatly enhanced protein flexibility and decreased internal coordination, characteristics that may enable the protein to overcome the energetic requirements necessary to access the closed conformational state and therefore could be classed as an allosteric activator of human Hsp90α. Indeed, the correlated global atomic motions observed for this complex are in agreement with our previous PRS study, in which external force perturbations at Site-2 resulted in global protein displacements towards the closed catalytic conformation [191], providing proof-of-concept for this approach through simulated ligand interactions.

In summary, our findings provide novel insights regarding the selective external modulation of Hsp90α conformational dynamics by small molecule interactions, providing a platform for future allosteric drug development studies for Hsp90. Additionally, this is the first study that leverages combined information from PRS coupled with DNR analysis to identify potential allosteric sites for novel allosteric inhibitor design. This proposed approach is applicable for identifying allosteric drug targeting sites in other proteins.

Concluding Remarks and Future Perspectives

The main focus of this thesis was to make a contribution to the understanding of allosteric regulation of conformational dynamics in the in the 70 kDa and 90 kDa heat shock proteins. This goal was achieved through the prediction and identification of allosteric control elements using dynamics based computational techniques. Given that there in as yet no “gold standard” for the prediction of protein allostery. The methodology described in this thesis contributes to the growing field of computational efforts for the investigation of allostery in proteins and presents PRS coupled with MD simulations and DRN analysis as a suitable strategy for the rapid identification allosteric sites in large multi-domain proteins that experience complex conformational rearrangements. The research described in each of the three data chapters was based on several studies that have been published in reputable peer-reviewed journals.

Chapter 2 described a study which set out to assess the relative effect of bound nucleotide and/or peptide substrate on the allosteric potential of Hsp70 to interconvert between two key functionally relevant structural conformations. An important aspect of this study was to assess the suitability of PRS as a technique for predicting allosteric hot residues in large multi-domain proteins that must undergo complex conformational rearrangements such as translational and rotational motions to interconvert between two known conformational states. The experimental approach used in this study included long-range all atom MD simulations that were important for (i) examining the effect of bound nucleotide and/or peptide substrate on the immediate conformational dynamics of the protein, and (ii) to accurately construct the underlying kernel used by PRS for predicting conformational displacements. *E. coli* DnaK was chosen as a model system in this study because it is a large multi-domain protein that involves complex structural rearrangements in response to the bound nucleotide/peptide. Furthermore its allosteric mechanism of action has been the focus of much research and been well described from a structural stand-point in DnaK. Thus, PRS was used to investigate the regulatory role of nucleotide/peptide in the ATPase activating closing transition and the substrate releasing opening transition.

PRS was shown to be able to quantify the overall allosteric potential of Hsp70 to spontaneously interconvert its open and closed states. Using an MD based kernel, PRS demonstrated sensitivity towards bound nucleotide or peptide and evidence was presented that supports the theory that simultaneous binding of ATP and peptide primes the complex for interstate conversion regardless of the transition direction. This allosterically ‘active’ state was found to be in agreement with an intermediate structure that must overcome energetic requirements before transition between the open and closed conformational states can be achieved [104]. MD simulations of this complex in the closed conformation provided structural evidence of such an intermediate structure in which the charge inter-domain linker is positioned in a binding cleft that only form in the presence of ATP at the NBD.

Detailed analysis of the ATP and peptide bound allosterically active intermediate structures (open and closed) revealed several select regions of the protein identified by PRS that are able to select conformational transitions towards the opposite state in response to random force perturbations at select sites. These findings were extensively cross-validated with literature and found to overlap with known functional sites that have been heavily implicated in allosteric signal propagation and functional regulation of the chaperone. Furthermore, PRS also identified several novel allosteric hot residues that have as yet no known function. Several accentuated residues were found to line critical sub-domain interfaces, suggesting a possible role in allosteric signal transduction from the NBD to the SBD. Residues located in the SBD-lid may be implicated in JDP and NEF binding and a hypothesis is proposed that natural interactions at these sites with either co-factor may provide the necessary allosteric trigger required for Hsp70 to overcome the energetic barrier posed by the intermediate state and thus drive interstate conversion.

Given the functional significance of the residues identified by PRS, it is proposed that the data presented in this study may present as a suitable platform for future biochemical and drug discovery studies aimed at probing the allosteric mechanisms of Hsp70. Furthermore, the MD based PRS methodology presented in this work was successfully implemented as a stand-alone tool in the MD-TASK software suite designed and published for command line ease-of-use by non-expert users [295].

Chapter 3 of this thesis describes a study that built on the experimental ground work laid out in Chapter 2, in which PRS was coupled with DRN analysis and all-atom MD simulations to evaluate allosteric modulation of conformational dynamics in the cytosolic isoform of human Hsp90. An important objective of this study was to address the complete lack experimental structural data for human Hsp90α through accurate calculation of homology models in two distinct conformations, one in the fully-closed (FC) catalytically active state and the second in a partially-open (PO) intermediate state. These structural models of the chaperone provides a start point for the first ever dynamics based analysis of allosteric modulation in the human isoform.

Long range all-atom MD simulations of the FC and PO conformational models in various nucleotide bound configurations (ATP/ADP/apo) were conducted to assess the effect of nucleotide on the conformational dynamics of the dimer. Bound ATP at the NTD was found to rigidify both conformations, causing the dimers to become more compact in favour of the catalytically active state. The presence of ADP on the other hand was found to relax both dimer conformations in preparation for the opening transition, and in the case of the PO complex, drove spatial separation of the protomers such that a fully-open (FO) "v-like" conformation was attained after 200 ns MD simulations. This FO structure provided the opportunity to assess the nucleotide dependent conformational dynamics of the chaperone in its most extended client loading conformational state, an assessment that is the first of its kind for eukaryotic Hsp90. As with the FC and PO conformations, ATP was found to limit conformational flexibility while ADP afforded greater flexibility allowing the dimer to sample a range of fully-open conformational states.

MD trajectory analysis of the functionally relevant FC and FO conformational complexes was extended with DRN analysis to assess the effect of bound nucleotide on the internal dynamics of the protein. Two network metrics were considered, average change in residue reachability (ΔL_i) and betweenness centrality (BC). The former metric informed on regions of the protein that are likely actively involved in intra-protein communication, while the latter metric singled out individual residue sites that for crucial communication hubs in the protein network. These results demonstrated the ability of both metrics to be capable of identifying functional sites on the protein that have been previously identified through experimental efforts, and provided much needed insight regarding the modulatory effect of bound nucleotide on the internal conformational dynamics of the chaperone.

A comparative assessment of the residue fluctuations (RMSF) and the aforementioned dynamic network metrics ΔL_i and BC confirmed previous reports of a proportional relationship between RMSF and ΔL_i in static experimental structures. However, this relationship was found to be sensitive to the distance cut-off used to construct the protein network and that for a dynamic system the magnitude of the distance fluctuations over time together with the relative compactness of the protein will greatly affect the proportionality of this relationship. Furthermore, this study shows for the first time that there exists an inverse relationship between BC and RMSF, ΔL_i. This relationship holds true regardless of the applied distance cut-off and magnitude of the distance fluctuations. Taken together, these findings suggest that since not all residues display small fluctuations act as communication bridges, BC and ΔL_i may be more accurate and sensitive quantities to follow for the identification of novel functional sites on proteins.

PRS analysis was utilised to assess the allosteric potential of the FC and FO conformations to interconvert to the opposite state in response to force perturbations at select residue sites. Allosteric hot residues identified in this analysis mirrored the results obtained from DRN analysis and overlapped with experimentally determined functional residues thought to be involved in co-chaperone binding and allosteric modulation. Of notable interest was the demonstration of NTD and CTD allosteric coupling particularly in the FO conformation. This finding resonates well with the current opinion that nucleotide binding and hydrolysis at the NTD of allosterically modulates functional conformational dynamics at the CTD. Thus, the identification of allosteric effector sites located at the CTD of the FO conformation present as a possible allosteric drug target sites for external modulation and ligand based tuning of Hsp90α conformational dynamics. Indeed, several residues throughout the protein including those at the CTD, were found to be capable of modulating the conformational dynamics of the protein in favour of the closed catalytic state when externally perturbed. These force perturbations could be naturally invoked through binding interactions with co-factors such as molecular co-chaperones or even small molecule ligands.

To date, computational approaches to investigating Hsp90 allostery and conformational dynamics have been largely based on the closed catalytically active conformation of yeast Hsp82. This study for the first time addressed these issues in the human cytosolic isoform of the chaperone, which is an attractive drug target for the treatment of cancer in humans. The data presented in this chapter for the first time sheds valuable insight regarding allosteric modulation of Hsp90α for both the fully-closed and fully-open conformers of the protein and may prove useful for future studies that probe in further detail the mechanisms at play in fine-tuning the functionality and activity of this complex molecular chaperone. Furthermore, given the effectiveness of PRS in identifying allosteric sites using force perturbations that in effect artificially represent forces involved in binding interactions, this study presents PRS and DRN analysis coupled with MD simulations as a viable strategy for screening allosteric proteins for putative allosteric targets sites that are capable of modulating conformational dynamics of the protein and thus fine-tuning protein functionality.

The final chapter of this thesis (*Chapter 4*), described a follow up study to the data presented in Chapter 3. In this chapter, the current status of Hsp90 inhibitor research was briefly reviewed and the attractiveness of allosteric inhibition and activation as a mechanism for fine-tuning Hsp90 functionality discussed. The research presented in this chapter built on the CTD allosteric effector sites identified by PRS analysis in Chapter 3. Computational molecular docking techniques were used to assess the druggability of the aforementioned CTD allosteric sites and a database of small molecule compounds indigenous to South Africa screened against two potential allosteric target sites that overlapped with several allosteric hot residues. MD based analysis techniques coupled with DRN and essential dynamics analyses were used to assess the allosteric potential of hit compounds at either site to modulate the conformational dynamics of the fully-open conformation of human Hsp90α.

The binding orientation of compounds that bound target Site-1 overlapped and were in agreement with the known allosteric inhibitor Novobiocin. These compounds were shown to demonstrate better binding affinity than Novobiocin for a CTD binding groove located above the four helix bundle at the CTD inter-protomer interface. Novobiocin, Cephalostatin 17 (SANC491) and 20(29)-Lupene-3β-isoferulate (SANC518) demonstrated similar allosteric modulatory characteristics that suggested conformational dynamics in favour of the open conformation and perhaps even protomer separation and thus Hsp90 inhibition. Identical small druggable sub-pockets adjacent to Site-1 located at the MiD:CTD interface of either protomer were identified as a second allosteric target site (Site-2).Only one compound 3'-Bromorubrolide F (SANC309) was reproducibly docked to Site-2 and demonstrated allosteric modulatory effects on the conformational dynamics of the protein that appeared to favour dimer closure – characteristics that suggest allosteric activation of the dimer complex.

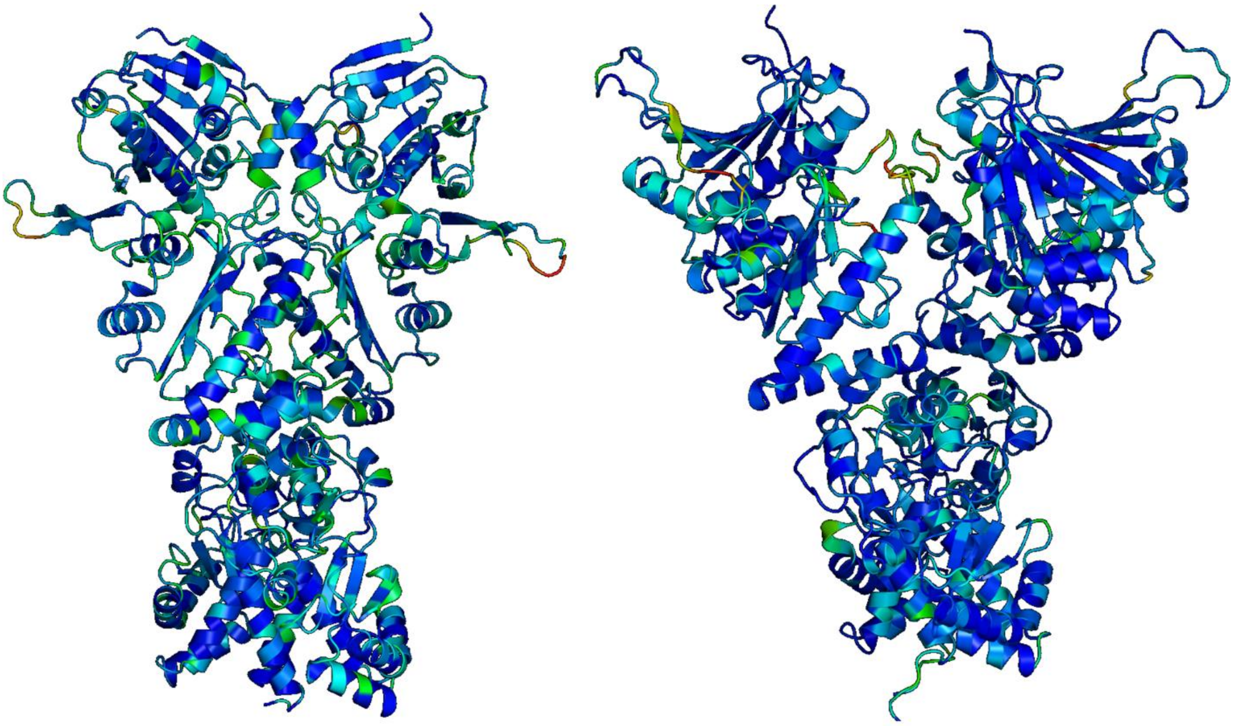
Taken together the data presented in this chapter confirmed the results obtained in Chapter 3, and demonstrated “proof-of-concept” for the use of PRS coupled with MD simulations as an accurate and efficient strategy for whole protein screening for putative allosteric target sites that are capable of modulating the conformational dynamics of the protein in favour of functionally relevant conformations. Evidence of both allosteric inhibition and activation demonstrates the sensitivity of the chaperone to binding interactions, supporting the current opinion of co-chaperone and co-factor mediated activity in Hsp90, and opens the possibility for the design and development of small molecule allosteric probes that could be useful for unearthing mechanistic insights as well as the potential selective treatment of Hsp90 associated diseases such as cancer.

Overall, this thesis has provided novel insights regarding the allosteric mechanisms that modulate conformational dynamics in both the Hsp70 and Hsp90 molecular chaperone systems. The data presented provides a platform for experimental studies to probe these allosteric mechanisms to further our understanding of chaperone function and allosteric drug discovery. Novel insights regarding NTD-CTD allosteric coupling in the open conformation of human Hsp90 coupled with CTD ligands that bind and probe allosteric effector residues provide crucial information regarding the viability of allosteric activation and inhibition of human Hsp90α as an attractive therapeutic avenue. Furthermore, the methodological approach for target driven allosteric discovery proposed in this study could be applicable for other protein systems.

To date the concept of allosteric activators has not yet obtained widespread attention, an observation that may reflect the relatively small number of structural and functional studies surrounding the topic and the lack of a systematic validated strategy to rationally design enzyme activators. It has been noted that the repertoire of enzyme agonists and antagonists can often overlap, leading to the hypothesis that small changes to the allosteric system could lead to switching from inhibition to activation. An improved understanding of these classes of allosteric modulators could offer new opportunities for probing and interrogating biological processes and pathways. The methodology presented in this thesis may pose as an interesting computational framework and diagnostic tool for differentiating between allosteric inhibitors and activators. The PRS methodology could also be expanded to include multiple simultaneous perturbations such that a combinatorial approach could be utilized to further examine the collective impact force perturbations at putative allosteric target sites. Furthermore, there is a growing body of evidence that suggests dynamics-driven allostery may be common to numerous protein systems and that allosteric mechanisms that are not limited to population-shift are mediated exclusively though structural transitions tat select and stabilize specific conformational states. PRS in combination with network analysis could potential provide the best strategy to classify and reconstruct diverse allosteric mechanisms in molecular chaperones and more broadly for all allosteric systems.

Appendices

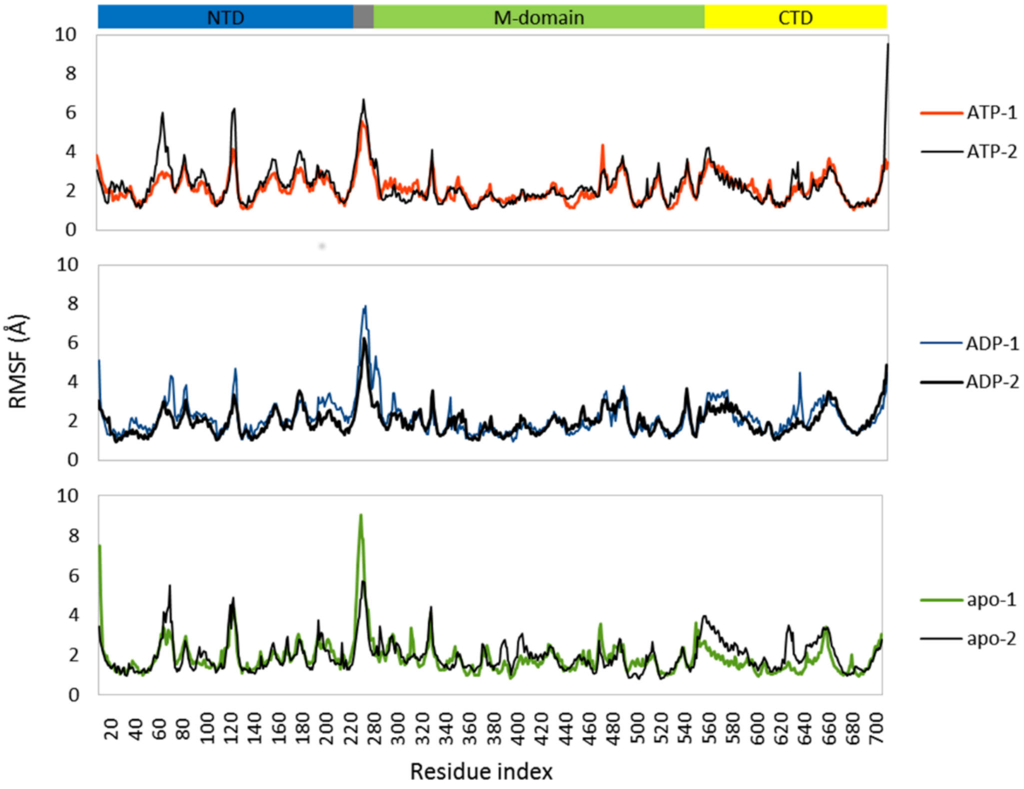
Appendix I:



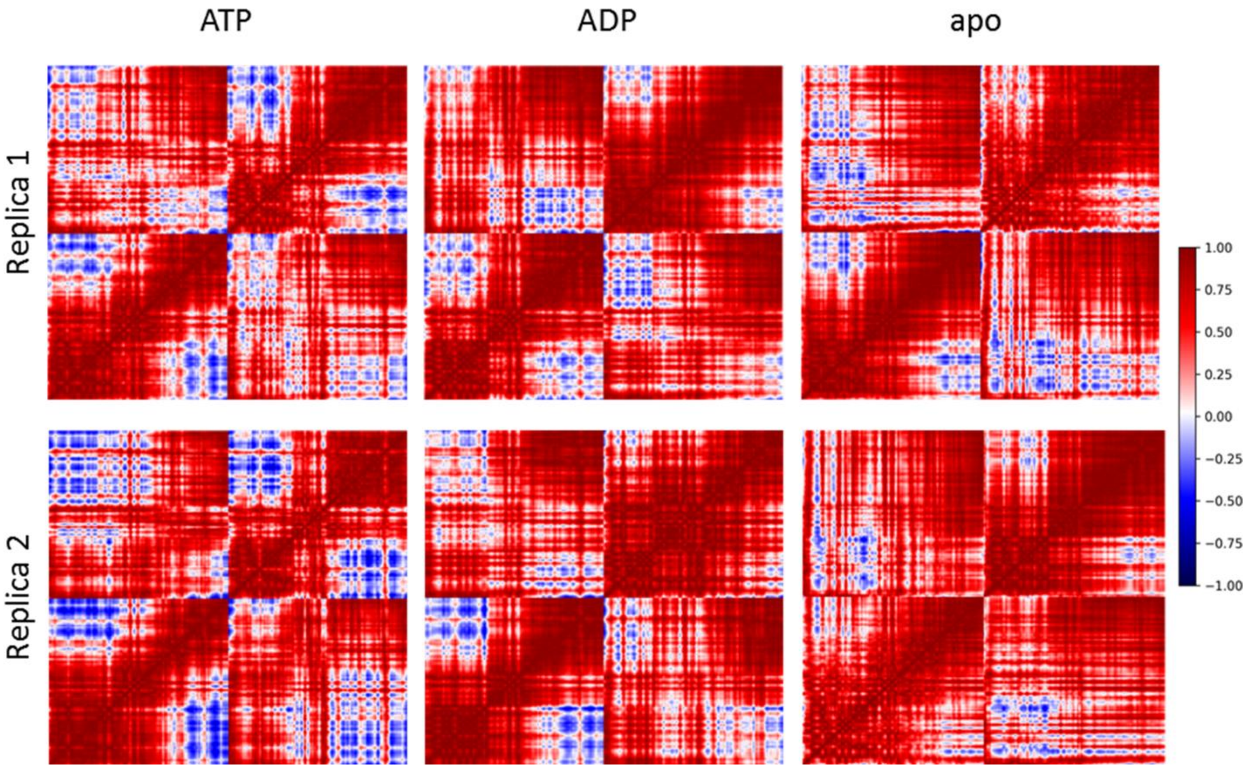
Fully-closed

Partially-open

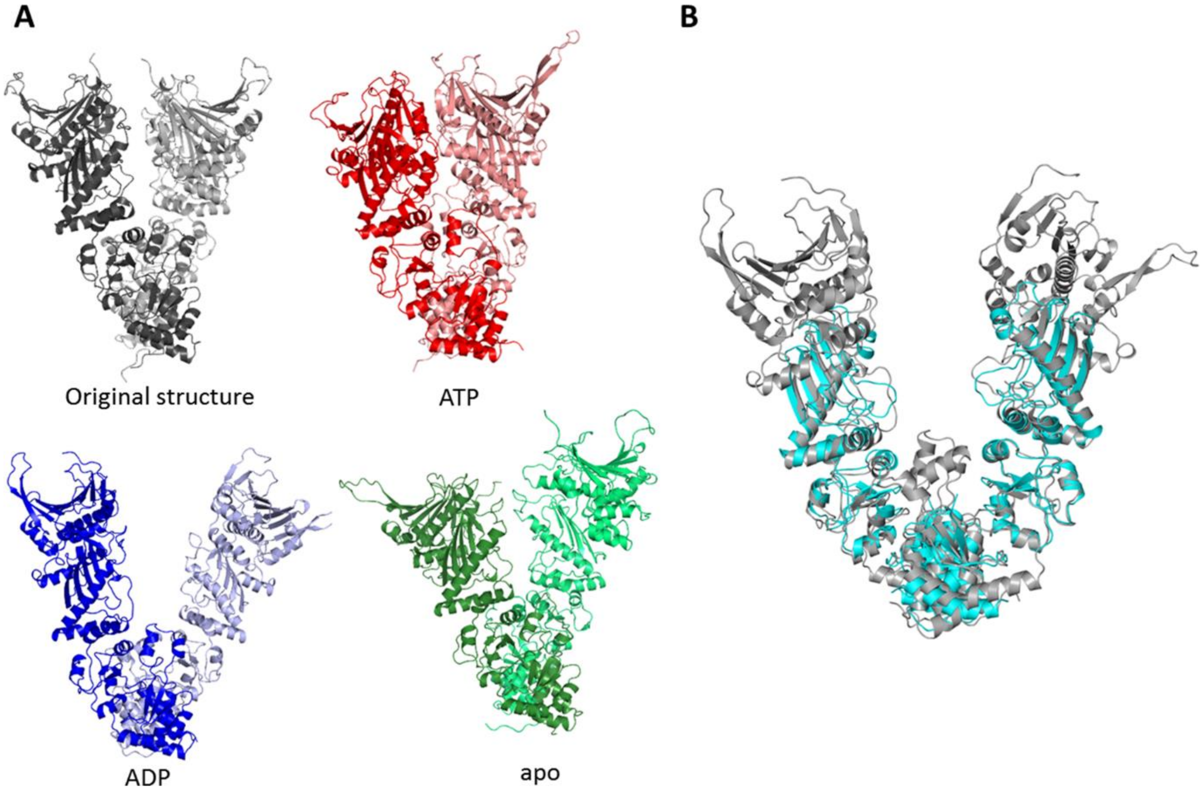
Appendix I-I: Illustration of the secondary structure qualities for the energy minimized homology models. (A) Closed conformation and **(B)** partially-open conformation. Quality colouring is from poor (red) to good (blue).



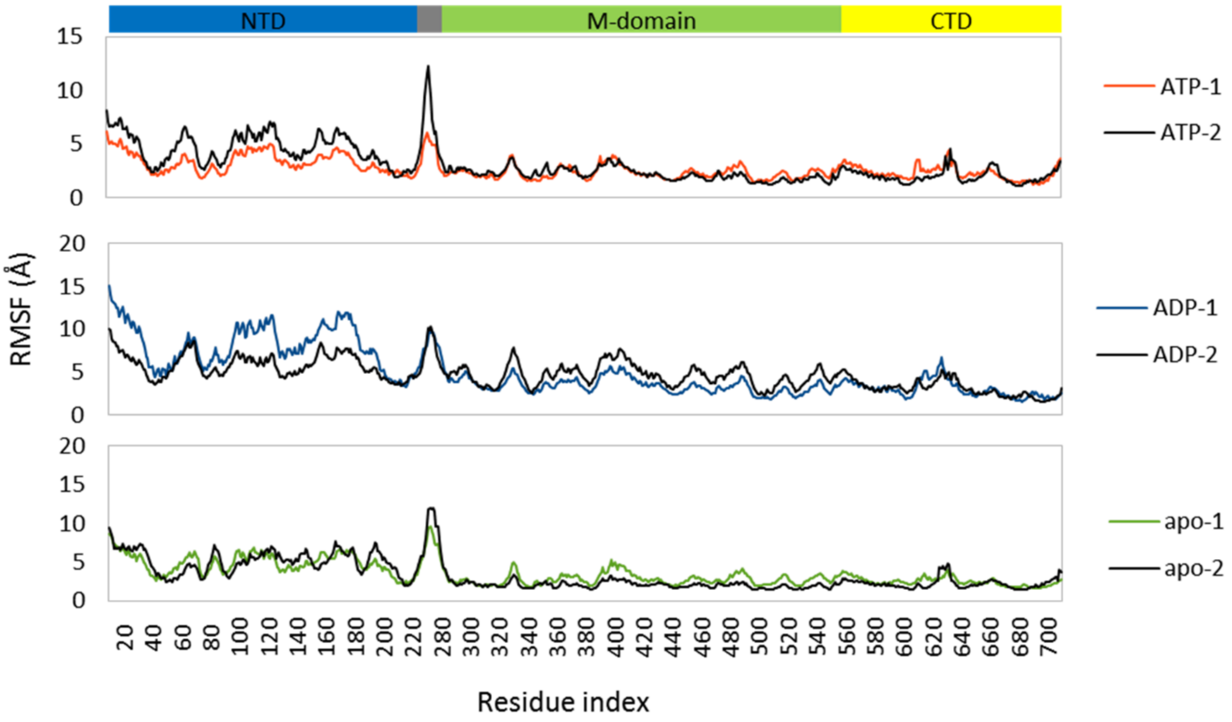
Appendix I-II: RMSF plots of closed conformation complexes. Protomer 1 is coloured by nucleotide configuration: ATP–red, ADP–blue, apo–green, and protomer 2 is in black. Penkler et. al 2018 [191]



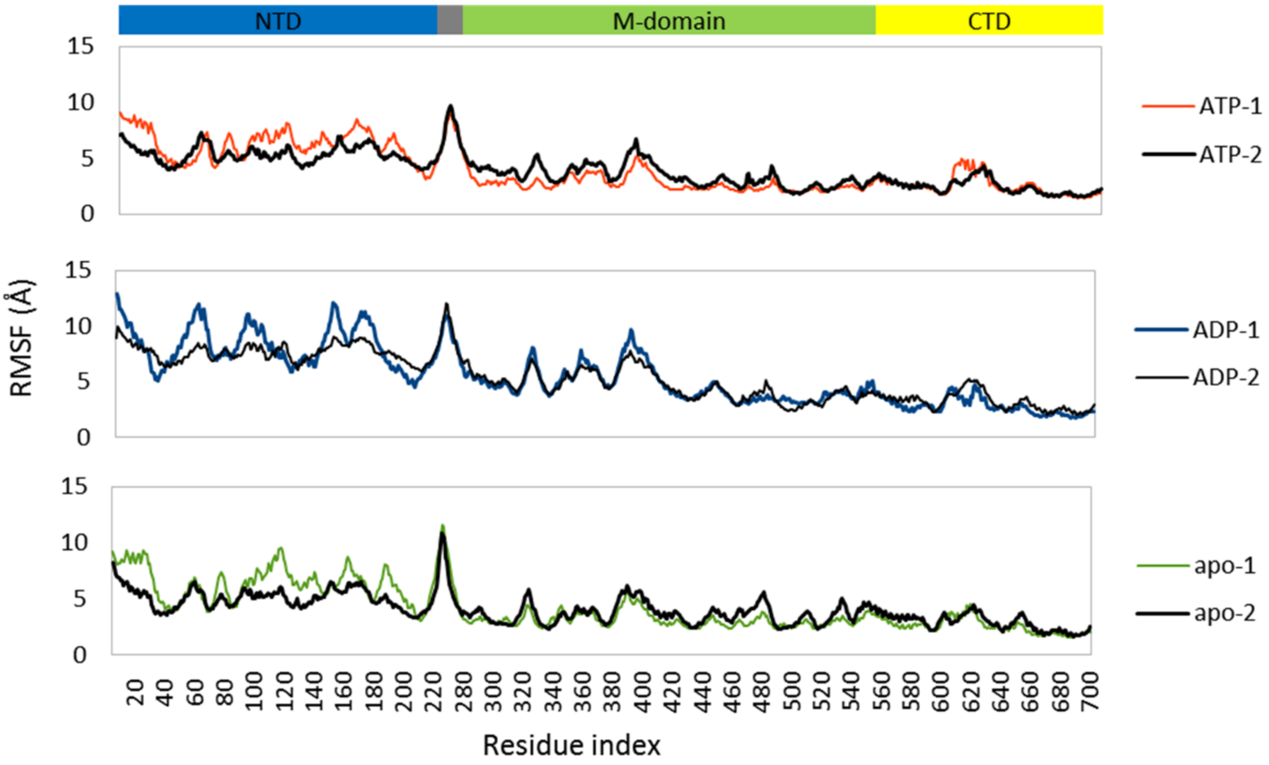
Appendix I-III: Comparison of replicate DCC heat maps for the closed conformation complexes. Penkler et. al 2018 [191]



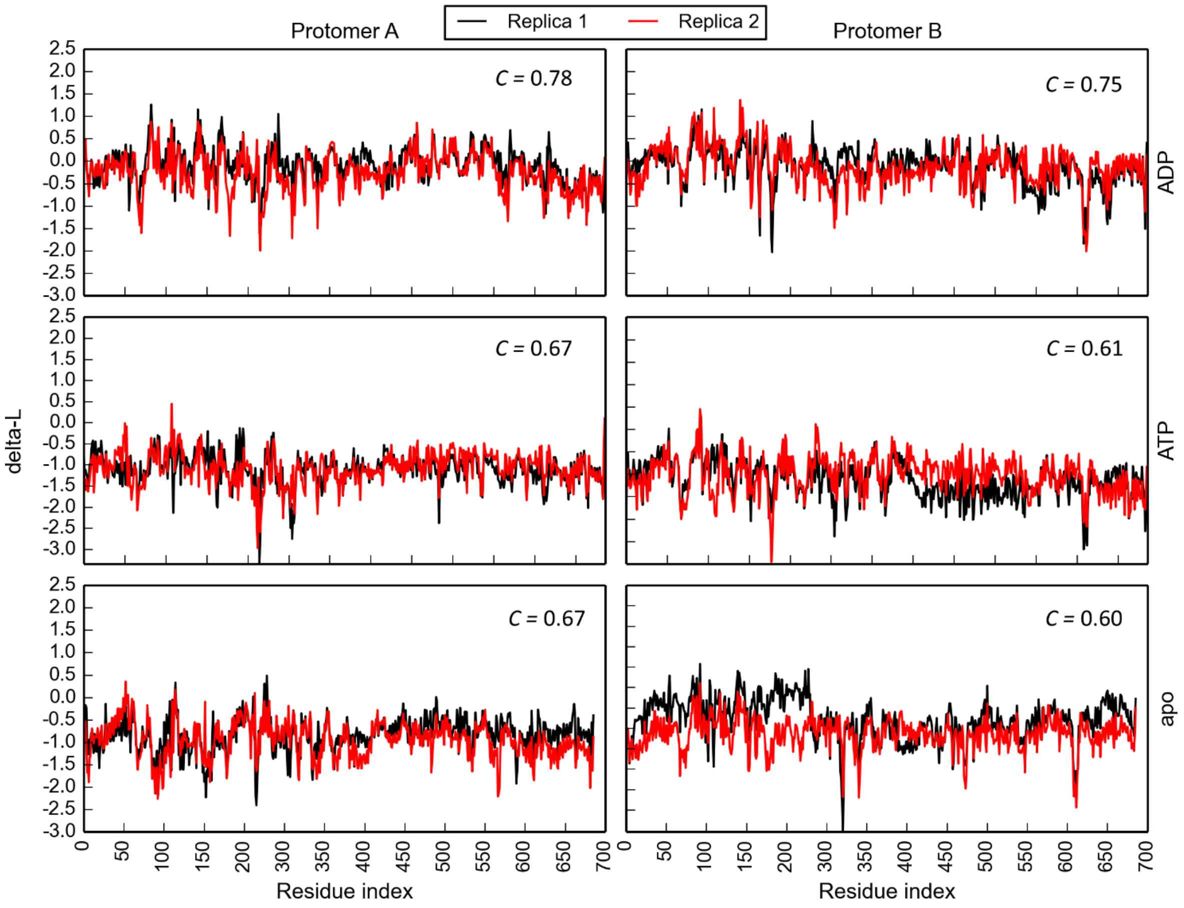
Appendix I-IV: Conformational restructuring of the PO complexes. (A) Final structures of the partially open complexes after 200 ns MD simulations, showing how the ADP structure opens compared to the ATP and apo complexes. **(B)** Bacterial HtpG structure (PDB 2IOQ) (cyan) superimposed on the 200 ns ADP complex structure (grey), showing the structural similarity. Penkler et. al 2018 [191]



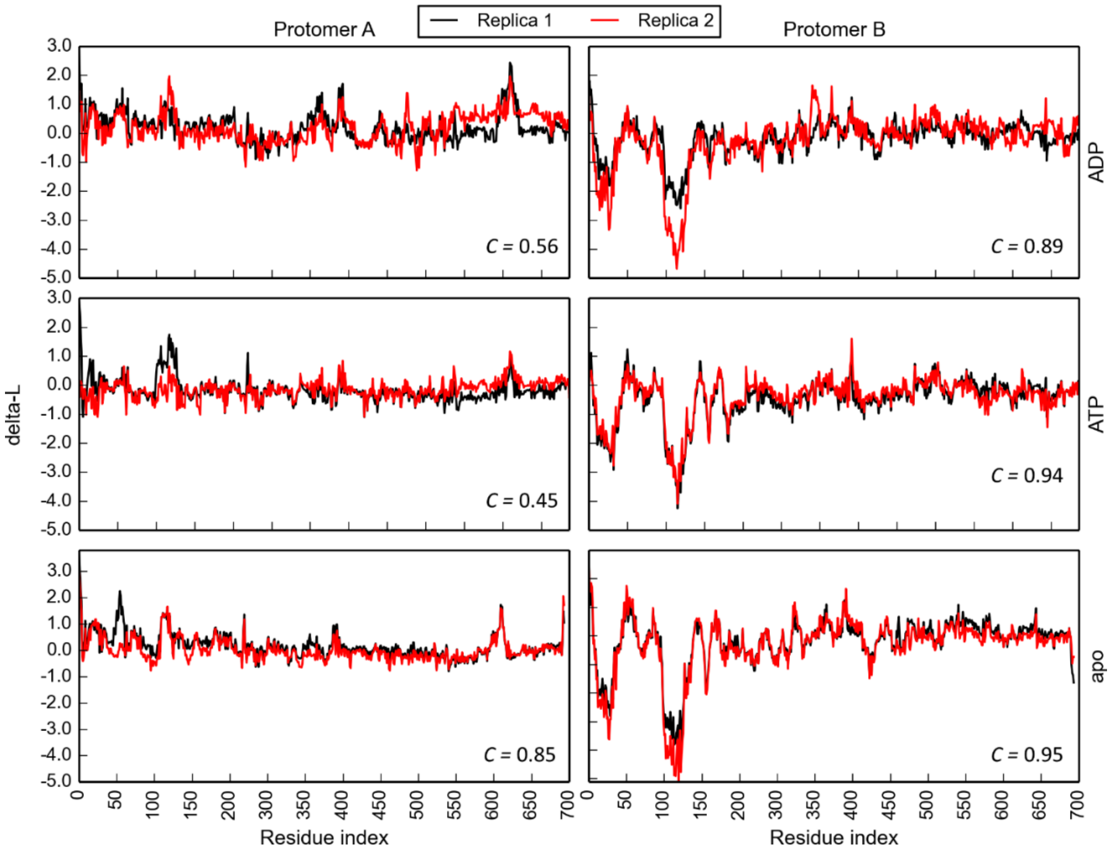
Appendix I-V: RMSF plots of the partially open complexes. Protomer 1 is coloured by nucleotide configuration: ATP–red, ADP–blue, apo–green, and protomer 2 is in black. Penkler et. al 2018 [191]



Appendix I-VI: RMSF plots for the fully-open complexes. Protomer 1 is coloured by nucleotide configuration: ATP–red, ADP–blue, apo–green, and protomer 2 is in black. Penkler et. al 2018 [191]



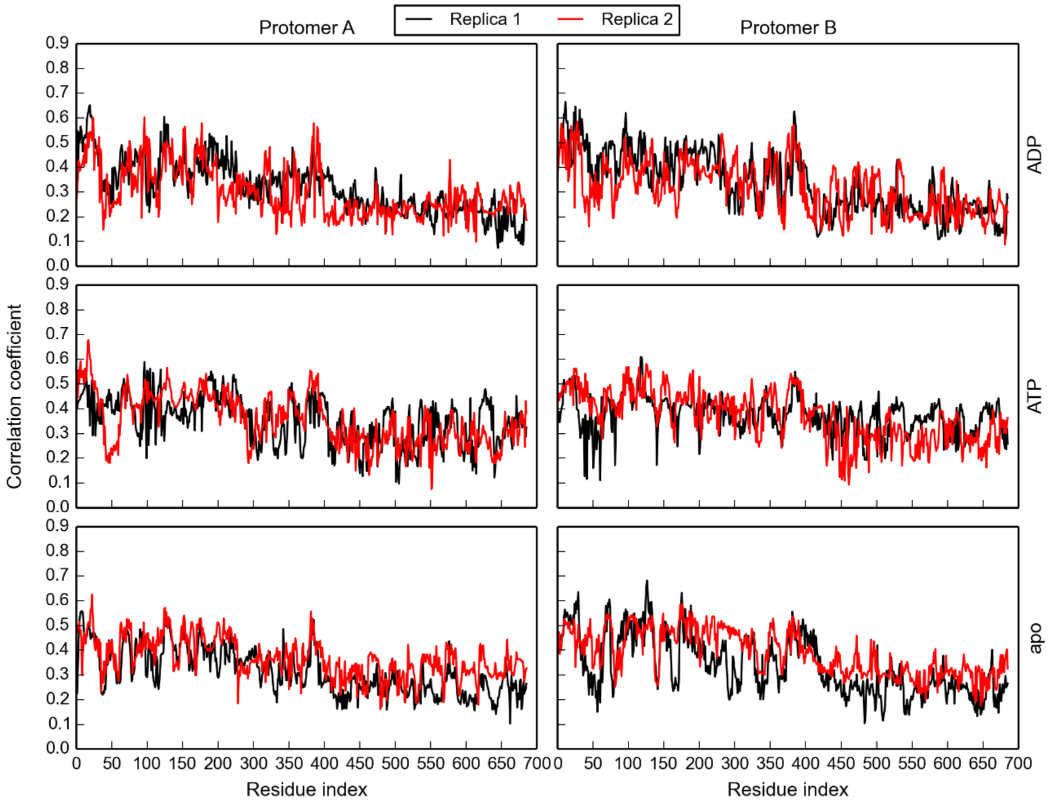
Appendix I-VII: Comparative *ΔLi* plots for the fully-closed replicate trajectories. *C* is the Pearson's correlation coefficient for each plot. Penkler et. al 2018 [191]



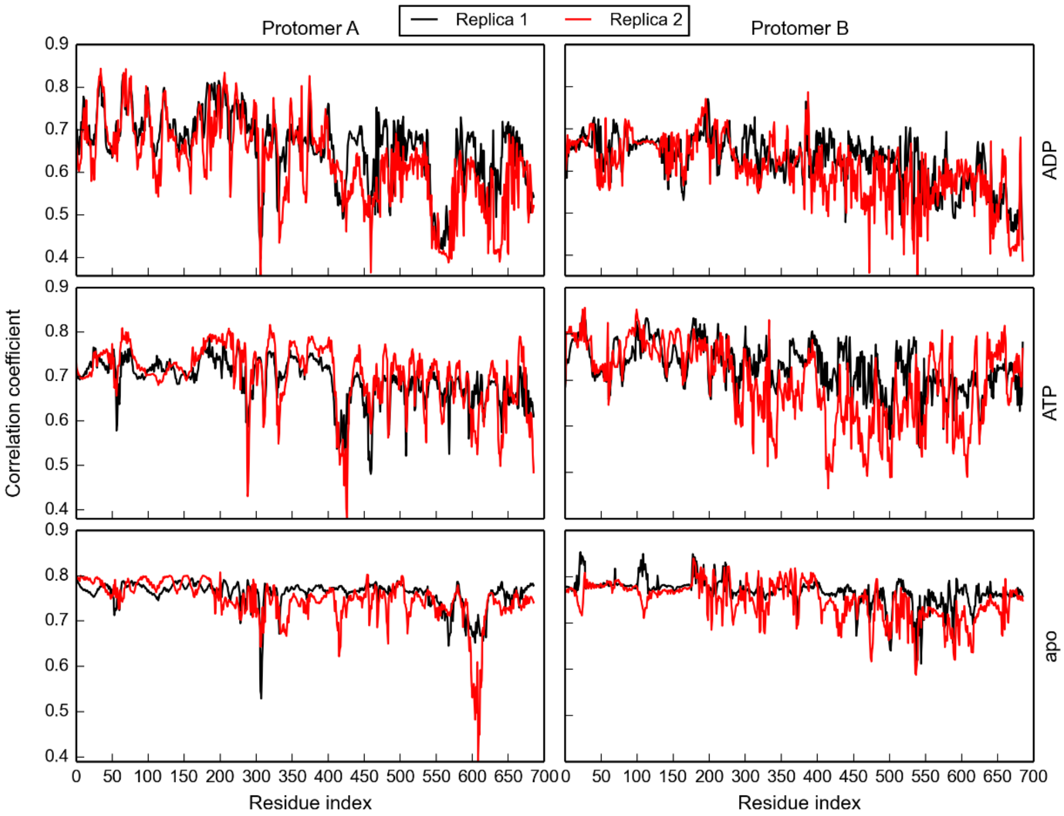
Appendix I-VIII: Comparative ΔLi plots for the fully open replicate trajectories. *C* is the Pearson's correlation coefficient for each plot. Penkler et. al 2018 [191]

			Protomer 1				Protomer 2			
			NTD	MiD	CTD	Whole	NTD	MiD	CTD	Whole
Fully-closed	BC vs L^{-1}	ADP	0.78	0.72	0.82	0.74	0.84	0.72	0.79	0.73
		ATP	0.84	0.73	0.76	0.74	0.75	0.68	0.80	0.70
		apo	0.76	0.72	0.81	0.73	0.82	0.72	0.78	0.73
	BC vs $RMSF^{-1}$	ADP	0.66	0.53	0.55	0.54	0.59	0.68	0.54	0.55
		ATP	0.63	0.56	0.5	0.55	0.58	0.52	0.55	0.50
		apo	0.59	0.31	0.54	0.42	0.51	0.39	0.60	0.45
	RMSF vs L	ADP	0.61	0.64	0.63	0.53	0.68	0.62	0.49	0.48
		ATP	0.69	0.69	0.64	0.60	0.55	0.70	0.52	0.56
		apo	0.87	0.75	0.76	0.77	0.69	0.68	0.78	0.72
Fully-open	BC vs L^{-1}	ADP	0.62	0.33	0.60	0.31	0.60	0.37	0.55	0.38
		ATP	0.58	0.34	0.57	0.30	0.62	0.33	0.56	0.38
		apo	0.59	0.34	0.56	0.32	0.57	0.36	0.54	0.37
	BC vs $RMSF^{-1}$	ADP	0.46	0.39	0.44	0.27	0.38	0.48	0.43	0.42
		ATP	0.46	0.45	0.29	0.29	0.51	0.45	0.42	0.44
		apo	0.49	0.45	0.40	0.36	0.39	0.57	0.31	0.43
	RMSF vs L	ADP	0.86	0.88	0.79	0.93	0.79	0.93	0.85	0.92
		ATP	0.76	0.78	0.72	0.84	0.90	0.88	0.89	0.94
		apo	0.86	0.76	0.86	0.88	0.82	0.73	0.67	0.74

Appendix I-IX: Pearson's correlation coefficient between BC, ΔL_i^{-1} and RMSF^{-1} for replicate FC-ATP and FO-ATP trajectories. Correlations were calculated for each protomer as a whole and for each separate domain (NTD, MiD, CTD). Penkler et. al 2018 [191]

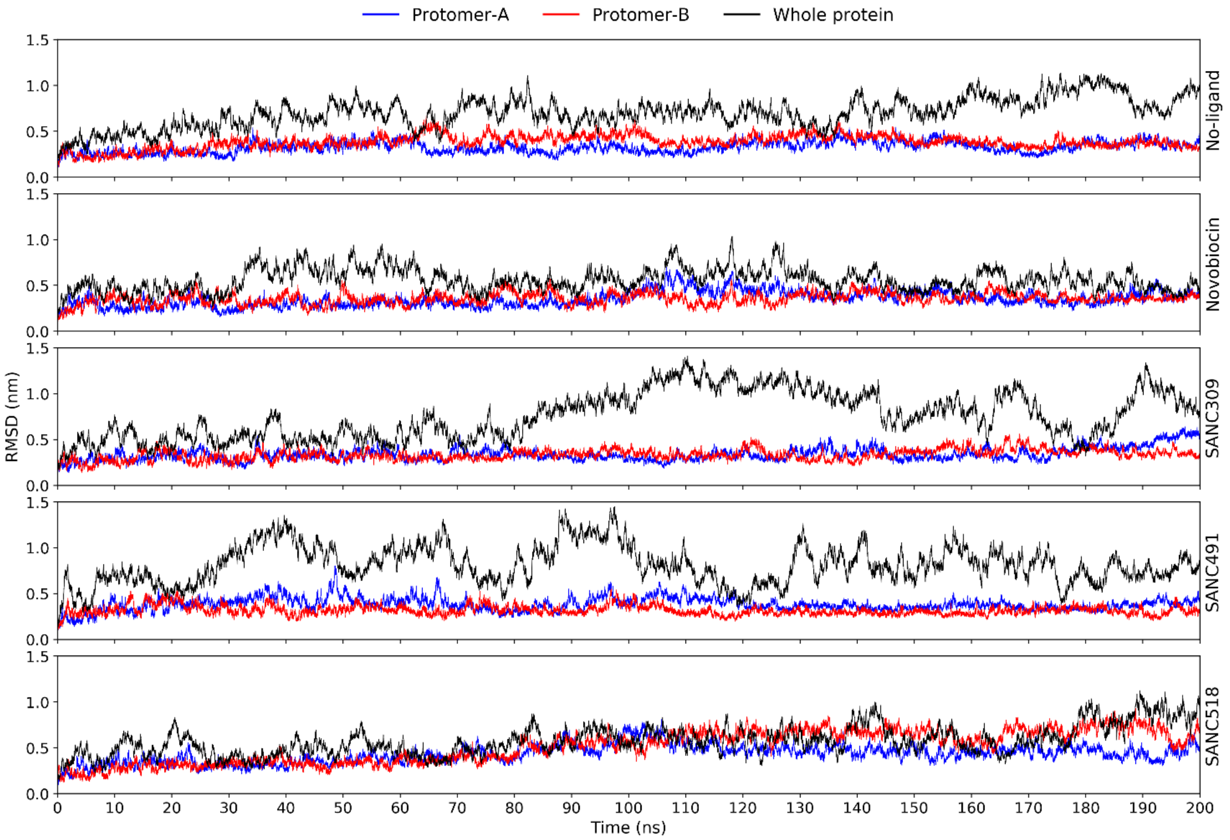


Appendix I-X: Comparative PRS profiles for separate duplicate MD trajectories of the opening transition experiments. *C* is the Pearson's correlation coefficient for each plot. Penkler et. al 2018 [191]



Appendix I-XI: Comparative PRS profiles for separate duplicate MD trajectories of the closing transition experiments. *C* is the Pearson's correlation coefficient for each plot. Penkler et. al 2018 [191]

Appendix II:



Appendix II-I: Backbone RMSD plots for each ligand bound complex. Black denotes the whole protein backbone RMSD, blue and red denote the backbone RMSD of each individual protomer. Reproduced with permission from Penkler et. al. 2018 [385]

	Mean	Variance	Std.	Wilcoxon	*p*	Z-stat	*p*	KS-stat	*p*
No-ligand	0.70	0.03	0.16	-	-	-	-	-	-
Novobiocin	0.56	0.02	0.12	-201.86	0.00	-219.90	0.00	0.16	0.51
Sanc309	0.76	0.07	0.26	33.21	0.00	55.48	0.00	0.18	0.36
Sanc491	0.81	0.04	0.21	120.36	0.00	134.08	0.00	0.08	1.00
Sanc518	0.57	0.03	0.16	-172.15	0.00	-177.47	0.00	0.10	0.95
No-ligand-A	0.32	0.00	0.06	-	-	-	-	-	-
Novobiocin-A	0.35	0.01	0.08	59.91	0.00	75.75	0.00	0.14	0.68
Sanc309-A	0.34	0.01	0.07	28.37	0.00	42.61	0.00	0.12	0.84
Sanc491-A	0.38	0.01	0.07	180.25	0.00	193.62	0.00	0.14	0.68
Sanc518-A	0.43	0.01	0.10	242.90	0.00	282.52	0.00	0.18	0.36
No-ligand-B	0.38	0.01	0.07	-	-	-	-	-	-
Novobiocin-B	0.35	0.00	0.06	-96.29	0.00	-91.41	0.00	0.08	1.00
Sanc309-B	0.33	0.00	0.06	-165.11	0.00	-165.15	0.00	0.10	0.95
Sanc491-B	0.32	0.00	0.05	-218.05	0.00	-227.32	0.00	0.12	0.84
Sanc518-B	0.51	0.03	0.18	150.36	0.00	214.71	0.00	0.20	0.24

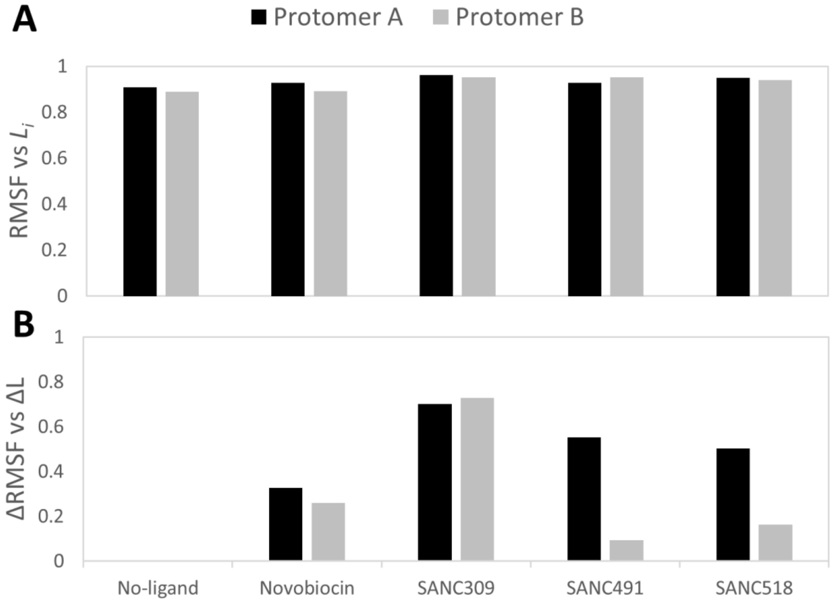
Appendix II-II: Whole protein and individual protomer RMSD distribution statistics. The means of each ligand bound complex was compared to the ligand-free system using the z-test statistic with α = 0.05 and a null hypothesis of H1 – H2 = 0. The two-sample KS-test statistic was computed to compare the difference in shape between the ligand un/bound systems. The Wilcoxon stat was used to assess the difference between the ligand un/bound distributions. Reproduced with permission from Penkler et. al. 2018 [385]

	Mean	Variance	Std.	Wilcoxon	*p*	Z-stat	*p*	KS-stat	*p*
No-ligand	7.00	0.35	0.59	-	-	-	-	-	-
Novobiocin	7.40	0.40	0.63	124.58	0.00	36.14	0.00	0.08	1.00
Sanc309	6.60	0.40	0.63	-136.46	0.00	36.14	0.00	0.18	0.36
Sanc491	8.09	0.49	0.70	288.06	0.00	36.14	0.00	0.14	0.68
Sanc518	7.12	0.66	0.81	39.26	0.00	36.14	0.00	0.10	0.95

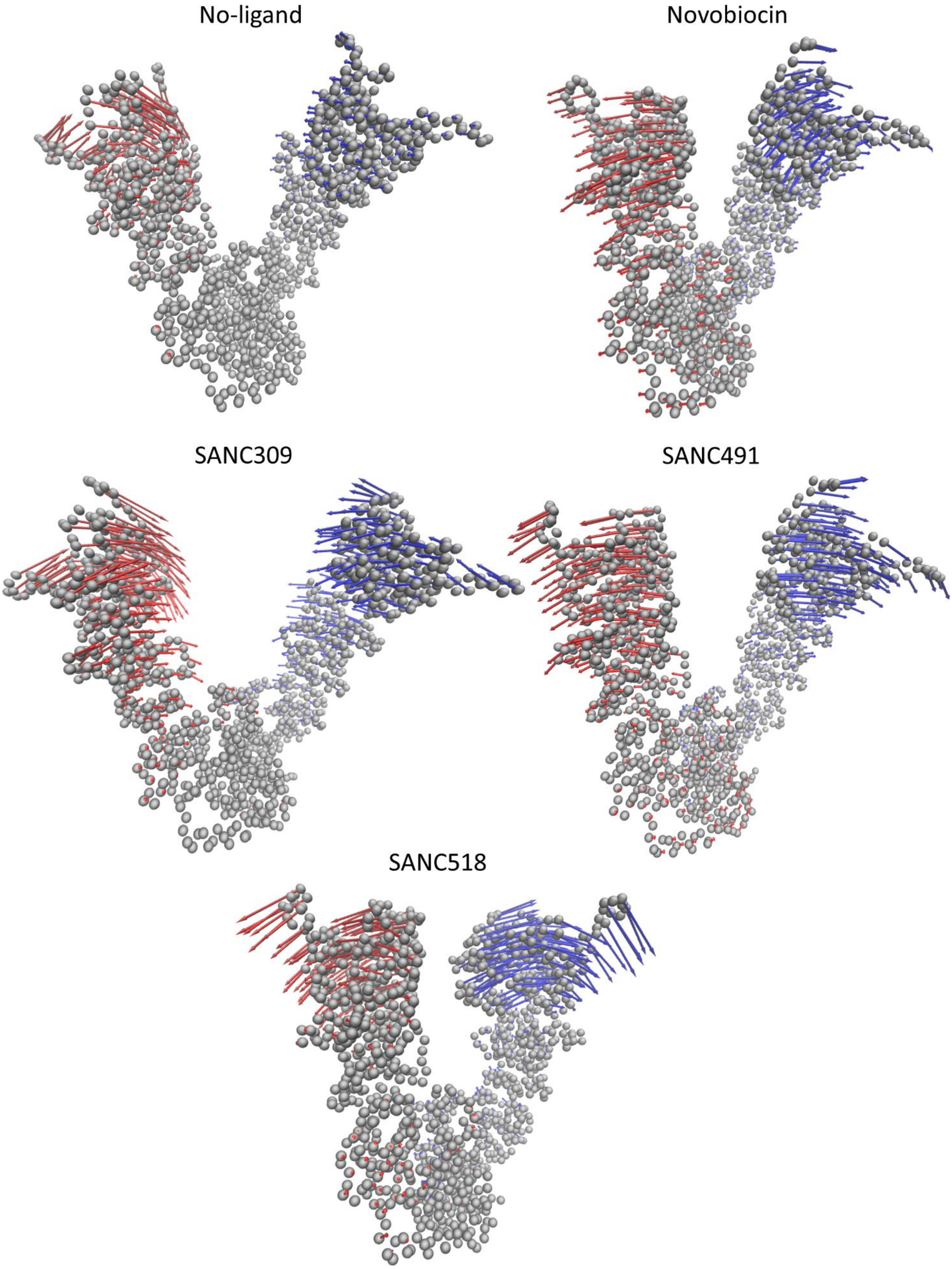
Appendix II-III: Inter-protomer distance distribution statistics. The means of each ligand bound complex was compared to the ligand-free system using the z-test statistic with α = 0.05 and a null hypothesis of H1 – H2 = 0. The two-sample KS-test statistic was computed to compare the difference in shape between the ligand un/bound systems. The Wilcoxon stat was used to assess the difference between the ligand un/bound distributions. Reproduced with permission from Penkler et. al. 2018 [385]

	Mean	Variance	Std.	Wilcoxon	*p*	Z-stat	*p*	KS-stat	*p*
No-ligand-A	8.21	0.06	0.24	-	-	-	-	-	-
Novobiocin-A	8.64	0.24	0.49	243.82	0.00	248.97	0.00	0.12	0.84
Sanc309-A	8.42	0.67	0.82	133.08	0.00	77.49	0.00	0.16	0.51
Sanc491-A	8.11	0.97	0.99	30.39	0.00	-30.98	0.00	0.26	0.06
Sanc518-A	7.90	1.49	1.22	6.06	0.00	-78.17	0.00	0.36	0.00
No-ligand-B	7.63	0.10	0.32	-	-	-	-	-	-
Novobiocin-B	7.75	0.09	0.29	83.65	0.00	85.86	0.00	0.20	0.24
Sanc309-B	8.01	1.01	1.01	211.95	0.00	112.62	0.00	0.14	0.68
Sanc491-B	8.06	0.14	0.37	267.11	0.00	277.49	0.00	0.18	0.36
Sanc518-B	7.91	0.21	0.46	170.17	0.00	156.74	0.00	0.12	0.84

Appendix II-IV: NTD-CTD distance distribution statistics. The means of each ligand bound complex was compared to the ligand-free system using the z-test statistic with α = 0.05 and a null hypothesis of H1 – H2 = 0. The two-sample KS-test statistic was computed to compare the difference in shape between the ligand un/bound systems. The Wilcoxon stat was used to assess the difference between the ligand un/bound distributions. Reproduced with permission from Penkler et. al. 2018 [385]



Appendix II-V: Pearson's correlation coefficients showing the relationship between: (A) L_i vs RMSF; **(B)** ΔRMSF vs ΔL_i Reproduced with permission from Penkler et. al. 2018 [385]



Appendix II-VI: Side view illustrating the atomic displacements of first eigenvector of each Hsp90 complex. The Hsp90 complex is represented by C_α atoms (grey spheres) and the arrows describe the relative direction and magnitude of the atomic displacements for protomer A (blue) and protomer B (red). Displacement arrows drawn for every 2nd C_α atom to simplify presentation. Reproduced with permission from Penkler et. al. 2018 [385]

Appendix II-VII: Movie illustration of the correlated motions of PC1 for the ligand-free complex. Depicting dimer closure as protomer A (green) approaches protomer B (cyan).

Appendix II-VIII: Movie illustration of the correlated motions of PC1 for the Novobiocin complex. Depicting dimer opening as protomer A (green) and protomer B (cyan) displace in opposite directions.

Appendix II-IX: Movie illustration of the correlated motions of PC1 for the SANC309 complex. Depicting dimer closure as protomer A (green) approaches protomer B (cyan).

Appendix II-X: Movie illustration of the correlated motions of PC1 for the SANC491 complex. Depicting dimer opening as protomer A (green) and protomer B (cyan) displace in opposite directions.

Appendix II-XI: Movie illustration of the correlated motions of PC1 for the SANC518 complex. Depicting dimer opening as protomer A (green) and protomer B (cyan) displace in opposite directions.

References

1. Monod J, Changeux J-P, Jacob F. Allosteric proteins and cellular control systems. J Mol Biol. 1963 Apr 1;6(4):306–29.

2. Tsai C-J, Nussinov R. A Unified View of “How Allostery Works.” Livesay DR, editor. PLoS Comput Biol. 2014 Feb 6;10(2):e1003394.

3. Kornev AP, Taylor SS. Dynamics-Driven Allostery in Protein Kinases. Trends Biochem Sci. 2015 Nov 1;40(11):628–47.

4. Pauling L. The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation. Proc Natl Acad Sci U S A. 1935 Apr;21(4):186–91.

5. Gerhart JC, Pardee AB. The enzymology of control by feedback inhibition. J Biol Chem. 1962 Mar;237:891–6.

6. Patte JC, Le Bras G, Loviny T, Cohen GN. Retro-inhibition and repression of the homoserine dehydrogenase of Escherichia coli. Biochim Biophys Acta. 1963 Jan 8;67:16–30.

7. Stadtman ER, Cohen GN, Le Bras G. Feedback inhibition and repression of aspartokinase activity in Escherichia coli. Ann N Y Acad Sci. 1961 Nov 2;94:952–9.

8. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. Nature. 2014 Apr 17;508(7496):331–9.

9. Monod J, Wyman J, Changeux J-P. On the nature of allosteric transitions: A plausible model. J Mol Biol. 1965 May 1;12(1):88–118.

10. Koshland DE, Némethy G, Filmer D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. Biochemistry. 1966 Jan;5(1):365–85.

11. Hilser VJ, Wrabl JO, Motlagh HN. Structural and Energetic Basis of Allostery. Annu Rev Biophys. 2012 Jun 9;41(1):585–609.

12. Cooper A, Dryden DT. Allostery without conformational change. A plausible model. Eur Biophys J. 1984;11(2):103–9.

13. Ptitsyn OB. Inter-domain mobility in proteins and its probable functional role. FEBS Lett. 1978

Sep 1;93(1):1–4.

14. Birktoft JJ, Kraut J, Freer ST. A detailed structural comparison between the charge relay system in chymotrypsinogen and in alpha-chymotrypsin. Biochemistry. 1976 Oct 5;15(20):4481–5.

15. Yonath A, Podjarny A, Honig B, Sielecki A, Traub W. Crystallographic studies of protein denaturation and renaturation. 2. Sodium dodecyl sulfate induced structural changes in triclinic lysozyme. Biochemistry. 1977 Apr 5;16(7):1418–24.

16. Levitt M, Warshel A. Computer simulation of protein folding. Nature. 1975 Feb 27;253(5494):694–8.

17. McCammon JA, Gelin BR, Karplus M, Wolynes PG. The hinge-bending mode in lysozyme. Nature. 1976 Jul 22;262(5566):325–6.

18. Honig B, Ray A, Levinthal C. Conformational flexibility and protein folding: rigid structural fragments connected by flexible joints in subtilisin BPN. Proc Natl Acad Sci U S A. 1976 Jun;73(6):1974–8.

19. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol. 2009 Nov 10;5(11):789–96.

20. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? Proteins Struct Funct Bioinforma. 2004 Jul 22;57(3):433–43.

21. Kar G, Keskin O, Gursoy A, Nussinov R. Allostery and population shift in drug discovery. Curr Opin Pharmacol. 2010 Dec;10(6):715–22.

22. Zhuravlev PI, Papoian GA. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. Q Rev Biophys. 2010 Aug 7;43(03):295–332.

23. Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science. 2001 Mar 23;291(5512):2429–33.

24. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. Nat Struct Mol Biol. 2006 Sep 13;13(9):831–8.

25. Gsponer J, Christodoulou J, Cavalli A, Bui JM, Richter B, Dobson CM, et al. A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction. Structure. 2008 May;16(5):736–46.

26. Tzeng S-R, Kalodimos CG. Protein dynamics and allostery: an NMR view. Curr Opin Struct Biol. 2011 Feb;21(1):62–7.

27. Tzeng S-R, Kalodimos CG. Dynamic activation of an allosteric regulatory protein. Nature. 2009 Nov 19;462(7271):368–72.

28. Dokholyan N V. Controlling Allosteric Networks in Proteins. Chemical Reviews. 2016.

29. Christopoulos A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. Nat Rev Drug Discov. 2002 Mar;1(3):198–210.

30. Szilágyi A, Nussinov R, Csermely P. Allo-network drugs: extension of the allosteric drug concept to protein- protein interaction and signaling networks. Curr Top Med Chem. 2013;13(1):64–77.

31. Lindsley CW. 2013 Philip S. Portoghese Medicinal Chemistry Lectureship: drug discovery targeting allosteric sites. J Med Chem. 2014 Sep 25;57(18):7485–98.

32. Wenthur CJ, Gentry PR, Mathews TP, Lindsley CW. Drugs for Allosteric Sites on Receptors. Annu Rev Pharmacol Toxicol. 2014 Jan 6;54(1):165–84.

33. Lu S, Ji M, Ni D, Zhang J. Discovery of hidden allosteric sites as novel targets for allosteric drug design. Drug Discov Today. 2018 Feb 1;23(2):359–65.

34. Nussinov R, Tsai C-J, Csermely P. Allo-network drugs: harnessing allostery in cellular networks. Trends Pharmacol Sci. 2011 Dec;32(12):686–93.

35. Dokholyan N V., Shakhnovich EI. Understanding hierarchical protein evolution from first principles 1 1Edited by J. Thornton. J Mol Biol. 2001 Sep 7;312(1):289–307.

36. Proctor EA, Kota P, Demarest SJ, Caravella JA, Dokholyan N V. Highly covarying residues have a functional role in antibody constant domains. Proteins Struct Funct Bioinforma. 2013 May;81(5):884–95.

37. Kannan N, Wu J, Anand GS, Yooseph S, Neuwald AF, Venter JC, et al. Evolution of allostery in the cyclic nucleotide binding module. Genome Biol. 2007;8(12):R264.

38. Shukla D, Meng Y, Roux B, Pande VS. Activation pathway of Src kinase reveals intermediate states as targets for drug design. Nat Commun. 2014 Dec 3;5(1):3397.

39. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. Chem Rev. 2016 Apr 13;116(11):6370–90.

40. Amaro RE. Toward Understanding “the Ways” of Allosteric Drugs. ACS Cent Sci. 2017 Sep 27;3(9):925–6.

41. Mittermaier A, Kay LE. New Tools Provide New Insights in NMR Studies of Protein Dynamics. Science (80-). 2006 Apr 14;312(5771):224–8.

42. Manley G, Loria JP. NMR insights into protein allostery. Arch Biochem Biophys. 2012 Mar 15;519(2):223–31.

43. Tzeng S-R, Kalodimos CG. Allosteric inhibition through suppression of transient conformational states. Nat Chem Biol. 2013 Jul 5;9(7):462–5.

44. Karplus M, Kuriyan J. Molecular dynamics and protein function. Proc Natl Acad Sci U S A. 2005 May 10;102(19):6679–85.

45. Vendruscolo M, Dokholyan N V., Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Phys Rev E. 2002 Jun 25;65(6):061910.

46. Atilgan AR, Akan P, Baysal C. Small-World Communication of Residues and Significance for Protein Dynamics. Biophys J. 2004 Jan;86(1):85–91.

47. Bagler G, Sinha S. Network properties of protein structures. Phys A Stat Mech its Appl. 2005 Feb 1;346(1–2):27–33.

48. Greene LH, Higman VA. Uncovering Network Systems Within Protein Structures. J Mol Biol. 2003 Dec 5;334(4):781–91.

49. Taylor NR. Small world network strategies for studying protein structures and binding. Comput Struct Biotechnol J. 2013;5(February):e201302006.

50. Guo J, Zhou H-X. Protein Allostery and Conformational Dynamics. Chem Rev. 2016 Jun 8;116(11):6503–15.

51. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci U S A. 2002 Jun 11;99(12):7821–6.

52. Atilgan C. Computational Methods for Efficient Sampling of Protein Landscapes and Disclosing Allosteric Regions. Adv Protein Chem Struct Biol. 2018 Jan 1;113:33–63.

53. Tasumi M, Takeuchi H, Ataka S, Dwivedi AM, Krimm S. Normal vibrations of proteins: Glucagon. Biopolymers. 1982 Mar;21(3):711–4.

54. Levitt M, Sander C, Stern PS. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. Int J Quantum Chem. 1983 Jun 19;24(10 S):181–99.

55. Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci U S A. 1983 Jun 1;80(12):3696–700.

56. Bahar I, Rader A. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol. 2005 Oct;15(5):586–92.

57. Ma J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. Structure. 2005 Mar 1;13(3):373–80.

58. Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. Phys Rev Lett. 1996 Aug 26;77(9):1905–8.

59. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins Struct Funct Genet. 1998 Nov 15;33(3):417–29.

60. Haliloglu T, Bahar I, Erman B. Gaussian Dynamics of Folded Proteins. Phys Rev Lett. 1997 Oct 20;79(16):3090–3.

61. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J. 2001 Jan;80(1):505–15.

62. VanWart AT, Eargle J, Luthey-Schulten Z, Amaro RE. Exploring Residue Component Contributions to Dynamical Network Models of Allostery. J Chem Theory Comput. 2012 Aug 14;8(8):2949–61.

63. Su JG, Qi LS, Li CH, Zhu YY, Du HJ, Hou YX, et al. Prediction of allosteric sites on protein surfaces with an elastic-network-model-based thermodynamic method. Phys Rev E. 2014 Aug 27;90(2):022719.

64. Chennubhotla C, Bahar I. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. Mol Syst Biol. 2006;2:36.

65. Zheng W, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. Proc Natl Acad Sci. 2006 May 16;103(20):7664–9.

66. Schueler-Furman O, Wodak SJ. Computational approaches to investigating allostery. Current Opinion in Structural Biology Elsevier Current Trends; Dec 1, 2016 p. 159–71.

67. Chiti F, Dobson CM. Protein Misfolding, Functional Amyloid, and Human Disease. Annu Rev Biochem. 2006 Jun;75(1):333–66.

68. M. van Noort J, Bugiani M, Amor S. Heat Shock Proteins: Old and Novel Roles in Neurodegenerative Diseases in the Central Nervous System.

69. Young JC, Agashe VR, Siegers K, Hartl FU. Pathways of chaperone-mediated protein folding in the cytosol. Nat Rev Mol Cell Biol. 2004 Oct 1;5(10):781–91.

70. Hartl F. Molecular Chaperones in the Cytosol: from Nascent Chainto Folded Protein. Science (80-). 2002;295(5561):1852.

71. Ellis J. Proteins as molecular chaperones. Nature. 1987 Jul 30;328(6129):378–9.

72. Latchman D. Heat shock proteins and human disease. J R Coll Physicians. 1991;(25):295–399.

73. Lindquist S, Craig EA. The Heat-Shock Proteins. Annu Rev Genet. 1988 Dec 28;22(1):631–77.

74. Bukau B, Horwich AL. The Hsp70 and Hsp60 chaperone machines. Cell. 1998 Feb 6;92(3):351–66.

75. Glover JR, Lindquist S. Hsp104, Hsp70, and Hsp40: A novel chaperone system that rescues previously aggregated proteins. Cell. 1998;

76. Mogk A, Tomoyasu T, Goloubinoff P, Rüdiger S, Röder D, Langen H, et al. Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. EMBO J. 1999 Dec 15;18(24):6934–49.

77. Mayer MP, Brehmer D, Gässler CS, Bukau B. Hsp70 chaperone machines. Advances in Protein Chemistry. 2001.

78. Mayer MP, Bukau B. Hsp70 chaperones: Cellular functions and molecular mechanism. Cell Mol Life Sci. 2005;62:670–84.

79. Chappell TG, Konforti BB, Schmid SL, Rothman JE. The ATPase core of a clathrin uncoating protein. J Biol Chem. 1987;

80. Ungewickell E. The 70-kd mammalian heat shock proteins are structurally and functionally related to the uncoating protein that releases clathrin triskelia from coated vesicles. EMBO J. 1985;

81. Voisine C, Craig EA, Zufall N, Von Ahsen O, Pfanner N, Voos W. The protein import motor of mitochondria: Unfolding and trapping of preproteins are distinct and separable functions of matrix Hsp70. Cell. 1999;

82. Lyman SK, Schekman R. Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. Cell. 1997;

83. Matlack KES, Misselwitz B, Plath K, Rapoport TA. BIP acts as a molecular ratchet during posttranslational transport of prepro-α factor across the ER membrane [2]. Cell. 1999.

84. Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A, Bukau B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature. 1999;

85. Goloubinoff P, Mogk A, Zvi APB, Tomoyasu T, Bukau B. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proc Natl Acad Sci. 1999;

86. Ben-Zvi AP, Goloubinoff P. Review: Mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. Journal of Structural Biology. 2001.

87. Teter SA, Houry WA, Ang D, Tradler T, Rockabrand D, Fischer G, et al. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell. 1999;

88. Flaherty KM, Wilbanks SM, DeLuca-Flaherty C, McKay DB. Structural basis of the 70-kilodalton heat shock cognate protein ATP hydrolytic activity. II. Structure of the active site with ADP or ATP bound to wild type and mutant ATPase fragment. J Biol Chem. 1994;

89. Harrison CJ, Hayer-Hartl M, Di Liberto M, Hartl F, Kuriyan J. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science. 1997 Apr 18;276(5311):431–5.

90. Zhang Y, Zuiderweg ERP. The 70-kDa heat shock protein chaperone nucleotide-binding domain

in solution unveiled as a molecular machine that can reorient its functional subdomains. Proc Natl Acad Sci U S A. 2004 Jul 13;101(28):10272–7.

91. Bertelsen EB, Chang L, Gestwicki JE, Zuiderweg ERP. Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. Proc Natl Acad Sci U S A. 2009 May 26;106(21):8471–6.

92. Gässler CS, Wiederkehr T, Brehmer D, Bukau B, Mayer MP. Bag-1M Accelerates Nucleotide Release for Human Hsc70 and Hsp70 and Can Act Concentration-dependent as Positive and Negative Cofactor. J Biol Chem. 2001;

93. Buchberger A, Theyssen H, Schroder H, McCarty JS, Virgallita G, Milkereit P, et al. Nucleotide-induced Conformational Changes in the ATPase and Substrate Binding Domains of the DnaK Chaperone Provide Evidence for Interdomain Communication. J Biol Chem. 1995 Jul 14;270(28):16903–10.

94. Bauer D, Merz DR, Pelz B, Theisen KE, Yacyshyn G, Mokranjac D, et al. Nucleotides regulate the mechanical hierarchy between subdomains of the nucleotide binding domain of the Hsp70 chaperone DnaK. Proc Natl Acad Sci U S A. 2015 Aug 18;112(33):10389–94.

95. Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, et al. Structural analysis of substrate binding by the molecular chaperone DnaK. Science. 1996 Jun 14;272(5268):1606–14.

96. Demand J, Lüders J, Höhfeld J. The carboxy-terminal domain of Hsc70 provides binding sites for a distinct set of chaperone cofactors. Mol Cell Biol. 1998;

97. Ballinger CA, Connell P, Wu Y, Hu Z, Thompson LJ, Yin LY, et al. Identification of CHIP, a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. Mol Cell Biol. 1999;

98. Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, et al. Structure of TPR domain-peptide complexes: Critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell. 2000;

99. Zhang M, Windheim M, Roe SM, Peggie M, Cohen P, Prodromou C, et al. Chaperoned ubiquitylation - Crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. Mol Cell. 2005;

100. Smock RG, Blackburn ME, Gierasch LM. Conserved, disordered C terminus of DnaK enhances cellular survival upon stress and DnaK in vitro chaperone activity. J Biol Chem. 2011;

101. Kityk R, Kopp J, Sinning I, Mayer MP. Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones. Mol Cell. 2012 Dec 28;48(6):863–74.

102. Qi R, Sarbeng EB, Liu Q, Le KQ, Xu X, Xu H, et al. Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP. Nat Struct Mol Biol. 2013 May 26;20(7):900–7.

103. Swain JF, Dinler G, Sivendran R, Montgomery DL, Stotz M, Gierasch LM. Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. Mol Cell. 2007 Apr 13;26(1):27–39.

104. Zhuravleva A, Clerico EM, Gierasch LM. An interdomain energetic tug-of-war creates the allosterically active state in Hsp70 molecular chaperones. Cell. 2012 Dec 7;151(6):1296–307.

105. Penkler D, Sensoy Ö, Atilgan C, Tastan Bishop Ö. Perturbation-Response Scanning Reveals Key Residues for Allosteric Control in Hsp70. J Chem Inf Model. 2017 Jun 26;57(6):1359–74.

106. Mapa K, Sikor M, Kudryavtsev V, Waegemann K, Kalinin S, Seidel CAM, et al. The Conformational Dynamics of the Mitochondrial Hsp70 Chaperone. Mol Cell. 2010;38(1):89–100.

107. Marcinowski M, Höller M, Feige MJ, Baerend D, Lamb DC, Buchner J. Substrate discrimination of the chaperone BiP by autonomous and cochaperone-regulated conformational transitions. Nat Struct Mol Biol. 2011;

108. Schlecht R, Erbse AH, Bukau B, Mayer MP. Mechanics of Hsp70 chaperones enables differential interaction with client proteins. Nat Struct Mol Biol. 2011;

109. Mayer MP, Rüdiger S, Bukau B. Molecular basis for interactions of the DnaK chaperone with substrates. Biol Chem. 2000;381(9–10):877–85.

110. Szabo A, Langer T, Schröder H, Flanagan J, Bukau B, Hartl FU. The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. Proc Natl Acad Sci U S A. 1994 Oct 25;91(22):10345–9.

111. Gamer J, Multhaup G, Tomoyasu T, McCarty JS, Rüdiger S, Schönfeld HJ, et al. A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the Escherichia coli

heat shock transcription factor sigma32. EMBO J. 1996;15(3):607–17.

112. Kampinga HH, Craig E a. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol. 2010;11(AuguST):579–92.

113. Flynn GC, Chappell TG, Rothman JE. Peptide binding and release by proteins implicated as catalysts of protein assembly. Science (80-). 1989;

114. Gao B, Emoto Y, Greene L, Eisenberg E. Nucleotide binding properties of bovine brain uncoating ATPase. J Biol Chem. 1993;

115. Jeung-Hoi H, McKay DB. ATPase Kinetics of Recombinant Bovine 70 kDa Heat Shock Cognate Protein and Its Amino-Terminal ATPase Domain. Biochemistry. 1994;

116. Jordan R, McMacken R. Modulation of the ATPase activity of the molecular chaperone DnaK by peptides and the DnaJ and GrpE heat shock proteins. J Biol Chem. 1995;

117. Karzai AW, McMacken R. A bipartite signaling mechanism involved in DnaJ-mediated activation of the Escherichia coli DnaK protein. J Biol Chem. 1996 May 10;271(19):11236–46.

118. McCarty JS, Buchberger A, Reinstein J, Bukau B. The role of ATP in the functional cycle of the DnaK chaperone system. J Mol Biol. 1995;

119. Theyssen H, Schuster HP, Packschies L, Bukau B, Reinstein J. The second step of ATP binding to DnaK induces peptide release. J Mol Biol. 1996;263(5):657–70.

120. Laufen T, Mayer MP, Beisel C, Klostermeier D, Mogk A, Reinstein J, et al. Mechanism of regulation of hsp70 chaperones by DnaJ cochaperones. Proc Natl Acad Sci U S A. 1999 May 11;96(10):5452–7.

121. Kelley WL. The J-domain family and the recruitment of chaperone power. Trends in Biochemical Sciences. 1998.

122. Mayer MP, Bukau B. Hsp70 chaperone systems: diversity of cellular functions and mechanism of action. Biol Chem. 1998;

123. Dekker PJT, Pfanner N. Role of mitochondrial GrpE and phosphate in the ATPase cycle of matrix Hsp70. J Mol Biol. 1997;

124. Liberek K, Marszalek J, Ang D, Georgopoulos C, Zylicz M. Escherichia coli DnaJ and GrpE heat

shock proteins jointly stimulate ATPase activity of DnaK. Proc Natl Acad Sci. 1991;

125. Miao B, Davis JE, Craig EA. Mge1 functions as a nucleotide release factor for Ssc1, a mitochondrial Hsp70 of Saccharomyces cerevisiae. J Mol Biol. 1997;

126. Bracher A, Verghese J. The nucleotide exchange factors of Hsp70 molecular chaperones. Front Mol Biosci. 2015;2(10):1–9.

127. Melero R, Moro F, Pérez-Calvo MÁ, Perales-Calvo J, Quintana-Gallardo L, Llorca O, et al. Modulation of the chaperone DnaK allosterism by the nucleotide exchange factor GrpE. J Biol Chem. 2015 Apr 17;290(16):10083–92.

128. Rauch JN, Gestwicki JE. Binding of Human Nucleotide Exchange Factors to Heat Shock Protein 70 (Hsp70) Generates Functionally Distinct Complexes in Vitro. J Biol Chem. 2014;289(3):1402–14.

129. Zuiderweg ERP, Bertelsen EB, Rousaki A, Mayer MP, Gestwicki JE, Ahmad A. Allostery in the Hsp70 chaperone proteins. Top Curr Chem. 2013 Jan;328:99–153.

130. Kityk R, Vogel M, Schlecht R, Bukau B, Mayer MP. Pathways of allosteric regulation in Hsp70 chaperones. Nat Commun. 2015 Sep 18;6:8308.

131. Mayer MP, Kityk R. Insights into the molecular mechanism of allostery in Hsp70s. Front Mol Biosci. 2015;

132. Shen Y, Liu J, Wang X, Cheng X, Wang Y, Wu N. Essential role of the first intron in the transcription of hsp90beta gene. FEBS Lett. 1997;

133. Zhang SL, Yu J, Cheng XK, Ding L, Heng FY, Wu NH, et al. Regulation of human hsp90α gene expression. FEBS Lett. 1999;

134. Chen B, Zhong D, Monteiro AA. Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. BMC Genomics. 2006;7(1):156.

135. Argon Y, Simen BB. GRP94, an ER chaperone with protein and peptide binding properties. Semin Cell Dev Biol. 1999;

136. Marzec M, Eletto D, Argon Y. GRP94: An HSP90-like protein specialized for protein folding and quality control in the endoplasmic reticulum. Biochim Biophys Acta. 2012 Mar;1823(3):774–87.

137. Felts SJ, Owen BAL, Nguyen PM, Trepel J, Donner DB, Toft DO. The hsp90-related protein TRAP1 is a mitochondrial protein with distinct functional properties. J Biol Chem. 2000;

138. Jakob U, Lilie H, Meyer I, Buchner J. Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase. Implications for heat shock in vivo. J Biol Chem. 1995;(270):7288–94.

139. Echeverría PC, Bernthaler A, Dupuis P, Mayer B, Picard D. An interaction network predicted from public data as a discovery tool: Application to the Hsp90 molecular chaperone machine. PLoS One. 2011;

140. Pratt WB, Toft DO. Steroid receptor interactions with heat shock protein and immunophilin chaperones. Endocr Rev. 1997;

141. Brugge JS, Erikson RL. Identification of a transformation-specific antigen induced by an avian sarcoma virus. Nature. 1977;

142. Picard D. Chaperoning steroid hormone action. Trends in Endocrinology and Metabolism. 2006.

143. Xu Y, Lindquist S. Heat-shock protein hsp90 governs the activity of pp60v-src kinase. Proc Natl Acad Sci. 1993;

144. Picard D, Khursheed B, Garabedian MJ, Fortin MG, Lindquist S, Yamamoto KR. Reduced levels of hsp90 compromise steroid receptor action in vivo. Nature. 1990;

145. Taipale M, Krykbaeva I, Koeva M, Kayatekin C, Westover KD, Karras GI, et al. Quantitative analysis of Hsp90-client interactions reveals principles of substrate recognition. Cell. 2012;

146. Grammatikakis N, Lin JH, Grammatikakis A, Tsichlis PN, Cochran BH. p50(cdc37) acting in concert with Hsp90 is required for Raf-1 function. Mol Cell Biol. 1999;

147. Boczek EE, Reefschläger LG, Dehling M, Struller TJ, Häusler E, Seidl A, et al. Conformational processing of oncogenic v-Src kinase by the molecular chaperone Hsp90. Proc Natl Acad Sci. 2015;

148. Kitagawa K, Skowyra D, Elledge SJ, Harper JW, Hieter P. SGT1 encodes an essential component of the yeast kinetochore assembly pathway and a novel subunit of the SCF ubiquitin ligase complex. Mol Cell. 1999;

149. Zhao R, Kakihara Y, Gribun A, Huen J, Yang G, Khanna M, et al. Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. J Cell Biol. 2008;

150. Lorenz OR, Freiburger L, Rutz DA, Krause M, Zierer BK, Alvira S, et al. Modulation of the Hsp90 chaperone cycle by a stringent client protein. Mol Cell. 2014 Mar 20;53(6):941–53.

151. Pratt WB, Dittmar KD. Studies with purified chaperones advance the understanding of the mechanism of glucocorticoid receptor-hsp90 heterocomplex assembly. Trends in Endocrinology and Metabolism. 1998.

152. Kirschke E, Goswami D, Southworth D, Griffin PR, Agard DA. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. Cell. 2014;

153. Toogun OA, DeZwaan DC, Freeman BC. The Hsp90 Molecular Chaperone Modulates Multiple Telomerase Activities. Mol Cell Biol. 2008;

154. DeZwaan DC, Toogun OA, Echtenkamp FJ, Freeman BC. The Hsp82 molecular chaperone promotes a switch between unextendable and extendable telomere states. Nat Struct Mol Biol. 2009;

155. Iwasaki S, Sasaki HM, Sakaguchi Y, Suzuki T, Tadakuma H, Tomari Y. Defining fundamental steps in the assembly of the Drosophila RNAi enzyme complex. Nature. 2015;

156. Echtenkamp FJ, Gvozdenov Z, Adkins NL, Zhang Y, Lynch-Day M, Watanabe S, et al. Hsp90 and p23 Molecular Chaperones Control Chromatin Architecture by Maintaining the Functional Pool of the RSC Chromatin Remodeler. Mol Cell. 2016;

157. Mayer MP, Le?Breton L, Le Breton L. Hsp90: Breaking the Symmetry. Mol Cell. 2015 Apr 2;58(1):8–20.

158. Wayne N, Bolon DN. Dimerization of Hsp90 is required for in vivo function: Design and analysis of monomers and dimers. J Biol Chem. 2007 Nov 30;282(48):35386–95.

159. Schopf FH, Biebl MM, Buchner J. The HSP90 chaperone machinery. Nat Rev Mol Cell Biol. 2017 Apr 21;18(6):345–60.

160. Mayer MP, Le Breton L. Hsp90: Breaking the Symmetry. Mol Cell. 2015 Apr;58(1):8–20.

161. Prodromou C, Roe SM, O’Brien R, Ladbury JE, Piper PW, Pearl LH. Identification and Structural

Characterization of the ATP/ADP-Binding Site in the Hsp90 Molecular Chaperone. Cell. 1997 Jul;90(1):65–75.

162. Meyer P, Prodromou C, Hu B, Vaughan C, Roe SM, Panaretou B, et al. Structural and Functional Analysis of the Middle Segment of Hsp90: Implications for ATP Hydrolysis and Client Protein and Cochaperone Interactions. Mol Cell. 2003 Mar;11(3):647–58.

163. Ali MMU, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, et al. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature. 2006 Apr 20;440(7087):1013–7.

164. Harris S, Shiau A, Agard D. The crystal structure of the carboxy- terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site. Structure. 2004;(12):1087–97.

165. Hainzl O, Lapina MC, Buchner J, Richter K. The charged linker region is an important regulator of Hsp90 function. J Biol Chem. 2009 Aug 21;284(34):22559–67.

166. Jahn M, Rehn A, Pelz B, Hellenkamp B, Richter K, Rief M, et al. The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function. Proc Natl Acad Sci U S A. 2014 Dec 16;111(50):17881–6.

167. Tsutsumi S, Mollapour M, Prodromou C, Lee C-T, Panaretou B, Yoshida S, et al. Charged linker sequence modulates eukaryotic heat shock protein 90 (Hsp90) chaperone activity. Proc Natl Acad Sci U S A. 2012 Feb 21;109(8):2937–42.

168. Buchner J. Hsp90 and Co. - A holding for folding. Trends in Biochemical Sciences. 1999.

169. Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci. 2000 Jan;25(1):24–8.

170. Roe SM, Ali MMU, Meyer P, Vaughan CK, Panaretou B, Piper PW, et al. The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). Cell. 2004 Jan 9;116(1):87–98.

171. Zhang W, Hirshberg M, McLaughlin SH, Lazar GA, Grossmann JG, Nielsen PR, et al. Biochemical and structural studies of the interaction of Cdc37 with Hsp90. J Mol Biol. 2004;

172. Stebbins C, Russo A, Schneider C, Rosen N, Hartl F, Pavletich N. Crystal structure of an Hsp90-

geldanamycin complex: targeting of a protein chaperone by an antitumor agent. Cell. 1997;(89):239–50.

173. Meyer P, Prodromou C, Liao C, Hu B, Roe SM, Vaughan CK, et al. Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. EMBO J. 2004 Mar 24;23(6):1402–10.

174. Huai Q, Wang H, Liu Y, Kim HY, Toft D, Ke H. Structures of the N-terminal and middle domains of E. coli Hsp90 and conformation changes upon ADP binding. Structure. 2005;

175. Minami Y, Kimura H, Kawasaki H, Suzuki K, Yahara I. The carboxy-terminal region of mammalian HSP90 is required for its dimerization and function in vivo. Mol Cell Biol. 1994;(14):1459–64.

176. Söti C, Rácz A, Csermely P. A Nucleotide-dependent molecular switch controls ATP binding at the C-terminal domain of Hsp90. N-terminal nucleotide binding unmasks a C-terminal binding pocket. J Biol Chem. 2002 Mar 1;277(9):7066–75.

177. Garnier C, Lafitte D, Tsvetkov PO, Barbier P, Leclerc-Devin J, Millot J-M, et al. Binding of ATP to heat shock protein 90: evidence for an ATP-binding site in the C-terminal domain. J Biol Chem. 2002 Apr 5;277(14):12208–14.

178. Marcu MG, Chadli A, Bouhouche I, Catelli M, Neckers LM. The Heat Shock Protein 90 Antagonist Novobiocin Interacts with a Previously Unrecognized ATP-binding Domain in the Carboxyl Terminus of the Chaperone. J Biol Chem. 2000 Nov 24;275(47):37181–6.

179. Prodromou C, Panaretou B, Chohan S, Siligardi G, O'Brien R, Ladbury JE, et al. The ATPase cycle of Hsp90 drives a molecular "clamp" via transient dimerization of the N-terminal domains. EMBO J. 2000 Aug 15;19(16):4383–92.

180. Shiau AK, Harris SF, Southworth DR, Agard D a. Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. Cell. 2006 Oct 20;127(2):329–40.

181. Southworth DR, Agard DA, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, et al. Species-dependent ensembles of conserved conformational states define the Hsp90 chaperone ATPase cycle. Mol Cell. 2008 Dec 5;32(5):631–40.

182. Southworth DR, Agard D a. Client-loading conformation of the Hsp90 molecular chaperone

revealed in the cryo-EM structure of the human Hsp90:Hop complex. Mol Cell. 2011 Jun 24;42(6):771–81.

183. Prodromou C. The “active life” of Hsp90 complexes. Biochim Biophys Acta. 2012 Mar;1823(3):614–23.

184. Scheibel T, Neuhofen S, Weikl T, Mayr C, Reinstein J, Vogel PD, et al. ATP-binding properties of human Hsp90. J Biol Chem. 1997 Jul 25;272(30):18608–13.

185. McLaughlin SH, Smith HW, Jackson SE. Stimulation of the weak ATPase activity of human hsp90 by a client protein. J Mol Biol. 2002 Jan 25;315(4):787–98.

186. Hellenkamp B, Wortmann P, Kandzia F, Zacharias M, Hugel T. Multidomain structure and correlated dynamics determined by self-consistent FRET networks. Nat Methods. 2016 Feb;14(2):174–80.

187. Schulze A, Beliu G, Helmerich DA, Schubert J, Pearl LH, Prodromou C, et al. Cooperation of local motions in the Hsp90 molecular chaperone ATPase mechanism. Nat Chem Biol. 2016 Jun 20;12(8):628–35.

188. Dollins DE, Warren JJ, Immormino RM, Gewirth DT. Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. Mol Cell. 2007 Oct 12;28(1):41–56.

189. Röhl A, Rohrberg J, Buchner J. The chaperone Hsp90: Changing partners for demanding clients. Vol. 38, Trends in Biochemical Sciences. 2013. p. 253–62.

190. Panaretou B, Siligardi G, Meyer P, Maloney A, Sullivan JK, Singh S, et al. Activation of the ATPase activity of hsp90 by the stress-regulated cochaperone aha1. Mol Cell. 2002 Dec;10(6):1307–18.

191. Penkler DL, Atilgan C, Tastan Bishop Ö. Allosteric Modulation of Human Hsp90α Conformational Dynamics. J Chem Inf Model. 2018 Jan 29;58(2):383–404.

192. Verba KA, Wang RYR, Arakawa A, Liu Y, Shirouzu M, Yokoyama S, et al. Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. Science (80-). 2016 Jun 24;352(6293):1542–7.

193. Prodromou C. Mechanisms of Hsp90 regulation. Biochem J. 2016;

194. Voellmy R, Boellmann F. Chaperone regulation of the heat shock protein response. Advances in Experimental Medicine and Biology. 2007.

195. Soroka J, Wandinger SK, Mäusbacher N, Schreiber T, Richter K, Daub H, et al. Conformational Switching of the Molecular Chaperone Hsp90 via Regulated Phosphorylation. Mol Cell. 2012 Feb 24;45(4):517–28.

196. Mollapour M, Neckers L. Post-translational modifications of Hsp90 and their contributions to chaperone regulation. Biochim Biophys Acta. 2012 Mar;1823(3):648–55.

197. Mollapour M, Tsutsumi S, Truman AW, Xu W, Vaughan CK, Beebe K, et al. Threonine 22 phosphorylation attenuates Hsp90 interaction with cochaperones and affects its chaperone activity. Mol Cell. 2011 Mar 18;41(6):672–81.

198. Zhao YG, Gilmore R, Leone G, Coffey MC, Weber B, Lee PWK. Hsp90 phosphorylation is linked to its chaperoning function: Assembly of the reovirus cell attachment protein. J Biol Chem. 2001;

199. Mollapour M, Tsutsumi S, Donnelly AC, Beebe K, Tokita MJ, Lee M-J, et al. Swe1Wee1-dependent tyrosine phosphorylation of Hsp90 regulates distinct facets of chaperone function. Mol Cell. 2010 Feb 12;37(3):333–43.

200. Bali P, Pranpat M, Bradner J, Balasis M, Fiskus W, Guo F, et al. Inhibition of histone deacetylase 6 acetylates and disrupts the chaperone function of heat shock protein 90: A novel basis for antileukemia activity of histone deacetylase inhibitors. J Biol Chem. 2005;

201. Scroggins BT, Robzyk K, Wang D, Marcu MG, Tsutsumi S, Beebe K, et al. An Acetylation Site in the Middle Domain of Hsp90 Regulates Chaperone Function. Mol Cell. 2007;

202. García-Cardeña G, Fan R, Shah V, Sorrentino R, Cirino G, Papapetropoulos A, et al. Dynamic activation of endothelial nitric oxide synthase by Hsp90. Nature. 1998;

203. Retzlaff M, Stahl M, Eberl HC, Lagleder S, Beck J, Kessler H, et al. Hsp90 is regulated by a switch point in the C-terminal domain. EMBO Rep. 2009 Oct;10(10):1147–53.

204. Johnson BD. Hop Modulates hsp70/hsp90 Interactions in Protein Folding. J Biol Chem. 1998 Feb 6;273(6):3679–86.

205. Wegele H, Wandinger SK, Schmid AB, Reinstein J, Buchner J. Substrate transfer from the

chaperone Hsp70 to Hsp90. J Mol Biol. 2006;

206. Li J, Richter K, Buchner J. Mixed Hsp90-cochaperone complexes are important for the progression of the reaction cycle. Nat Struct Mol Biol. 2011;

207. Richter K, Muschler P, Hainzl O, Reinstein J, Buchner J. Sti1 is a non-competitive inhibitor of the Hsp90 ATPase. Binding prevents the N-terminal dimerization reaction during the ATPase cycle. J Biol Chem. 2003;

208. Prodromou C, Siligardi G, O'Brien R, Woolfson DN, Regan L, Panaretou B, et al. Regulation of Hsp90 ATPase activity by tetratricopeptide repeat (TPR)-domain co-chaperones. EMBO J. 1999 Feb 1;18(3):754–62.

209. Vaughan CK, Mollapour M, Smith JR, Truman A, Hu B, Good VM, et al. Hsp90-Dependent Activation of Protein Kinases Is Regulated by Chaperone-Targeted Dephosphorylation of Cdc37. Mol Cell. 2008;

210. Bose S, Weikl T, Bügl H, Buchner J. Chaperone function of Hsp90-associated proteins. Science (80-). 1996;

211. Freeman BC, Toft DO, Morimoto RI. Molecular chaperone machines: Chaperone activities of the cyclophilin Cyp-40 and the steroid aporeceptor-associated protein p23. Science (80-). 1996;

212. Siligardi G, Panaretou B, Meyerc P, Singh S, Woolfson DN, Piper PW, et al. Regulation of Hsp90 ATPase activity by the Co-chaperone Cdc37p/p50 cdc37. J Biol Chem. 2002;

213. Li J, Richter K, Reinstein J, Buchner J. Integration of the accelerator Aha1 in the Hsp90 co-chaperone cycle. Nat Struct Mol Biol. 2013 Feb 10;20(3):326–31.

214. Richter K, Walter S, Buchner J. The co-chaperone Sba1 connects the ATPase reaction of Hsp90 to the progression of the chaperone cycle. J Mol Biol. 2004;

215. Johnson JL, Toft DO. A novel chaperone complex for steroid receptors involving heat shock proteins, immunophilins, and p23. J Biol Chem. 1994;

216. Smith DF. Chaperones in progesterone receptor complexes. Seminars in Cell and Developmental Biology. 2000.

217. Patricia Hernández M, Chadli A, Toft DO. HSP40 binding is the first step in the HSP90

chaperoning pathway for the progesterone receptor. J Biol Chem. 2002;

218. Smith DF, Toft DO. Steroid receptors and their associated proteins. Mol Endocrinol. 1993;

219. Young JC. Polypeptide release by Hsp90 involves ATP hydrolysis and is enhanced by the co-chaperone p23. EMBO J. 2000;

220. Whitesell L, Lindquist SL. HSP90 and the chaperoning of cancer. Nat Rev Cancer. 2005 Sep 20;5(10):761–72.

221. Jhaveri K, Taldone T, Modi S, Chiosis G. Advances in the clinical development of heat shock protein 90 (Hsp90) inhibitors in cancers. Biochim Biophys Acta. 2012 Mar;1823(3):742–55.

222. Taldone T, Ochiana SO, Patel PD, Chiosis G. Selective targeting of the stress chaperome as a therapeutic strategy. Trends Pharmacol Sci. 2014;35(11):592–603.

223. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011.

224. Neckers L. Heat shock protein 90: The cancer chaperone. J Biosci. 2007;

225. Sidera K, Patsavoudi E. Extracellular HSP90. Conquering the cell surface. Cell Cycle. 2008.

226. Mileo AM, Fanuele M, Battaglia F, Scambia G, Benedetti-Panici P, Mancuso S, et al. Selective over-expression of mRNA coding for 90 KDa stress-protein in human ovarian cancer. Anticancer Res. 1990;

227. Nanbu K, Konishi I, Komatsu T, Mandai M, Yamamoto S, Kuroda H, et al. Expression of heat shock proteins HSP70 and HSP90 in endometrial carcinomas: Correlation with clinicopathology, sex steroid receptor status, and p53 protein expression. Cancer. 1996;

228. Scaltriti M, Dawood S, Cortes J. Molecular pathways: Targeting Hsp90 - Who benefits and who does not. Clin Cancer Res. 2012;

229. De Mattos-Arruda L, Cortes J. Breast cancer and HSP90 inhibitors: Is there a role beyond the HER2-positive subtype? Breast. 2012;

230. Franzén B, Linder S, Alaiya AA, Eriksson E, Fujioka K, Bergman AC, et al. Analysis of polypeptide expression in benign and malignant human breast lesions. In: Electrophoresis. 1997.

231. Gallegos Ruiz MI, Floor K, Roepman P, Rodriguez JA, Meijer GA, Mooi WJ, et al. Integration of gene dosage and gene expression in non-small cell lung cancer, identification of HSP90 as

potential target. PLoS One. 2008;

232. Li CF, Huang WW, Wu JM, Yu SC, Hu TH, Uen YH, et al. Heat shock protein 90 overexpression independently predicts inferior disease-free survival with differential expression of the α and β isoforms in gastrointestinal stromal tumors. Clin Cancer Res. 2008;

233. Solit DB, Scher HI, Rosen N. Hsp90 as a Therapeutic Target in Prostate Cancer. Seminars in Oncology. 2003.

234. Workman P, Burrows F, Neckers L, Rosen N. Drugging the cancer chaperone HSP90: Combinatorial therapeutic exploitation of oncogene addiction and tumor stress. In: Annals of the New York Academy of Sciences. 2007.

235. Pearl L, Prodromou C, Workman P. The Hsp90 molecular chaperone: an open and shut case for treatment. J Biochem. 2008;(410):439–53.

236. Ung PM-U, Thompson AD, Chang L, Gestwicki JE, Carlson HA. Identification of key hinge residues important for nucleotide-dependent allostery in E. coli Hsp70/DnaK. PLoS Comput Biol. 2013 Jan 21;9(11):10032–79.

237. Vogel M, Bukau B, Mayer MP. Allosteric regulation of Hsp70 chaperones by a proline switch. Mol Cell. 2006 Feb 3;21(3):359–67.

238. Chiappori F, Merelli I, Milanesi L, Colombo G, Morra G, Mayer MP, et al. An atomistic view of Hsp70 allosteric crosstalk: from the nucleotide to the substrate binding domain and back. Sci Rep. 2016 Mar 30;6:23474.

239. Nicolaï A, Delarue P, Senet P. Decipher the mechanisms of protein conformational changes induced by nucleotide binding through free-energy landscape analysis: ATP binding to Hsp70. PLoS Comput Biol. 2013 Jan 12;9(12):10033–79.

240. Montgomery DL, Morimoto RI, Gierasch LM. Mutations in the substrate binding domain of the Escherichia coli 70 kDa molecular chaperone, DnaK, which alter substrate affinity or interdomain coupling. J Mol Biol. 1999 Feb 26;286(3):915–32.

241. Zhuravleva A, Gierasch LM. Substrate-binding domain conformational dynamics mediate Hsp70 allostery. Proc Natl Acad Sci. 2015 Jun 2;112(22):2865–73.

242. Burkholder WF, Zhao X, Zhu X, Hendrickson WA, Gragerov A, Gottesman ME. Mutations in the

C-terminal fragment of DnaK affecting peptide binding. Proc Natl Acad Sci U S A. 1996 Oct 1;93(20):10632–7.

243. Clerico EM, Tilitsky JM, Meng W, Gierasch LM. How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. J Mol Biol. 2015 Apr 10;427(7):1575–88.

244. Nicolaï A, Senet P, Delarue P, Ripoll DR. Human Inducible Hsp70: Structures, Dynamics, and Interdomain Communication from All-Atom Molecular Dynamics Simulations. J Chem Theory Comput. 2010 Aug 10;6(8):2501–19.

245. Gołaś E, Maisuradze GG, Senet P, Ołdziej S, Czaplewski C, Scheraga HA, et al. Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. J Chem Theory Comput. 2012 May 8;8(5):1750–64.

246. Nicolaï A, Delarue P, Senet P. Conformational dynamics of full-length inducible human Hsp70 derived from microsecond molecular dynamics simulations in explicit solvent. J Biomol Struct Dyn. 2013 Oct;31(10):1111–26.

247. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011 Jul 21;475(7356):324–32.

248. Chiappori F, Merelli I, Colombo G, Milanesi L, Morra G. Molecular mechanism of allosteric communication in Hsp70 revealed by molecular dynamics simulations. PLoS Comput Biol. 2012 Jan 27;8(12):10028–44.

249. Smock RG, Rivoire O, Russ WP, Swain JF, Leibler S, Ranganathan R, et al. An interdomain sector mediating allostery in Hsp70 molecular chaperones. Mol Syst Biol. 2010 Sep 21;6:414.

250. General IJ, Liu Y, Blackburn ME, Mao W, Gierasch LM, Bahar I. ATPase subdomain IA is a mediator of interdomain allostery in Hsp70 molecular chaperones. PLoS Comput Biol. 2014 May;10(5):e1003624.

251. Stetz G, Verkhivker GGM, Mayer M, Brehmer D, Gässler C, Bukau B, et al. Dancing through Life: Molecular Dynamics Simulations and Network-Centric Modeling of Allosteric Mechanisms in Hsp70 and Hsp110 Chaperone Proteins. Zheng J, editor. PLoS One. 2015 Nov 30;10(11):e0143752.

252. Stetz G, Verkhivker GM. Probing Allosteric Inhibition Mechanisms of the Hsp70 Chaperone

Proteins Using Molecular Dynamics Simulations and Analysis of the Residue Interaction Networks. J Chem Inf Model. 2016 Aug;1490–517.

253. Whitten ST, Garcia-Moreno E. B, Hilser VJ. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. Proc Natl Acad Sci. 2005;

254. England JL. Allostery in protein domains reflects a balance of steric and hydrophobic effects. Structure. 2011;

255. Atilgan C, Atilgan AR. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. PLoS Comput Biol. 2009 Oct;5(10):10005–44.

256. Atilgan C, Gerek ZN, Ozkan SB, Atilgan a. R. Manipulation of conformational change in proteins by single-residue perturbations. Biophys J. 2010;99(3):933–43.

257. Yilmaz LS, Atilgan AR. Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts. J Chem Phys. 2000 Sep 8;113(10):4454.

258. Ikeguchi M, Ueno J, Sato M, Kidera A. Protein structural change upon ligand binding: Linear response theory. Phys Rev Lett. 2005 Feb 24;94(7):078102.

259. Atilgan C, Okan OB, Atilgan AR. Network-based models as tools hinting at nonevident protein functionality. Annu Rev Biophys. 2012 Jan 11;41:205–25.

260. Atilgan AR, Aykut AO, Atilgan C. Subtle pH differences trigger single residue motions for moderating conformations of calmodulin. J Chem Phys. 2011;

261. Aykut AO, Atilgan AR, Atilgan C. Designing Molecular Dynamics Simulations to Shift Populations of the Conformational States of Calmodulin. PLoS Comput Biol. 2013;

262. Guven G, Atilgan AR, Atilgan C. Protonation states of remote residues affect binding-release dynamics of the ligand but not the conformation of apo ferric binding protein. J Phys Chem B. 2014;

263. Abdizadeh H, Guven G, Atilgan AR, Atilgan C. Perturbation response scanning specifies key regions in subtilisin serine protease for both function and stability. J Enzyme Inhib Med Chem. 2015;00(00):1–8.

264. Abdizadeh H, Atilgan C. Predicting long term cooperativity and specific modulators of receptor

interactions in human transferrin from dynamics within a single microstate. Phys Chem Chem Phys. 2016 Mar 9;18(11):7916–26.

265. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995 Sep;91(1–3):43–56.

266. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013 Apr 1;29(7):845–54.

267. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1:19–25.

268. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ(1) and χ(2) dihedral angles. J Chem Theory Comput. 2012 Sep 11;8(9):3257–73.

269. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998 Apr;102(18):3586–616.

270. Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem. 2004 Aug;25(11):1400–15.

271. Okan OB, Atilgan AR, Atilgan C. Nanosecond motions in proteins impose bounds on the timescale distributions of local dynamics. Biophys J. 2009 Oct 7;97(7):2080–8.

272. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr Sect A. 1976 Sep 1;32(5):922–3.

273. Song G, Jernigan RL. An enhanced elastic network model to represent the motions of domain-swapped proteins. Proteins Struct Funct Bioinforma. 2006 Jan 30;63(1):197–209.

274. Hinsen K, Thomas A, Field MJ. Analysis of domain motions in large proteins. Proteins. 1999 Feb 15;34(3):369–82.

275. Tina KG, Bhadra R, Srinivasan N. PIC: Protein Interactions Calculator. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W473-6.

276. Vogel M, Mayer MP, Bukau B. Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. J Biol Chem. 2006 Dec 15;281(50):38705–11.

277. Jiang J, Maes EG, Taylor AB, Wang L, Hinck AP, Lafer EM, et al. Structural basis of J cochaperone binding and regulation of Hsp70. Mol Cell. 2007 Nov 9;28(3):422–33.

278. Kumar DP, Vorvis C, Sarbeng EB, Cabra Ledesma VC, Willis JE, Liu Q. The Four Hydrophobic Residues on the Hsp70 Inter-Domain Linker Have Two Distinct Roles. J Mol Biol. 2011;411(5):1099–113.

279. Gerek ZN, Bolia A, Ozkan SB. Perturbation Response Scanning Method for Identifying Allosteric Transitions and Utilizing in Flexible Docking. Biophys J. 2011 Feb 2;100(3):372–3.

280. Mayer MP. Hsp70 chaperone dynamics and molecular mechanism. Trends Biochem Sci. 2013 Oct 1;38(10):507–14.

281. Sikor M, Mapa K, von Voithenberg LV, Mokranjac D, Lamb DC, Andrec M, et al. Real-time observation of the conformational dynamics of mitochondrial Hsp70 by spFRET. EMBO J. 2013 May 29;32(11):1639–49.

282. Rist W, Graf C, Bukau B, Mayer MP. Amide hydrogen exchange reveals conformational changes in hsp70 chaperones important for allosteric regulation. J Biol Chem. 2006 Jun 16;281(24):16493–501.

283. Gässler CS, Buchberger A, Laufen T, Mayer MP, Schröder H, Valencia A, et al. Mutations in the DnaK chaperone affecting interaction with the DnaJ cochaperone. Proc Natl Acad Sci U S A. 1998 Dec 22;95(26):15229–34.

284. Suh WC, Burkholder WF, Lu CZ, Zhao X, Gottesman ME, Gross CA. Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. Proc Natl Acad Sci U S A. 1998 Dec 22;95(26):15223–8.

285. Gao X-C, Zhou C-J, Zhou Z-R, Wu M, Cao C-Y, Hu H-Y. The C-terminal helices of heat shock protein 70 are essential for J-domain binding and ATPase activation. J Biol Chem. 2012 Feb 17;287(8):6044–52.

286. Krebs WG, Gerstein M. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. Nucleic Acids Res. 2000 Apr 15;28(8):1665–75.

287. Flores S, Echols N, Milburn D, Hespenheide B, Keating K, Lu J, et al. The Database of Macromolecular Motions: new features added at the decade mark. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D296-301.

288. Chiappori F, Fumian M, Milanesi L, Merelli I. DnaK as Antibiotic Target: Hot Spot Residues Analysis for Differential Inhibition of the Bacterial Protein in Comparison with the Human HSP70. PLoS One. 2015 Jan 23;10(4):1245–63.

289. Ozbaykal G, Rana Atilgan A, Atilgan C. In Silico Mutational Studies of Hsp70 Disclose Sites with Distinct Functional Attributes. Proteins. 2015 Nov;83(11):2077–90.

290. Rüdiger S, Buchberger A, Bukau B. Interaction of Hsp70 chaperones with substrates. Nat Struct Biol. 1997 May;4(5):342–9.

291. Mayer MP. Gymnastics of molecular chaperones. Mol Cell. 2010 Aug 13;39(3):321–31.

292. Bukau B, Mayer MP, Schröder H, Rüdiger S, Paal K, Laufen T. Multistep mechanism of substrate binding determines chaperone activity of Hsp70. Nat Struct Biol. 2000 Jul 1;7(7):586–93.

293. Moro F, Fernández-Sáiz V, Muga A. The lid subdomain of DnaK is required for the stabilization of the substrate-binding site. J Biol Chem. 2004 May 7;279(19):19600–6.

294. Slepenkov S V., Patchen B, Kenneth M. Peterson A, Witt SN. Importance of the D and E Helices of the Molecular Chaperone DnaK for ATP Binding and Substrate Release. 2003;42:5867–76.

295. Brown DK, Penkler DL, Sheik Amamuddy O, Ross C, Atilgan AR, Atilgan C, et al. MD-TASK: a software suite for analyzing molecular dynamics trajectories. Bioinformatics. 2017 May 31;17:2768–71.

296. Colombo G, Morra G, Meli M, Verkhivker G. Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution. Proc Natl Acad Sci U S A. 2008 Jun 10;105(23):7976–81.

297. Morra G, Verkhivker GG, Colombo G. Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer.

Jacobson MP, editor. PLoS Comput Biol. 2009 Mar 20;5(3):e1000323.

298. Verkhivker GM, Dixit A, Morra G, Colombo G. Structural and computational biology of the molecular chaperone Hsp90: from understanding molecular mechanisms to computer-based inhibitor design. Curr Top Med Chem. 2009 Jan;9(15):1369–85.

299. Morra G, Neves MAC, Plescia CJ, Tsustsumi S, Neckers L, Verkhivker G, et al. Dynamics-Based Discovery of Allosteric Inhibitors: Selection of New Ligands for the C-terminal Domain of Hsp90. J Chem Theory Comput. 2010 Sep 14;6(9):2978–89.

300. Vettoretti G, Moroni E, Sattin S, Tao J, Agard DA, Bernardi A, et al. Molecular Dynamics Simulations Reveal the Mechanisms of Allosteric Activation of Hsp90 by Designed Ligands. Sci Rep. 2016 Jan 1;6:23830.

301. Dixit A, Verkhivker GM. Probing molecular mechanisms of the Hsp90 chaperone: biophysical modeling identifies key regulators of functional dynamics. PLoS One. 2012 Jan;7(5):e37605.

302. Morra G, Potestio R, Micheletti C, Colombo G. Corresponding functional dynamics across the Hsp90 chaperone family: Insights from a multiscale analysis of MD simulations. Livesay DR, editor. PLoS Comput Biol. 2012 Mar 22;8(3):e1002433.

303. Seifert C, Gräter F, Dyson HJ, al. et. Force distribution reveals signal transduction in E. coli Hsp90. Biophys J. 2012 Nov 21;103(10):2195–202.

304. Simunovic M, Voth GA. Molecular and Thermodynamic Insights into the Conformational Transitions of Hsp90. Biophys J. 2012 Jul;103(2):284–92.

305. Blacklock K, Verkhivker GM. Computational Modeling of Allosteric Regulation in the Hsp90 Chaperones: A Statistical Ensemble Analysis of Protein Structure Networks and Allosteric Communications. Briggs JM, editor. PLoS Comput Biol. 2014 Jun 12;10(6):e1003679.

306. Stetz G, Tse A, Verkhivker GM. Dissecting Structure-Encoded Determinants of Allosteric Cross-Talk between Post-Translational Modification Sites in the Hsp90 Chaperones. Sci Rep. 2018;

307. Blacklock K, Verkhivker GM, Reuter N, Jr RD, Evanseck J. Differential Modulation of Functional Dynamics and Allosteric Interactions in the Hsp90-Cochaperone Complexes with p23 and Aha1: A Computational Study. Levy YK, editor. PLoS One. 2013 Aug 19;8(8):e71936.

308. Blacklock K, Verkhivker GM. Allosteric regulation of the Hsp90 dynamics and stability by client

recruiter cochaperones: protein structure network modeling. Najmanovich RJ, editor. PLoS One. 2014 Jan 20;9(1):e86547.

309. Blacklock K, Verkhivker GM. Experimentally Guided Structural Modeling and Dynamics Analysis of Hsp90–p53 Interactions: Allosteric Regulation of the Hsp90 Chaperone by a Client Protein. J Chem Inf Model. 2013 Nov 25;53(11):2962–78.

310. Johnson JL. Evolution and function of diverse Hsp90 homologs and cochaperone proteins. Biochim Biophys Acta. 2012 Mar;1823(3):607–13.

311. Sali A, Blundell T. Comparitive Protein Modelling by Satisfaction of Spacial Restraints. J Mol Biol. 1993;234:799–815.

312. Eswar N, Webb B, Marti-Renom M a, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative Protein Structure Modeling Using MODELLER. In: Current Protocols in Protein Science. John Wiley & Sons, Inc.; 2006. p. 1–5.

313. Pei J, Kim B-H, Grishin N V. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 2008 Apr;36(7):2295–300.

314. Šali a. MODELLER: A Program for Protein Structure Modeling Release 9.12, r9480. Rockefeller Univ. 2013;779–815.

315. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997;277:396–404.

316. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993 Sep;2(9):1511–9.

317. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007 Jul;35:W407–10.

318. Li J, Sun L, Xu C, Yu F, Zhou H, Zhao Y, et al. Structure insights into mechanisms of ATP hydrolysis and the activation of human heat-shock protein 90. Acta Biochim Biophys Sin (Shanghai). 2012 Apr 1;44(4):300–6.

319. Obermann WM, Sondermann H, Russo AA, Pavletich NP, Hartl FU. In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. J Cell Biol. 1998 Nov 16;143(4):901–10.

320. Dijkstra EW, W. E. A note on two problems in connexion with graphs. Numer Math. 1959

Dec;1(1):269–71.

321. Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006;2507–24.

322. Eramian D, Eswar N, Shen M. How well can the accuracy of comparative protein structure models be predicted ? 2008;1881–93.

323. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science. 1991 Jul 12;253(5016):164–70.

324. Luethy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992 Mar 5;356(6364):83–5.

325. Lobanov MY, Bogatyreva NS, Galzitskaya O V. Radius of gyration as an indicator of protein structure compactness. Mol Biol. 2008 Aug 10;42(4):623–8.

326. Flynn JM, Mishra P, Bolon DNA. Mechanistic Asymmetry in Hsp90 Dimers. J Mol Biol. 2015;427(18):2904–11.

327. Krukenberg KA, Förster F, Rice LM, Sali A, Agard DA. Multiple Conformations of E. coli Hsp90 in Solution: Insights into the Conformational Dynamics of Hsp90. Structure. 2008 May;16(5):755–65.

328. Krukenberg KA, Street TO, Lavery LA, Agard DA. Conformational dynamics of the molecular chaperone Hsp90. Q Rev Biophys. 2011 May;44(2):229–55.

329. Liu Y, Bahar I. Toward understanding allosteric signaling mechanisms in the ATPase domain of molecular chaperones. Pac Symp Biocomput. 2010 Jan;269–80.

330. Liu R, Hu J, Thornton J, Singh M, Funkhouser T. Computational Prediction of Heme-Binding Residues by Exploiting Residue Interaction Network. Uversky VN, editor. PLoS One. 2011 Oct 3;6(10):e25560.

331. del Sol A, Fujihashi H, O'Meara P. Topology of small-world networks of protein-protein complex structures. Bioinformatics. 2005 Apr 15;21(8):1311–5.

332. Cheng TM-K, Blundell TL, Fernandez-Recio J. pyDock: Electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. Proteins Struct Funct Bioinforma. 2007 Apr 19;68(2):503–15.

333. Brown DK, Sheik Amamuddy O, Tastan Bishop Ö. Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex. Glob Heart. 2017 Mar 13;12(2):121–32.

334. Brown DK, Tastan Bishop Ö. Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis. Glob Heart. 2017 Jun;12(2):151–61.

335. Dollins DE, Immormino RM, Gewirth DT. Structure of Unliganded GRP94, the Endoplasmic Reticulum Hsp90: basis for nucleotide induced conformational change. J Biol Chem. 2005 Jul 13;280(34):30438–47.

336. Pearl LH, Prodromou C. Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. Annu Rev Biochem. 2006 Jun;75(1):271–94.

337. Richter K, Reinstein J, Buchner J. N-terminal Residues Regulate the Catalytic Efficiency of the Hsp90 ATPase Cycle. J Biol Chem. 2002 Nov 22;277(47):44905–10.

338. Lavery LA, Partridge JR, Ramelot TA, Elnatan D, Kennedy MA, Agard DA. Structural Asymmetry in the Closed State of Mitochondrial Hsp90 (TRAP1) Supports a Two-Step ATP Hydrolysis Mechanism. Mol Cell. 2014;53(2):330–43.

339. Tsutsumi S, Mollapour M, Graf C, Lee C-T, Scroggins BT, Xu W, et al. Hsp90 charged-linker truncation reverses the functional consequences of weakened hydrophobic contacts in the N domain. Nat Structual Mol Biol. 2009 Nov 18;16(16):1141–7.

340. Cunningham CN, Krukenberg KA, Agard DA. Intra- and intermonomer interactions are required to synergistically facilitate ATP hydrolysis in Hsp90. J Biol Chem. 2008 Jul 25;283(30):21170–8.

341. Neckers L. Hsp90 inhibitors as novel cancer chemotherapeutic agents. Trends Mol Med. 2002 Jan;8(4 Suppl):S55-61.

342. Vaughan CK, Piper PW, Pearl LH, Prodromou C. A common conformationally coupled ATPase mechanism for yeast and human cytoplasmic HSP90s. FEBS J. 2009 Jan;276(1):199–209.

343. Hessling M, Richter K, Buchner J. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. Nat Struct Mol Biol. 2009 Mar 22;16(3):287–93.

344. Xu W, Mollapour M, Prodromou C, Wang S, Scroggins B, Palchick Z, et al. Dynamic Tyrosine

Phosphorylation Modulates Cycling of the HSP90-P50CDC37-AHA1 Chaperone Machine. Mol Cell. 2012 Aug;47(3):434–43.

345. Siligardi G, Hu B, Panaretou B, Piper PW, Pearl LH, Prodromou C. Co-chaperone regulation of conformational switching in the Hsp90 ATPase cycle. J Biol Chem. 2004 Dec 10;279(50):51989–98.

346. Rehn A, Moroni E, Zierer BK, Tippel F, Morra G, John C, et al. Allosteric Regulation Points Control the Conformational Dynamics of the Molecular Chaperone Hsp90. J Mol Biol. 2016;428(22):4559–71.

347. Street TO, Zeng X, Pellarin R, Bonomi M, Sali A, Kelly MJS, et al. Elucidating the Mechanism of Substrate Recognition by the Bacterial Hsp90 Molecular Chaperone. J Mol Biol. 2014;426(12):2393–404.

348. Genest O, Reidy M, Street TO, Hoskins JR, Camberg JL, Agard DA, et al. Uncovering a Region of Heat Shock Protein 90 Important for Client Binding in E. coli and Chaperone Function in Yeast. Mol Cell. 2013;49(3):464–73.

349. Panaretou B. ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone invivo. EMBO J. 1998 Aug 17;17(16):4829–36.

350. Retzlaff M, Hagn F, Mitschke L, Hessling M, Gugel F, Kessler H, et al. Asymmetric activation of the hsp90 dimer by its cochaperone aha1. Mol Cell. 2010 Feb 12;37(3):344–54.

351. Sattin S, Tao J, Vettoretti G, Moroni E, Pennati M, Lopergolo A, et al. Activation of Hsp90 Enzymatic Activity and Conformational Dynamics through Rationally Designed Allosteric Ligands. Chem - A Eur J. 2015 Sep 21;21(39):13598–608.

352. Schmid AB, Lagleder S, Gräwert MA, Röhl A, Hagn F, Wandinger SK, et al. The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. EMBO J. 2012 Mar 21;31(6):1506–17.

353. Hatherley R, Blatch GL, Bishop OT. Plasmodium falciparum Hsp70-x: a heat shock protein at the host-parasite interface. J Biomol Struct Dyn. 2013;(January 2015):37–41.

354. Li J, Buchner J. Structure, function and regulation of the hsp90 machinery. Biomed J. 2013 Jan 1;36(3):106–17.

355. Verkhivker GM. Computational Studies of Allosteric Regulation in the Hsp90 Molecular Chaperone: From Functional Dynamics and Protein Structure Networks to Allosteric Communications and Targeted Anti-Cancer Modulators. Isr J Chem. 2014 Aug;54(8–9):1052–64.

356. Guarnera E, Berezovsky IN. On the perturbation nature of allostery: sites, mutations, and signal modulation. Current Opinion in Structural Biology. 2019.

357. Guarnera E, Berezovsky IN. Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. PLoS Comput Biol. 2016;

358. Whitesell L, Mimnaugh E, De Costa B, Myers C, Neckers L. Inhibition of heat shock protein HSP90-pp 60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. Proc Natl Acad Sci U S A. 1994;(91):8324–8.

359. Schulte TW, Akinaga S, Soga S, Sullivan W, Stensgard B, Toft D, et al. Antibiotic radicicol binds to the N-terminal domain of Hsp90 and shares important biologic activities with geldanamycin. Cell Stress Chaperones. 1998;

360. Sidera K, Patsavoudi E. HSP90 inhibitors: current development and potential in cancer therapy. Recent Pat Anticancer Drug Discov. 2014 Jan;9(1):1–20.

361. Olotu F, Adeniji E, Agoni C, Bjiji I, Khan S, Elrashedy A, et al. An update on the discovery and development of selective heat shock protein inhibitors as anti-cancer therapy. Expert Opin Drug Discov. 2018 Sep 12;1–16.

362. Garg G, Khandelwal A, Blagg BSJ. Anticancer Inhibitors of Hsp90 Function: Beyond the Usual Suspects. Adv Cancer Res. 2016 Jan 1;129:51–88.

363. Pimienta G, Herbert KM, Regan L. A Compound That Inhibits the HOP–Hsp90 Complex Formation and Has Unique Killing Effects in Breast Cancer Cell Lines. 2011;

364. Marcu MG, Schulte TW, Neckers L. Novobiocin and related coumarins and depletion of heat shock protein 90-dependent signaling proteins. J Natl Cancer Inst. 2000 Feb 2;92(3):242–8.

365. Allan RK, Mok D, Ward BK, Ratajczak T. Modulation of chaperone function and cochaperone interaction by novobiocin in the C-terminal domain of Hsp90: evidence that coumarin antibiotics disrupt Hsp90 dimerization. J Biol Chem. 2006 Mar 17;281(11):7161–71.

366. Moroni E, Zhao H, Blagg BSJ, Colombo G. Exploiting Conformational Dynamics in Drug Discovery: Design of C-Terminal Inhibitors of Hsp90 with Improved Activities. J Chem Inf Model. 2014 Jan 27;54(1):195–208.

367. Ferraro M, D'Annessa I, Moroni E, Morra G, Paladino A, Rinaldi S, et al. Allosteric modulators of Hsp90 and Hsp70: Dynamics meets Function through Structure-Based Drug Design. J Med Chem. 2018 Jul 26;acs.jmedchem.8b00825.

368. Roy SS, Kapoor M. In silico identification and computational analysis of the nucleotide binding site in the C-terminal domain of Hsp90. J Mol Graph Model. 2016 Nov 1;70:253–74.

369. Goode KM, Petrov DP, Vickman RE, Crist SA, Pascuzzi PE, Ratliff TL, et al. Targeting the Hsp90 C-terminal domain to induce allosteric inhibition and selective client downregulation. Biochim Biophys Acta - Gen Subj. 2017 Aug;1861(8):1992–2006.

370. Davenport J, Manjarrez JR, Peterson L, Krumm B, Blagg BSJ, Matts RL. Gambogic acid, a natural product inhibitor of Hsp90. J Nat Prod. 2011;

371. Vasko RC, Rodriguez RA, Cunningham CN, Ardi VC, Agard DA, McAlpine SR. Mechanistic studies of Sansalvamide A-amide: an allosteric modulator of Hsp90. ACS Med Chem Lett. 2010 Jan 1;1(1):4–8.

372. Kunicki JB, Petersen MN, Alexander LD, Ardi VC, McConnell JR, McAlpine SR. Synthesis and evaluation of biotinylated sansalvamide A analogs and their modulation of Hsp90. Bioorganic Med Chem Lett. 2011;

373. Alexander LD, Partridge JR, Agard DA, McAlpine SR. A small molecule that preferentially binds the closed conformation of Hsp90. Bioorganic Med Chem Lett. 2011;

374. Zorn JA, Wells JA. Turning enzymes on with small molecules. Nature Chemical Biology. 2010.

375. Yokoyama Y, Ohtaki A, Jantan I, Yohda M, Nakamoto H. Goniothalamin enhances the ATPase activity of the molecular chaperone Hsp90 but inhibits its chaperone activity. J Biochem. 2015 Mar 1;157(3):161–8.

376. Zierer BK, Weiwad M, Rübbelke M, Freiburger L, Fischer G, Lorenz OR, et al. Artificial Accelerators of the Molecular Chaperone Hsp90 Facilitate Rate-Limiting Conformational Transitions. Angew Chemie - Int Ed. 2014;

377. Zierer BK, Rübbelke M, Tippel F, Madl T, Schopf FH, Rutz DA, et al. Importance of cycle timing for the function of the molecular chaperone Hsp90. Nat Struct Mol Biol. 2016;

378. Roe MS, Wahab B, Török Z, Horváth I, Vigh L, Prodromou C. Dihydropyridines Allosterically Modulate Hsp90 Providing a Novel Mechanism for Heat Shock Protein Co-induction and Neuroprotection. Front Mol Biosci. 2018;

379. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide Folding: When Simulation Meets Experiment. Angew Chemie Int Ed. 1999 Jan 15;38(1–2):236–40.

380. Hatherley R, Brown DK, Musyoka TM, Penkler DL, Faya N, Lobb KA, et al. SANCDB: a South African natural compound database. J Cheminform. 2015 Jun 19;7(1):29.

381. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009 Dec;30(16):2785–91.

382. Chennubhotla C, Bahar I, Maritan A, McIntosh N, Ziser L. Signal Propagation in Proteins and Relation to Equilibrium Fluctuations. PLoS Comput Biol. 2007;3(9):e172.

383. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins Struct Funct Genet. 1993 Dec;17(4):412–25.

384. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc. 2015 Apr 9;10(5):733–55.

385. Penkler DL, Bishop OT. Modulation of Human Hsp90α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain. bioRxiv. 2018 Aug 8;386755.

386. Ishida R, Takaoka Y, Yamamoto S, Miyazaki T, Otaka M, Watanabe S, et al. Cisplatin differently affects amino terminal and carboxyl terminal domains of HSP90. FEBS Lett. 2008 Nov 26;582(28):3879–83.

387. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature. 2007 Dec 13;450(7172):964–72.

388. Atilgan AR, Turgut D, Atilgan C. Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. Biophys J. 2007 May

1;92(9):3052–62.

389. Pettit GR, Xu J-P, Schmidt JM, Boyd MR. Isolation and structure of the exceptional Pterobranchia human cancer inhibitors cephalostatins 16 and 17. Bioorg Med Chem Lett. 1995 Sep 7;5(17):2027–32.

390. Pettit GR, Xu J-P, Chapuis J-C, Melody N. The Cephalostatins. 24. Isolation, Structure, and Cancer Cell Growth Inhibition of Cephalostatin 20. J Nat Prod. 2015 Jun 26;78(6):1446–50.