UNIVERSITÀ DEGLI STUDI DI PAVIA

DOTTORATO IN SCIENZE CHIMICHE E FARMACEUTICHE E INNOVAZIONE INDUSTRIALE
(XXXVIII Ciclo)

Coordinatore: Chiar.mo Prof. Giorgio Colombo

Development of Biochemical and Biophysical Computational Methods to Investigate Protein Complexes

Tesi di dottorato

Giorgio Bonollo

Anno Accademico 2024/2025

Tutor
Chiar.mo Prof. Giorgio Colombo

Co-Tutor
Chiar.mo Prof. Stefano Artin Serapian

Contents

Abstract **v**

Abstract (Italian) **vii**

1 Introduction **1**
1.1 Proteins, folding and chaperones . . . 1
1.2 Understanding main open questions on protein dynamics . . . 2
1.3 Molecular dynamics: a microscope for proteins . . . 4
1.4 Explored systems . . . 5
1.4.1 GR maturation cycle . . . 5
1.4.2 Hsp90 PTMs . . . 7
1.4.3 Kinesin-5 ATPase cycle . . . 8
1.4.4 Capturing allosteric events . . . 8
1.5 Aim of the thesis . . . 9
1.5.1 Code and data availability . . . 10
1.6 List of relevant publications . . . 11

2 Materials and Methods **13**
2.1 Key principles in Molecular Dynamics . . . 13
2.1.1 Statistical Thermodynamics . . . 13
2.1.2 Equations of motions . . . 15
2.1.3 Coupling to external bath . . . 16
2.1.4 Forces evaluation . . . 21
2.2 Non equilibrium Molecular Dynamics . . . 23
2.3 Performing Molecular Dynamics Simulations . . . 24
2.4 Classical Analysis of Molecular simulations . . . 26
2.4.1 Deviations and Fluctuations . . . 26
2.4.2 Principal Component Analysis . . . 27
2.4.3 Distance Fluctuations Analysis (DF) . . . 28
2.4.4 Shortest Path Map (SPM) . . . 29
2.5 Novel Analysis on Molecular Dynamics Simulations . . . 30
2.5.1 Tensor of Gyration (TOG) Analysis . . . 30
2.5.2 Analysis of internal dynamics from D-NEMD . . . 32
2.6 Concluding remarks . . . 34

3 Ligand Control in the Hsp90 Chaperone Cycle of GR **35**
3.1 Introduction . . . 35
3.2 Systems modeling . . . 38

3.2.1 Loading−DEX and Loading−DEX complexes . . . 38
3.2.2 Maturation+DEX and Maturation−DEX complexes . . . 40
3.2.3 Post-maturation-FKBP51+DEX and Post-maturation-FKBP51−DEX complexes . . . 40
3.2.4 Post-maturation-FKBP52+DEX and Post-maturation-FKBP52−DEX complexes . . . 40
3.2.5 Preparation, simulations and analyses . . . 41
3.2.6 D-NEMD Simulations . . . 41
3.2.7 Tensor of Gyration (TOG) Analysis . . . 42
3.2.8 Contact Analysis and Contact Probability Distributions . . . 43
3.3 Results . . . 43
3.3.1 Loading−DEX vs. Loading+DEX . . . 44
3.3.2 Maturation+DEX vs. Maturation−DEX . . . 47
3.3.3 FKBP51+DEX vs. FKBP51−DEX . . . 49
3.3.4 FKBP52+DEX vs. FKBP52−DEX . . . 53
3.3.5 FKBP51+DEX vs. FKBP52+DEX . . . 57
3.3.6 MiD–GR Distances . . . 59
3.3.7 Comparison of Simulation Results with Experimental Characterization of Points of Regulation . . . 60
3.4 Discussion . . . 61
3.5 Conclusions . . . 63

4 State-Dependent Effects of Hsp90β Phosphorylation 65
4.1 Introduction . . . 65
4.2 Modelling, preparation and simulations . . . 67
4.3 Analyses and results . . . 68
4.3.1 Phosphorylation reshapes hydrogen bonds network . . . 68
4.3.2 Phosphorylation affects coordination in Hsp90 . . . 73
4.3.3 Collective motions analysis . . . 76
4.4 Conclusions . . . 79

5 Allosteric regulation of Eg-5 motor domain 83
5.1 Introduction . . . 83
5.2 Molecular dynamics and protein preparation . . . 88
5.3 Results . . . 89
5.3.1 General results for Distance Fluctuation (DF) analysis . . . 89
5.3.2 General results for SPM . . . 92
5.3.3 Formation of the L5 binding pocket . . . 94
5.3.4 Loop 5 is allosterically connected to the ATP binding site . . . 95
5.3.5 Trp127 is crucial for the functional interaction between L5 and helix α3 . . . 96
5.3.6 Allosteric inhibitors lock L5 and increase the ADP binding affinity . . . 98
5.4 Conclusions . . . 99

6 Protein Conformational Motion: FAN Method 101
6.1 Introduction . . . 101
6.1.1 Method Overview . . . 103
6.2 Theoretical and Methodological Framework . . . 103
6.2.1 Statistical analysis of descriptors . . . 105

6.2.2 Minimal Frustration Defines an Optimal Network . . . 107
6.2.3 Determination of time evolution of the collective ensemble . . . 108
6.2.4 Finding the Allosteric Impact . . . 109
6.2.5 Signalling pathways . . . 111
6.2.6 Molecular Dynamics Simulations . . . 112
6.3 Results . . . 112
6.3.1 Ubiquitin . . . 112
6.3.2 Second PDZ Domain . . . 114
6.3.3 Checkpoint kinase . . . 117
6.4 Discussion . . . 119

7 Conclusion and Perspectives **121**

A Supplementary files for Chapter 3 **125**

B Supplementary files for Chapter 5 **139**

Abstract

Protein function depends critically on the acquisition and maintenance of a correct three-dimensional structure. Incorrect folding can lead to the formation of misfolded states and protein aggregation, contributing to the development of severe diseases, including cancer and neurodegenerative disorders. Molecular chaperones play a central role in maintaining cellular proteostasis, assisting non-native proteins in reaching functional conformations and preventing their aggregation. Among these, the Heat Shock Protein (Hsp) family, particularly Hsp90, Hsp70 and Hsp40, regulates protein maturation through ATP-dependent conformational cycles under allosteric control. Despite extensive structural and biochemical characterisation, the molecular mechanisms coordinating chaperone dynamics, allosteric regulation and client protein processing remain incompletely understood.

This thesis uses molecular dynamics (MD) simulations to investigate the conformational mechanisms underlying Hsp90 function, with particular focus on glucocorticoid receptor (GR) maturation. Equilibrium MD simulations of structurally characterised intermediates of the Hsp90–GR cycle were employed to analyse the molecular determinants of client maturation and release, as well as the role of co-chaperones in regulating alternative maturation pathways. The simulations identified specific interactions and conformational changes underlying these different outcomes.

Non-equilibrium MD simulations were also used to characterise the structural response of Hsp90 to ATP hydrolysis. The results revealed pronounced asymmetric displacements between the two Hsp90 monomers, providing evidence for a previously underappreciated mechanism that may contribute to functional asymmetry during the chaperone cycle and to the coupling between nucleotide state and client processing.

The effects of phosphorylation of two serine residues within the charged loop of Hsp90β were also investigated. Comparative MD simulations revealed changes in conformational flexibility, hydrogen-bonding patterns, inter-domain communication, and interactions with clients and co-chaperones. These results provide a molecular-level description of how phosphorylation can alter the Hsp90 conformational ensemble and modulate its functional states.

Finally, a novel computational framework was developed to investigate protein allostery from a conformational ensemble perspective. Unlike approaches based on a fixed structural reference, the proposed method combines the ensemble model of allostery with a Gaussian description of shifts in equilibrium distributions, enabling reference-free comparison of conformational fluctuations across different regions of the protein. Preliminary applications demonstrate the ability of the method to identify residues and communication pathways involved in ligand recognition and long-range signalling. Overall, this thesis integrates equilibrium and non-equilibrium MD simulations with advanced computational analyses to provide a mechanistic description of Hsp90-mediated protein maturation, ATP-dependent conformational regulation, phosphorylation-induced modulation, and allosteric communication. The results highlight the intrinsically dynamic and collective nature of protein function and demonstrate the value of physics-based computational approaches for investigating complex biomolecular mechanisms.

Abstract (Italian)

La funzione delle proteine dipende in modo cruciale dall'acquisizione e dal mantenimento di una corretta struttura tridimensionale. Un ripiegamento errato può portare alla formazione di stati misfolded e all'aggregazione proteica, contribuendo allo sviluppo di patologie gravi, tra cui il cancro e le malattie neurodegenerative. Gli chaperoni molecolari svolgono un ruolo centrale nel mantenimento della proteostasi cellulare, assistendo le proteine non native nel raggiungimento della conformazione funzionale e prevenendone l'aggregazione. Tra questi, la famiglia delle Heat Shock Proteins (Hsp), in particolare Hsp90, Hsp70 e Hsp40, regola la maturazione proteica attraverso cicli conformazionali ATP-dipendenti e sottoposti a controllo allosterico. Nonostante l'ampia caratterizzazione strutturale e biochimica, i meccanismi molecolari che coordinano dinamica degli chaperoni, regolazione allosterica e processamento delle proteine client non sono ancora completamente chiariti.

Questa tesi utilizza simulazioni di dinamica molecolare (MD) per studiare i meccanismi conformazionali alla base della funzione di Hsp90, con particolare attenzione alla maturazione del recettore dei glucocorticoidi (GR). Simulazioni MD all'equilibrio degli intermedi strutturalmente caratterizzati del ciclo Hsp90–GR sono state impiegate per analizzare i determinanti molecolari della maturazione e del rilascio del client, nonché il ruolo dei co-chaperoni nella regolazione di percorsi alternativi di maturazione. Le simulazioni hanno permesso di identificare specifiche interazioni e variazioni conformazionali alla base di questi differenti esiti.

Simulazioni MD di non equilibrio sono state inoltre utilizzate per caratterizzare la risposta strutturale di Hsp90 all'idrolisi dell'ATP. I risultati hanno evidenziato marcati spostamenti asimmetrici tra i due monomeri di Hsp90, fornendo evidenze a supporto di un meccanismo finora poco considerato che potrebbe contribuire all'asimmetria funzionale durante il ciclo dello chaperone e al collegamento tra stato nucleotidico e processamento del client.

Sono stati inoltre studiati gli effetti della fosforilazione di due serine presenti nel charged loop di Hsp90β. Simulazioni MD comparative hanno evidenziato variazioni nella flessibilità conformazionale, nei pattern di legame a idrogeno, nella comunicazione interdominio e nelle interazioni con client e co-chaperoni. Questi risultati forniscono una descrizione a livello molecolare di come la fosforilazione possa modificare l'ensemble conformazionale di Hsp90 e modularne gli stati funzionali.

Infine, è stato sviluppato un nuovo framework computazionale per lo studio dell'allosteria proteica secondo una prospettiva basata sull'ensemble conformazionale. A differenza degli approcci fondati su un riferimento strutturale fisso, il metodo proposto combina il modello ensemble dell'allosteria con una descrizione gaussiana degli spostamenti nelle distribuzioni di equilibrio, consentendo un confronto reference-free delle fluttuazioni conformazionali in diverse regioni della proteina. Le applicazioni preliminari dimostrano la capacità del metodo di identificare residui e percorsi di comunicazione coinvolti nel riconoscimento dei ligandi e nella segnalazione a lunga distanza.

Nel complesso, questa tesi integra simulazioni MD al (non) equilibrio con analisi computazionali avanzate per fornire una descrizione meccanicistica della maturazione proteica mediata da Hsp90, della regolazione conformazionale dipendente dall'ATP, della modulazione indotta dalla fosforilazione e della comunicazione allosterica. I risultati evidenziano la natura intrinsecamente dinamica e collettiva della funzione proteica e dimostrano il valore degli approcci computazionali basati sulla fisica nello studio di complessi meccanismi biomolecolari.

Chapter 1

Introduction

1.1 Proteins, folding and chaperones

Proteins are one of the constitutive components of life: synthesized at the ribosome following the translation of genetic information encoded in DNA, they regulate and control a wide variety of cellular processes [1, 2]. They are constituted by 20 different types of building blocks, aminoacids, whose chemical diversity enables a large degree of variation in the complexity of structural configurations and biological functions [1, 2].
The mechanism by which, once synthesized, proteins reach a stable 3D structure has long fascinated scientists. As first highlighted by Levinthal through the famous paradox, the available conformational space for proteins is enormous [3]. For instance, if a protein is composed of 100 aminoacids and each residue can adopt only two possible conformations, the available conformational space is made up of, at least 2^{100} possible 3D shapes. Even assuming that the sampling time of a single conformation occurs in the timescale of 10^{-12} s, the time to sample all the available conformations and identify the most stable one would be in the order of 10^5 years, which is impossible, since experimental observations show that proteins fold in the ms-s range [3].
As suggested by Zwanzig later, the exploration mechanism of the entire conformational space must be therefore guided by a physicochemical bias that drives the protein towards the correct, native state [4]. This idea, supported by a simple mathematical framework, demonstrated how proteins can efficiently explore conformational space, reaching the folded state within experimentally accessible times. However, not all the proteins follow the simple two-state, unfolded-to-folded transition [5]. Many proteins instead display a more complex behaviour, with the presence of multiple intermediate states before the correct functional shape is reached. The ability of proteins to move from an initially vast conformational space to a progressively more restricted one, possibly through a pathway involving several intermediates, is commonly explained by the concept and pictorial representation of a rough, funnel-shaped energy landscape, that guides the folding process [6–8]. The roughness of the energy surface may also lead to kinetic trapping of proteins in possibly wrong, misfolded states (**Figure 1.1**).
The acquisition of the correct three-dimensional structure is essential for correct protein functioning [9, 10]. The number of protein sequences encoded in the human genome is extremely large and difficult to quantify precisely, and its magnitude reflects the vast diversity of molecular tasks performed within the cellular environment [11]. For the correct functioning of all cellular activities, the proper folding of all proteins is therefore of paramount importance. High concentration of misfolded proteins can have deleterious consequences on cell life [12]. Indeed, several diseases are strongly associated with the accumulation of misfolded or aggregated protein species, such as amyloid structures. Cells must therefore prevent the formation of such aggregates by promoting the refolding of those misfolded states, thereby restoring protein networks and maintaining proteostasis [13]. This task is performed by molecular chaperones, proteins that can bind misfolded partners, restricting the conformational search toward optimal structures closer to the native states. The Heat Shock Protein family is a key

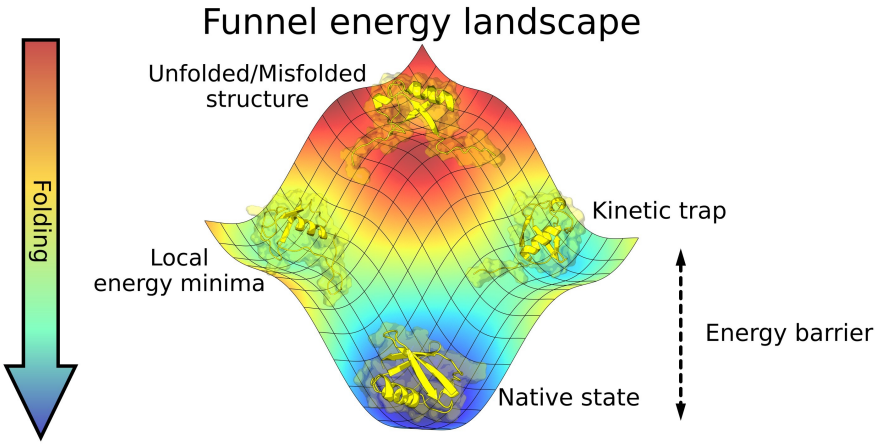

Figure 1.1: Protein on an Energy Landscape. *Simple representation of a protein moving onto an energy surface, passing from an initially high-energy, unstructured state to a final native, biologically active, minimum.*

representative of chaperones, which will be discussed later in this thesis.
Understanding how proteins fold, how the folding energy landscape is organised, and which intermediates are crucial for the correct functioning of each molecule is becoming increasingly important in recent years from many points of view. The ability of predict functional structures, manipulate folding pathways, or modulate the activity of chaperones are only a few examples of the many active areas of research in protein chemistry [14].

1.2 Understanding main open questions on protein dynamics

Once folded, proteins are complex molecular machineries capable of performing a wide variety of functions and activities within cells [1, 2]. Evolutionary pressure tuned both their amino-acid sequences and their three-dimensional structures to optimize their efficiency and functional roles throughout cellular life [15]. Understanding which functions are associated with each protein, and how these functions are carried out at the molecular level, remains a long-standing scientific challenge that is not yet fully resolved [16]. Because proteins can accomplish many different tasks, ranging from enzymatic catalysis to protein-protein interaction, molecular transport, and ion translocation, numerous investigations have focused on elucidating their structural properties as a starting point for understanding their mechanisms of action.
With the first protein structures obtained through X-ray crystallography, researchers began to associate a specific three-dimensional shapes with precise biological functions inside organism, giving rise to the well-known structure-function relationship paradigm [17]. Early studies on haemoglobin soon revealed, however, that this static view of proteins was incomplete [18]. It became clear that proteins are inherently dynamical entities, that can undergo conformational changes essential for accomplishing their physicochemical roles in cells.
To describe structural transition within proteins, often triggered by the binding of an external ligand or other perturbations, the concept of allostery was introduced. The term derives from the Greek words *allos* (other) and *stereos* (solid or structure), indicating a structural change occurring at a distal site, different from the one where the perturbation takes place. In their seminal paper of 1965, Monod, Wyman and Changeux gave birth to the first formal model of allosteric transitions, commonly known as the MWC model [18]. In this framework, the kinetic mechanism underlying oxygen binding in haemoglobin, together with other proteins, was analysed by assuming that a protein can exist in two distinct states, typically described as active and inactive. Ligand binding shifts the equilibrium

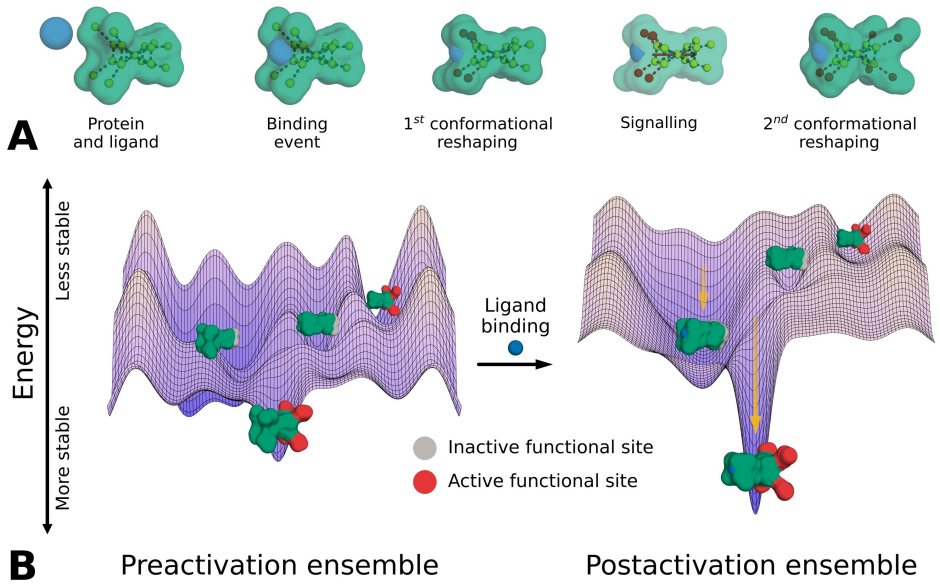

Figure 1.2: Allosteric processes. *Comparison of two main theories for allosteric processes:* ***(A)*** *Pictorial representation of KNF-like model, where a binding event at one site triggers conformational reshaping at the distal, allosteric site.* ***(B)*** *Two representation of ensemble theory of allostery, where binding event reshapes energy landscape lowering the minima of ligand-bound conformation (adapted from [19]).*

population between these states, favouring the active form and thereby enhancing protein function following the binding event. In this view, the equilibrium between states is regulated by kinetic constants that govern both binding efficiency and conformational transitions.

This simple yet powerful model profoundly changed the prevailing paradigm of protein function, opening a new perspective on the mechanism underlying biological activity. Shortly after the introduction of MWC model, another complementary allosteric theory was proposed: the Koshland-Némethy-Filmer (KNF) model [20]. In contrast to the predefined equilibrium between active and inactive states assumed in the MWC framework, the KNF model suggests that ligand binding itself induces the conformational change in the protein, promoting a structural rearrangement that propagates through the molecule. This complementary view introduced the concept of signal propagation within proteins, leading to a broader interpretation of proteins as dynamic systems capable of transmitting physicochemical signals from one region to another. In this framework, proteins are no longer regarded as static entities, but rather as molecular machines able to transduce structural and energetic changes across different sites [21, 22] (**Figure 1.2A**).

As knowledge of biological systems expanded, additional models and theories were proposed to explain the mechanisms of allostery and protein dynamics, stimulating scientific community to question whether simple models could fully capture the complexity of protein behaviour. In 1984, Cooper proposed that allosteric effects could occur even in the absence of large structural rearrangements in the three-dimensional structure of a protein [23]. This idea shifted the attention toward subtle signalling mechanisms and the role of entropic contributions that can propagate throughout the structure [24–26]. More recently, Hilser proposed a framework that emphasizes the ensemble nature of proteins. In his 2010 work, he suggested that the protein energy landscape cannot be described as a single funnel-shaped minimum leading uniquely to a native state [19]. Instead, native proteins can populate a wide range of conformational ensembles, and ligand binding shifts the equilibrium among these different states (**Figure 1.2B**). Within this perspective, proteins continuously sample multiple conformations, with entropy and fluctuations (particularly in flexible regions like loops) playing a crucial role in shaping their functional behaviour [27–30].

In this context, the fluctuating nature of proteins has increasingly been recognized as a key feature of their function, together with their ability to transmit signals following perturbations such as ligand binding. The existence of multiple loop conformations within proteins can be viewed as a restricted and localized analogue of the folding problem discussed in the previous section [31–36]. An important concept introduced to characterize such behaviour is that of frustration. Originally developed in the context of spin-glass physics to describe systems in which many nearly degenerate energy minima compete with one another, the concept of frustration has also been applied to protein dynamics. Wolynes and coworkers demonstrated explicitly that frustration plays an important role in protein behaviour [37]. According to this view, evolutionary processes have selected proteins that possess an appropriate degree of frustration, enabling them to balance structural stability with the flexibility required to perform their biological functions [38, 39].
Moreover, a comprehensive understanding of protein mechanisms must also consider the role of signalling and non-equilibrium processes. One of the most relevant events in protein dynamics is the transition between different conformational ensembles, which inherently represents a non-equilibrium process. The mechanisms underlying such transitions are therefore essential for a complete understanding of protein activity, including ligand binding and conformational rearrangements. Recent computational advances in this area have been achieved by Mulholland and coworkers, who extended approaches originally developed for heterogeneous inorganic systems undergoing non-equilibrium perturbations and applied the framework introduced by Ciccotti to protein systems [40].
The description of protein motions must also account for their cooperative nature: understanding how these molecular machines have been shaped by evolution to transmit signals between distant sites within the structure is thus a central problem in the study of protein communication networks. In this context, network analysis and correlation-based approaches applied to protein simulations have proven to be powerful tools for identifying pathways of information transfer and for explaining the behaviour of such complex systems [41].
Despite the large number of theories proposed over the years to explain protein behaviour, a comprehensive, exhaustive picture of protein dynamics is still lacking. Since the early 80s, the development of computer simulations has played a crucial role in addressing many of these questions, often providing valuable insights into molecular mechanisms. Today, computational approaches have become an indispensable tool for investigating and capturing the fundamental nature of molecular motions [42, 43].

1.3 Molecular dynamics: a microscope for proteins

The intrinsically high-dimensional nature of proteins, and biomolecules in general, makes it practically impossible to derive and solve analytical equations capable of describing their conformational motions. However, the knowledge of basic thermodynamic principles, together with the steady increase in computational power, has supported and stimulated the development of the first computer simulations of proteins.
Since the early 70s, simple coarse-grain models have been used to investigate folding dynamics [44]. Although these models were highly simplified, representing aminoacids as beads connected by springs and interacting through harmonic potentials, they were nevertheless able to capture the essential physical features of the folding process. Starting from the first Go-like models, which could simulate only a few hundred reduced particles, the progressive development of more accurate and transferable forcefields has led to the current generation of all-atom simulations, in which large protein assemblies can be studied in explicit solvent environments [45, 46].
The continuous growth in computational capabilities has enabled the full atomistic description of molecular motions occurring at equilibrium under the classical laws of dynamics [47]. Although this approach still represents an approximation, as quantum mechanical effects are generally not explicitly included, molecular dynamics force-fields have proven capable of reproducing many important features of biologically relevant systems. Applications of molecular dynamics range from the sampling

of local loop conformations to the investigation of large-scale rearrangements of protein domains and the study of active-site dynamics [9]. Molecular dynamics simulations also proved able to provide a glimpse into the mechanisms of protein folding, albeit to the use of a dedicated specifically designed hardware [33]. Nevertheless, these simulations remain limited by computational cost and by the need to sample sufficiently long timescales, especially when large proteins or (physiologically significant) complex assemblies are considered. To overcome these limitations and increase the portion of conformational phase space that can be explored, enhancing sampling techniques, such as umbrella sampling or metadynamics, have become widely used tools for investigating protein behaviour [48, 49].
More recently, advances in GPU computing and the rapid development of machine learning techniques have introduced new strategies for studying protein structure and dynamics. In particular, the artificial intelligence revolution initiated by the introduction of Alphafold in 2020 has demonstrated the remarkable potential of machine learning algorithms for predicting protein structures directly from aminoacid sequences [50]. Since then, an increasing number of machine learning approaches have been proposed to predict protein conformations or to generate structural ensembles. In addition, models trained on molecular dynamics data have been developed to generate protein conformations without the need to perform computationally expensive simulations [51].
Despite these impressive advances, structure prediction alone does not fully capture the dynamical nature of proteins. Structural datasets used to train AlphaFold-like models are derived from X-ray crystallography experiments, which intrinsically provide static snapshots of protein structures [52]. However, many key processes in protein biophysics (such as ligand binding, signal transmission, and conformational transitions) are inherently dynamical and occur through continuous exploration of conformational space. For this reason, machine learning models alone cannot yet replace atomistic simulations when the goal is to understand the mechanisms underlying protein motion.
Even with the large amount of structural and dynamical data currently available, a unified picture of the general principles governing protein conformational dynamics is still lacking. Identifying suitable descriptors of motion that can be related to biologically relevant functional activity therefore remains an important challenge. In this context, molecular dynamics simulations represent a powerful computational microscope that allows the detailed investigation of protein motions across different spatial and temporal scales.

1.4 Explored systems

During this work, several biological systems were investigated through molecular dynamics simulations to understand the key mechanisms underlying protein-protein interactions and allosteric communication. In particular, the role of molecular chaperones and their mechanisms of action were extensively studied, with particular attention devoted to the complex and often subtle nature of their constitutive protein-protein interactions. The main focus of this thesis is the Hsp90 chaperone machinery, which has been studied both under equilibrium and non-equilibrium conditions, also examining the effects of post-translational modifications (PTMs) on its biological activity [53–55]. In addition, the nature of allosteric communication and conformational equilibria in proteins has been investigated using different model systems, with the aim of developing new methods to characterize protein motions and ligand-binding effects in a reference-free perspective. In the following, a brief overview of the biological systems explored in this thesis is presented.

1.4.1 GR maturation cycle

Hsp90 is a molecular chaperone belonging to the Heat Shock Protein family and is present, with different paralogs, in nearly all eukaryotic cells [53–56]. Its primary role is to assist other proteins,

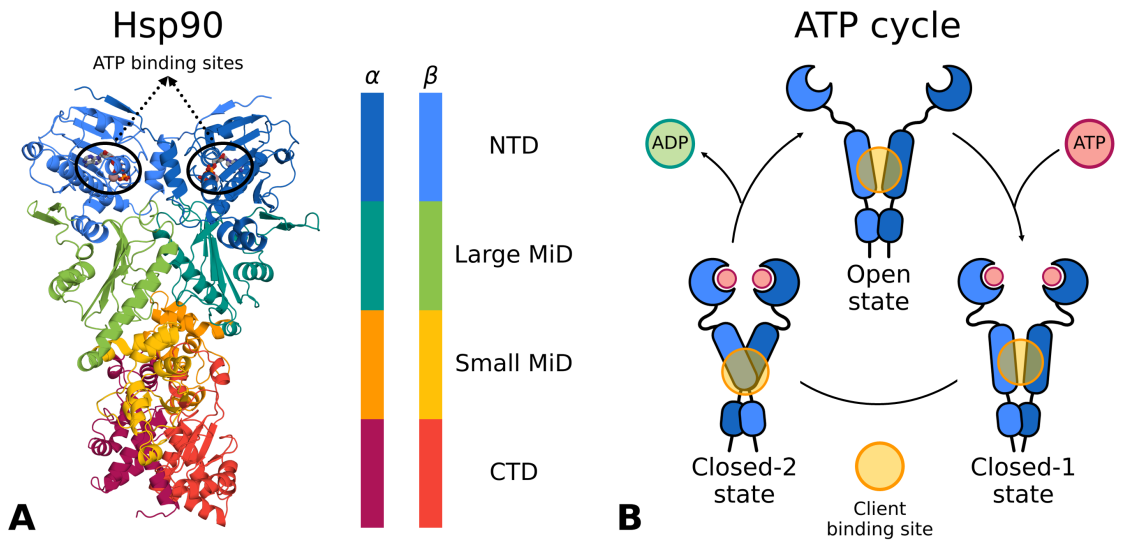

Figure 1.3: Hsp90 chaperone machinery. *(A) Hsp90 dimer is represented with different domains coloured respectively.* ***(B)*** *Hsp90 ATPase conformational cycle is sketched.*

commonly referred to as clients, in achieving their correct folded and biologically active conformations [57]. Proteins may become partially unfolded or misfolded for several reasons, and a precise regulation of misfolded protein concentration is essential for proper cellular function. Molecular chaperones were originally discovered because their expression levels increase significantly in response to cellular stress, such as heat shock [58].

From a structural point of view, Hsp90 functions as a homodimer composed of two identical subunits. Each monomer consists of three main domains: the N-terminal domain (NTD), the middle domain (MD), and C-terminal domain (CTD) [53–58] (**Figure 1.3**). A highly charged loop of approximately one hundred residues connects the NTD and the MD and plays a crucial role in regulating conformational dynamics, client maturation, and interactions with co-chaperones [59, 60]. Functionally, Hsp90 acts as a molecular clamp that captures partially folded client proteins within a central binding cavity, often referred to as lumen. In doing so, the chaperone restricts the conformational space accessible to the client protein, promoting sampling of structures closer to the native functional state [61, 62].

The functional cycle of Hsp90 is driven by an ATPase mechanism [63, 64]. The cycle begins with the open conformation of the dimer in the absence of ATP. Upon ATP binding at the N-terminal domains, and in the presence of a client protein, the two subunits undergo a conformational change that leads to the closure of the dimer, effectively trapping the client within the central cavity. Through a series of allosteric rearrangements, the protein progressively approaches its mature functional state. ATP hydrolysis subsequently triggers conformational changes that promote reopening of the dimer and release of the client protein. Finally, the release of ADP, restores the initial open conformation of Hsp90, allowing the system to begin a new functional cycle [65] (**Figure 1.3**).

Given its central role in cellular homeostasis, Hsp90 is one of the most extensively studied molecular chaperones [66, 67]. A large number of proteins are known to interact with Hsp90 as clients, including kinases, transcription factors, and steroid hormone receptors. Among the latter, one particularly important client is the Glucocorticoid Receptor (GR) [68, 69], which has been widely studied and only recently characterized in detail from a structural perspective [70–72]. GR is a steroid-hormone-activated transcription factor that strongly depends on Hsp90 for its activity and plays key roles in organ development, metabolic regulation, and the cellular responses to stress and inflammation [68]. Structurally, GR is composed of three domains: the N-terminal domain (NTD), the DNA-binding domain (DBD) and the ligand binding domain (LBD). While DBD and LBD are well-structured and responsible for DNA recognition and the hormone binding, the NTD is largely

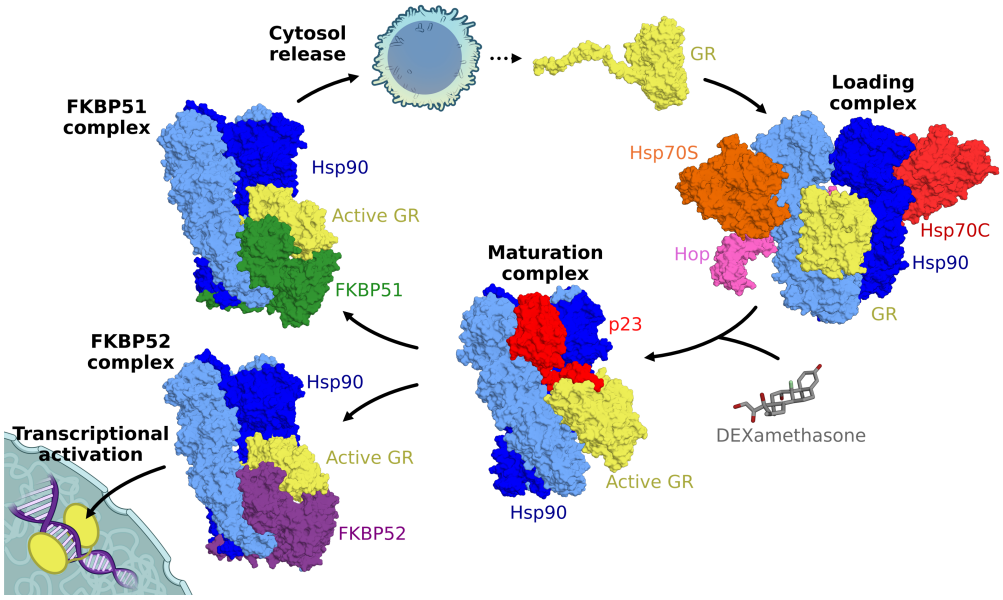

Figure 1.4: GR Maturation cycle. *Glucocorticoid receptor maturation cycle is represented with all the complexes investigated during this work.*

flexible and plays a crucial role in the recruitment of transcriptional co-factors [69].
Because the GR sequence contains significant intrinsically disordered regions, its maturation requires the assistance of the Hsp90 chaperone machinery [70–72]. Through this process, GR transitions from an inactive and partially disordered conformation to its active and functional state. The maturation cycle of GR has therefore been extensively investigated from both structural and kinetic perspectives to clarify the mechanisms regulating its activation following the binding of steroid hormones such as cortisol. Despite its three known domains, due to its largely disordered sequence, only the LBD was correctly characterized in the Cryo-EM studies from Agard and co-workers [70–72].
The GR maturation cycle can be divided into several main stages. Initially, the partially folded GR is recognized by the Hsp70 co-chaperone and subsequently transferred to the Hsp90 machinery together with the co-chaperone Hop [70]. This stage, known as the Loading complex, has been recently characterized by Agard and coworkers using cryo-electron microscopy. After the loading stage, the associated co-cochaperones are released and Hsp90 adopts a closed conformation in which the unfolded tail of GR is trapped inside the central lumen. At this stage, the co-chaperone p23 binds near the N-terminal domains of Hsp90, stabilizing the complex and preventing premature structural rearrangements. This configuration is typically referred to as the maturation complex [71] (**Figure 1.4**).
Following this stage, two alternative regulatory pathways can occur involving the binding of immunophilin co-chaperones. If FKBP51 binds to the complex after the release of p23, the system tends to dissociate, leading to the release of GR into the cytosol. Conversely, if FKBP52 binds to the Hsp90–GR complex, the interaction becomes more stable, and the complex can associate with the dynein motor protein [72]. This interaction promotes the transport of the complex toward the cell nucleus, where the active GR can be released and act as a transcription factor. This complex chaperone–client system has been investigated in this thesis using atomistic simulations in order to provide insights into the interactions between Hsp90 and its client proteins [73, 74].

1.4.2 Hsp90 PTMs

Another important regulator of Hsp90 activity is the presence of post-translational modifications (PTMs) [75]. PTMs are chemical modifications that occur after protein synthesis and involve the

covalent attachment of chemical groups to specific amino acid residues. Common examples include phosphorylation, glycosylation, acylation, and alkylation. These modifications expand the functional diversity of proteins by altering their physicochemical properties and can significantly affect protein behaviour, including ligand binding, enzymatic activity, and responsiveness to allosteric regulation [2].
Because Hsp90 plays a central role in proteostasis and signal transduction, PTMs affecting this protein can strongly influence its ATPase activity, dimerization, interactions with co-chaperones, and conformational equilibrium, thereby modulating client maturation [75]. Among the various PTMs identified for Hsp90, phosphorylation is one of the most frequently observed, with more than a dozen phosphorylation sites distributed across the different domains of the protein. In many cases, phosphorylation has been shown to reduce chaperone activity and slow down the conformational cycle of the protein.
In this thesis, the effects of phosphorylation were investigated by comparing molecular dynamics simulations of three functional states of the Hsp90 cycle—namely a semi-closed state, a fully closed state, and a client–co-chaperone bound state—in both the unmodified and phosphorylated forms [76, 77]. Two serine residues located in the charged loop (S226 and S255) were selected as phosphorylation sites because experimental studies have suggested a possible link between these modifications and the development of certain carcinomas. Through molecular dynamics simulations, it was possible to propose mechanistic hypotheses regarding the impact of these phosphorylation events on different stages of the Hsp90 functional cycle.

1.4.3 Kinesin-5 ATPase cycle

In addition to chaperone systems, other proteins involved in large-scale cellular processes were investigated. In particular, members of the kinesin superfamily are motor proteins associated with microtubules that play crucial roles in intracellular transport, mitosis and meiosis, regulation of microtubule dynamics, and signal transduction [78–81]. Because of their key role in controlling the proper formation of the bipolar mitotic spindle, kinesin motors represent important targets for anticancer therapies.
Kinesin motor domains couple cycles of ATP hydrolysis to cycles of microtubule binding and conformational changes that generate directional force and movement along microtubules [78] (**Figure 1.5**). Although the general principles of this mechanochemical coupling are well established, many atomistic details of the underlying allosteric mechanisms remain poorly understood. This lack of knowledge limits our understanding of how mutations located at distal sites can alter motor activity and also hampers the development of new inhibitory molecules [81–83].
In particular, several inhibitors are known to bind to an allosteric pocket formed by loop 5 (L5), a key structural element conserved among kinesin-5 homologues across different organisms. A deeper understanding of the allosteric signalling mechanisms triggered by inhibitor binding within the motor domain could therefore support the development of new therapeutic strategies and provide insights into the structural determinants regulating ligand binding. For this reason, the kinesin motor domain was simulated in several states corresponding to different stages of its ATPase cycle, including the apo form, the ATP-bound state, the ADP+P_i state, and the ADP–inhibitor bound form. The differences among these states were then analyzed to characterize the subtle allosteric mechanisms governing kinesin-5 motor activity and to better understand the molecular basis of inhibitor action.

1.4.4 Capturing allosteric events

The understanding of equilibrium motions of proteins is one of the main challenges for computational life sciences. Signal propagation following conformational changes and responses to non-equilibrium

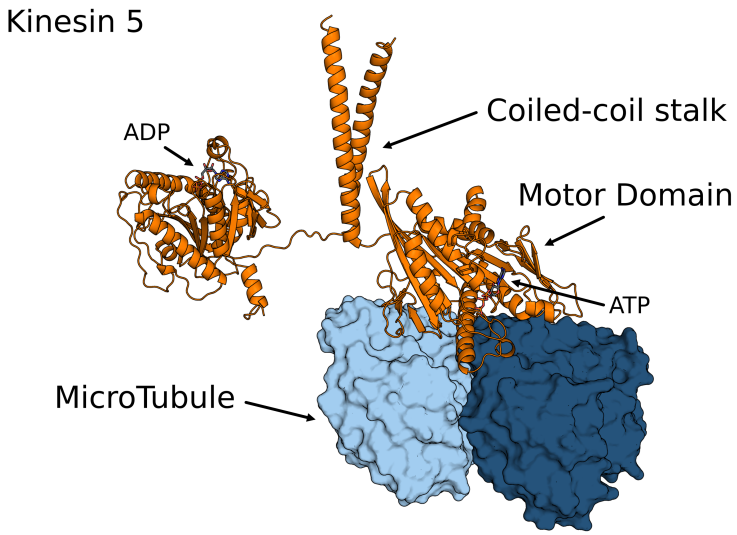

Figure 1.5: Kinesin 5 structure. *Representation of Kinesin-5 dimer (in orange) with motor domain attached to the microtubule (in blue).*

perturbations are fundamental processes that allow proteins to perform their biological functions. To investigate these phenomena, several small model systems were studied to understand how small conformational changes during equilibrium simulations can propagate through the protein structure, potentially revealing pathways of allosteric communications [32, 42].
Known allosteric targets were selected as test cases to study ligand or substrate binding events, including second PDZ domain and checkpoint kinases. Simulation of both apo and ligand-bound forms were performed in order to detect subtle communication pathways and propose theoretical mechanism through which protein interaction network adapted to achieve specific functional responses.

1.5 Aim of the thesis

Despite the long-standing understanding of the fundamental physicochemical principles governing many biological processes, a clear and comprehensive picture of how complex biomolecular systems behave at equilibrium is still lacking. The development of molecular dynamics has provided a powerful framework for investigating microscopic mechanisms in protein science, significantly expanding our knowledge of the structural and dynamical properties of many biological systems. More recently, advances in artificial intelligence have further enhanced our ability to predict protein structures and explore conformational landscapes.
However, despite their remarkable success and ease of use, data-driven approaches are inherently dependent on the availability and quality of experimental datasets. As such, they are not always able to capture the full complexity of biological systems or to provide general, physics-based principles capable of explaining the subtle mechanisms underlying protein function.
The aim of this work is therefore to develop novel methods for the analysis of molecular dynamics simulations, with the goal improving the connection between the large amount of data generated by simulations and the mechanistic insights that can be extracted from them. In particular, this thesis focuses on the design of new computational tools grounded in physical principles that can help to reduce the effective complexity of the systems under investigation, thereby facilitation the interpretation of simulations results.
Within this framework, the methodologies developed in this work are intended to:

- describe and characterize unstable or highly fluctuating regions of proteins under equilibrium

conditions

- improve the identification and characterization of allosteric signalling processes, bridging ensemble-based and signal-propagation perspectives of allostery
- reduce the complexity associated with the description of multiproteic complexes through simplified representations of simulated conformational states
- improve the characterization of non-equilibrium response of proteins following instantaneous perturbations

This thesis therefore combines methodological developments with applications to biologically relevant systems, aiming to provide new insights into the mechanisms governing protein dynamics, allosteric communication, and chaperone-mediated regulation. Following chapters describe the theoretical background, computational methodologies, and the results obtained for the biological systems introduced in the previous sections.

1.5.1 Code and data availability

In this work various Python 3 scripts were developed for the analyses. The scripts associated with published articles are already publicly available from the research group's GitHub repository: https://github.com/colombolab. Instead, the scripts developed in the present work not yet published will be made publicly available in the same repository upon publication of the article. All code is, or will be, released under the MIT licence. The molecular dynamics simulation data and/or data generated during the analyses are available from different Zenodo repositories reported in the related chapters, or are available upon request, as well as the input files and analysis scripts required to reproduce the results presented in this work.

1.6 List of relevant publications

- Castelli, M., Magni, A., Bonollo, G., Pavoni, S., Frigerio, F., Oliveira, A. S. F., Cinquini, F., Serapian, S. A., Colombo, G., Molecular mechanisms of chaperone-directed protein folding: Insights from atomistic simulations. *Prot. Sci.* **33**, e4880 (2024)

- Guarra, F., Sciva, C., Bonollo, G., Chiranjeevi, P., Chiosis, G., Moroni, E., Colombo, G., Cracking the chaperone code through the computational microscope. *Cell Stress Chaperones* **29**, 626–640 (2024).

- Mansoor, S., Frasnetti, E., Cucchi, I., Magni, A., Bonollo, G., Serapian, S. A., Pavarino, L. F., Colombo, G., Large-scale energy decomposition for the analysis of protein stability. *Cell Stress Chaperones* **30**, 57–68 (2025).

- Bonollo, G.*, Trèves, G., Komarov, D., Mansoor, S., Moroni, E., Colombo, G., Advancing Molecular Simulations: Merging Physical Models, Experiments, and AI to Tackle Multiscale Complexity. *J. Phys. Chem. Lett.* **16**, 3606–3615 (2025).

- Xie, S., Zuo, K., De Rubeis, S., Bonollo, G., Colombo, G., Ruggerone, P., Carloni, P., Impact of Genetic Variants Associated with Neurodevelopmental Disorders on the WAVE Regulatory Complex. *J. Chem. Inf. Model.* **65**, 7399–7405 (2025).

- Runge, S., Bilgen, E., Magni, A., Bonollo, G., Colombo, G., Itzen, A., Lamb, D. C., The functional dynamics of FicD's TPR domain are modulated by the interaction with ATP and BiP. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2500079122 (2025).

- Xie, S., Zuo, K., De Rubeis, S., Bonollo, G., Colombo, G., Ruggerone, P., Carloni, P., Common characteristics of variants linked to autism spectrum disorder in the WAVE regulatory complex. *Front. Comput. Neurosci.* **19**, (2025).

- Magni, A., Bonollo, G., Guarra, F., Torielli, L., Moroni, E., Serapian, S. A., Colombo, G., *Investigating allostery with atomistic biomolecular simulations* in *Reference Module in Life Sciences* (Elsevier, 2025)

- Rodríguez-Santos, G., Bonollo, G., Sciva, C., Colombo, G., Pérez-Melero, C., Serapian, S. A., Molecular dynamics simulations provide further insights into the allosteric regulation of the kinesin-5 motor domain by Loop 5, *JCIM* **66**, 4101-4115 (2026).

- Magni, A., Bonollo, G.*, Trèves, G., Frigerio, F., Cinquini, F., Pavoni, S., Oliveira, A. S. F., Mulholland, A. J., Serapian, S. A., Colombo, G., Ligand-driven modulation of chaperone–cochaperone networks shapes proteostasis outcomes, *Prot. Sci.* **35**, e70543 (2026)

- Bonollo, G.*, Roncati, B., Torielli, L., Wang, S., Pasala, C., Pavoni, S., Frigerio, F., Cinquini, F., Chiosis, G., Serapian, S. A., Colombo, G., Post-Translational Modification as an Allosteric Switch in Hsp90: How Dual Phosphorylation Locks Chaperone Complexes into Hyperstabilized States, *J. Phys. Chem. Lett.* **17**, 8098-8104 (2026)

* First author or co-first author.

Chapter 2

Materials and Methods

In this chapter, we will present the main theoretical and methodological foundations of this work. Beginning with a brief introduction to the fundamental principles of statistical thermodynamics, the discussion will move to an overview of the most important operating principles underlying molecular dynamics simulations. The standard protocol employed throughout this work will be described, together with the analytical methods adopted for the various investigations.

2.1 Key principles in Molecular Dynamics

Computer simulations were developed in the early 1950s as a tool capable of bridging the gap between the theoretical understanding of a given system and the macroscopic experimental measurements of its larger-scale counterpart [84, 85]. Despite being grounded in well-established physical principles, numerical simulations inherently rely on several approximations, which may reduce the degree of correspondence between the real-world scenarios and their *in silico* representations. Nevertheless, results obtained over the years demonstrated that these computational approaches are capable of revealing key microscopic molecular processes that are often inaccessible through experimental techniques.
The applicability of molecular dynamics simulations spans a wide range of fields, from materials science to drug discovery. More recent applications have achieved the capability to model nearly all-atom *in silico* representations of cellular environments and to compare them directly with in vivo cellular dynamics.
Different computational approaches and algorithms can be employed to simulate a system, depending on the specific properties or phenomena under investigation. The method most extensively used in this work is Molecular Dynamics (MD), a brief overview of which is presented in the following sections.

2.1.1 Statistical Thermodynamics

Molecular dynamics is a computational technique used to simulate the time evolution of a system under equilibrium conditions. The underlying concept is analogous to that of a classical experimental approach: the value of a physical observable can be estimated by performing a statistically significant number of measurements and subsequently averaging the obtained results. Similarly, by initiating multiple simulations and allowing them to evolve over time, it is possible, under the ergodic hypothesis, to obtain ensemble averages that represent plausible equilibrium values of the observable of interest [86].
A given system can, indeed, be uniquely described by a set of $3N$ canonical coordinates, denoted as $q_1, q_2, \ldots, q_{3N}$, and a corresponding set of $3N$ canonical momenta, denoted as $p_1, p_2, \ldots, p_{3N}$. System dynamics are determined by a function of these coordinates, namely the Hamiltonian $\mathcal{H}(p, q)$, from which the equations of motion can be derived [86]:

$$\dot{q}_i = \frac{\partial \mathcal{H}(p,q)}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}(p,q)}{\partial q_i} \tag{2.1}$$

It can therefore be stated that all q_i, p_i lie within a $3N$ dimensional space, referred to as the phase space Γ. Imposing the condition $\mathcal{H}(p,q) = E$, where E corresponds to the energy of the system, defines a hypersurface in Γ on which the positions and momenta uniquely determine the total energy of the system [86].
In both real and simulated systems, the primary interest does not lie in identifying the exact positions and momenta of every atom individually. Rather, the focus is on understanding the collective properties of the ensemble composed of infinitely many copies of the system satisfying the same thermodynamic conditions.
To characterize this distribution of points in phase space, the density function $\rho(p,q,t)$ is introduced, such that the infinitesimal volume element in Γ:

$$\rho(p,q,t) d^{3N}q d^{3N}p \tag{2.2}$$

is equal to the number of points contained within the volume element $d^{3N}q d^{3N}p$ in Γ at time t. From the Liouville's theorem, together with the Hamiltonian definitions of $\dot{q}_i$, $\dot{p}_i$ it follows that [86]:

$$\frac{\partial \rho}{\partial t} + \sum_{i=1}^{N} \left(\frac{\partial \rho}{\partial q_i} \frac{\partial \mathcal{H}(p,q)}{\partial p_i} - \frac{\partial \mathcal{H}(p,q)}{\partial q_i} \frac{\partial \rho}{\partial p_i} \right) = 0 \tag{2.3}$$

which corresponds to stating that, within the phase space Γ, the representative points evolve analogously to an incompressible fluid. Restricting the discussion to the case in which ρ is time independent — namely, when the system is at equilibrium and time-dependent effects are neglected — and further assuming that ρ depends on p, q only through the Hamiltonian, it follows that:

$$\frac{\partial}{\partial t} \rho(p,q,t) = 0 \tag{2.4}$$

This implies that $\rho(p,q)$ remains constant over time under equilibrium conditions. These assumptions allow the definition of a density function through which ensemble averages can be evaluated. In particular, given a measurable quantity of the system, $f(q,p)$, the corresponding ensemble average can be defined as:

$$\langle f \rangle_{\text{ens}} \equiv \frac{\int d^{3N}q \, d^{3N}p \, \rho(p,q) \, f(p,q)}{\int d^{3N}q \, d^{3N}p \, \rho(p,q)} \tag{2.5}$$

This definition establishes a connection between a macroscopical observable, $f(q,p)$, and the microscopic statistical properties of the system, represented by $\rho(q,p)$, thereby linking theoretical descriptions with experimentally measurable quantities [86].
Although the importance of the ensemble density ρ has now been established, its analytical form has not yet been introduced. In the thermodynamic description of a system, it is common to approximate it as two interacting subsystems characterized by Hamiltonians $\mathcal{H}_1(p_1,q_1)$ and $\mathcal{H}_2(p_2,q_2)$, containing N_1 and N_2 particles, respectively, with $N_2 \gg N_1$, while both systems remain macroscopically large. This framework can be interpreted as a thermodynamically small system in contact with an effectively infinite reservoir, such as an ideal heat bath.
Under these assumptions, the resulting ensemble is referred to as the canonical ensemble, and it can be demonstrated that the ensemble density may be expressed as:

$$\rho(p_1,q_1) = e^{-\frac{\mathcal{H}_1(p_1,q_1)}{k_B T}} \tag{2.6}$$

where $\mathcal{H}_1$ is the Hamiltonian of the first subsystem, k_B is the Boltzmann constant, and T is the temperature. The factor $\frac{1}{k_B T}$ is commonly denoted as β. It is then possible to calculate the volume

element of the phase space Γ occupied by the ensemble, commonly referred to as the partition function:

$$Q_N(V,T) \equiv \frac{1}{N!\,h^{3N}} \int d^{3N}q\, d^{3N}p\, e^{-\beta\mathcal{H}(p,q)} \tag{2.7}$$

Here, the constant h is introduced to render the partition function dimensionless, while the factor $N!$ accounts for the correct counting of indistinguishable states, according to Boltzmann's statistical interpretation linking the probability of phase-space partitioning to macroscopic thermodynamic observables. The partition function can subsequently be related to macroscopic physical quantities through the following expression:

$$Q_N(V,T) = e^{-\beta A(V,T)} \tag{2.8}$$

where $A(V,T)$ denotes the Helmoltz free energy. It can be demonstrated that A is an extensive thermodynamic quantity and that it is directly related to both the internal energy and the entropy of the system. By applying standard thermodynamic derivations, additional fundamental thermodynamic quantities can also be obtained [86].

2.1.2 Equations of motions

The fundamental objective of molecular dynamics is to reproduce, through computer simulations, the ideal conditions of statistical mechanics described in the previous section and to exploit them for the investigation of a wide range of physical phenomena. Through computational simulations, it is possible to relate time-averaged quantities obtained from the simulations to macroscopic properties of the system, such as heat transport through convection processes or conformational motions in protein loops [84, 85].
Unlike Monte Carlo simulations, in which the system is assumed to remain at equilibrium and particle movements are generated individually according to predefined displacement rules, molecular dynamics simulations allow the system to evolve naturally from a set of equilibrium initial conditions, thereby exploring the equilibrium phase space over time [84, 85]. In the ideal case, the entire phase space Γ is sampled, enabling the identification of microscopic behaviours that manifest as measurable macroscopic equilibrium properties.
Nevertheless, these simulations inevitably require approximations of certain physical aspects of the system, together with the implementation of theoretical laws into computationally efficient algorithms designed to minimise numerical errors. The central principle of molecular dynamics therefore consists in assigning initial position $\vec{r}(t)$ and velocities $\vec{v}(t)$ to the particles of the system and subsequently solving the equations of motions iteratively [84, 85]:

$$\ddot{\vec{r}}_i(t) = \frac{\vec{F}_i(t)}{m_i} \tag{2.9}$$

where $\vec{r}_i$ represents the position vector of atom i, m_i its mass, and $\vec{F}_i$ the force acting on the i-th atom. Several algorithms have been developed over the years to propagate the system in time. A key requirement for these algorithms is time-reversal invariance, together with the conservation of the phase-space volume element $d\Gamma$, which, as previously discussed constitutes a fundamental property of systems at equilibrium. Among the most widely employed integration schemes is the Velocity-Verlet algorithm. By performing a Taylor expansion around the particle position at time t, the following expressions are obtained [84, 85, 87, 88]:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\frac{\vec{F}_i(t)}{m_i}\Delta t^2 + \mathcal{O}(t^3) \tag{2.10}$$

Instead, by expanding around position $t+\Delta t$ but backward in time we get:

$$\vec{r}_i(t) = \vec{r}_i(t+\Delta t) - \vec{v}_i(t)\Delta t + \frac{1}{2}\frac{\vec{F}_i(t+\Delta t)}{m_i}\Delta t^2 + \mathcal{O}(t^3) \tag{2.11}$$

By then combining the two equations, we get for the velocity:

$$\vec{v}_i(t+\Delta t) = \vec{v}_i(t) + \frac{1}{2}\Delta t\frac{\vec{F}_i(t) + \vec{F}_i(t+\Delta t)}{m_i} + \mathcal{O}(t^3) \tag{2.12}$$

It can be demonstrated that this algorithm satisfies the required properties of time-reversal symmetry and conservation of phase-space volume. Starting from the initial positions and velocities, $\vec{r}_i(t_0)$ and $\vec{v}_i(t_0)$, together with the corresponding initial forces, it is first possible to compute the particle positions at time $t_0 + \Delta t$. The forces $\vec{F}_i(t_0 + \Delta t)$ can then be evaluated at the updated positions, allowing the calculation of the new velocities. By iteratively repeating this procedure, the system evolves over time, and equilibrium conformations can be sampled to evaluate ensemble properties. Another commonly employed integration scheme is the Leap-frog algorithm, in which the velocities are initially defined as follows [84, 85]:

$$\vec{v}_i\left(t + \frac{\Delta t}{2}\right) \equiv \frac{\vec{r}_i(t+\Delta t) - \vec{r}_i(t)}{\Delta t}, \qquad \vec{v}_i\left(t - \frac{\Delta t}{2}\right) \equiv \frac{\vec{r}_i(t) - \vec{r}_i(t - +\Delta t)}{\Delta t} \tag{2.13}$$

From the first equation we get that:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \Delta t\vec{v}_i\left(t + \frac{\Delta t}{2}\right) \tag{2.14}$$

And, for the update of velocities one can use the equation:

$$\vec{v}_i\left(t + \frac{\Delta t}{2}\right) = \vec{v}_i\left(t - \frac{\Delta t}{2}\right) + \Delta t\frac{\vec{F}_i}{m_i} \tag{2.15}$$

2.1.3 Coupling to external bath

However, the results discussed so far were derived for the microcanonical ensemble which, recalling **Equation 2.7**, describes systems in which the number of particles N, the volume V and the energy E remain constant (this ensemble is commonly referred to as the NVE ensemble). In many practical applications, however, these conditions may not provide the most suitable description of the system under investigation. In particular, for biological systems, the NPT ensemble is generally more appropriate, since pressure and temperature are maintained constant rather than volume and energy, thereby allowing volume fluctuations and local energy variations [89].
The adoption of the NPT ensemble requires the implementation of algorithms capable of controlling such fluctuations without introducing numerical instabilities that could affect the calculated ensemble averages. Algorithms designed to regulate temperature and pressure are commonly denoted as thermostats and barostats, respectively, and several different approaches are available depending on the nature of the simulated system. In the following sections, the main methods developed to control temperature and pressure will be discussed [90–96].
One of the earliest approaches for temperature control is the Langevin thermostat [97], in which the simulated particles are coupled to an external thermal bath composed of lighter particles. To account for collisions with bath particles, the equations of motions become:

$$m_i\ddot{\vec{r}}_i = -\frac{\delta\mathcal{H}}{\delta x} - m_i\gamma_i\vec{v}_i + \vec{R}_i(t) \tag{2.16}$$

This formulation describes the coupling of the simulated ensemble to an external heath bath at temperature T_0. The interaction is governed by the damping constant γ and by a stochastic force $\vec{R}(t)$ which has zero mean and satisfies $\langle\vec{R}_i(t)\vec{R}_j(t+\tau)\rangle = 2m_i\gamma k_B T_0\delta_{ij}\delta(\tau)$. This condition ensures

that the random force is delta-correlated in time and satisfies the fluctuation-dissipation theorem. The introduction of frictional and stochastic forces modifies the conservation properties of the system. Indeed, the time derivative of the total energy E becomes:

$$\frac{dE}{dt} = \sum_{i=1}^{N} \left(\frac{\delta \mathcal{H}}{\delta p_i} \frac{\delta p_i}{\delta t} + \frac{\delta \mathcal{H}}{\delta r_i} \frac{\delta r_i}{\delta t} \right) = - \sum_{i=1}^{N} \left(m_i \gamma \, \vec{v}_i^{\,2} - \vec{v}_i \cdot \vec{R}_i(t) \right) \tag{2.17}$$

Therefore, the total energy is not conserved. However, averaging over sufficiently short time intervals allows the stochastic contribution to vanish because the random force has zero mean, while positions and velocities remain essentially unchanged in the limit $\Delta t \to 0$. Under these assumptions, one obtains:

$$\frac{dE}{dt} = -\gamma \sum_{i=1}^{N} \left(m_i \vec{v}_i^{\,2}(t) - k_B T \right) = -\gamma \left(2E_k - k_B T \right) \tag{2.18}$$

Thus, although the total energy is not strictly conserved, the perturbation introduced by the thermostat remains small if $\frac{1}{\gamma} \ll \tau_c$ (with τ_c timescale that characterizes system's dynamics). Then the energy can be considered as conserved since fluctuation have effects on times smaller than that of the system's dynamics.
Despite its conceptual simplicity, the Langevin thermostat presents several drawbacks, including the absence of a conserved quantity and the difficulties associated with time integration. Motivated by the stochastic framework described above, Berendsen proposed an alternative approach in which the instantaneous temperature T is coupled to an external target temperature T_0 according to:

$$\left(\frac{dT}{dt} \right)_{\text{bath}} = 2\gamma \left(T_0 - T \right) \tag{2.19}$$

This equation defines the temperature coupling employed in the Berendsen thermostat [93]. The collision constant can alternatively be expressed in terms of the temperature coupling time $\tau_T = (2\gamma)^{-1}$. Such coupling condition is equivalent to introducing an additional thermostat force into the equations of motion, which ultimately is corresponds to a velocity scaling of a factor $\vec{v}(t+\Delta t) = \lambda \vec{v}(t)$, where:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right) \right]^{1/2} \tag{2.20}$$

Such choice conserves the Maxwellian shape of velocities distribution. Nevertheless, despite its simplicity and computational efficiency, the Berendsen thermostat is associated with neither a conserved quantity nor a well-defined thermodynamic ensemble.
To overcome these limitations, the stochastic velocity-rescaling thermostat was developed [98]. The underlying idea is that the particle velocities can be rescaled by a factor α so that the instantaneous kinetic energy K approaches the target average kinetic energy $\bar{K}$. However, deterministic rescaling alone suppresses the natural temperature fluctuations required for proper canonical sampling. To restore these fluctuations, the target kinetic energy is assumed to be drawn from the canonical equilibrium distribution:

$$\bar{P}(K_t)\, dK_t \propto K_t^{\left(\frac{N_f}{2}-1\right)} e^{-\beta K_t}\, dK_t \tag{2.21}$$

A corresponding Fokker–Planck equation can then be derived for the time evolution of the kinetic energy:

$$dK = \left(D(K) \frac{\partial \log \bar{P}}{\partial K} + \frac{\partial D(K)}{\partial K} \right) dt + \sqrt{2D(K)}\, dW \tag{2.22}$$

where $D(K)$ is an arbitrary positive function that acts as a diffusion coefficient and dW is a Wiener process. By imposing $D(K) = \frac{2K\bar{K}}{N_f\tau}$ with τ coupling parameter, the stochastic evolution equation becomes:

$$dK = \frac{K - \bar{K}}{\tau}\, dt + 2\sqrt{\frac{K\,\bar{K}}{N_f}}\,\frac{dW}{\sqrt{\tau}} \tag{2.23}$$

With this approach, the canonical ensemble is sampled correctly under ergodic conditions [98]. In the absence of the stochastic term, the formulation reduces to the Berendsen thermostat. Analogous approaches have been developed for pressure control through the introduction of barostat algorithms. One of the most widely used methods is the Berendsen barostat, in which pressure relaxation is imposed through:

$$\left(\frac{dP}{dt}\right)_{\text{bath}} = \frac{P_0 - P}{\tau_P} \tag{2.24}$$

where τ_P is the pressure-coupling time. The pressure can be expressed as:

$$P = \frac{2}{3V}\,(E_k - \Xi) \tag{2.25}$$

where $\Xi = -\frac{1}{2}\sum_{i<j}\vec{r}_{ij}\cdot\vec{F}_{ij}$ is the internal virial for pair additive potentials. Then a pressure change can be achieved by scaling inter-particle distances, changing the virial [99]. Simple linear coordinate scaling changes the positions as:

$$\dot{\vec{r}}_i(t) = \vec{v}_i(t) + \alpha\vec{r}_i \tag{2.26}$$

While for the volume it stands $\dot{V} = 3\alpha V$. Using standard equation for pressure derivative with respect to time and using the isothermal compressibility β, the scaling factor for the box can be obtained:

$$\mu = 1 - \frac{\beta\Delta t}{3\tau_P}(P_0 - P) \;\simeq\; \left[1 - \frac{\beta\Delta t}{\tau_P}(P_0 - P)\right]^{1/3} \tag{2.27}$$

As in the case of the Berendsen thermostat, this algorithm does not generate the correct equilibrium distribution. To address this issue, Bussi and co-workers developed a stochastic cell-rescaling algorithm capable of sampling the isobaric ensemble correctly [100]. The target distribution is:

$$\bar{P}(p, x, V) \propto \exp\!\left(-\frac{K + U + P_0 V}{k_B T}\right) \tag{2.28}$$

With P_0 denotes the target pressure. Following the same procedure used for the thermostat, the volume is assumed to evolve according to:

$$dV = D(V)\,\frac{\partial \log\!\big(\bar{D}(V)\,\bar{P}\big)}{\partial V}\,dt + \sqrt{2D(V)}\,dW \tag{2.29}$$

And $D(V)$ is a positive diffusion coefficient. Choosing $D(V) = \frac{\beta_T V k_B T}{\tau_P}$, with β_T denoting an estimate of the isothermal compressibility and τ_P coupling time, one can obtain the equation for the stochastic evolution of the volume. The simplest formulation can be obtained by rewriting such differential equation by using the strain variable $\varepsilon = \log(V/V_0)$, with V_0 reference volume needed for the correct dimensionality:

$$d\varepsilon = -\frac{\beta_T}{\tau_P}\,(P_0 - P_{\text{int}})\,dt + \sqrt{\frac{2k_B T\beta_T}{V\tau_P}}\,dW \tag{2.30}$$

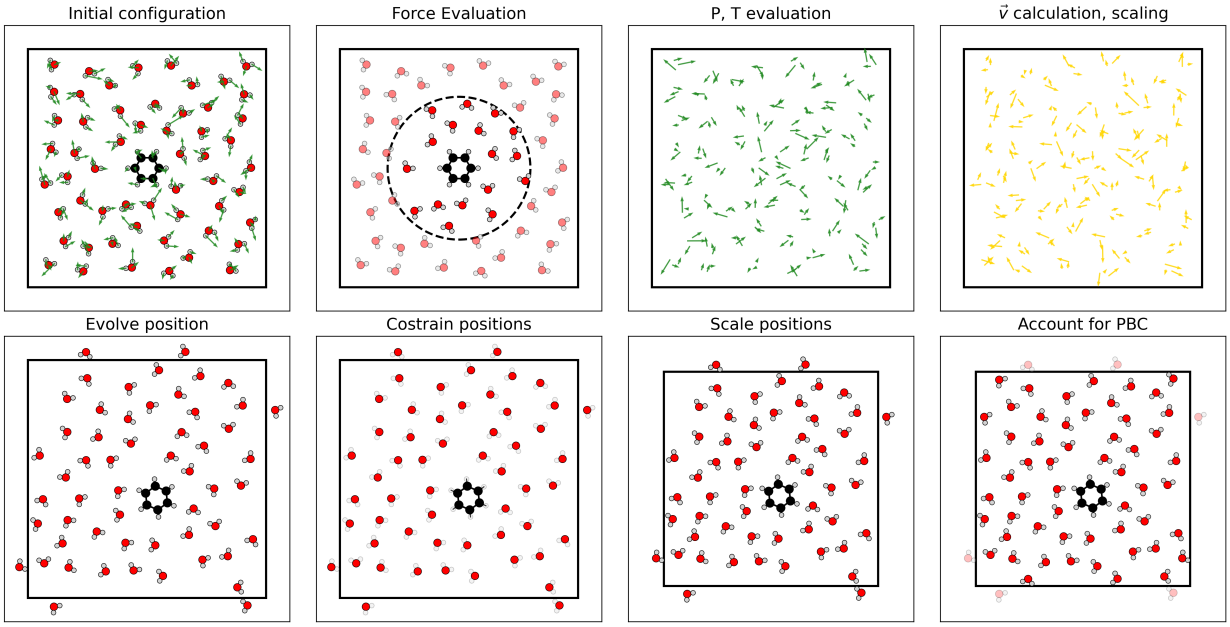

Figure 2.1: MD key steps. *Schematic view of main steps of a MD algorithm described in the text in a simple benzene in water system; from left to right, first row: an initial configuration is assigned velocities from Maxwellian distribution at a given temperature, forces are evaluated (here a simple cutoff is shown for benzene only), then temperature and pressure terms are evaluated and velocities are evolved. In the second row, positions are evolved, and constraints on atom positions are employed (for examplehere, hydrogen-involved bonds are restrained, and hydrogen are shown as transparent), then positions are scaled according to barostat and PBCs are taken into account.*

Where P_0 is the target pressure and $P_{\text{int}} = \frac{2K}{3V} - \frac{\partial U}{\partial V}$. As seen before, removing the noise term, the classical Berendsen barostat can be reobtained [100]. Combining **Equation 2.22** and **Equation 2.29** the correct sampling of the isothermal isobaric ensemble is ensured. In cases where the system is not isotropic, previous equations can be reformulated and extended into a vectorial form.

Therefore, simulating a system in the NPT ensemble generally requires the integration of the equations of motion together with the iterative application of corrections accounting for temperature and pressure fluctuations. Furthermore, for complex systems, it is common practice to remove selected degrees of freedom that can be considered negligible by introducing appropriate constraints. In protein simulations, for example, such constraints are typically applied to hydrogen atoms in order to neglect high-frequency vibrational motions during the integration steps. This procedure is implemented through the so-called SHAKE algorithm, which introduces an additional constraint force such that [101, 102]:

$$\sigma_k(\{\vec{r}(t)\}) \equiv (\vec{r}_j(t) - \vec{r}_i(t))^2 - d_{ij}^2 = 0, (k = 1, \ldots, l) \tag{2.31}$$

This equation imposes a constraint ensuring that the distance between two atoms remains fixed throughout each step of the simulation. Such a constraint is mathematically equivalent to introducing an additional force into the system that corrects the position of the constrained atoms following each integration step, leading to the modified equation of motion $m_i\dot{\vec{v}}_i(t) = \vec{F}_i(t) + \vec{G}_i(t)$, where $\vec{G}_i(t)$ represents the constraint force. Consequently, a general molecular dynamics simulation algorithm may proceed according to a Leap-frog integration scheme, beginning with:

1. Generate the set of atomic coordinates $\vec{r}(t)$ and velocities $\vec{v}\left(t - \frac{\Delta t}{2}\right)$
2. Evaluate unconstrained forces $\vec{F}(t)$ on all atoms taking periodic images into account (see below)

3. Evaluate the pressure correction. This procedure begins with the calculation of pressure terms in **Equation 2.25** for molecule i. For the virial contribution associated with intermolecular interactions, one obtains:

$$\Xi(t) = -\frac{1}{2}\sum_{\alpha<\beta}\vec{R}_{\alpha\beta}(t)\cdot\vec{F}_{\alpha\beta}(t) = -\frac{1}{2}\sum_{\alpha<\beta}\left(\vec{R}_\alpha - \vec{R}_\beta\right)\cdot\sum_{i_\alpha,j_\beta}\vec{F}_{i_\alpha - j_\beta} \tag{2.32}$$

where α, β are the molecular indices, $\vec{R}$ represents the centre of mass of the molecule, and $\vec{F}_{i_\alpha - j_\beta}$ is the force acting between atom i of molecule α and the atom j of molecule β. Then, the kinetic energy is calculated as:

$$E_k\left(t - \frac{\Delta t}{2}\right) = \frac{1}{2}\sum_{\alpha} M_\alpha\,\vec{V}_\alpha^{\,2}\left(t - \frac{\Delta t}{2}\right) \tag{2.33}$$

where M_α is the total mass of molecule α and V_α denotes the molecule velocity, computed as the total atomic momentum normalized by the molecular mass. Using these quantities, one can obtain the instantaneous pressure as $P(t) = \frac{1}{3V(t)}\left[E_k\left(t - \frac{\Delta t}{2}\right) - \Xi(t)\right]$ and the pressure scaling factor:

$$\mu = \left[1 + \frac{\beta\Delta t}{\tau_P}\left(P(t) - P_0\right)\right]^{1/3} \tag{2.34}$$

errors arising from the time difference between the evaluations of virial and E_k can generally be neglected, since pressure coupling constant is much larger than Δt.

4. Evaluation of temperature according to:

$$T\left(t - \frac{\Delta t}{2}\right) = \frac{2k}{3N - M - 3}E_k\left(t - \frac{\Delta t}{2}\right) = \frac{2k}{3N - M - 3}\sum_i m_i\,v_i^2\left(t - \frac{\Delta t}{2}\right) \tag{2.35}$$

and calculation of the corresponding temperature scaling factor:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T\left(t - \frac{\Delta t}{2}\right)} - 1\right)\right]^{1/2} \tag{2.36}$$

5. New velocities calculation

$$\vec{v}\left(t + \frac{\Delta t}{2}\right) = \vec{v}\left(t - \frac{\Delta t}{2}\right) + \vec{a}(t)\,\Delta t = \vec{v}\,'\left(t + \frac{\Delta t}{2}\right) \tag{2.37}$$

6. Velocity scaling via $\vec{v}\left(t + \frac{\Delta t}{2}\right) = \lambda\,\vec{v}\,'\left(t + \frac{\Delta t}{2}\right)$

7. New position calculation:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}\left(t + \frac{\Delta t}{2}\right)\Delta t = \vec{r}\,'(t + \Delta t) \tag{2.38}$$

8. Application of constraints on selected coordinates via SHAKE$(\vec{r}_{\text{uncost}}(t + \Delta t)) = \vec{r}_{\text{cost}}(t + \Delta t)$

9. Compute constrained velocities with $\vec{v}_{\text{rest}}\left(t + \frac{\Delta t}{2}\right) = \frac{\vec{r}_{\text{rest}}(t+\Delta t) - \vec{r}_{\text{rest}}(t)}{\Delta t}$

10. Apply pressure scaling to coordinates with $\vec{r}(t + \Delta t) = \mu\,\vec{r}\,'(t + \Delta t)$

11. Reinsert particles that have moved outside the simulation box according to the adopted periodic boundary conditions (for a schematic view, see **Figure 2.1**)

The scaling of coordinates may be applied either directly to individual atomic coordinates or to the centres of mass of the molecules. In the latter case, an additional transformation is subsequently performed to restore the internal molecular geometry and orientation.

2.1.4 Forces evaluation

Following the definition of temperature and pressure coupling schemes, another fundamental aspect of molecular dynamics simulations concerns the description of the forces governing interactions between the simulated particles. These interactions are specified through the definition of a force field. In the following discussion, only the classical AMBER force field will be considered [45, 103, 104], as it is the sole force field employed throughout this thesis; nevertheless, several alternative force fields are available in the literature.
At the microscopic level, molecular behaviour is fundamentally governed by quantum mechanics. However, the direct solution of the Schrödinger equation for large systems such as proteins remains computationally prohibitive. Consequently, the use of classical expressions to describe interatomic interactions represents a practical and widely adopted approximation.
Within the molecular dynamics framework, the force acting on an atom is obtained from the gradient of the system potential energy, and is defined as:

$$\begin{aligned} V_{\text{AMBER}}(\vec{r}^{\,N}) = & \sum_{i \in \text{bond}} k_{b_i} \left(l_i - l_i^0\right)^2 \\ & + \sum_{i \in \text{ang}} k_{a_i} \left(\theta_i - \theta_i^0\right)^2 \\ & + \sum_{i \in \text{dih}} \sum_{n \in \text{dih}_i} \frac{V_i^n}{2} \left[1 + \cos(n\omega_i - \gamma_i)\right] \\ & + \sum_{i<j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \\ & + \sum_{i<j} \frac{q_i q_j}{\varepsilon r_{ij}^2} \end{aligned} \tag{2.39}$$

The first three terms correspond to the intramolecular contributions to the potential energy that arise when a quantum-mechanical system, such as a molecule, is represented using a classical description. Within this approximation, a molecule is modelled as a collection of harmonic springs describing bond stretching and angle bending, together with periodic torsional potentials that account for dihedral rotations. The latter are particularly important for maintaining ring planarity and restricting energetically unfavourable conformations [84, 85].
Each contribution is associated with specific molecular properties. In the bond-stretching term, k_{b_i} denotes the force constant of bond i, while l_i^0 represents its equilibrium bond length. Similarly, k_{a_i} and θ_i^0 correspond to the force constant and equilibrium value of angle i, respectively. Dihedral interactions are generally represented by a cosine potential, where V_i^n denotes the energetic contribution of a given torsional term, γ_i is the phase angle, and n defines the periodicity of the potential.
The fourth term accounts for short-range van der Waals interactions and is commonly described using the classical Lennard–Jones 12-6 potential. The parameters A_{ij} and B_{ij} depend on the atom types and their hybridisation states. In most molecular dynamics simulations, van der Waals interactions are evaluated using a cut-off scheme, whereby only atoms located within a specified distance of atom i contribute to the force calculation. Under this approximation, long-range van der Waals interactions arising from periodic images are generally neglected.
The final term represents the electrostatic contribution to the total potential energy. Unlike van der Waals interactions, which decay rapidly with distance and become negligible at long range, Coulombic interactions remain significant over much larger distances. Consequently, an appropriate treatment of periodic boundary conditions (PBCs) becomes essential.
Periodic boundary conditions eliminate artificial surface effects by effectively replicating the simulation box in all spatial directions, thereby generating an infinite lattice of identical periodic images. As a result, atoms located near one boundary of the simulation box interact with atoms located near the opposite boundary. This approach is particularly important for the accurate treatment of

electrostatic interactions.
Since Coulombic forces scale according to $\vec{F}_{\text{Coul}} \sim \frac{f(r)}{r^2}$, the total electrostatic force acting on an atom cannot be accurately evaluated by considering only interactions within a finite cut-off radius. Such an approximation would introduce substantial errors capable of significantly affecting the dynamical behaviour of the system. To overcome this limitation, electrostatic interactions in periodic systems are commonly evaluated using the Ewald summation method, which provides a rigorous and computationally efficient framework for treating long-range interactions under periodic boundary conditions.
The Ewald approach separates the periodic space surrounding a given atom into two complementary regions: a short-range region comprising the atom and its local environment within a specified cut-off radius, and a long-range region consisting of the remainder of the infinite periodic system. For a periodic system containing N charged particles, the electrostatic energy contribution can therefore be expressed as:

$$E_{\text{Coul}} = \frac{1}{2} \sum_{i,j=1}^{N} \sum_{\vec{n} \in \mathbb{Z}^3} \frac{q_i q_j}{\|\vec{r}_{ij} + \vec{n}L\|} \tag{2.40}$$

where q_i, q_j are the charges associated with particles i and j respectively, r_{ij} denotes their pairwise separation vector, $\vec{n}$ is the lattice vector accounting for the periodic images, and L represents the simulation box length. For simplicity, this formulation assumes a cubic simulation box, although straightforward generalisations exist for other box geometries. However, the explicit summation over all periodic images does not guarantee convergence of the electrostatic energy series. This limitation motivates the central approximation underlying the Ewald summation method, namely the decomposition of the Coulombic $(1/r)$ dependence into two complementary contributions [84, 105]. Specifically, the Ewald method introduces and subtracts a Gaussian charge distribution around each point charge, thereby decomposing the Coulomb interaction into a rapidly convergent short-range contribution in real space and a smoothly varying long-range contribution in reciprocal space:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \tag{2.41}$$

This decomposition enables the evaluation of the Coulombic energy within the two subregions introduced above. The short-range contribution, represented by the term $\frac{f(r)}{r}$, decays rapidly and can therefore be evaluated directly in real space using a finite cut-off radius. Conversely, the complementary long-range contribution is computed in reciprocal space, where the relevant $\vec{k}$ vectors are restricted to a finite and computationally manageable domain. Thus, by means of Fourier transform, one can write that [106–108]:

$$f(\vec{r}) = \frac{1}{V} \sum_{\vec{k}=-\infty}^{+\infty} \tilde{f}(\vec{k})\, e^{i\vec{k}\cdot\vec{r}}, \tilde{f}(\vec{k}) = \int_V d\vec{r}\, f(\vec{r})\, e^{-i\vec{k}\cdot\vec{r}} \tag{2.42}$$

where V denotes the volume of the simulation box, l its characteristic length, and $\vec{k} = \frac{2\pi}{L}\vec{l}$ is the reciprocal-space wave vector associated with the lattice vector $\vec{l}$. In this formulation, long-range contributions corresponding to large values of $\vec{l}$ in real space are represented by low $\vec{k}$ components in reciprocal space, allowing an efficient evaluation of long-range electrostatic interactions. By combining the real-space and reciprocal-space contributions, together with the appropriate self-interaction corrections, the final expression for the Coulombic energy within the Ewald summation formalism can be written as:

$$E_{\text{Coul}} = E^r + E^k + E^s + E^d \tag{2.43}$$

where E^r represents the real-space contribution, E^k the reciprocal-space contribution, E^s the self-interaction correction term, and E^d the dipole correction [84, 106–108].

Although the Ewald summation method provides an accurate evaluation of electrostatic interactions, its computational cost scales approximately as $N^{\frac{3}{2}}$, where N is the number of atoms in the system. For large molecular systems, such as proteins composed of several hundred residues, this computational burden becomes substantial. Consequently, a variety of approximations to the Ewald formalism have been developed to improve computational efficiency while maintaining a high level of accuracy [109–111].
Among these approaches, the Particle Mesh Ewald (PME) method is one of the most widely employed in molecular dynamics simulations [108]. PME relies on the discretisation of the simulation box into a finite three-dimensional lattice, known as the particle mesh. Through interpolation schemes, the charge associated with each atom is distributed among neighbouring mesh points, thereby generating a discretised charge density. The reciprocal-space component of the Ewald summation is then evaluated directly on this mesh using fast Fourier transform techniques, whereas the short-range real-space interactions are calculated directly. As a result, the computational complexity is reduced from approximately $N^{\frac{3}{2}}$ to $N\log(N)$, at the expense of a small and controllable approximation in the calculated electrostatic forces, which depends on the mesh spacing, interpolation order, and Ewald parameters.
The accurate evaluation of all the terms described above enables molecular dynamics simulations to provide a computationally efficient and physically meaningful approximation of conformational phase-space sampling. Recent advances in computational power and algorithm development have extended the applicability of these methods to increasingly complex systems, ranging from simulations of entire cellular environments to millisecond-scale simulations of protein dynamics. However, the accurate representation of the physical forces governing protein motion constitutes only the first stage of the overall problem. Once confidence has been established that the simulation protocol provides a reliable approximation of real molecular behaviour, it becomes essential to analyse the resulting trajectories. Such analyses allow the identification of dynamic and structural features emerging from the simulations, which may provide valuable insights into the behaviour and functional properties of proteins in their biological environment.

2.2 Non equilibrium Molecular Dynamics

As discussed in the previous section, molecular dynamics simulations are founded on a robust theoretical framework that enables the characterisation of protein motions and conformational changes. However, these simulations generally probe fluctuations around equilibrium states, and the resulting observations are therefore confined to the equilibrium basin explored by the system. This limitation can become significant when investigating the molecular mechanisms underlying protein function, as many biologically relevant processes—including ATP hydrolysis, chemical reactions, and signal transduction—occur under non-equilibrium conditions.
To address this challenge, Dynamical Non-Equilibrium Molecular Dynamics (D-NEMD) was introduced by Ciccotti [112] and co-workers as a methodology capable of quantitatively characterising the response of a system subjected to an external perturbation. By driving the system away from equilibrium, D-NEMD enables the investigation of the structural and dynamical changes associated with non-equilibrium processes [40, 113–115].
The method was initially applied to relatively simple physical systems, such as the perturbation induced by the passage of an electric charge through a homogeneous medium or the relaxation dynamics of two immiscible fluids following the removal of a separating membrane. Subsequently, these approaches were extended to biomolecular simulations, where many functionally relevant phenomena inherently occur under non-equilibrium conditions. In the biological context, D-NEMD has been employed primarily to investigate signal transmission and allosteric communication in proteins following sudden perturbative events, such as bond cleavage, ligand dissociation, or other rapid changes in the chemical state of the system [115–117].
The theoretical foundation of this approach is provided by the Onsager principle of regressive

fluctuations, which states that:

$$O(\vec{x},t) = \left\langle \hat{O}(\vec{x},\Gamma), S^{\dagger}(t)w(\Gamma) \right\rangle \equiv \left\langle S(t)\hat{O}(\vec{x},\Gamma), w(\Gamma) \right\rangle \equiv \int d\Gamma\, \hat{O}(\vec{x},\Gamma(t,\Gamma))\, w(\Gamma) \tag{2.44}$$

where $O(\vec{x},t)$ represents the value of the observable at time t and position $\vec{x}$ following the application of the perturbation. Its value is obtained as the ensemble average of the observable operator $\hat{O}(\vec{x},\Gamma)$, defined in the initial phase space Γ, weighted by the time-evolved probability distribution function. In this formalism, $S^{\dagger}(t)$ denotes the adjoint of the time-evolution operator, which propagates the observable in time. The average response is then obtained by evaluating the observable over the equilibrium distribution $w(\Gamma)$ of initial configurations. Consequently, the observable at time t can be evaluated by propagating the probability distribution in time while averaging over the initial equilibrium ensemble.
This formulation implies that the system response following a perturbation can be determined by performing a thermodynamic average over the equilibrium probability distribution generated from standard molecular dynamics simulations. In practical terms, this corresponds to applying the perturbative event to a large number of independent configurations sampled from the equilibrium trajectory and subsequently evaluating the average response over the resulting ensemble of non-equilibrium trajectories.

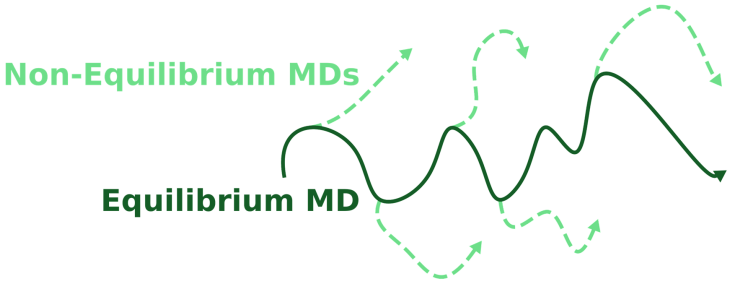

Figure 2.2: Schematic view of D-NEMD approach. *An equilibrium Molecular Dynamics is simulated (dark green), evolving in the phase space. Then, starting from equilibrium conditions, slight, non-equilibrium modifications are introduced in the system at different times (light green trajectories). This enables a correct sampling of the non-equilibrium phase space distribution.*

Such an approach enables the quantitative characterisation of the molecular processes triggered by events such as ATP hydrolysis, while also providing a framework for identifying signalling pathways and communication networks that propagate from the active site to distant allosteric regions of the protein. By analysing the collective response of the system to a well-defined perturbation, it becomes possible to distinguish statistically significant dynamical effects from the intrinsic fluctuations of the equilibrium ensemble.

2.3 Performing Molecular Dynamics Simulations

Throughout this thesis, a general protocol for protein molecular dynamics simulations was applied consistently across all studied systems. The overall procedure is described in this section, while any system-specific modifications will be discussed in the corresponding chapters where relevant.
Protein structures were obtained from the Protein Data Bank. Unless otherwise specified, crystallographic ions, solvent molecules, and co-crystallised cofactors not directly relevant to the investigated biological process were removed prior to system preparation. In cases where structural regions were missing from the experimental model, two different approaches were adopted depending on the length of the missing segment:

- For short missing loops (typically 5–15 residues), structural reconstruction was performed using the Maestro Protein Preparation Wizard

- For longer missing regions that could not be reliably modelled using the Maestro Protein Preparation Wizard, the MODELLER software package was employed [118]. Multiple loop conformations were generated, and the model exhibiting the lowest energy score, as evaluated by MODELLER, was selected for subsequent simulations.

Once missing regions had been reconstructed, a combination of software tools was employed to complete the structural preparation of each system. Missing side chains were added using the *pdb4amber* utility, the most favourable tautomeric states of residues were assigned using *Reduce*, and protonation states were assessed using *PROPKA* [119–121]. The resulting protein models were subsequently combined with the crystallographic water molecules, when present, and solvated using the tleap utility from the AMBER package [120].
Simulations were performed in either cubic or truncated octahedral periodic boxes using the TIP3P water model. To minimise artefacts arising from periodic boundary conditions and avoid interactions between protein periodic images, the dimensions of the simulation box were chosen such that a minimum distance of 11 Å separated any protein atom from the box boundary. The net charge of each system was neutralised by the addition of the appropriate number of Na^+/Cl^- counterions placed at randomly selected solvent positions. Following these preparation steps, the systems were ready for molecular dynamics simulations.
All simulations employed a consistent set of force-field parameters. Proteins were described using the ff14SB force field [103], while the parameters developed by Joung and Cheatham were used for Na^+/Cl^- [122], ions, and those proposed by Allnér and Nilsson were adopted for Mg^{2+} [123]. ATP and ADP molecules, when present, were parametrised using the polyphosphate parameters developed by Meagher and co-workers [124].
Additional ligands present in specific systems were parametrised using quantum-mechanical calculations for the assignment of atomic partial charges. Molecular geometries were first optimised in the gas phase using the Gaussian16 package [125], employing density functional theory (DFT) at the B3LYP/6-31G(d) level of theory. Electrostatic potential (ESP) charges were subsequently calculated at the Hartree–Fock/6-31G(d) level [126] using ten sampling shells per atom and a density of 17 grid points per square Bohr. Following RESP fitting [127], atomic partial charges were assigned using the *antechamber* utility within AMBER, which was also employed to generate the remaining bonded and non-bonded parameters according to the GAFF force field [128].
All molecular dynamics simulations were performed using the AMBER molecular simulation suite [119]. Energy minimisation and the initial equilibration stages were carried out using the CPU-based sander engine included in AmberTools, whereas the GPU-accelerated *pmemd* engine was employed for the later equilibration stages and production simulations [119, 120]. The simulation protocol consisted of five consecutive stages: pre-minimisation, minimisation, solvent equilibration, heating, and equilibration. The pre-minimisation stage was designed to relieve structural strain introduced during system preparation and to remove any steric clashes involving protein, ligand, or solvent hydrogen atoms. A total of 300 minimisation steps was performed, consisting of 10 steps of steepest-descent minimisation followed by 290 steps of conjugate-gradient minimisation. During this stage, only protein and ligand hydrogen atoms were allowed to relax, while all remaining heavy atoms were restrained using a harmonic potential with a force constant of 5 kcal mol^{-1} $Å^{-2}$ was applied.
The solvent equilibration stage was introduced to reduce the artificial ordering of water molecules generated during the *tleap* solvation procedure. This stage consisted of a 9 ps simulation in the NVT ensemble using a 1 fs integration time step and harmonic positional restraints on all solute atoms with a force constant of 10 kcal mol^{-1} $Å^{-2}$. The protocol was divided into three consecutive 3 ps simulations: heating from 25 K to 400 K with a temperature-coupling constant of 0.2 ps, equilibration at 400 K with the same coupling constant, and a final cooling stage from 400 K to 25 K, during which the coupling constant was progressively relaxed from 0.2 ps to 1.0 ps.

The heating stage increased the system temperature from 25 K to 300 K over a 20 ps NVT simulation using the Langevin thermostat [129]. A 2 fs integration time step was employed. To limit large structural fluctuations during heating, harmonic positional restraints were applied to all C_α atoms with a force constant of 5 kcal mol^{-1} Å^{-2}. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms [130]. At this stage, a new set of initial velocities was assigned to each replica, while all subsequent stages inherited the velocities generated in the immediately preceding simulation.
Equilibration was performed in the NPT ensemble using a Langevin thermostat maintained at 300 K [97, 129] and a Berendsen barostat set to 1 atm [93]. Hydrogen-containing bonds were constrained using the SHAKE algorithm [130] and a 2 fs integration time step was adopted throughout. The equilibration protocol consisted of three consecutive stages. First, a 20 ps simulation was carried out with positional restraints on C_α atoms reduced to 3.75 kcal mol^{-1} Å^{-2}. This was followed by a second 20 ps simulation with the restraint force constant further reduced to 1.75 kcal mol^{-1} Å^{-2}. Finally, a 2 ns unrestrained simulation was performed to ensure complete equilibration prior to production runs.
Production simulations were conducted in the NPT ensemble at 300 K and 1 atm pressure. Temperature was controlled using the Langevin thermostat [97, 129], while pressure was maintained using the Berendsen barostat [93]. A 2 fs integration time step was employed; long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) [108], whereas a cut-off distance of 8 Å was applied to Lennard–Jones interactions. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms [130]. Each system was simulated using at least four independent replicas, each consisting of a 1 µs production trajectory.

2.4 Classical Analysis of Molecular simulations

In the context of molecular dynamics simulations, a fundamental objective is the identification and characterisation of the molecular processes underlying protein motions and conformational changes. Over the years, numerous computational approaches have been developed to facilitate the analysis of protein dynamics, providing valuable insights into collective motions, conformational transitions, and communication pathways within biomolecular systems. In this section, a brief overview of the principal analytical methods employed throughout this work is presented,together with a discussion of their main limitations.

2.4.1 Deviations and Fluctuations

Among the most widely used metrics for assessing the structural stability and flexibility of proteins during molecular dynamics simulations are the Root Mean Square Deviation (RMSD) and the Root Mean Square Fluctuation (RMSF), defined as:

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\vec{x}_i - \vec{x}_i^{\,\mathrm{ref}}\right\|^2}, \mathrm{RMSF}_i = \sqrt{\frac{1}{t}\sum_{t'=1}^{t}\left\|\vec{x}_i(t') - \overline{\vec{x}}_i\right\|^2} \tag{2.45}$$

The RMSD is calculated between two structures composed of N atoms: a reference structure, in which atom i is located at position $\vec{x}_i^{\,\mathrm{ref}}$, and a target structure, in which the same atom occupies position $\vec{x}_i$. The quantity $\left\|\vec{d}\right\|$ denotes the norm of the displacement vector $\vec{d}$ between the corresponding atomic positions after optimal structural superposition, obtained through a roto-translational alignment that minimises the RMSD value.
Whereas RMSD provides a single global measure of the structural difference between two conformations, the Root Mean Square Fluctuation (RMSF) quantifies the flexibility of individual atoms

throughout a trajectory. RMSF is calculated as the square root of the average squared displacement of atom i with respect to its mean position $\overline{\vec{x}}_i$ over the course of the simulation. Consequently, RMSF provides an estimate of the amplitude of the fluctuations experienced by a given atom around its average location.
Both descriptors are commonly employed to assess the structural stability of proteins during molecular dynamics simulations. For example, RMSD is frequently used to monitor the deviation of a protein from its initial conformation, thereby providing an indication of the extent of conformational change occurring during the simulation. Conversely, RMSF identifies regions exhibiting enhanced flexibility and can therefore highlight residues or domains that are more prone to conformational rearrangements.
Despite their widespread use, both RMSD and RMSF possess inherent limitations. In particular, they rely on comparison with either a reference structure or an average conformation, thereby reducing the complexity of protein dynamics to relatively simple geometric descriptors. As a consequence, they are generally unable to capture collective motions, correlated displacements, or other higher-order dynamical features that often play a key role in protein function. Nevertheless, these metrics remain valuable qualitative tools for obtaining a preliminary characterisation of protein behaviour and conformational stability during molecular dynamics simulations.

2.4.2 Principal Component Analysis

The search for more informative and quantitative descriptors of protein dynamics led Berendsen and co-workers to formally introduce theapplication of Principal Component Analysis (PCA) to the study of slow, collective motions in proteins [131, 132]. PCA is a widely used statistical technique for the analysis of high-dimensional datasets and is primarily aimed at reducing the dimensionality of multicomponent systems while retaining the most relevant information.
In the context of protein dynamics, if only atomic positions are considered, the conformational state of a system composed of N atoms can be represented within a $(3N)$-dimensional configurational space. Owing to the highly interconnected nature of protein structures, however, not all $3N$ degrees of freedom contribute equally to the observed dynamics. In many cases, a relatively small number of collective motions accounts for a substantial fraction of the total conformational variance.
Given a trajectory described by the set of atomic coordinates $\{\vec{x}_k\}$, where $\vec{x}_k$ denotes the instantaneous atomic positions at a given time frame, the roto-translational degrees of freedom are first removed by fitting all structures to a common reference conformation. Subsequently, the covariance matrix of atomic positional fluctuations can be calculated as:

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N}\left(\vec{x}_i^k - \langle\vec{x}_i\rangle\right)\left(\vec{x}_j^k - \langle\vec{x}_j\rangle\right) \tag{2.46}$$

Where $\langle\vec{x}_i\rangle$ denotes the average value of coordinate $\vec{x}_i$ calculated over the trajectory. The covariance matrix therefore quantifies the degree to which pairs of atomic coordinates fluctuate in a correlated manner during the simulation. Once the covariance matrix $\boldsymbol{C}$, has been constructed, it can be diagonalised according to $\boldsymbol{R^T C R = \Lambda}$, where $\boldsymbol{R}$ is the matrix whose columns are the eigenvectors of $\boldsymbol{C}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues.
Diagonalisation of the covariance matrix is equivalent to solving the eigenvalue problem:

$$\boldsymbol{C}\vec{v}_n = \lambda_n\vec{v}_n \tag{2.47}$$

where λ_n and $\vec{v}_n$ denote n-th eigenvalue and eigenvector, respectively. The eigenvectors define the principal modes of motion of the system and form the basis of the so-called Essential Dynamics approach. Each eigenvector identifies a direction in conformational space along which the variance of the molecular motion is maximised, while the corresponding eigenvalue quantifies the magnitude of the fluctuations occurring along that direction.

From a physical perspective, these principal components describe collective motions involving large groups of atoms moving in a correlated manner. Such motions often correspond to biologically relevant conformational changes, including domain rearrangements, loop opening and closure, hinge-bending motions, or allosteric transitions. Consequently, the identification of the dominant principal components can provide valuable insight into the relationship between protein dynamics and biological function.
Because each eigenvector contains $3N$ components, corresponding to the Cartesian coordinates of all atoms included in the analysis, it can be mapped onto the protein structure to visualise the direction and relative amplitude of the collective motion. A common representation is the so-called hedgehog plot, in which a vector is associated with each atom to illustrate its contribution to a given principal mode.
Furthermore, each frame of a molecular dynamics trajectory can be projected onto one or more principal components, allowing the conformational sampling explored during the simulation to be represented in a reduced-dimensionality space. The projection onto a principal component is obtained through the following expression:

$$PC_i^k = \vec{v}_i \cdot \Delta\vec{x}^k \tag{2.48}$$

This projection enables the representation of a molecular dynamics trajectory within a reduced-dimensional conformational space, facilitating the visualisation of the regions explored by the system during the simulation. In this reduced space, conformational clusters may emerge, corresponding to distinct dynamical regimes or structural states sampled by the protein. The analysis of these clusters can provide valuable insights into conformational transitions and the relationships between different functional states.
To assess the significance of each principal component, the corresponding eigenvalue is typically normalised by the sum of all eigenvalues. The resulting quantity represents the fraction of the total positional variance explained by that principal component. Consequently, components associated with larger eigenvalues account for a greater proportion of the observed dynamics and are generally considered the most relevant for describing the collective behaviour of the system.
In summary, Principal Component Analysis provides an effective framework for reducing the complexity of molecular dynamics trajectories while retaining the most important dynamical information. By identifying the dominant collective motions and quantifying their contribution to the overall conformational variability, PCA facilitates the interpretation of protein dynamics and often enables the association of specific structural motions with biologically relevant functions.

2.4.3 Distance Fluctuations Analysis (DF)

When analysing protein dynamics from molecular dynamics trajectories, it is often desirable to move beyond the characterisation of global conformational motions and investigate the relationships between different regions of the protein [133]. In particular, one may be interested in identifying groups of residues that move in a coordinated manner, as well as the communication pathways through which structural and dynamical information is transmitted across the protein structure.
To address these questions, several complementary analytical approaches have been developed. Coupled and correlated motions can be investigated through Distance Fluctuation (DF) analysis, whereas the identification of putative allosteric communication pathways can be achieved using the Shortest Path Map (SPM) method. Together, these approaches provide valuable information regarding the dynamical organisation of proteins and the mechanisms underlying long-range intramolecular communication.
The DF method is based on the calculation of the distance fluctuation matrix, defined as:

$$DF_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle \tag{2.49}$$

where d_{ij} represents the distance between two atoms (typically C_α atoms) at a given frame of the trajectory, and $\langle\cdot\rangle$ denotes the time-average over the simulation.
The visualisation of the distance fluctuation (DF) matrix provides direct insight into the mechanical coupling between different regions of the protein. In particular, residue pairs exhibiting low distance fluctuations correspond to strongly correlated, mechanically coupled motions, whereas large fluctuations indicate weak coupling and higher degrees of relative flexibility. In this way, DF analysis enables the identification of structurally stable regions, as well as segments that are more prone to conformational rearrangements.
An additional useful application of DF analysis is the computation of the difference DF matrix, (ΔDF), obtained by subtracting the DF matrix of a reference state (e.g. the wild-type protein) from that of a perturbed system (e.g. a mutant). This differential analysis highlights the regions most affected by the perturbation, allowing the identification of local rigidification or increased flexibility induced by mutations, and thereby revealing subtle effects on the equilibrium dynamics of the protein.

2.4.4 Shortest Path Map (SPM)

Proteins are often conceptualised as networks that transmit signals or information between distant sites in a coordinated, allosteric manner. The investigation of such communication networks has led Osuna and co-workers to develop methodologies for identifying and quantifying allosteric pathways in proteins [134, 135]. In this framework, molecular dynamics simulations are typically performed and reduced to C_α representations. After removing roto-translational motion (e.g. by superimposing all frames onto an average or reference structure), the covariance matrix (C_{ij}) is computed as:

$$C_{ij} = \frac{\langle \Delta\vec{r}_i \cdot \Delta\vec{r}_j \rangle}{\sqrt{(\langle \Delta\vec{r}_i \rangle^2 \langle \Delta\vec{r}_j \rangle^2)}} \tag{2.50}$$

which represents the degree of (anti)correlation between each pair of selected atoms. In this formulation, each atom is interpreted as a node in a network representation of the protein, while the interaction strength between two nodes is defined on the basis of their dynamical correlation.
In particular, the effective edge length (or distance) between two nodes is typically defined as a function of the covariance (or correlation) between their motions, for example:

$$l_{ij} = -\log\left(|C_{ij}|\right) \tag{2.51}$$

This definition assigns a large edge length to pairs of atoms that are weakly (anti)correlated, whereas strongly (anti)correlated pairs are assigned a small edge length. In this way, the protein can be represented as a weighted network in which nodes correspond to residues (or atoms) and edges encode the strength of their dynamical coupling.
Once the network has been constructed, the Dijkstra algorithm [136] can be employed to determine the shortest paths between any pair of nodes. By systematically evaluating all $i - j$ pairs, it is possible to extract two relevant quantities: (i) the node importance, defined as the frequency with which a given node is visited along all shortest paths connecting source and target sites, and (ii) the edge importance, defined as the frequency with which a given edge is included in the set of optimal paths.
Mapping these quantities onto the three-dimensional protein structure enables a direct visualisation of communication pathways. By applying appropriate thresholds to node and edge importance, and representing them as spheres and bonds whose sizes are proportional to their respective weights, it becomes possible to highlight the most frequently used routes for dynamical information transfer within the protein.
Such representations are particularly useful for identifying how signals originating at specific functional sites, such as catalytic or ligand-binding regions, propagate through the protein matrix to distal sites. In this context, the combined use of Distance Fluctuation (DF) analysis and Shortest Path

Map (SPM) methods provides a powerful framework for identifying putative allosteric pathways and for predicting regions of the protein that may act as allosteric communication hubs or pockets coupled to the active site [116].

2.5 Novel Analysis on Molecular Dynamics Simulations

2.5.1 Tensor of Gyration (TOG) Analysis

One of the principal aims of this thesis is to develop methods to characterise the interactions between the proteins involved in the multiprotein complexes and to quantify their response to different induced perturbations. To this end, we developed a Tensor of Gyration (TOG) analysis framework designed to describe large-scale domain motions and inter-protein rearrangements in a reduced-dimensional representation. To the best of our knowledge, this approach has not previously been applied to the analysis large protein assemblies.
For a given group of atoms, the gyration tensor is defined as:

$$S_{mn} = \frac{1}{M} \sum_{i=1}^{M} \left(r_m^i - r_m^{\mathrm{COM}}\right) \left(r_n^i - r_n^{\mathrm{COM}}\right) \tag{2.52}$$

where S_{mn} denotes the (m, n) element of the gyration tensor, M is the number of atoms in the selected group, r_m^i is the m-th Cartesian coordinate of atom i, and r_m^{COM} is the corresponding coordinate of the centre of mass. All tensor calculations were performed in Python 3 using the MDTraj package [137].
The TOG framework approximates protein domains as ellipsoidal bodies, thereby reducing the dimensionality of the system while retaining its principal geometrical features. Each protein was initially divided into one or more biologically relevant subdomains according to the architecture of the starting structure (**Figure 2.3A**). This coarse partitioning was subsequently refined using information derived from Distance Fluctuation (DF) analysis (see above).
To identify residues that best represent the rigid structural core of each domain, a DF score was calculated for every residue:

$$DF_{\mathrm{score}}^i = \frac{1}{M} \sum_{j=1}^{M} DF_{ij} \tag{2.53}$$

where DF_{ij} is the distance fluctuation between residues i and j; M is the total number of amino residues. Residues with low scores correspond to highly coordinated and structurally stable regions and were therefore considered representative of the domain architecture.
Following the initial domain assignment, an iterative filtering procedure was applied. At each iteration, residues with $DF_{\mathrm{score}} > 3\,\text{Å}$ are removed, were removed and the scores recalculated. The process was repeated until all remaining residues fell below the threshold, yielding a compact set of residues representative of each domain (**Figure 2.3B**).
The procedure was applied independently to all proteins present in various multiprotein assemblies, retaining different domains. For each filtered domain, the gyration tensor was calculated for every simulation frame and subsequently diagonalised. Diagonalisation yields the principal axes of the ellipsoid (eigenvectors) and their corresponding magnitudes (eigenvalues):

$$\mathrm{TOG}_p^s = \left\{ \mathrm{COM}^s;\ \mathrm{COM}^s \pm \sum_{i=1}^{3} \lambda_i^s\, \vec{v}_i^s \right\} \tag{2.54}$$

For each subdomain s, seven characteristic points were defined: the centre of mass and the six end-points corresponding to the positive and negative directions of the three principal axes. These points

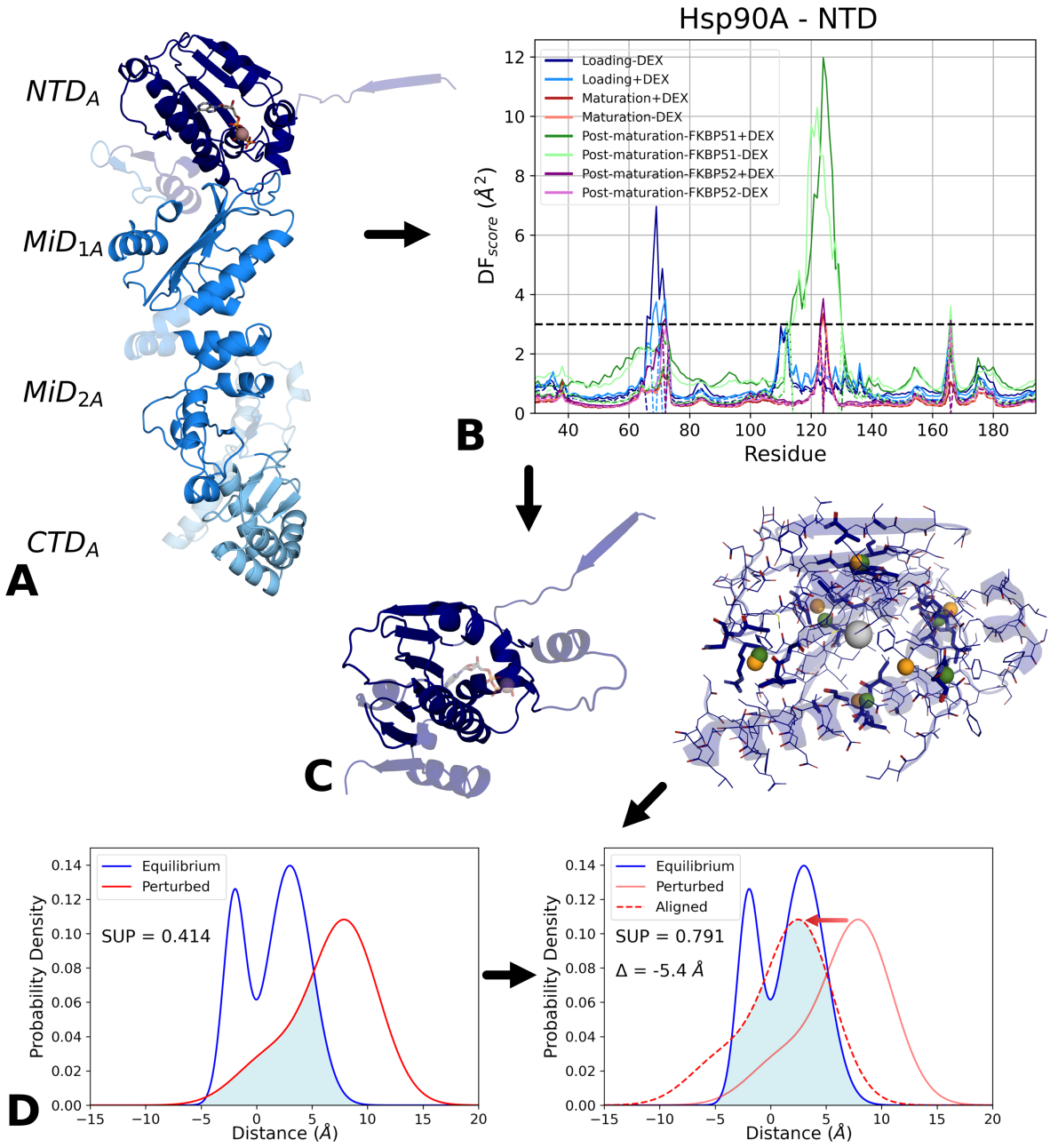

Figure 2.3: Representation of TOG analysis on the $\mathrm{NTD_A}$ ***of Hsp90.*** ***(A)*** *Hsp90 protomer A (see* ***Chapter 1*** *and* ***Chapter 3****) is coloured by its coarse domain division.* ***(B)*** DF_{score} *plots of the* $\mathrm{NTD_A}$ *for each simulated system.* DF_{score} *in coloured solid lines; horizontal black dotted line represents the threshold to iteratively discard highly flexible amino acids. Coloured dotted lines represent the* DF_{score} *at the end of the selection process.* ***(C)*** *On the right, selected amino acids are highlighted on the* $\mathrm{NTD_A}$ *structure. The results of TOG diagonalization are reported: orange dummy atoms correspond to the positions of geometrical semiaxes of the ellipsoid while green ones represent their closest physical counterpart; green spheres are extracted from the centre of mass of the atoms in the surroundings of geometrical points for at least 25% of the simulation (represented as sticks). On the left, transparent representation highlights the common Hsp90 residues that are discarded during both filtering processes.* ***(D)*** *Superposition and alignment procedure for a pair of distance distributions (before and after the perturbation). On the left, raw distance distributions from the (un)perturbed dynamics (poor superposition, 41.4%). On the right, the perturbed distance distribution is shifted by a* Δ *value to the maximum degree of overlap (improved to 79.1%).*

provide a reduced geometric representation of the domain shape and orientation. The endpoints were obtained as the sum of the coordinates of the centre of mass and the product of the eigenvectors $\vec{v}_i$ and eigenvalues λ_i along the three dimensions.
To facilitate trajectory analysis, each geometric point was represented by a dummy atom. However, direct use of eigenvector-derived coordinates introduced artefacts associated with eigenvector sign inversion and axis swapping. To overcome these issues, each dummy atom was anchored to physically meaningful regions of the protein. Atoms located within a spherical neighbourhood of a given dummy atom for at least 25% of the trajectory were identified, and hierarchical clustering of their average pairwise distances was performed to define stable atomic groups (**Figure 2.3C**). The centres of mass of these clusters were then used as physically anchored reference points, thereby eliminating artefacts arising from eigenvector degeneracy and improving the interpretability of the resulting motions.
The resulting TOG representation provides a dimensionally reduced description of protein dynamics, in which complex atomic motions are approximated by fluctuations of a small number of geometrically and structurally meaningful reference points.
Principal component analysis (PCA) was subsequently performed on the dummy-atom trajectories to identify the dominant collective motions sampled by each system.
To compare large-scale motions between systems, distributions of inter-dummy-atom distances were analysed using overlap integral calculations. For each distance distribution pair, one distribution was translated in increments of 0.5 Å until maximal overlap was achieved (**Figure 2.3D**). Two matrices were generated from this procedure. The first reports the overlap before and after alignment, thereby quantifying the similarity of the distributions. The second reports the magnitude of the translation required to maximise the overlap. Because translations were applied to the perturbed systems, negative values indicate an increase in inter-domain separation, whereas positive values correspond to a decrease in separation.
Overall, TOG analysis provides a robust framework for reducing complex protein motions to a set of interpretable geometric descriptors, enabling quantitative comparison of domain-level dynamics across distinct multiprotein assemblies.

2.5.2 Analysis of internal dynamics from D-NEMD

To analyse the D-NEMD trajectories, we developed a complementary analysis framework that extends the conventional methodology commonly employed in D-NEMD studies [40, 113, 115]. In the standard approach, hereafter referred to as the *external* analysis, the response to a perturbation is quantified by comparing the positions of individual C_α atoms in equilibrium and non-equilibrium trajectories. Specifically, the average displacement of each atom is calculated at fixed times following the perturbation, providing a direct measure of the structural response (**Figure 2.4A**).
While this approach efficiently captures large-scale conformational changes, it does not explicitly describe how perturbations affect the internal organisation of the protein assembly. To address this limitation, we additionally computed complete pairwise C_α distance matrices for both equilibrium and non-equilibrium trajectories at equivalent time points following the perturbation (**Figure 2.4B**). This internal analysis characterises changes in the relative positions of residue pairs and therefore provides direct information on coordinated structural rearrangements within the complex, such as local remodelling within active sites or long-range allosteric responses.
Because the number of pairwise distances scales quadratically with the number of residues, direct interpretation of the resulting datasets is challenging. Inspection of the ensemble-averaged distance trajectories indicated that, for most residue pairs, the early-time response could be adequately described by a linear model. This observation is consistent with the assumptions of linear response theory, which predicts that systems perturbed sufficiently close to equilibrium exhibit a predominantly first-order response, thus the temporal evolution of each $C_\alpha - C_\alpha$ distance was approximated as:

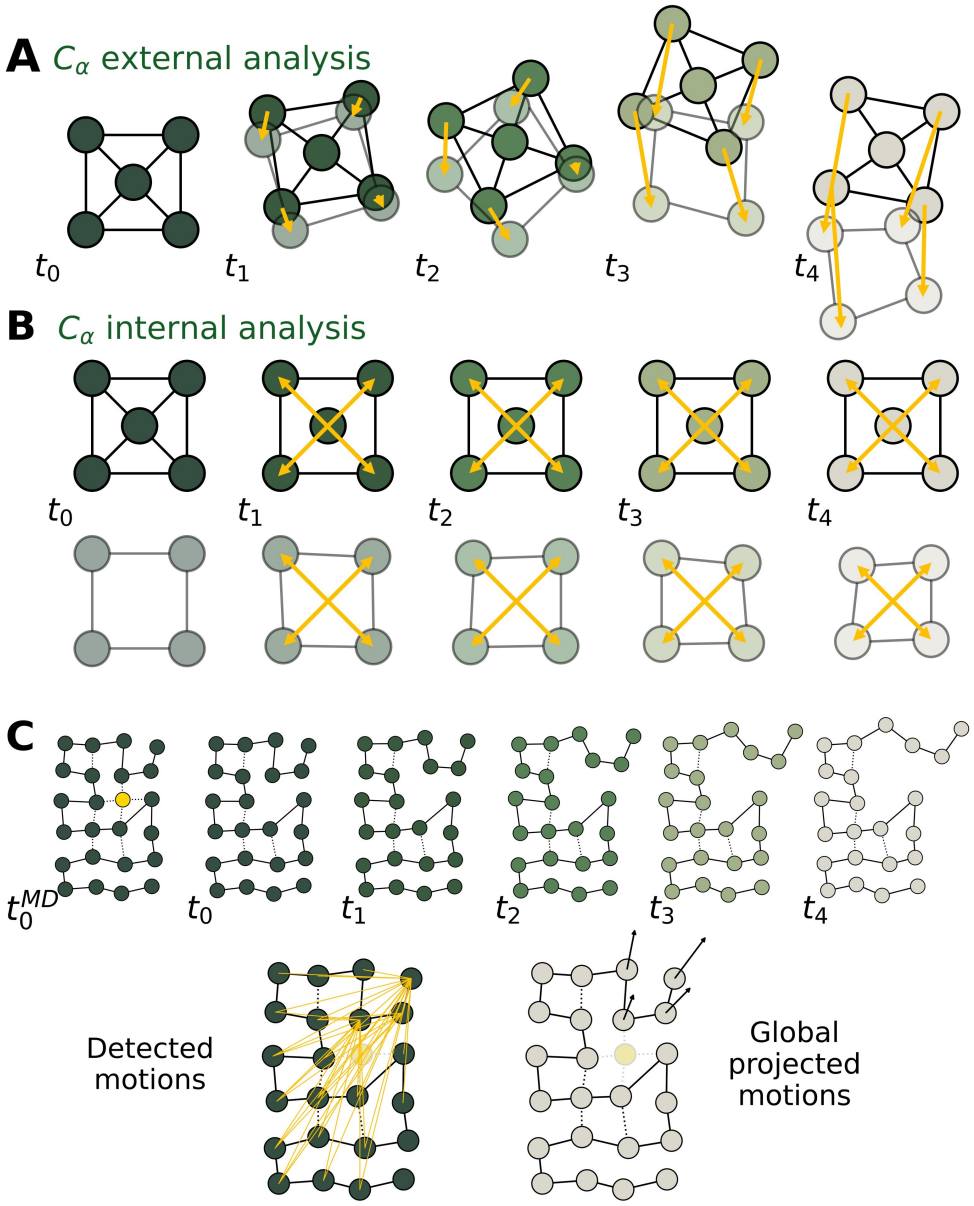

Figure 2.4: Schematic view of different D-NEMD analyses to characterize the systems' internal dynamics. ***(A)*** *An illustrative representation of the external method to analyse the effect of the perturbation on the system, it relies on the measure of the average deviation from the equilibrium simulation of corresponding atoms.* ***(B)*** *An illustrative representation of the newly implemented internal method to explore dynamical differences within D-NEMD, which focuses on the differences between internal distances.* ***(C)*** *Top lane: schematic representation of the time-dependent D-NEMD perturbation effect, caused by ligand removal (yellow circle) in a protein-like model (green circles). Bottom left: a representation of the residue pairs involved in the D-NEMD analysis using pairwise distances (yellow lines), which correspond to the residues affected by the perturbation (selected by a threshold on the pairwise distances). Bottom right, vector directions and norms encode the direction and entity of the most significant residue displacements from the D-NEMD analysis.*

$$C_\alpha^{i-j}(t) = a_{C_\alpha^{i-j}} + b_{C_\alpha^{i-j}} + \mathcal{O}(t^2) \tag{2.55}$$

where C_α^{i-j} is the distance between the α atoms of residue i and j at time t, $a_{C_\alpha^{i-j}}$ describes the displacement at $t = 0$, $b_{C_\alpha^{i-j}}$ is the corresponding linear response coefficient, and $\mathcal{O}(t^2)$ represents higher-order contributions neglected within the linear approximation.
With the exception of residues located in the immediate vicinity of the perturbed active sites, where the response occasionally displayed non-linear behaviour, the linear approximation provided a satisfactory description of the observed distance changes (see **Chapter 3** for different examples of perturbation-induced effects). The fitted coefficients were therefore used to estimate the conformational response of the system at selected times following the perturbation.
To identify the regions most strongly affected by the perturbation, residue pairs displaying an absolute linearised distance variation greater than 0.3 Å were first selected. This value was adopted as a visibility threshold. For each residue, a frequency score was then calculated by counting the number of pairwise interactions involving that residue that exceeded the threshold. Residues whose frequency exceeded a predefined cut-off were classified as affected residues and considered to participate significantly in the structural response.
For each affected residue i, an overall displacement vector was calculated by summing the unit vectors corresponding to all $i - j$ directions, weighted by the associated distance variation $C_\alpha^{i-j}(t)$. The resulting vector provides an approximate description of the net direction of the local structural response relative to the remainder of the system. For visualisation purposes, the vectors were scaled by a constant factor and projected onto a reference structure corresponding to the first frame of the first equilibrium replica of each system (**Figure 2.4C**).
Although qualitative in nature, this representation facilitates interpretation of the perturbation-induced motions. For example, residues belonging to an opening loop are expected to display coordinated increases in their distances from neighbouring regions, generating a resultant vector oriented along the direction of the opening motion.
Compared with the conventional external analysis, the internal approach exploits a substantially larger set of structural observables and is therefore more sensitive to subtle coordinated rearrangements and local conformational changes that may not be apparent from atomic displacements alone.

2.6 Concluding remarks

The introduction of the molecular dynamics formalism and the underlying physical principles provides the theoretical framework necessary for understanding the simulations presented throughout this thesis. Furthermore, the overview of the principal analytical techniques employed for the investigation of molecular dynamics trajectories establishes the methodological basis for the interpretation of the results discussed in the following chapters.
In addition to established approaches, two recently developed methodologies—TOG and internal D-NEMD analysis—have been introduced. Their application to the study of large protein complexes will enable the investigation of the mechanistic and dynamical consequences of biologically relevant perturbations, providing further insight into the relationship between molecular motions and protein function.
The following chapters will therefore focus on the characterisation of different protein functional cycles, with particular attention to the identification of communication pathways, allosteric effects, and the molecular mechanisms underlying protein activity.

Chapter 3

Ligand Control in the Hsp90 Chaperone Cycle of GR

3.1 Introduction

Cells are highly dynamic systems in which numerous biological processes occur continuously [138–140]. Among the most fundamental of these processes is the synthesis of polypeptides, which must adopt the correct three-dimensional conformation in order to become biologically functional. Conversely, misfolded proteins must be efficiently recognised and degraded [141–143]. Molecular chaperones play a central role in maintaining this balance between protein folding and degradation by ensuring that client proteins attain their native conformations and thereby preserving proteome integrity [144–151].

Among these chaperones, Heat Shock Protein 90 (Hsp90) occupies a particularly important position. Unlike many other molecular chaperones, Hsp90 functions at the later stages of the protein folding pathway. Moreover, its dysregulation has been implicated in a wide range of pathological conditions, including cancer and neurodegenerative diseases [53, 152–155].

Hsp90 is a homodimeric protein comprising two identical protomers (protomer A and protomer B), each organised into three distinct domains [53, 156–161]. The N-terminal domain (NTD) contains the ATP-binding site and mediates ATPase activity through an asymmetric mechanism involving ATP hydrolysis in one protomer followed by hydrolysis in the second. The middle domain (MiD) is responsible for binding client proteins within the client-binding cavity, commonly referred to as the lumen. Finally, the C-terminal domain (CTD) maintains dimerisation and also contributes to client recognition (**Figure 3.1A**).

The activity of Hsp90 is tightly regulated through interactions with a diverse range of task-specific proteins known as cochaperones, which selectively modulate the dynamics of chaperone–client interactions. The chaperoning function of Hsp90 is driven by an ATP-dependent conformational cycle involving coordinated structural rearrangements of the chaperone, its cochaperones, and client proteins. These conformational transitions ultimately determine client-specific remodelling and activation.

During the Hsp90 chaperone cycle, ATP binding to both protomers shifts the conformational ensemble towards a closed, high-energy state characterised by an asymmetric organisation. Subsequent ATP hydrolysis initiates a series of conformational transitions, beginning with a change in symmetry that leads to the formation of a symmetric closed dimer. Closure of the Hsp90 protomers is accompanied by extensive remodelling of the client-binding site, thereby facilitating the structural rearrangements required for client maturation. Hydrolysis of the second ATP then drives the chaperone into a transient compact ADP-bound state from which client proteins are released. Finally, dissociation of ADP restores the apo state, completing the cycle and returning Hsp90 to its initial conformation [159].

Biochemical and biophysical studies, together with advances in cryo-electron microscopy (cryo-EM),

have provided high-resolution structural insights into the molecular assemblies that govern the principal stages of glucocorticoid receptor (GR) activation [70, 71, 162–167].

GR mediates the cellular effects of glucocorticoids, a class of steroid hormones whose dysregulation is associated with disorders such as Cushing's syndrome and Addison's disease [69]. The receptor comprises three domains; however, experimental structural information is currently available only for the ligand-binding domain (LBD) and a portion of the interdomain linker region. Functional maturation of GR is orchestrated by Hsp90 through a multistep chaperone cycle that converts the receptor from an inhibited, partially unfolded state into its active, ligand-bound conformation (**Figure 3.1B**).

GR is initially recruited and stabilised by Hsp40 and Hsp70 before being transferred to Hsp90 together with Hop and a second Hsp70 molecule to form the so-called Loading Complex [70]. In this state, Hsp90 adopts a semi-closed conformation that encloses the receptor's disordered tail within its central lumen, while GR remains inactive and partially unfolded. Dissociation of Hsp70 and Hop promotes the transition to the Maturation Complex, [71] which is characterised by incorporation of the cochaperone p23. In this assembly, Hsp90 adopts a fully closed, ATP-hydrolysis-competent conformation that enables remodelling of the GR ligand-binding domain, ultimately leading to receptor activation. This mechanism is exemplified by cryo-EM structures showing GR bound to the synthetic glucocorticoid dexamethasone (DEX).

Subsequent late-stage remodelling is mediated through mutually exclusive interactions with the immunophilins FKBP51 and FKBP52 [72] which displace p23 and bind to the C-terminal region of Hsp90 while simultaneously contacting GR near its ligand-binding pocket (**Figure 3.1B**). Despite adopting remarkably similar binding modes, these cochaperones exert opposing functional effects. FKBP52 stabilises the interaction between GR and the chaperone complex and promotes dynein-mediated nuclear translocation of the entire assembly, thereby enhancing glucocorticoid signalling and downstream transcriptional activity. In contrast, FKBP51 acts antagonistically by retaining GR in the cytosol and facilitating receptor release [72] Notably, in both immunophilin-bound complexes, GR remains ligand-bound and signalling competent, highlighting the finely balanced nature of chaperone–cochaperone regulation and suggesting that ligand occupancy itself may contribute to determining the outcome of GR maturation and release.

Despite the wealth of structural and biochemical information currently available, several key questions remain regarding the molecular mechanisms governing GR maturation and function. These include the role of dynamic communication between the components of the multiprotein assemblies, the structural organisation of GR within distinct chaperone complexes, and the relationship between ligand binding and the molecular determinants underlying the divergent functional effects of the highly homologous immunophilins FKBP51 and FKBP52. In this work, these questions are addressed using unbiased equilibrium and non-equilibrium atomistic molecular dynamics (MD) simulations. The primary objective is to identify the molecular determinants governing the dynamics of the various chaperone complexes and GR ligand-bound states, and to relate these features to the experimentally observed functional roles of the chaperone machinery in regulating client folding pathways. By combining extensive equilibrium MD simulations with dynamical non-equilibrium molecular dynamics (D-NEMD) approaches and advanced analysis methodologies [40, 114–117, 133, 168–170], we investigate how internal dynamics and intercomponent communication govern progression through the chaperone cycle (**Figure 3.1C**).

Our results demonstrate how the Hsp90 machinery primes GR for steroid binding, while also suggesting that the bound steroid itself may act as a signalling element capable of modulating the stability and dynamics of distinct chaperone assemblies. These findings establish a direct connection between the atomic-scale conformational dynamics of GR–chaperone complexes and the chemical state of their binding sites. Furthermore, our analyses reveal long-range allosteric effects arising from subtle structural perturbations, whether induced by changes in the nucleotide state of Hsp90 or by steroid binding and unbinding within GR. These perturbations propagate across domains and protein interfaces, ultimately influencing the functional behaviour of the entire assembly.

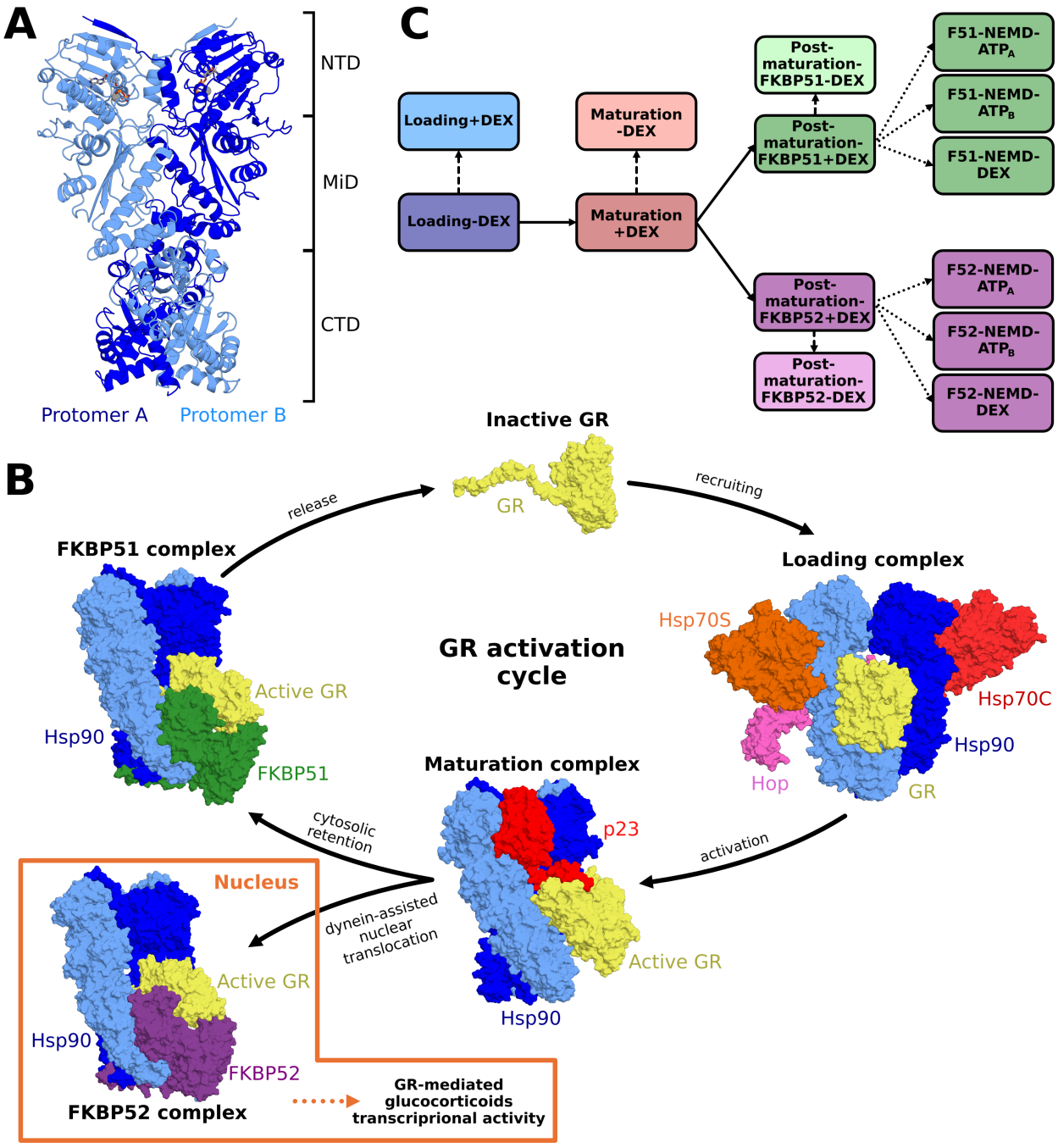

Figure 3.1: Structural organization and chaperone cycle of the Hsp90 machinery. *(A) The structural organization of Hsp90: protomer A in blue; protomer B in light blue.* ***(B)*** *Schematic representation of the GR activation cycle, showing Cryo-EM resolved structures of multiprotein complexes at different stages. Individual proteins are defined in the main text.* ***(C)*** *A schematic view of the different equilibrium and non-equilibrium simulations performed for this study.*

Table 3.1: *On the top, a summary of the equilibrium classical MD models and simulations featured in this work, with the third column specifying whether the simulated system is in its **U**nperturbed or **P**erturbed state. On the bottom, a summary of the non-equilibrium simulations (D-NEMD).*

Model	Components	U/P	PDB ID	MD name	Length (µs)
Loading	Hsp90:Hsp70C$_{ADP}$: Hsp70S$_{ADP}$:Hop:GR	U	7KRW	L	4
Loading+DEX	Hsp90:Hsp70C$_{ADP}$: Hsp70S$_{ADP}$:Hop:GR$_{DEX}$	P	7KRW	L+DEX	5
Maturation	Hsp90$_{ATP}$:p23:GR$_{DEX}$	U	7KRJ	M	4
Maturation−DEX	Hsp90$_{ATP}$:p23:GR	P	7KRJ	M-DEX	5
FKBP51	Hsp90$_{ATP}$:FKBP51:GR$_{DEX}$	U	8FFW	F51	9
FKBP51−DEX	Hsp90$_{ATP}$:FKBP51:GR	P	8FFW	F51-DEX	5
FKBP52	Hsp90$_{ATP}$:FKBP52:GR$_{DEX}$	U	8FFV	F52	9
FKBP52−DEX	Hsp90$_{ATP}$:FKBP52:GR	P	8FFV	F52-DEX	5

Model	D-NEMD name	Perturbation
FKBP51	• F51-NEMD-ATP • F51-NEMD-DEX	ATP hydrolysis DEX removal
FKBP52	• F52-NEMD-ATP • F52-NEMD-DEX	ATP hydrolysis DEX removal

3.2 Systems modeling

The simulation systems considered in this work and their corresponding labels are listed in **Table 3.1**. Given the size and complexity of the multiprotein assemblies investigated, the modelling strategy adopted here differs slightly from that described in **Chapter 2**. Accordingly, the system preparation protocol employed prior to molecular dynamics (MD) simulations is described in detail below. The glucocorticoid receptor (GR), which is present together with Hsp90 in all simulated stages, has been structurally characterized by cryo-EM only for its ligand-binding domain (LBD). Therefore, throughout the Results section, the term GR refers specifically to the GR-LBD, unless otherwise stated.

3.2.1 Loading−DEX and Loading−DEX complexes

The Loading−DEX complex was modelled using the cryo-EM structure deposited under PDB entry 7KW7 [70]. The Hsp90α sequence was derived from UniProt entry P07900 and comprised residues E16–D699; the disordered N- and C-terminal tails were not included. Owing to its high flexibility, the charged linker region (residues 233–283) is unresolved in both Hsp90 protomers in the experimental structure. To preserve backbone continuity, this region was replaced in each protomer by a shortened linker generated using the Crosslink Proteins utility implemented in Maestro (Schrödinger Release 2021-4; Schrödinger, LLC, New York, NY, USA), following the procedure described previously [171]. Specifically, the resolved protein segments were connected through an EKEK loop. Although omitting this loop reduces the computational complexity of the simulations, this was not the primary motivation for our modelling choice. Rather, because this region is unresolved in the available cryo-EM structures, any reconstruction would require additional modelling assumptions and could introduce structural artifacts. Furthermore, owing to its expected conformational flexibility, modelling the loop would likely increase the conformational variability of the system, potentially hindering the structural differences associated with the processes investigated in this Chapter. We acknowledge that omitting the loop represents a limitation of the structural model and may influence the dynamics

of the complex. However, since this study does not aim to investigate the role of this region, we do not expect its omission to qualitatively affect the conclusions drawn from the simulations.
Hsp90 was modelled in its apo state, whereas both Hsp70 molecules—the scaffolding Hsp70 (Hsp70S) and catalytic Hsp70 (Hsp70C)—contained one ADP molecule and one Mg^{2+} ion bound within their respective nucleotide-binding sites.
The Loading+DEX complex was generated from the Loading−DEX model by introducing dexamethasone (DEX) into the GR ligand-binding pocket. As the inactive receptor adopts a conformation with a partially occluded ligand-binding site, induced-fit docking (IFD) was employed to account for local protein flexibility during ligand accommodation.
To identify receptor conformations compatible with ligand binding, 4 µs equilibrium MD trajectories of the Loading−DEX and Maturation+DEX complexes were analysed. For each frame of the Loading−DEX simulations, the backbone root mean square deviation (RMSD) of the folded GR core (residues 559-769) was calculated against all frames of the Maturation+DEX trajectories, excluding the extended tail loop. Several Loading−DEX conformations exhibited close structural similarity to the Maturation+DEX ensemble, with minimum RMSD values of approximately 1.7 Å.
The resulting two-dimensional RMSD matrix was used to identify Loading−DEX conformations most closely resembling the ligand-bound receptor observed in the Maturation+DEX complex. Four representative structures corresponding to the lowest RMSD values were selected and subjected to induced-fit docking using the Maestro Schrödinger Suite version 2024-3 (Schrödinger Release 2024-3; Schrödinger, LLC, New York, NY, USA). This strategy enabled the identification of ligand-bound receptor conformations within the Loading complex that closely resembled the active GR state sampled in the Maturation+DEX simulations, thereby providing consistent starting structures for subsequent MD simulations.
The four selected Loading−DEX conformations were prepared using the Protein Preparation Wizard implemented in Maestro. Dexamethasone (DEX) was prepared using LigPrep, with protonation states assigned at pH 7.0 using Epik [172]. Induced-fit docking (IFD) was subsequently performed using the Maestro IFD workflow [173].
The docking region was defined by a 10 $Å^3$ cubic inner box centred on the centroid of residues located within 4 Å of DEX in the reference Maturation+DEX structure. These residues included M560, L563, N564, L566, G567, Q570, M601, M604, A605, L608, F623, Q642, M646, L732, Y735, C736, T739, I747, F749 and L753. Inspection of the ligand-binding pocket revealed that several side chains substantially restricted the available volume and could potentially generate steric clashes during ligand placement. To alleviate these constraints, selected residues lining the binding pocket (L563, N564, L566, M604, L608 and Y735) were temporarily truncated to alanine during the docking calculations and subsequently restored during the refinement stage.
Following docking, all residues within 7.0 Å of the ligand poses were subjected to local optimisation. Side-chain reconstruction and energy minimisation were carried out using Prime to restore the original residue identities and refine protein–ligand interactions. In the final stage of the protocol, 20 representative poses generated by the IFD workflow were redocked using Glide in Standard Precision (SP) mode.
The GR structures from Maturation+DEX simulations were then used as references for evaluating the docking poses. Indeed, docking scores were not used as the primary selection criterion because the highest-scoring poses frequently displayed an incorrect orientation of DEX within the GR binding pocket, indicating a poor correlation between docking score and pose quality. Therefore, poses with an incorrect ligand orientation were first discarded, retaining only those consistent with the reference binding mode. The final poses were chosen as the ones with the lowest RMSD with respect to the DEX conformation in the corresponding Maturation+DEX reference structure. The five poses exhibiting the closest agreement with the reference ligand orientation were selected as starting structures for the subsequent equilibrium MD simulations (**Figure A.1**).

3.2.2 Maturation+DEX and Maturation−DEX complexes

The Maturation+DEX complex was modelled using the cryo-EM structure deposited under PDB entry 7KRJ, comprising Hsp90, the cochaperone p23, and the client protein glucocorticoid receptor (GR) [71] (**Figure 3.1B**). As in the Loading−DEX model, the charged linker region of Hsp90 is unresolved in both protomers and was therefore reconstructed as described above to maintain backbone continuity between the N-terminal (NTD) and middle (MiD) domains.
In addition to the unresolved Hsp90 linker, the cryo-EM structure lacks four N-terminal GR residues (SIVP) relative to the Loading−DEX complex. These residues form part of the GR tail threaded through the Hsp90 lumen between the MiDs of protomers A and B and may therefore contribute to chaperone–client interactions. To obtain a complete model of the resolved GR segment, residues 519–522 (UniProt entry P04150) were reconstructed using the PyMOL Molecular Graphics System (Version 1.8, Schrödinger, LLC, New York, NY, USA). The resulting GR model comprised residues 519–777, encompassing the ligand-binding domain and the associated tail region.
The cochaperone p23 was retained as resolved in the experimental structure. Each Hsp90 protomer contained one ATP molecule and one Mg^{2+} ion bound within the NTD nucleotide-binding site, while dexamethasone (DEX) occupied the GR ligand-binding pocket.
The corresponding Maturation−DEX model was generated by removing DEX from the Maturation+DEX structure, while retaining all other components of the assembly unchanged.

3.2.3 Post-maturation-FKBP51+DEX and Post-maturation-FKBP51−DEX complexes

The Post-maturation-FKBP51+DEX complex was modelled using the cryo-EM structure deposited under PDB entry 8FFW [72]. As in the other systems investigated, the charged linker region of Hsp90 is not fully resolved in either protomer and was therefore replaced with a shortened EKEK loop using the Crosslink Proteins utility implemented in the Schrödinger Maestro Suite (Schrödinger Release 2021-4, Schrödinger, LLC, New York, NY, USA), following the protocol described above.
Unlike the other cryo-EM structures considered in this work, PDB entry 8FFW contains a partially resolved segment of the C-terminal domain of protomer B (CTD_B); residues 724–732, which forms direct contacts with FKBP51. However, this segment is separated from the remainder of the CTD by an unresolved loop region. To obtain a continuous protein chain, the missing residues 704–723 (TADDTSAAVTEEMPPLEGDD) were reconstructed using the same crosslinking procedure, thereby connecting the resolved CTD_B fragments into a single continuous structure.
The cryo-EM model contains a threonine residue at position 519 of GR, whereas the wild-type (WT) sequence contains a serine at the corresponding position. Accordingly, the experimentally resolved residue was reverted to the WT sequence by introducing a T519S mutation. No additional modifications were made to FKBP51, which was retained as resolved in the experimental structure. The final assembly contained one ATP molecule and one Mg^{2+} ion bound within the nucleotide-binding site of each Hsp90 N-terminal domain, while dexamethasone (DEX) occupied the GR ligand-binding pocket. The corresponding Post-maturation-FKBP51−DEX model was generated by removing DEX from the Post-maturation-FKBP51+DEX structure, while retaining all other components of the assembly unchanged.

3.2.4 Post-maturation-FKBP52+DEX and Post-maturation-FKBP52−DEX complexes

The Post-maturation-FKBP52+DEX complex was modelled using the cryo-EM structure deposited under PDB entry 8FFV [72]. System preparation followed the protocol described for the Post-maturation-FKBP51+DEX complex, including reconstruction of unresolved regions and restoration of the wild-type GR sequence through the T519S mutation.

The final assembly contained one ATP molecule and one Mg^{2+} ion bound within the nucleotide-binding site of each Hsp90 N-terminal domain, while dexamethasone (DEX) occupied the GR ligand-binding pocket.
The corresponding Post-maturation-FKBP52−DEX model was generated by removing DEX from the ligand-binding pocket, while retaining all other components unchanged.

3.2.5 Preparation, simulations and analyses

All pre-processed models were prepared and simulated according to the protocol described in **Chapter 2**. The only exception concerned the Loading+DEX systems, for which a modified minimisation procedure was employed prior to the production simulations. Specifically, two additional energy minimisation steps were introduced. In the first step, hydrogen atoms, ligand atoms, and protein side chains were minimised while applying positional restraints to the protein backbone. In the second step, restraints were applied exclusively to the heavy atoms of DEX, allowing the ligand-binding pocket to undergo further structural relaxation. These steps were performed to locally minimize the structure following the IFD procedure, which involved side-chain trimming, thereby reducing the risk of numerical instabilities during the subsequent MD simulations. All initial structures used in this study are publicly available at https://doi.org/10.5281/zenodo.18267647.
Each system was simulated using multiple independent replicas of 1 µs duration. The cumulative simulation time for each system is reported in **Table 3.1**. As shown in **Table 3.1**, the different systems were simulated with different numbers of replicas. This reflects differences in the available computational resources and the sizes of the simulated complexes. In particular, because ATP hydrolysis in the immunophilin system was one of the primary focuses of this study, additional simulations were performed to obtain a statistically meaningful number of frames for the D-NEMD analysis (see below).
Structural and dynamical analyses were performed on the concatenated production trajectories obtained by combining all replicas corresponding to a given system. Trajectory stability was assessed by monitoring the C_α root mean square deviation (RMSD) of both the complete assemblies and their individual components (**Figure A.2**).
Changes in internal coordination were characterised using Distance Fluctuation (DF) analysis, as described in **Chapter 2**. Comparisons between perturbed and reference systems were quantified using the difference in distance fluctuation (ΔDF) metric. C_α root mean square fluctuations (RMSF) were calculated using the atomicfluct module implemented in CPPTRAJ, employing the equilibrated trajectories and using the average structure derived from the complete production ensemble as the reference structure [174].
Unless otherwise stated, all Cartesian distances reported in this work were calculated using Python 3 and the MDTraj analysis package.

3.2.6 D-NEMD Simulations

The D-NEMD methodology described in **Chapter 2** [40, 113–115, 168] was employed to investigate the response of the different chaperone complexes to specific external perturbations, namely ATP hydrolysis and dexamethasone (DEX) dissociation.
To model ATP hydrolysis within a classical molecular mechanics framework, a perturbation protocol was designed to reproduce the principal chemical changes associated with the reaction:

$$\mathrm{GLU47} + \mathrm{ATP} + \mathrm{H_2O} \rightarrow \mathrm{GLH47} + \mathrm{ADP} + \mathrm{HPO_4^{2-}} \tag{3.1}$$

following an approach similar to that previously applied to investigate nucleotide hydrolysis in the mitochondrial Hsp90 paralogue TRAP1 [175]. The protocol was initiated by protonation of Glu47, yielding GLH47. The reactive water molecule was identified as the water molecule closest to the ATP γ-phosphate oxygen that was not coordinated to the Mg^{2+} ion. Conversion of ATP into ADP

and inorganic phosphate (HPO_4^{2-}) was performed following the procedure described by Castelli et al. [117]. This involved cleavage of the bond between the β- and γ-phosphate groups and inversion of the phosphate tetrahedral geometry according to the direction of the cleaved bond. The resulting inorganic phosphate and protonated glutamate residues were constructed to match the equilibrium geometries of the force field.
Parameters for GLH47 were taken from the Amber ff14SB force field, whereas parameters for inorganic phosphate were obtained from the work of Kashefolgheta and Vila-Verde [176]. ATP and ADP parameters were taken from the parameter set developed by Meagher and co-workers [124].
For the DEX-removal perturbation, the perturbed structures were generated by removing dexamethasone from the GR ligand-binding pocket without introducing any additional structural modifications. Starting from configurations extracted every 5 ns from the equilibrium MD trajectories, ATP hydrolysis or DEX removal was applied instantaneously, thereby driving the systems out of equilibrium. For the ATP-hydrolysis perturbation only, a brief energy minimisation of the newly generated inorganic phosphate and GLH47 residues was performed while applying positional restraints to all remaining atoms. This additional relaxation step was necessary to relieve local strain introduced by the perturbation and to ensure numerical stability during subsequent MD simulations.
Following perturbation, each system was simulated for 20 ns, allowing the structural response of the complexes to be monitored as they relaxed towards a new equilibrium state. The resulting non-equilibrium trajectories were analysed using the Kubo–Onsager formalism [112], whereby differences between equilibrium and non-equilibrium ensembles were used to characterise the time-dependent response of the system to the applied perturbation (**Figure 2.4A**). The large number of independent replicas generated for each perturbation enabled robust estimation of the equilibrium probability distribution, $w(\Gamma)$ (**Equation 2.44**), and of the corresponding time-dependent non-equilibrium observables.
D-NEMD simulations were performed on the Post-maturation-FKBP51+DEX and Post-maturation-FKBP52+DEX systems using identical protocols. Configurations were extracted every 5 ns from the equilibrium trajectories obtained across all nine replicas and used as starting points for the non-equilibrium simulations. ATP hydrolysis was applied independently to each Hsp90 protomer, generating separate perturbation datasets corresponding to hydrolysis within protomer A and protomer B. This procedure yielded 1773 starting configurations for each protomer.
For each perturbed structure, a 20 ns non-equilibrium simulation was performed. In a fraction of the ATP-hydrolysis simulations, the abrupt perturbation generated large local forces that resulted in numerical instabilities and failure of energy convergence. Consequently, only successfully completed trajectories were included in the subsequent D-NEMD analyses. For the Post-maturation-FKBP51+DEX system, 1009 and 1114 trajectories converged successfully following ATP hydrolysis in protomers A and B, respectively. In contrast, DEX removal did not introduce significant numerical instabilities, and all simulations completed successfully.
For the Post-maturation-FKBP52+DEX system, 1254 and 1155 trajectories converged successfully following ATP hydrolysis in protomers A and B, respectively. Analysis of D-NEMD simulations were performed with the internal analysis method, as explained in **Chapter 2**.

3.2.7 Tensor of Gyration (TOG) Analysis

To characterize differences induced by ligands on the chaperone complex, we employed Tensor of Gyration (TOG) analysis framework designed to describe large-scale domain motions and inter-protein rearrangements in a reduced-dimensional representation (**Chapter 2**).
The domain selection procedure was applied independently to all proteins present in the various chaperone assemblies. Because Hsp90 and GR occur in multiple complexes, slightly different residue selections were obtained for each system. To ensure consistency across the analyses, the filtered residue sets generated for each complex were merged to produce a common domain representation applicable to all models.
The final partitioning comprised four subdomains for each Hsp90 protomer (NTD, MiD_1, MiD_2,

CTD) and a single domain corresponding to the GR core. Hsp70C was divided into nucleotide-binding (NBD) and substrate-binding (SBD) domains, whereas only the NBD of Hsp70S was considered. Hop was partitioned into three domains (Hop_1–Hop_3), p23 into its core region and C-terminal α-helix, and FKBP51/FKBP52 into four domains (FK1, FK2, TPR1 and TPR2). The residue ranges defining each domain are reported in **Table A.1**.
Principal component analysis (PCA) were calculated as described in **Chapter 2** and the resulting trajectories are available at https://doi.org/10.5281/zenodo.18267647. To compare large-scale motions between systems, distributions of inter-dummy-atom distances were analysed using overlap integral calculations (see **Chapter 2**).

3.2.8 Contact Analysis and Contact Probability Distributions

To characterise the interactions between the different protein components of the chaperone assemblies and their reorganisation following DEX-induced perturbations, residue contact analysis was performed. For each frame of every MD trajectory, pairwise residue interactions were evaluated by calculating all heavy-atom distances. Two residues were considered to be in contact when at least one heavy atom from each residue was separated by a distance of less than 4.5 Å.
Application of this procedure to the complete trajectories enabled the construction of residue–residue contact probability maps, in which each element represents the fraction of simulation frames in which a given contact was present. To identify the interactions most strongly affected by the perturbation, residue pairs exhibiting an absolute change in contact probability of at least 25% were retained for further analysis. This threshold was chosen to retain the pairs exhibiting a difference in contact probability that could be considered as significant. Indeed, the threshold splits between the distribution of low difference in contact probability (due to stochastic fluctuation around the same equilibrium position) and the distribution of high difference (probably due to structural rearrangements caused by the perturbation).
For visualisation purposes, significant contact changes were projected onto the corresponding protein structure by placing a sphere at the midpoint between the C_α atoms of the interacting residue pair. This representation provides an intuitive description of the regions most affected by perturbation-induced contact rearrangements.

3.3 Results

To investigate the molecular determinants governing GR activation, we analysed both the original cryo-EM assemblies and strategically perturbed models in which the steroid ligand dexamethasone (DEX) was either removed from, or introduced into, the GR ligand-binding pocket. Specifically, DEX was removed from complexes originally resolved in the ligand-bound state, whereas ligand-bound models were generated for complexes whose cryo-EM structures were determined in the absence of DEX. These perturbations were designed to probe how ligand occupancy influences the conformational landscape of GR and, consequently, modulates the dynamics and organisation of the associated multiprotein assemblies.
The biological relevance of ligand-free states is supported by in vivo cross-linking experiments demonstrating that GR–FKBP51 and GR–FKBP52 interactions persist in the absence of ligand and are largely indistinguishable from those observed in cryo-EM structures [177]. Furthermore, available cryo-EM data suggest that apo assemblies are sampled within the experimental ensembles, although they appear to exhibit reduced structural order compared with fully ligand-bound complexes [70–72]. Together, these observations support the exploration of both ligand-bound and ligand-free states as physiologically relevant conformations of the chaperone machinery.
Given the number of systems investigated and the diversity of their functional states, the results are first presented through pairwise comparisons of individual complexes, followed by a broader discussion of the mechanistic implications emerging from the collective analyses. Accordingly, we

begin by comparing each complex in its experimentally resolved state with its corresponding DEX-perturbed counterpart. For the Loading complex, this involves comparison of the original ligand-free assembly with models in which DEX was introduced into the GR binding pocket. Conversely, for the Maturation+DEX, Post-maturation-FKBP51+DEX, and Post-maturation-FKBP52+DEX complexes, DEX was removed from the receptor and the resulting trajectories were compared with those of the corresponding ligand-bound assemblies. Finally, the Post-maturation-FKBP51+DEX and Post-maturation-FKBP52+DEX complexes are compared directly to investigate the molecular basis underlying the distinct functional roles of these highly homologous cochaperones.
The presence or absence of DEX constitutes a controlled perturbation aimed at elucidating the relationship between ligand occupancy and the dynamic regulation of the GR-processing machinery. We hypothesise that, despite its relatively small size compared with the complete assembly, the steroid exerts a substantial influence on the conformational dynamics of the chaperone complexes. Comparative analysis of the resulting trajectories therefore provides insights into the long-range communication pathways and dynamic cross-talk that coordinate the functional behaviour of the different components of the maturation cycle.
For each pair of systems, we first analyse the Distance Fluctuation (DF) matrices, which identify regions of the assemblies displaying mechanically coordinated motions (**Chapter 2**). In general, low DF values between distal residues or domains, including residues belonging to different proteins, are indicative of coordinated dynamical behaviour and potential allosteric communication. DF analysis is therefore used to identify both conserved and perturbation-dependent networks of dynamic interactions that may underlie the functional motions of the various complexes.
DF analysis is complemented by root mean square fluctuation (RMSF) calculations to characterise local flexibility. To investigate collective and large-scale conformational changes, these analyses are further integrated with the Tensor of Gyration (TOG) framework introduced in **Chapter 2**, together with residue–residue contact probability analysis.
Finally, equilibrium simulation data are combined with results from dynamical non-equilibrium molecular dynamics (D-NEMD) simulations. D-NEMD enables characterisation of the propagation of structural responses following defined perturbations, thereby providing complementary information on allosteric communication within the complexes. Two perturbation schemes are considered where applicable: ATP hydrolysis within either Hsp90 protomer and DEX removal from the GR ligand-binding pocket. Comparison of the resulting responses provides a direct means of assessing how local chemical events are transmitted through the multiprotein assemblies.
For consistency, the presentation of the results follows the same general structure for all systems examined: (i) effects on GR, (ii) effects on the GR–Hsp90 dynamic network, (iii) effects on Hsp90, and (iv) effects on the cochaperone network.

3.3.1 Loading−DEX vs. Loading+DEX

In the cryo-EM structure of the Loading complex (L−DEX; **Figure 3.1B**), GR is captured in a ligand-free state. To investigate the consequences of steroid recognition at this stage of the chaperone cycle, a DEX-bound model (L+DEX) was generated using an induced-fit docking (IFD) protocol (see above). The rationale behind this perturbation was to assess whether ligand binding could influence the dynamic organisation of the Loading assembly and potentially facilitate progression towards subsequent maturation states.
Notably, the IFD procedure identified receptor conformations within the Loading ensemble that closely resemble those sampled by ligand-bound GR in the Maturation complex. In particular, several conformations of the folded GR core exhibited backbone RMSD values of approximately 1.7 Å relative to the mature ligand-bound state. Although qualitative, this observation suggests that partially folded, ligand-competent conformations are already accessible within the Loading ensemble and may therefore be stabilised by steroid binding, reshaping interactions in the chaperone machinery.
Comparison of the DF matrices reveals that DEX binding substantially remodels the internal dy-

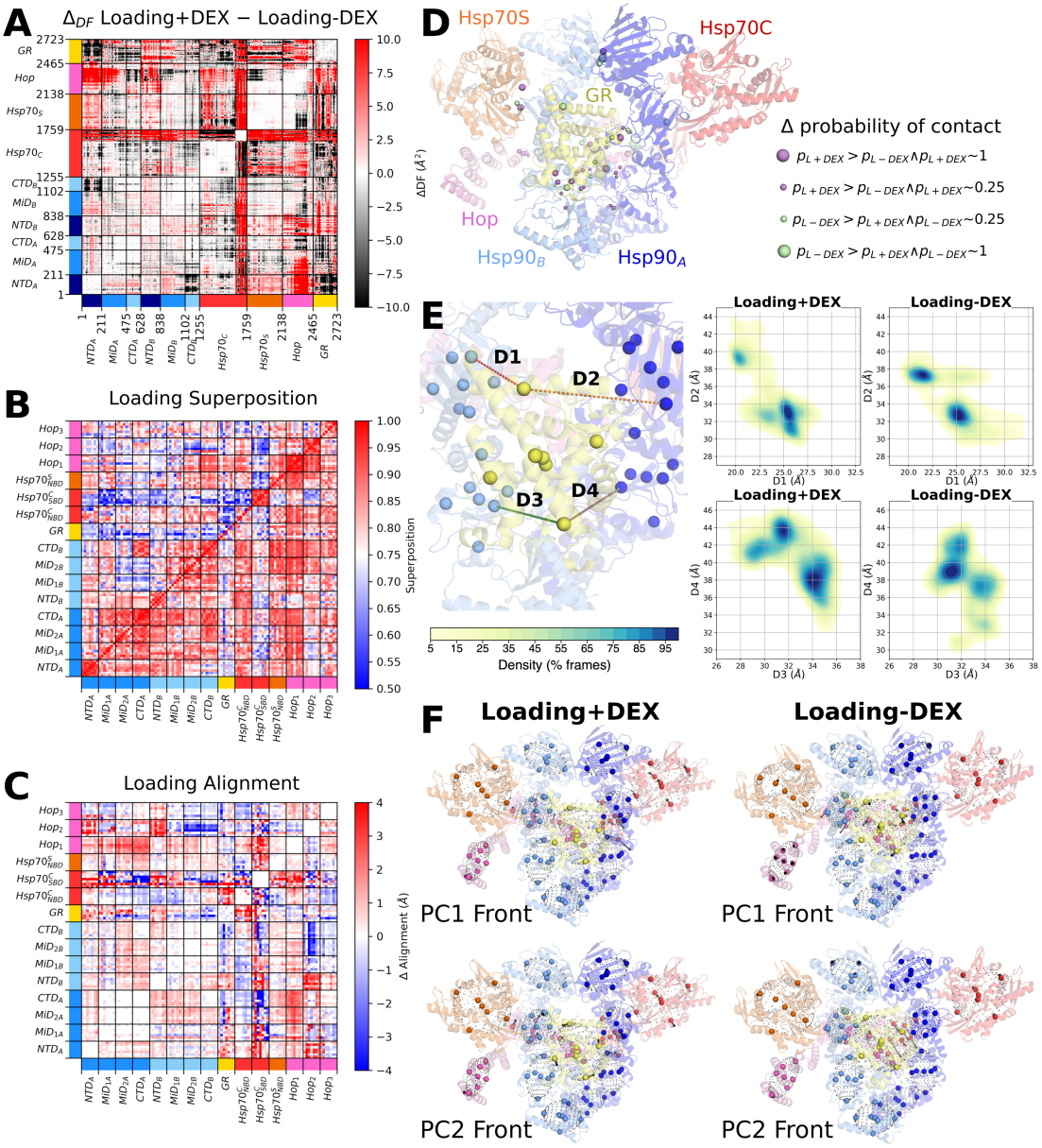

Figure 3.2: Dynamic analysis of the Loading complex. ***(A)*** *ΔDF matrix obtained by subtracting equilibrium matrix (L−DEX) from the perturbed one (L+DEX). In black regions DF value is higher in the equilibrium matrix (coordination gain after DEX insertion), while red areas mean the opposite.* ***(B)*** *In superposition map the overlap integral values for each pairwise distance distribution, upper diagonal depicts values before superposition, while the lower represents the values after superposition. Red values indicate a similar dynamical behaviour between the two dynamics, while white/blue values underlie a marked dynamical difference.* ***(C)*** *Alignment map represents the Δ value pertaining to the translation with the highest degree of overlap for each reduced point pairwise distance distribution between trajectories. Blue spots ($\Delta < 0$ Å) mean that the perturbation has increased the distance between reduced points, while red spots ($\Delta > 0$ Å) mean the opposite.* ***(D)*** *Projection of differences in probability of contact onto the Loading complex structure. Spheres indicate statistically significant modified interaction between residue pairs in the unperturbed/perturbed contact matrices.* ***(E)*** *On the right, 2D histograms of two key distance pairs of TOG reduced points (D1-D2 and D3-D4); On the left, their representation on the structure of the Loading complex.* ***(F)*** *Projection of the first two PCA eigenvectors onto structures of the Loading complexes, together with TOG points.*

namic organisation of the complex (**Figure 3.2A**). While the overall block structure of the matrices remains largely preserved, reflecting the domain architecture of the constituent proteins, the ΔDF analysis highlights significant changes in the coordination network. The most pronounced effects occur within GR itself. Upon ligand binding, the folded receptor core becomes more mechanically coordinated, whereas the unfolded tail region displays increased dynamical decoupling from the remainder of the protein. This behaviour is consistent with RMSF analysis, which indicates a general rigidification of the receptor upon DEX binding (**Figure A.3A,B**).

Importantly, DEX insertion causes visible modifications in the TOG superposition map and alignment (**Figure 3.2B,C**), where, even after the alignment procedure, intra-protein points display distinct behaviours within GR.

The effects of ligand binding extend beyond GR and propagate to its interactions with Hsp90. DEX induces enhanced coordination between GR and protomer A, whereas coordination with most regions of protomer B decreases, with the notable exception of CTD_B (**Figure 3.2A**). Consistent with these findings, contact analysis reveals substantial remodelling of the GR–Hsp90 interface (**Figure 3.2D** and **Figure A.4**). Several contacts involving the upper region of MiD_B are weakened (between Q404 of Hsp90 and K703 of GR), whereas new interactions emerge at the interfaces between GR and both protomers, particularly in regions surrounding the MiD_A domain (in the region 319-355 of Hsp90A with 580-590 of GR) and the CTD_B–GR interface (whole CTD_B domain GR tail and loop around residue 625). The agreement between contact and DF analyses indicates that ligand binding alters both the structural and dynamical coupling between client and chaperone.

These changes are accompanied by large-scale rearrangements of the Loading assembly. TOG analysis reveals a global tendency towards a clamp-like closure of the Hsp90 dimer following DEX binding, characterised by a reduction in inter-domain distances between the middle and C-terminal regions of the two protomers (**Figure 3.2B,C**). Contact analysis supports this observation, showing the emergence of interactions consistent with the initial stages of protomer closure. The widespread reduction in overlap between TOG distance distributions further indicates that DEX substantially alters the collective dynamics of the complex.

To characterise these rearrangements in greater detail, distances between the centre of mass of GR and the Hsp90 middle domains were analysed (**Figure A.5A–C**). In the absence of ligand, GR occupies a relatively restricted conformational space with respect to the Hsp90 lumen. Upon DEX binding, this distribution broadens considerably, indicating enhanced mobility of the receptor. At the same time, GR shifts towards conformations in which it is positioned closer to Hsp90, particularly to MiD_A. Two-dimensional distance distributions reveal that this transition is associated with a rotational displacement of the receptor relative to the middle domains, enabling exploration of conformations not sampled in the ligand-free ensemble (**Figure 3.2E**).

Ligand-induced remodelling also affects the cochaperones responsible for client loading. In the absence of DEX, strong dynamical coupling is observed between GR, the substrate-binding domain of Hsp70C ($Hsp70C_{SBD}$), and Hop (**Figure 3.2A**). Upon ligand binding, this coordination is substantially weakened. The effect is particularly evident for $Hsp70C_{SBD}$, whose dynamical behaviour differs markedly between the two systems, and for the Hop_1 domain, which undergoes extensive reorganisation of its contact network. In the ligand-free state, Hop_1 forms interactions with the GR tail region and maintains stronger coupling with Hsp90 and Hsp70C. Following DEX binding, these interactions are partially lost and replaced by increased contacts with MiD_B, accompanied by a displacement of Hop_1 towards the Hsp90 lumen.

Principal component analysis of the TOG trajectories further supports the existence of ligand-dependent collective motions (**Figure 3.2F** and **Figure A.6**). Movies for the first and second principal components (PC1 and PC2) are provided at the link: https://doi.org/10.5281/zenodo.18267647. In the L−DEX ensemble, the dominant motions involve extensive rearrangements of Hsp70C and Hop_2, consistent with an active loading platform. In the L+DEX complex, the movement of Hsp70CNBD disappears, replaced by the introduction of a rotational movement of GR around its disordered tail axis. In PC2, this rotational movement is also observed in the L-DEX complex,

together with some sparse motions on Hsp90A NTD and Hsp70CNBD. Again, Hop_2 domain displays wide motions. In the L+DEX complex, instead, larger movements involve only $Hsp70C_{SBD}$.
Overall, the results indicate that DEX binding triggers long-range remodelling of the dynamic communication network within the Loading complex. The perturbation propagates from the GR ligand-binding pocket to distal interfaces involving Hsp90, Hsp70C and Hop, altering both local interactions and global collective motions. Particularly striking is the loss of coordination between GR and the $Hsp70C_{SBD}$/Hop module following ligand binding. Given the central role of these cochaperones in stabilising the Loading complex, their dynamical decoupling may represent an early step in the transition towards the Maturation state. More generally, these findings suggest that steroid binding is not merely a consequence of receptor maturation, but may actively contribute to directing progression through the Hsp90 chaperone cycle by reshaping the allosteric communication network of the assembly.

3.3.2 Maturation+DEX vs. Maturation−DEX

Unlike the Loading complex, the cryo-EM structure of the Maturation+DEX assembly captures GR in a ligand-bound, near-native conformation. To investigate the contribution of ligand occupancy to the mature chaperone state, DEX was removed from the receptor binding pocket and the resulting trajectories were compared with those of the unperturbed Maturation+DEX system. The DF matrices of the Maturation+DEX and Maturation−DEX systems retain the overall block-like organisation expected from the domain architecture of the constituent proteins, indicating that the global structural framework of the assembly remains largely intact. However, analysis of the ΔDF matrix reveals a measurable reduction in the internal coordination of GR following ligand removal, manifested as a mild increase in pairwise distance fluctuations throughout the receptor (**Figure 3.3A**). RMSF analysis is consistent with this interpretation, showing a modest increase in receptor flexibility in the absence of DEX (**Figure A.3**). Despite these dynamical changes, TOG analysis indicates that the overall three-dimensional organisation of GR is largely preserved, with strong overlap between the ligand-bound and ligand-free ensembles for intra-GR distances (**Figure 3.3B,C**). The primary effects of ligand removal are observed in the interaction network connecting GR to Hsp90. Upon DEX removal, GR loses coordination with most regions of the Hsp90 dimer, while retaining or slightly increasing its coupling to NTD_B and MiD_B **Figure 3.3A**). Contact analysis reveals substantial remodelling of the GR–Hsp90 interfaces (**Figure 3.3D**). In particular, the ligand-free state shows a general reduction in contacts between the GR terminal tail and luminal residues of Hsp90, together with pronounced reorganisation of the GR–Hsp90B interface. One region involves the partially folded α-helical segment of the GR tail (residues 533–539) and two portions of the Hsp90B client-binding surface near CTD_B and MiD_B (residues 530–534 and 605–637). Consistent with these contact changes, TOG analysis indicates a global displacement of GR away from CTD_B upon ligand removal **Figure 3.3C**).
A second remodelling event is observed at the interface between residues 453–454 of Hsp90B and residues 699–703 of GR. In this region, DEX removal increases the contact probability between GR and MiD_{1B}, and TOG analysis reveals a corresponding shift of GR towards the MiD_{1B}/NTD_B region. Overall, the ligand-free receptor adopts conformations that are positioned closer to the Hsp90 N-terminal side of the lumen and farther from CTD_B, accompanied by a tighter and less heterogeneous GR–MiD_B distance distribution (**Figure A.5C**).
The perturbation also affects communication between the two Hsp90 protomers. The ΔDF matrix shows widespread changes in inter-protomer coordination, and multiple contact probabilities are altered throughout the dimer (**Figure 3.3A,D**). Nevertheless, TOG analysis detects only minor changes in the relative positions of the major Hsp90 subdomains, indicating that the globally closed conformation of Hsp90 remains largely preserved.
Thus, the mature chaperone stays in an activated closed state even after DEX removal. The cochaperone p23 is also sensitive to ligand occupancy. In the absence of DEX, coordination between

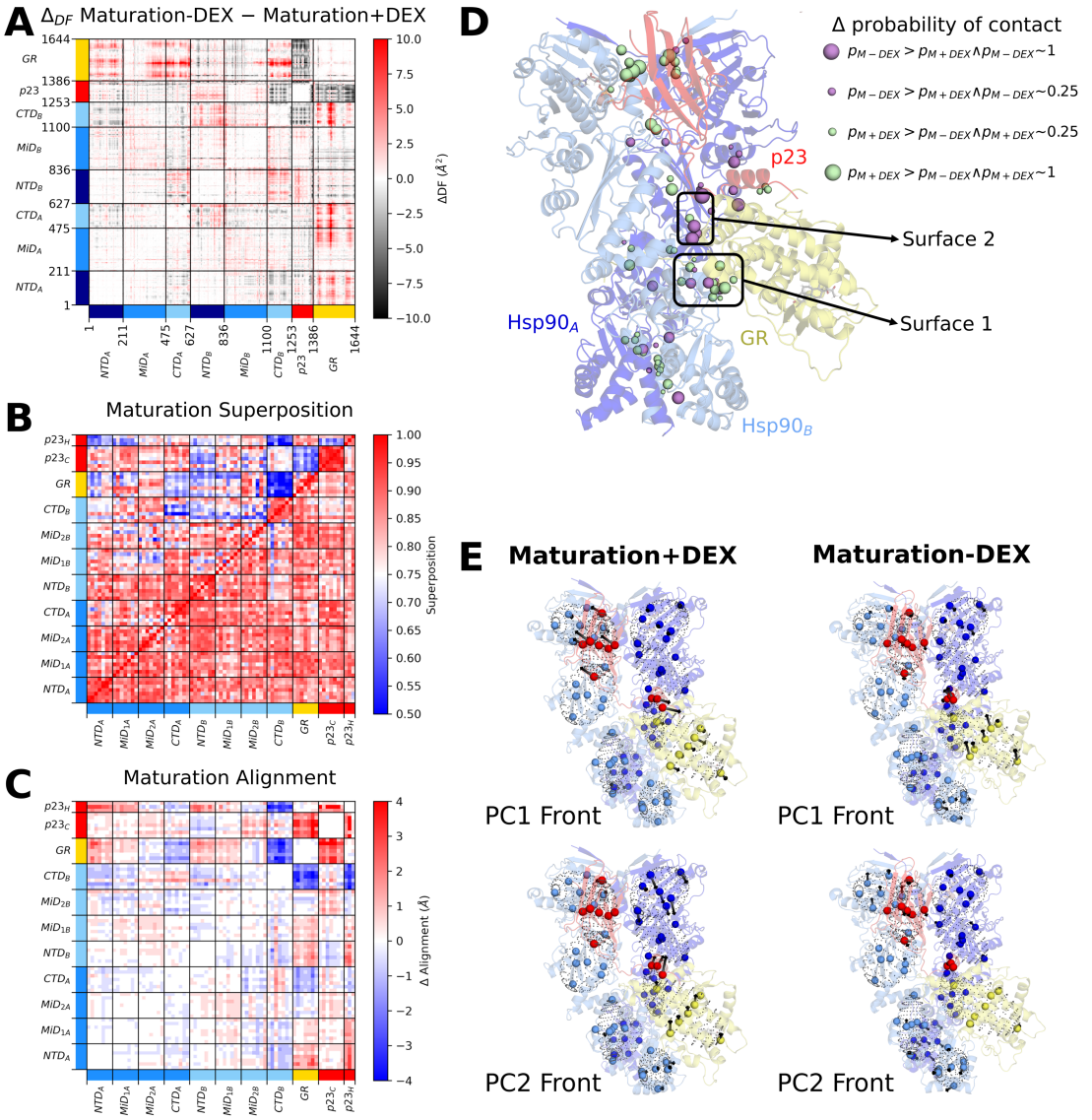

Figure 3.3: Dynamic analysis of the Maturation complex. ***(A)*** *ΔDF matrix obtained by subtracting equilibrium matrix (M+DEX) from the perturbed one (M−DEX).* ***(B)*** *In superposition map the overlap integral values for each pairwise distance distribution, upper diagonal depicts values before superposition, while the lower represents the values after superposition. Red values indicate a similar dynamical behaviour between the two dynamics, while white/blue values underlie a marked dynamical difference.* ***(C)*** *Alignment map represents the Δ value pertaining to the translation with the highest degree of overlap for each reduced point pairwise distance distribution between trajectories. Blue spots ($\Delta < 0$ Å) mean that the perturbation has increased the distance between reduced points, while red spots ($\Delta > 0$ Å) mean the opposite.* ***(D)*** *Projection of differences in probability of contact onto the Maturation complex structure. Spheres indicate statistically significant modified interaction between residue pairs in the unperturbed/perturbed contact matrices. Black rectangles depict the points corresponding to the surfaces discussed in the main text.* ***(E)*** *Projection of the first two PCA eigenvectors onto structures of the Maturation complexes, together with TOG points.*

p23 and the directly associated NTD_B decreases, accompanied by a reduction in p23–NTD contact frequency and a TOG-detected displacement of the p23 core away from NTD_B (**Figure 3.3A,C,D**). In contrast, coordination between p23 and GR increases. This is reflected by closer TOG distances between the GR and p23 cores and by enhanced contacts between GR and the p23 α-helical tail. Notably, the GR loop residing in the Hsp90 lumen and the p23 α-helical tail become more dynamically coupled to the assembly in the ligand-free state, whereas both regions are comparatively decoupled in the DEX-bound complex.
Principal component analysis of the TOG trajectories supports these observations (**Figure 3.3E** and **Figure A.8**). The dominant motion in the Maturation−DEX ensemble involves a tilting of GR towards the Hsp90 N-terminal region, which propagates to the p23 α-helical tail and, to a lesser extent, to the p23 core and Hsp90A-NTD.
Instead, the dominant motion of the ligand-bound complex is characterised by displacement of GR towards the Hsp90 CTDs together with partial detachment of p23. Thus, DEX occupancy appears to bias the mature assembly towards a CTD-oriented configuration, whereas ligand removal favours an NTD-oriented state. Taken together, these results indicate that DEX contributes primarily to the stabilisation and organisation of the mature GR–Hsp90–p23 assembly rather than to maintenance of the receptor fold itself. Ligand removal leaves the overall closed architecture of Hsp90 intact but weakens the dynamic coupling between p23 and Hsp90 while strengthening p23–GR interactions. This redistribution of interactions suggests that steroid occupancy modulates the balance of forces stabilising the maturation complex and may therefore influence the lifetime and functional competence of the active GR–chaperone assembly.

3.3.3 FKBP51+DEX vs. FKBP51−DEX

As in the Maturation+DEX complex, GR is resolved in a ligand-bound state in the cryo-EM structure of the Post-maturation-FKBP51+DEX assembly. To assess the contribution of ligand occupancy to the stability and dynamics of this late-stage complex, DEX was removed from the GR binding pocket and the resulting trajectories were compared with those of the unperturbed system.
Analysis of the ΔDF matrix indicates that ligand removal remodels the internal coordination network of GR, resulting in a modest but widespread reduction in mechanical coupling throughout the receptor (**Figure 3.4A**). Conversely, DEX binding promotes a more coordinated dynamical state, consistent with the increased rigidity observed in the RMSF profiles (**Figure A.3A,B**). Despite these differences in internal dynamics, TOG analysis reveals only minor changes in the overall three-dimensional organisation of GR, indicating that the perturbation primarily affects dynamical coupling rather than the global receptor architecture (**Figure 3.4B,C**).
The effects of DEX removal extend to the GR–Hsp90 interface. ΔDF analysis reveals a general loss of coordination between GR and the chaperone, which is accompanied by substantial remodelling of the contact network (**Figure 3.4A,D**). Two interaction surfaces are particularly affected. The first is located within the Hsp90 lumen, where contacts between the GR tail loop and luminal residues become less frequent in the absence of ligand. The second involves the interface between the GR core and the MiD_{1A} domain of Hsp90, where ligand removal increases both the number and persistence of contacts. Together with the TOG results, these observations indicate a rigid-body displacement of GR towards the N-terminal and MiD_1 regions of Hsp90, accompanied by partial disengagement from the CTD and MiD_2 regions (**Figure 3.4C**).
The perturbation also influences the internal dynamics of Hsp90. Although the overall pattern is heterogeneous, changes in coordination are particularly evident around the nucleotide-binding sites of both protomers (**Figure 3.4A**). These local rearrangements are reflected by corresponding modifications in contact probabilities and TOG-derived distance distributions, indicating that the effects of ligand removal propagate throughout the chaperone scaffold.
The most pronounced response, however, is observed for FKBP51. In the absence of DEX, the immunophilin displays a marked reduction in mechanical coordination with virtually all components of the assembly (**Figure 3.4A**). This loss of coupling is accompanied by a decrease in GR–FKBP51

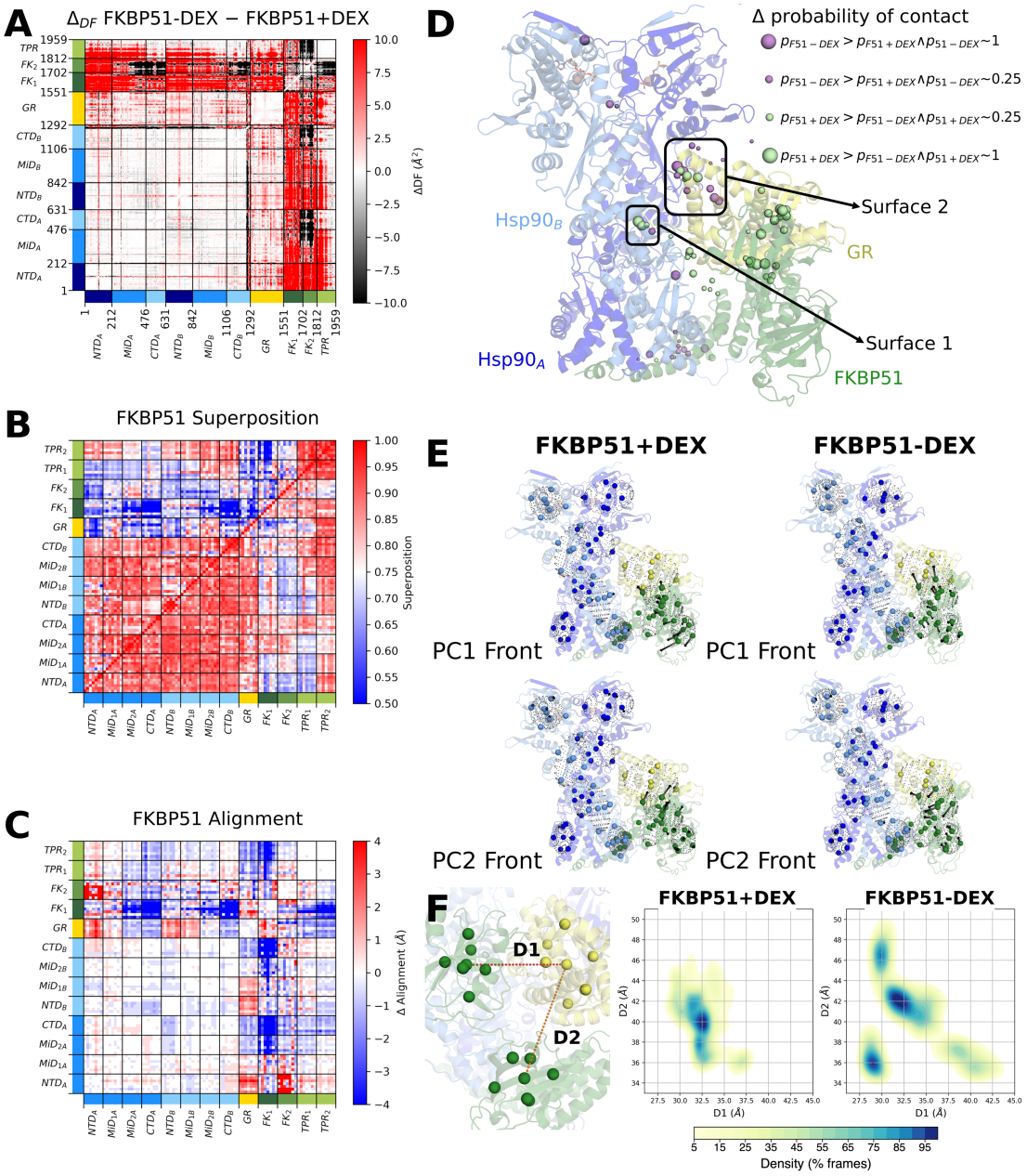

Figure 3.4: Dynamic analysis of the Post-maturation-FKBP51 complex. ***(A)*** *ΔDF matrix obtained by subtracting equilibrium matrix (F51+DEX) from the perturbed one (F51−DEX).* ***(B)*** *In superposition map the overlap integral values for each pairwise distance distribution, upper diagonal depicts values before superposition, while the lower represents the values after superposition. Red values indicate a similar dynamical behaviour between the two dynamics, while white/blue values underlie a marked dynamical difference.* ***(C)*** *Alignment map represents the Δ value pertaining to the translation with the highest degree of overlap for each reduced point pairwise distance distribution between trajectories. Blue spots ($\Delta < 0$ Å) mean that the perturbation has increased the distance between reduced points, while red spots ($\Delta > 0$ Å) mean the opposite.* ***(D)*** *Projection of differences in probability of contact onto the FKBP51 complex structure. Spheres indicate statistically significant modified interaction between residue pairs in the unperturbed/perturbed contact matrices.* ***(E)*** *Projection of the first two PCA eigenvectors onto structures of the FKBP51 complexes, together with TOG points.* ***(F)*** *On the right, KDE plot of distances between the centre of mass of GR core and centres of mass of FK1 (D1) and FK2 (D2) domains respectively. On the left, the representation of the investigated distances onto the structure of FKBP51 complex.*

contact probability and by substantial changes in the dynamical behaviour of all FKBP51 domains (**Figure 3.4B,D**). TOG analysis reveals poor overlap between the ligand-bound and ligand-free ensembles, particularly for the FK1 and FK2 domains, indicating extensive remodelling of immunophilin dynamics. More generally, FKBP51 exhibits a tendency to move away from the Hsp90–GR complex, accompanied by enhanced fluctuations and partial loss of intermolecular contacts.

The interaction between FKBP51 and Hsp90 is likewise affected. Although contacts involving the CTD_B C-terminal loop and the TPR2 domain remain stable or become slightly more frequent, the overall contact network linking the immunophilin to CTD_B is weakened (**Figure 3.4D**). Consistent with this behaviour, the TPR2 domain displays the highest degree of structural conservation among FKBP51 domains, whereas the remaining domains undergo substantial displacements relative to the rest of the assembly. These observations suggest that TPR2 acts as a persistent anchoring point while the remainder of the immunophilin becomes increasingly decoupled from the complex.

Principal component analysis of the reduced TOG trajectories further supports this interpretation (**Figure 3.4E** and **Figure A.9**). In the ligand-free ensemble, the dominant collective motions are centred on FKBP51 and involve coordinated displacements of the FK1, FK2 and TPR1 domains. These motions are largely absent in the ligand-bound simulations and indicate extensive reorganisation of the interaction network connecting FKBP51 to GR and Hsp90. The consequences of this remodelling are particularly evident in the two-dimensional KDE representation of the distances between GR and the FK1/FK2 domains (**Figure 3.4F**). Whereas the ligand-bound system samples a relatively restricted conformational ensemble, DEX removal substantially broadens the distribution and gives rise to additional conformational states, reflecting increased mobility and reduced coupling between GR and the immunophilin.

Overall, these results demonstrate that DEX contributes significantly to the stabilisation of the Post-maturation-FKBP51 assembly. Ligand removal weakens the dynamic coupling between FKBP51, GR and Hsp90, promotes enhanced immunophilin mobility, and favours partial disengagement of FKBP51 from the complex. Notably, this increased conformational heterogeneity is consistent with the lower structural order observed for apo GR:Hsp90:FKBP51 particles in cryo-EM datasets. Together, these findings suggest that ligand occupancy plays an important role in maintaining the structural and dynamical integrity of the FKBP51-bound post-maturation complex.

Next, D-NEMD simulations were employed to characterise the dynamical response of the complex to ATP hydrolysis and, independently, to DEX removal (**Figure 3.5**).

The internal D-NEMD analysis indicates that both perturbations induce relatively localised structural responses. In the case of ATP hydrolysis, the largest effects are confined to the vicinity of the nucleotide-binding site, where the perturbation promotes active-site remodelling consistent with the early stages of phosphate release (**Figure 3.5A-C**). Notably, the two Hsp90 protomers exhibit distinct responses to hydrolysis. Hydrolysis within protomer A produces a relatively limited structural rearrangement, primarily affecting the active-site loop (residues 113–133), which undergoes a subtle displacement towards the core of the NTD (**Figure 3.5A**). In contrast, hydrolysis within protomer B elicits a broader response involving not only the corresponding active-site loop but also the neighbouring helix spanning residues 54–64 and the modelled charged-linker region around residue 220 (**Figure 3.5D-F**). These motions collectively result in partial opening of the nucleotide-binding pocket, which subsequently propagates to the charged linker, suggesting a more extensive redistribution of local strain following ATP hydrolysis in protomer B.

Removal of DEX from the GR ligand-binding pocket produces a distinct response. The perturbation rapidly induces closure of the ligand-binding site, primarily through displacement of the loop comprising residues 635–640 (**Figure 3.5G-I**). Compared with ATP hydrolysis, the structural response associated with DEX removal is both faster and larger in magnitude. This behaviour is evident from the corresponding distance profiles, which display a pronounced initial exponential variation following ligand removal (**Figure 3.5C,F,I**). These observations indicate that the GR ligand-binding pocket is highly responsive to changes in ligand occupancy and undergoes rapid local reorganisation upon steroid dissociation, whereas the nucleotide-binding sites of Hsp90 exhibit a

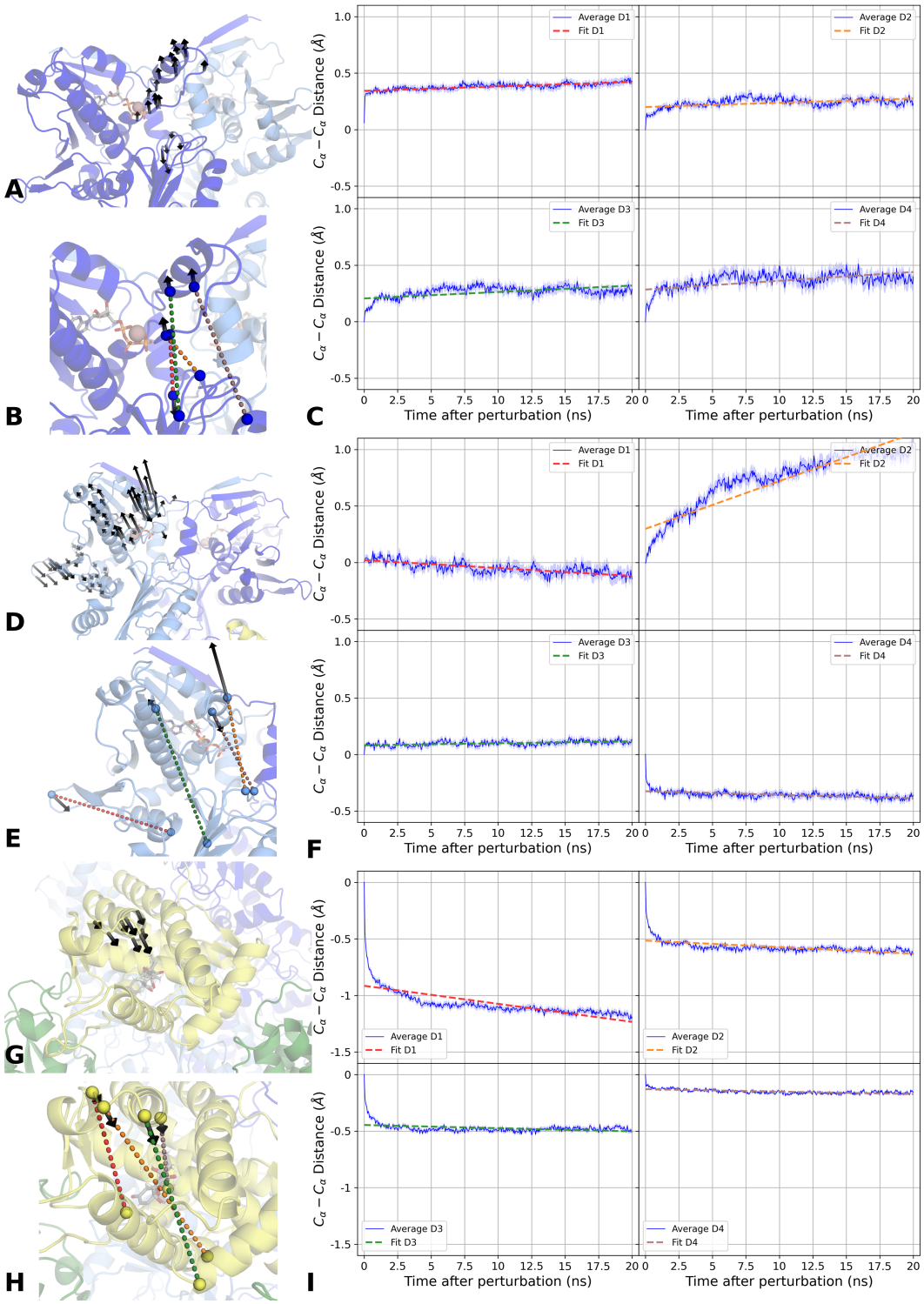

Figure 3.5: Nonequilibrium dynamics of the FKBP51 complex. *D-NEMD responses to ATP hydrolysis within the* $\mathrm{NTD_A}$, $\mathrm{NTD_B}$ *or DEX removal the FKBP51+DEX complex (panels* ***(A)***, ***(D)***, ***(G)*** *respectively) after 20* ns *from the perturbation instant. Arrows represent the extent and the direction of residues affected by the perturbation. In* ***(B)***, ***(E)***, ***(H)*** *panels, the region of interested is highlighted, with 4 affected and non-affected residue pairs. The direction of these pairs is chosen to be along the one found with internal D-NEMD analysis method. In* ***(C)***, ***(F)***, ***(I)*** *panels the averaged value of the distance differences between equilibrium and non-equilibrium dynamics are plotted as function of time after perturbation, together with linear fit.*

more gradual structural adaptation following ATP hydrolysis.

3.3.4 FKBP52+DEX vs. FKBP52−DEX

The Post-maturation-FKBP52+DEX complex represents an alternative branch of the GR maturation pathway relative to the FKBP51-containing assembly. In the experimentally resolved structure, dexamethasone (DEX) is bound within the GR ligand-binding pocket. To investigate the contribution of ligand occupancy to the organization and dynamics of the complex, we compared equilibrium simulations of the native DEX-bound assembly with simulations in which the ligand was removed. Unlike the FKBP51-containing system, DEX removal from the FKBP52 complex primarily affects the internal dynamics of GR rather than its global architecture. The ΔDF matrix reveals a reduction in internal mechanical coordination throughout the receptor, indicating that ligand binding contributes to stabilizing coordinated motions within the GR core (**Figure 3.6A**). This trend is supported by RMSF analysis, which shows enhanced local flexibility in the absence of DEX (**Figure A.3A,B**). In contrast, TOG analysis detects only minor changes in the relative positioning of GR reduced points (**Figure 3.6A-C**), suggesting that ligand removal perturbs internal coordination without inducing substantial large-scale conformational rearrangements.

The GR–Hsp90 interaction network is only modestly affected by the perturbation. The most evident changes occur within the Hsp90 lumen, where the GR tail reorganizes its interaction pattern with the chaperone (**Figure 3.6D**). Additional contact remodelling is observed between the GR core and two residues of the CTD_B domain (residues 602 and 624), whose interaction probabilities increase following ligand removal. Despite these local rearrangements, both TOG and ΔDF analyses indicate that the overall GR–Hsp90 interface remains largely preserved (**Figure 3.6A-C**). Similarly, the coordination patterns between the two Hsp90 protomers show only minor alterations, consistent with the limited differences observed in contact probability and reduced-point distance distributions (**Figure 3.6C-D**).

More pronounced effects emerge in the interaction network involving FKBP52. In the DEX-bound state, GR and FKBP52 exhibit weaker coordination than in the ligand-free system. Upon DEX removal, the ΔDF matrix reveals an overall strengthening of the mechanical coupling between the receptor and the FK domains of the immunophilin (**Figure 3.6A**). Contact analysis identifies four major regions of interaction remodelling, two involving GR–FKBP52 interfaces and two involving contacts between FKBP52 and the Hsp90 CTD_B loop (**Figure 3.6D**).

The first GR–FKBP52 interface is located between the receptor and the FK1 domain. Here, DEX removal produces a generalized increase in contact probability, accompanied by a reduction in the corresponding TOG distances and an increase in mechanical coordination (**Figure 3.6A-C**). A second interaction surface involves the loop connecting the FK1 and FK2 domains and the GR core. Strengthening of these contacts brings the FK2 domain closer to the receptor and further reinforces the GR–FKBP52 interaction network.

In contrast, the TPR domains progressively disengage from Hsp90 following ligand removal. Contact analysis shows a decrease in interactions between the terminal helix of TPR2 and the lower portion of CTD_B, accompanied by the formation of alternative contacts with the terminal segment of the CTD_B loop. Consistently, TOG analysis indicates a displacement of both TPR domains away from the MiD_2 and CTD regions of Hsp90, while ΔDF highlights a generalized loss of coordination between these domains and the rest of the assembly (**Figure 3.6A-C**).

Interestingly, although DEX binding stabilizes the internal organization of GR, it is associated with weaker coordination between the receptor and FKBP52. The ligand-bound simulations display reduced GR–FKBP52 mechanical coupling and lower internal coordination within FKBP52 itself, excluding the TPR region (**Figure 3.6A**). At the same time, TOG analysis reveals a tendency for the FK domains to remain closer to the Hsp90 NTDs in the presence of DEX, indicating a distinct dynamical organization of the complex.

Principal component analysis of the TOG-reduced trajectories further highlights the effect of

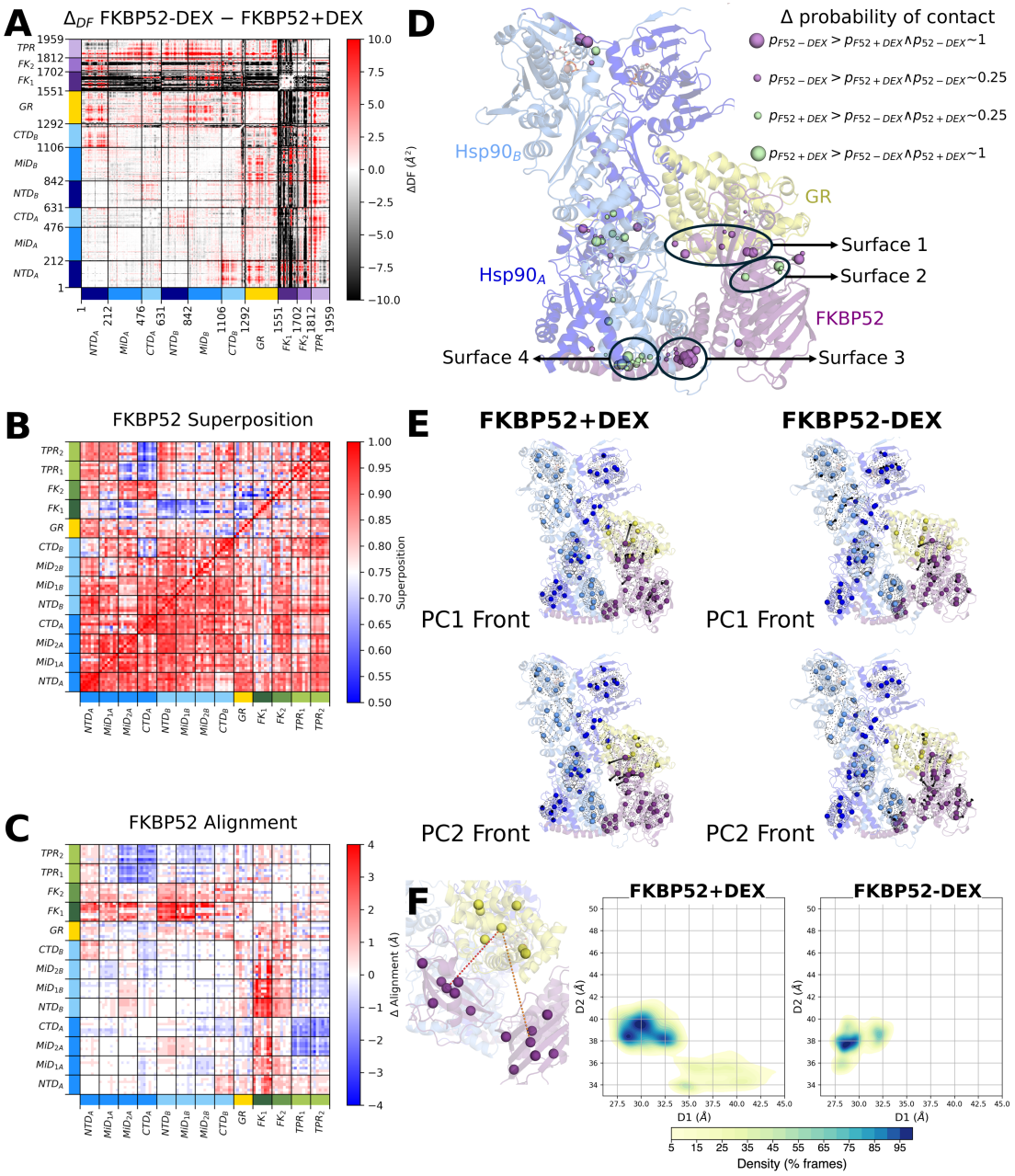

Figure 3.6: Dynamic analysis of the Post-maturation-FKBP52 complex. ***(A)*** *ΔDF matrix obtained by subtracting equilibrium matrix (F52+DEX) from the perturbed one (F52−DEX).* ***(B)*** *In superposition map the overlap integral values for each pairwise distance distribution, upper diagonal depicts values before superposition, while the lower represents the values after superposition. Red values indicate a similar dynamical behaviour between the two dynamics, while white/blue values underlie a marked dynamical difference.* ***(C)*** *Alignment map represents the Δ value pertaining to the translation with the highest degree of overlap for each reduced point pairwise distance distribution between trajectories. Blue spots ($\Delta < 0$ Å) mean that the perturbation has increased the distance between reduced points, while red spots ($\Delta > 0$ Å) mean the opposite.* ***(D)*** *Projection of differences in probability of contact onto the FKBP52 complex structure. Spheres indicate statistically significant modified interaction between residue pairs in the unperturbed/perturbed contact matrices.* ***(E)*** *Projection of the first two PCA eigenvectors onto structures of the Loading complexes, together with TOG points.* ***(F)*** *On the right, the KDE plot of distances between the centre of mass of GR core and centres of mass of FK1 (D1) and FK2 (D2) domains respectively. On the left, the representation of the investigated distances onto the structure of the Post-maturation-FKBP52 complex.*

ligand occupancy on collective motions (**Figure 3.6E** and **Figure A.10**). The dominant principal components differ substantially between the DEX-bound and ligand-free systems. In the native complex, the FK1 domain predominantly moves toward GR, whereas in the absence of DEX it undergoes a rotational motion around the receptor. Furthermore, the ligand-free simulations display a concerted closing movement involving NTD_A and MiD_{1B}, a feature absent in the DEX-bound state. The second principal component of the DEX-bound system resembles the dominant motion observed in the ligand-free simulations, while PC2 of the latter is characterized by more dispersed fluctuations involving the FK and TPR1 domains.

The overall effect of ligand removal is summarized by the two-dimensional KDE analysis of the distances between the GR centre of mass and those of the FK1 and FK2 domains (**Figure 3.6F**). The increased coordination observed in the ΔDF and contact analyses translates into a progressive closure of the FK domains around the receptor. As a consequence, FKBP52 adopts a more constrained arrangement relative to GR, reducing its overall mobility while strengthening its association with the client.

Taken together, these results indicate that DEX occupancy exerts only a limited influence on the global architecture of the Post-maturation-FKBP52 complex. Instead, ligand removal selectively reinforces the interaction network between GR and the FK domains of FKBP52 while weakening the association of the TPR region with Hsp90. This behaviour contrasts sharply with that observed in the FKBP51-containing complex, where DEX removal promotes dissociation of the immunophilin from the receptor. The distinct responses of FKBP51 and FKBP52 to the same perturbation underscore the fundamentally different dynamical roles played by the two cochaperones during late stages of the GR maturation cycle.

As for the complexes described above, D-NEMD simulations were employed to characterize the structural response of the FKBP52-containing assembly to ATP hydrolysis at each Hsp90 nucleotide-binding site and, separately, to DEX removal from GR (**Figure 3.7**).

The internal analysis of the nonequilibrium trajectories indicates that ATP hydrolysis generates predominantly local responses confined to the vicinity of the perturbed active site. Hydrolysis within NTD_A produces an extensive rearrangement of the nucleotide-binding region, involving most of the domain and promoting a rotational opening motion that facilitates HPO_4^{2-} release (**Figure 3.7A-C**). The largest responses are observed in the active-site loop (residues 113–133), the α-helix contributing to ATP binding (residues 54–64), and the modelled charged linker region (around residue 220), all of which move away from their equilibrium positions. Although weaker, the perturbation also propagates across the dimer interface, inducing detectable structural responses within protomer B (**Figure 3.7D-F**), with maximum average displacements approaching 1 Å after 20 ns. Hydrolysis within NTD_B elicits a qualitatively similar response (**Figure 3.7A-D**). The same active-site structural elements participate in the rearrangement, while the modelled charged linker exhibits an even stronger displacement than that observed following NTD_A hydrolysis. In addition, perturbation of protomer B affects the MiD_B lumen-facing loop (residues 394–405), which displays a net displacement toward CTD_B. Overall, the response patterns indicate that ATP hydrolysis primarily drives local remodelling events associated with active-site opening and product release, while only limited signal propagation is detected at longer distances.

The response to instantaneous DEX removal closely resembles that observed in the FKBP51-containing complex (**Figure 3.5G-I** and **Figure 3.7G-I**). In both systems, the dominant effect is a rapid closure of the ligand-binding pocket, mediated by the inward movement of the active-site loop comprising residues 635–640. Compared with ATP hydrolysis, the structural response induced by ligand removal is more localized and largely restricted to the receptor binding site.

Taken together, the equilibrium and nonequilibrium analyses indicate that the Post-maturation-FKBP52+DEX complex occupies a highly coordinated and dynamically stable state. In contrast to the FKBP51-containing assembly, ligand removal does not weaken the interaction network between the cochaperone and GR; rather, it promotes tighter coupling between the receptor and the FK domains of FKBP52. Furthermore, the limited effect of DEX removal on the global Hsp90 dynamics

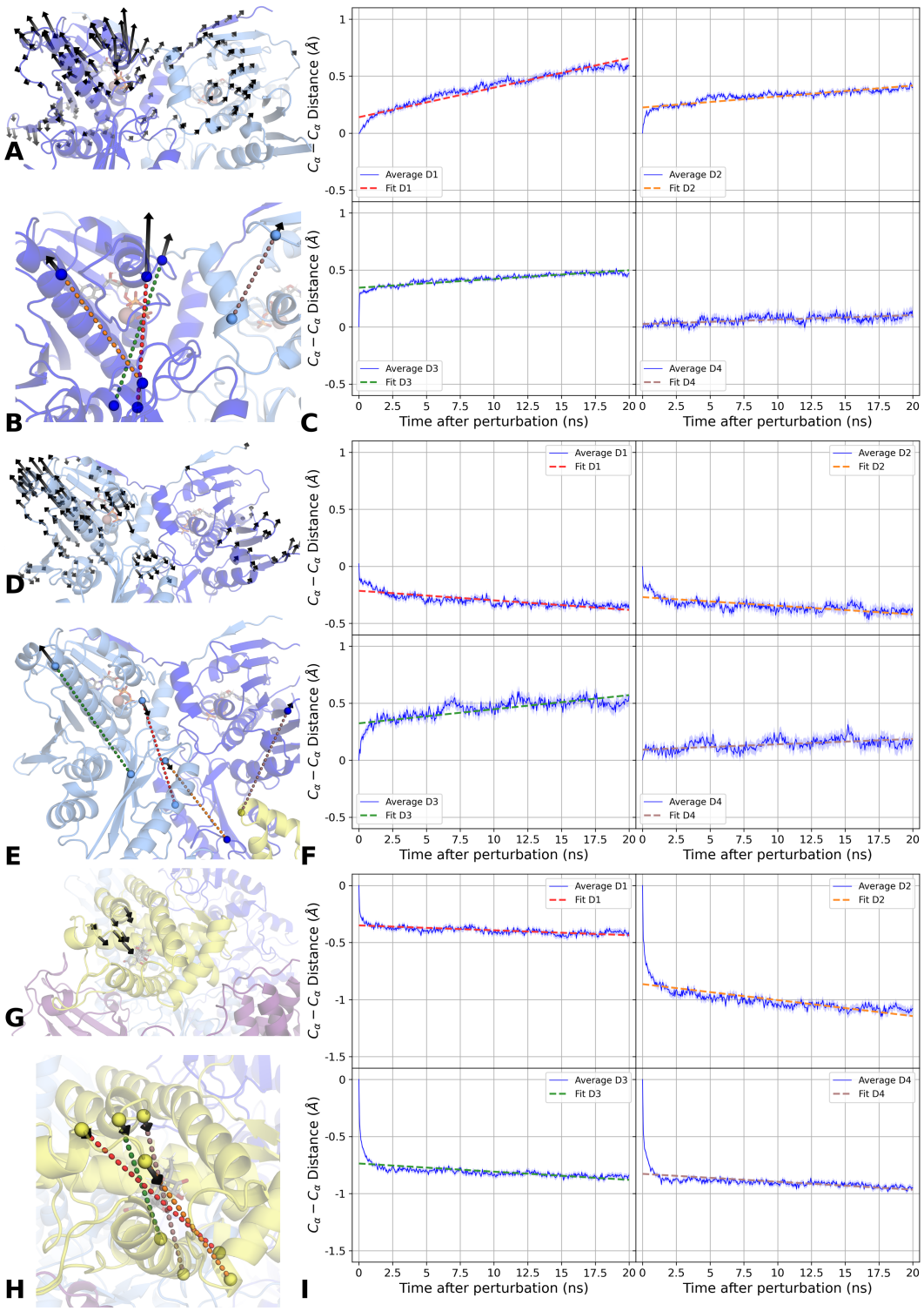

Figure 3.7: Nonequilibrium dynamics of the FKBP21 complex. *D-NEMD responses to ATP hydrolysis within the* $\mathrm{NTD_A}$, $\mathrm{NTD_B}$ *or DEX removal the FKBP52+DEX complex (panels* ***(A)****,* ***(D)****,* ***(G)*** *respectively) after 20* ns *from the perturbation instant. Arrows represent the extent and the direction of residues affected by the perturbation. In* ***(B)****,* ***(E)****,* ***(H)*** *panels, the region of interested is highlighted, with 4 affected and non-affected residue pairs. The direction of these pairs is chosen to be along the one found with internal D-NEMD analysis method. In* ***(C)****,* ***(F)****,* ***(I)*** *panels the averaged value of the distance differences between equilibrium and non-equilibrium dynamics are plotted as function of time after perturbation, together with linear fit.*

suggests that the chaperone remains in a stable activated conformation independently of ligand occupancy.
These observations support a model in which FKBP52 favours a receptor configuration that remains competent for steroid recognition and binding. The extensive dynamical coupling between FKBP52 and GR may contribute to maintaining a ligand-responsive conformation of the receptor while simultaneously stabilizing the mature chaperone complex. Such behaviour is consistent with the established role of FKBP52 in promoting productive GR activation and subsequent nuclear translocation. At the same time, the pronounced sensitivity of FKBP52 to both equilibrium and nonequilibrium perturbations highlights the dynamic nature of immunophilin regulation, supporting the idea that rapid exchange between FKBP family members represents an important mechanism for adapting GR signalling to changing cellular conditions.

3.3.5 FKBP51+DEX vs. FKBP52+DEX

The molecular mechanisms through which FKBP51 and FKBP52 regulate GR maturation and progression through the Hsp90 chaperone cycle remain incompletely understood [72]. To investigate how the two immunophilins differentially influence the dynamics of the mature GR assembly, we compared the equilibrium behaviour of the Post-maturation-FKBP51+DEX and Post-maturation-FKBP52+DEX complexes through DF, TOG, contact probability, PCA, and D-NEMD analyses (**Figure 3.8**).
The comparison of the DF matrices reveals that the two complexes share a common overall organization but differ substantially in the distribution of mechanical coordination throughout the assembly (**Figure 3.8A**). Interpretation of the TOG analysis requires particular caution because FKBP51 and FKBP52 are not identical proteins. Their sequence differences lead to distinct residue selections during the TOG filtering procedure and, consequently, to partially different reduced representations of the systems. Nevertheless, meaningful comparisons can still be drawn for shared structural components, particularly GR and Hsp90 (**Figure 3.8B,C**).
Despite the presence of different cochaperones, GR adopts broadly similar global conformations in both complexes, as indicated by the substantial overlap of intra-GR distance distributions (**Figure 3.8B,C**). However, the receptor exhibits markedly different coordination patterns. In the FKBP51-containing assembly, GR displays enhanced internal coordination throughout most of its structure relative to the FKBP52 system (**Figure 3.8A**), indicating a more mechanically integrated receptor. At the same time, TOG analysis reveals distinct positioning of GR within the Hsp90 lumen. In the presence of FKBP51, the receptor tends to remain closer to the CTD regions while moving away from $\mathrm{MiD_B}$, whereas FKBP52 favours an alternative spatial arrangement (**Figure 3.8B,C**). These differences are accompanied by corresponding changes in GR–Hsp90 coordination and by remodelling of contact networks involving the GR tail and lumen-facing regions of Hsp90 (**Figure 3.8D**).
The influence of the two immunophilins extends to Hsp90 itself. The FKBP51-containing complex exhibits a general increase in Hsp90 internal coordination relative to the FKBP52 assembly (**Figure 3.8A**). TOG analysis indicates an overall tendency of the two protomers to approach one another, although the MiD domains move slightly apart, producing a subtle deformation of the client-binding region (**Figure 3.8B,C**). In contrast, the nucleotide-binding site of protomer A displays a distinctive behaviour in the FKBP51 complex, where it remains poorly coordinated with multiple regions of the assembly, including GR. This observation suggests that the two immunophilins differentially modulate communication pathways linking the Hsp90 active sites to distal regions of the complex.
The most pronounced differences are observed within the immunophilin domains themselves. The ΔDF matrix highlights a striking redistribution of coordination between the FK1 and FK2 domains depending on which cochaperone is present (**Figure 3.8A**). In the FKBP51-containing complex, FK1 exhibits stronger coordination with both Hsp90 and GR, whereas FK2 contributes more prominently

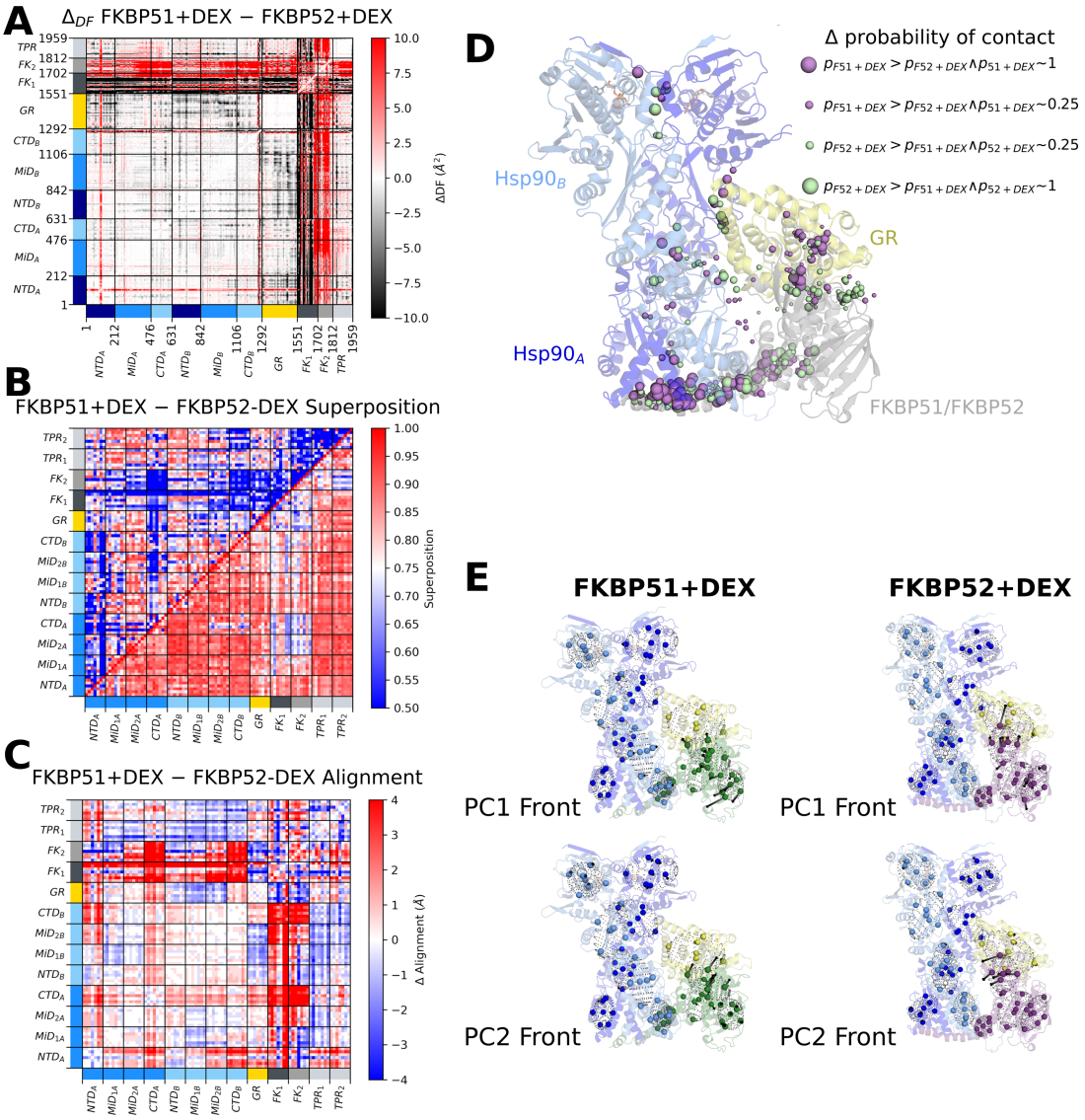

Figure 3.8: Differential dynamics of the two DEX-containing immunophilin complexes. *(**A**) ΔDF matrix obtained by subtracting the Post-maturation-FKBP52+DEX DF matrix from the Post-maturation-FKBP51+DEX. (**B**) In superposition map the overlap integral values for each pairwise distance distribution, upper diagonal depicts values before superposition, while the lower represents the values after superposition. Red values indicate a similar dynamical behaviour between the two dynamics, while white/blue values underlie a marked dynamical difference. (**C**) Alignment map represents the Δ value pertaining to the translation with the highest degree of overlap for each reduced point pairwise distance distribution between trajectories. Blue spots ($\Delta < 0$ Å) mean that the perturbation has increased the distance between reduced points, while red spots ($\Delta > 0$ Å) mean the opposite. (**D**) Projection of contact probability differences (FKBP51+DEX vs FKBP52+DEX) onto the Post-maturation-FKBP52 complex structure.Spheres indicate statistically significant modified interaction between residue pairs in the unperturbed/perturbed contact matrices. (**E**) Projection of the first two PCA eigenvectors onto structures of the immunophilin-bound complexes, together with TOG points.*

to the interaction network in the FKBP52 assembly. This redistribution of mechanical coupling may underlie the distinct functional outcomes associated with the two immunophilins.
The TOG analysis further supports this interpretation. Distance distributions involving FK domains show limited overlap even after alignment, indicating substantially different dynamical behaviors (**Figure 3.8B,C**). Contact probability analysis reveals that FKBP51 establishes stronger interactions between its FK1 domain and GR, whereas FKBP52 preferentially stabilizes contacts involving the FK2 domain (**Figure 3.8D**). In the latter case, an extended interaction patch at the GR–FK2 interface appears to contribute to the increased coupling between the two proteins. The TPR domains also display distinct interaction patterns with the long CTD_B loop of Hsp90. Although the detailed contact network is complex, FKBP52 shows a general tendency to remain more closely associated with the CTD_B region than FKBP51.
Principal component analysis of the TOG-reduced trajectories confirms that the dominant collective motions differ substantially between the two assemblies (**Figure 3.8E** and **Figure A.11**). The largest discrepancies involve the FK domains, whose movements differ both in amplitude and direction. In both systems, FK1 displays a tendency to approach GR; however, in the FKBP52 complex it simultaneously moves toward MiD_B. FK2 undergoes broader and more extensive motions in the FKBP51-containing assembly, reflecting the distinct dynamical organization imposed by the two immunophilins.
D-NEMD simulations provide further evidence for differential regulation of Hsp90 dynamics. Following ATP hydrolysis, the nucleotide-binding site of protomer A exhibits a weaker structural response in the FKBP51-containing complex than in the FKBP52 assembly, suggesting that the active site resides in a more pre-organized state prior to perturbation. Consistently, analysis of the equilibrium coordination patterns around NTD_A (**Figure A.12**) indicates distinct dynamical environments in the two systems. Together, these observations support the existence of different allosteric communication networks linking the Hsp90 catalytic machinery to the client and cochaperone. Rather than simply stabilizing alternative conformations, FKBP51 and FKBP52 appear to reshape the internal dynamical architecture of the assembly in distinct ways, thereby modulating how information propagates between the Hsp90 active sites, GR, and the immunophilin domains.
Overall, the results indicate that FKBP51 and FKBP52 do not merely represent interchangeable binding partners of the mature GR complex. Instead, each immunophilin establishes a characteristic pattern of mechanical coordination, intermolecular contacts, and collective motions. These distinct dynamical fingerprints provide a plausible structural basis for the different biological functions associated with the two cochaperones during late stages of the GR maturation cycle.

3.3.6 MiD–GR Distances

To characterise how cochaperone binding influences receptor positioning throughout the chaperone cycle, we compared the distances between the centre of mass of the Hsp90 middle domains (MiDs) and that of the folded GR core, excluding the unfolded tail loop. The resulting MiD_B–GR distance distributions (**Figure A.5C**) provide a coarse-grained yet biologically informative measure of the spatial relationship between the client and the chaperone.
Analysis of the simulations initiated from the original cryo-EM structures reveals a clear progression along the maturation pathway. The Loading−DEX complex displays the shortest average MiD_B–GR distance, followed by the Maturation+DEX complex and, subsequently, the Post-Maturation+DEX assemblies. This gradual increase in separation is consistent with the sequential remodelling of the complex during the chaperone cycle. In the Loading state, GR remains closely associated with the Hsp90 lumen, whereas maturation is accompanied by stabilisation of the folded receptor through Hsp90 closure and interactions with the p23 tail helix. In the Post-maturation assemblies, insertion of the immunophilin FK1 domain between Hsp90B and GR further alters receptor positioning, increasing the average separation between the receptor and the chaperone core.
The two Post-Maturation+DEX complexes display distinct behaviours that are consistent with their

proposed biological roles. The Post-maturation-FKBP51+DEX complex exhibits the largest mean MiD_B–GR distances, indicative of a weaker spatial coupling between GR and Hsp90. In contrast, the Post-maturation-FKBP52+DEX complex samples a broader distribution centred at lower distances, partially overlapping with those observed for the Maturation+DEX state. This behaviour suggests that FKBP52 maintains a tighter association between GR and the chaperone machinery, even after acquisition of the folded receptor conformation.
Perturbation of ligand occupancy further supports these observations. Introduction of DEX into the Loading−DEX complex broadens the distance distribution and promotes sampling of conformations more similar to those observed in the Maturation+DEX ensemble. Conversely, removal of DEX from the Maturation+DEX complex produces a narrowing of the distribution and a shift towards shorter distances, resembling those characteristic of the Loading state. Comparable ligand-dependent effects are observed in both Post-maturation complexes, where DEX removal alters the relative positioning of GR and Hsp90 in a manner consistent with the changes identified by DF, TOG, and contact analyses.
Overall, these results demonstrate that ligand occupancy is reflected in the spatial organisation of the GR–Hsp90 interface. The relative positioning of GR with respect to the Hsp90 middle domains emerges as a sensitive descriptor of the conformational state of the assembly, capturing both the progression of the chaperone cycle and the distinct dynamical effects exerted by the different cochaperones.

3.3.7 Comparison of Simulation Results with Experimental Characterization of Points of Regulation

The simulations presented above identify ligand-dependent changes in dynamic states that remodel biologically important regions of the GR-processing assemblies. Strikingly, many of the regions highlighted by our analyses coincide with experimentally characterised regulatory sites, including key post-translational modifications (PTMs) and disease-associated mutations. These observations provide an important link between the dynamic networks identified computationally and known mechanisms of functional regulation. PTMs introduce an additional layer of control over chaperone activity, contributing to the so-called chaperone code, which expands the capacity of folding assemblies to respond to changing cellular conditions and stress signals. Advances in proteomic technologies have revealed numerous PTMs on chaperones, including phosphorylation, acetylation, methylation, SUMOylation and ubiquitination [75, 155].
One of the most prominent findings of our analyses is the ligand-dependent remodelling of the GR–Hsp90 interface. In several complexes, DEX binding or removal alters contacts involving the region surrounding K703, the only currently known SUMOylation site within the GR ligand-binding domain (LBD), which has been implicated in receptor degradation [178]. Notably, this region is strongly remodelled in the Maturation and Post-maturation-FKBP51 complexes, whereas little perturbation is observed in the Post-maturation-FKBP52 assembly. Since FKBP51-containing complexes are associated with cytoplasmic retention and recycling of GR, while FKBP52 promotes nuclear translocation of the mature receptor, these observations suggest that differential modulation of the K703 region may contribute to the distinct functional outcomes of the two pathways. A more dynamic interface may facilitate receptor turnover and degradation, whereas tighter protection of the LBD may preserve receptor activity during transport to the nucleus [179].
A second regulatory hotspot emerges at the NTD–NTD interface of Hsp90. Phosphorylation of K38 in yeast Hsp90 is known to influence cell growth, steroid receptor maturation, and interactions with the cochaperones Hop and p23 [180]. In all systems analysed here except the Post-maturation-FKBP52 complex, perturbations originating at the GR ligand-binding pocket induce measurable remodelling of the Hsp90 NTD interface, where K38 is located. These observations suggest the existence of long-range communication pathways linking the GR active site to the N-terminal dimerization interface of Hsp90, potentially influencing cochaperone recruitment and chaperone activity.

Further evidence of stage-specific allosteric communication is provided by K615 of human Hsp90α (equivalent to K594 in yeast Hsp90), a residue whose methylation has been shown to affect ATP hydrolysis, p23 binding and GR regulation [181]. Interestingly, little structural remodelling is observed around K615 in the Loading complex following DEX insertion. In contrast, ligand removal from the Maturation complex induces a clear response at the Hsp90A–Hsp90B interface surrounding this residue. Similar effects are observed in the Post-maturation-FKBP51 complex, where perturbations propagate to the region encompassing K615 and the GR segment occupying the Hsp90 lumen [180]. These findings suggest that K615 becomes progressively integrated into the allosteric network as the maturation cycle advances, acquiring functional relevance only in the later stages of GR processing.
In addition to PTMs, several disease-associated GR mutations map to regions identified here as important communication hubs. The T556I substitution disrupts a critical hydrogen-bonding network and impairs ligand binding, nuclear translocation and transcriptional activation [182]. Similar functional defects have been reported for the I559N [183] and F737L [184] variants. All three residues cluster within the LBD, close to the steroid-binding pocket and opposite the main Hsp90 interaction surface. Notably, our analyses consistently identify this region as a central node of the allosteric network. In the Loading complex, DEX binding induces substantial remodelling of contacts between the region surrounding T556 and the MiD domain of Hsp90 protomer B. Furthermore, D-NEMD simulations reveal pronounced propagation of structural responses towards this region following both ATP hydrolysis and ligand removal, highlighting its role as a focal point for Hsp90–GR communication.
Another particularly relevant mutation is T773C, which has been associated with impaired ligand binding and signal transduction [185]. Although located near the C-terminus of GR, this residue lies adjacent to the p23 tail helix in the Maturation complex. Consistent with its functional importance, our analyses reveal extensive remodelling of the GR–p23 interface in this region following perturbation. D-NEMD simulations further show that structural responses initiated at the Hsp90 ATP-binding sites propagate towards the C-terminal region of GR across multiple complexes, identifying this area as another major communication hub within the chaperone machinery.
Taken together, these observations demonstrate that the dynamic networks uncovered by our simulations converge on experimentally validated regulatory sites. The correspondence between allosteric communication pathways, PTM hotspots and disease-associated mutations provides strong support for the biological relevance of the identified mechanisms and suggests that chaperone-mediated regulation of GR function emerges from the integration of ligand-dependent dynamics with multiple layers of post-translational and genetic control.

3.4 Discussion

Multicomponent protein assemblies are increasingly recognised as fundamental determinants of biological function [140, 162, 186, 187]. Within these systems, individual components may fulfil distinct roles, acting as enzymes, kinetic modulators, ligand-binding proteins, adaptors that regulate substrate selection, or signalling factors controlling localisation and turnover. Function therefore emerges from the coordinated action of all assembly members, while regulatory control depends on their composition, architecture and dynamic organisation.
Many biological assemblies can be viewed as modular interaction networks in which hub proteins integrate signals originating from distinct pathways. The Hsp90 chaperone machinery represents a paradigmatic example of this principle. Hsp90 acts as a central hub whose interaction landscape continuously changes through the binding and release of cochaperones and clients, generating context-dependent functional states adapted to specific cellular conditions [188]. Conformational dynamics play a central role in this process, enabling Hsp90 to interact with a diverse range of partners and to transition between different functional states throughout the chaperone cycle [53, 161, 189–191]. Such adaptability is a common feature of major cellular assemblies, including the ribosome [192],

proteasome [193], and spliceosome [194], where function emerges from regulated structural plasticity. A major challenge in structural and chemical biology is understanding how local perturbations, such as ligand binding, nucleotide hydrolysis or changes in assembly composition, are propagated through large molecular machines to generate coordinated functional responses [14, 22]. In this work, we addressed this problem by investigating the molecular mechanisms governing folding and activation of the glucocorticoid receptor (GR) [70–72, 162, 165, 195], one of the best-characterised Hsp90 clients. By combining equilibrium molecular dynamics and dynamical nonequilibrium molecular dynamics (D-NEMD) simulations across multiple stages of the GR maturation cycle, we examined how changes in ligand occupancy and cochaperone composition influence the dynamic organisation of the Hsp90 machinery.

To characterise these complex systems, we introduced new analysis approaches capable of reducing atomistic simulations into interpretable descriptions of domain-level motions while retaining the underlying dynamic information. Together, these analyses reveal substantial plasticity within all assemblies and demonstrate that ligand binding, cochaperone composition and nucleotide state act as interconnected regulators of a shared allosteric network spanning the entire chaperone machinery.

A central finding emerging from this work is that ligand occupancy of GR functions as an active regulator of the Hsp90 cycle rather than merely reflecting the folding state of the client. In the Loading complex, the artificial introduction of DEX into the ligand-binding pocket induces widespread remodelling of interaction networks involving GR, Hsp90, Hsp70 and Hop. These changes favour closure of the Hsp90 dimer, weaken the coordination of Hsp70 and Hop with the assembly, and promote dynamic states resembling those observed in later stages of the cycle. Importantly, the Loading complex already samples conformations compatible with ligand binding, indicating that the partially unfolded client explores steroid-competent states before complete maturation. Ligand binding therefore appears capable of biasing the conformational landscape towards progression along the folding pathway.

The Maturation complex provides further evidence that ligand occupancy contributes directly to assembly stability. DEX stabilises the dynamic coupling between GR, Hsp90 and p23, while ligand removal weakens this network and promotes reorganisation of contacts at the GR–Hsp90 interface. Particularly notable is the response of the p23 α-helical tail, whose interactions with GR are strongly modulated by ligand occupancy. These observations suggest that p23 may act as a sensor of client maturation state, preferentially stabilising ligand-engaged receptors while destabilising conformations that are not competent for activation. Such behaviour would place p23 at a critical regulatory node where information concerning client folding and ligand binding is integrated to determine progression through the chaperone cycle.

The comparison between Post-maturation complexes containing FKBP51 or FKBP52 highlights how cochaperone composition determines the interpretation of these ligand-dependent signals. Although structurally similar, the two immunophilins generate markedly different dynamic responses. FKBP51 promotes stronger coupling between specific FK domains and GR, favours configurations associated with receptor release, and exhibits greater sensitivity to perturbations arising from ATP hydrolysis or ligand removal. By contrast, FKBP52 stabilises interactions with GR and maintains conformations compatible with receptor transport and signalling. These observations are fully consistent with their established biological roles, whereby FKBP51 promotes cytoplasmic retention and recycling, whereas FKBP52 facilitates nuclear translocation of active GR.

Importantly, our analyses indicate that the functional differences between FKBP51 and FKBP52 do not arise from large structural rearrangements but rather from subtle rewiring of dynamic communication pathways. The FK domains emerge as major allosteric hubs capable of coupling signals originating from both the GR ligand-binding pocket and the Hsp90 nucleotide-binding sites. In this view, immunophilins do not merely accompany receptor maturation but actively reshape the communication network of the assembly, thereby influencing client fate.

The combined equilibrium and nonequilibrium simulations support a model in which information flows bidirectionally throughout the assembly. ATP hydrolysis within Hsp90 propagates structural

signals towards the client and associated cochaperones, whereas ligand binding within GR simultaneously remodels communication pathways extending towards Hsp90 and its regulatory partners. Rather than behaving as independent functional modules, the different components therefore operate as an integrated allosteric network whose dynamic state emerges from the combined influence of client occupancy, cochaperone composition and nucleotide state.

This integrated view is particularly evident in the asymmetric behaviour observed between the two Hsp90 protomers. Asymmetric responses emerge consistently from equilibrium dynamics, D-NEMD analyses and interaction network remodelling. The extent and nature of this asymmetry depend strongly on both the bound cochaperone and ligand conditions. Such behaviour is consistent with the known sequential nature of ATP hydrolysis within Hsp90 and suggests that protomer asymmetry represents an additional layer of regulation exploited by the chaperone machinery. The preferential association of GR and cochaperones with one protomer may therefore contribute to directing information flow through the assembly and controlling progression along the maturation cycle.

Interestingly, the degree of asymmetry differs substantially between FKBP51- and FKBP52-containing complexes. FKBP51 promotes a more asymmetric organisation and stabilises conformations associated with a catalytically competent nucleotide-binding site, whereas FKBP52 maintains a more balanced dynamic regime. These differences may contribute directly to their divergent biological functions, with FKBP51 favouring ATP turnover and receptor release, and FKBP52 maintaining a ligand-competent state conducive to nuclear transport.

Taken together, our results support a model in which asymmetry cooperates with modular assembly to expand the repertoire of dynamic states available to Hsp90. By combining different cochaperones, clients and ligand conditions, the chaperone machinery generates a spectrum of allosterically connected conformations that provide both selectivity and directionality to client processing.

A limitation of the present work concerns the representation of ATP hydrolysis using classical force fields. Although our simulations identify clear differences in nucleotide-dependent dynamic states and reveal distinct responses following hydrolysis, they do not allow definitive assignment of the protomer that hydrolyses ATP first. Addressing this question will require future QM/MM studies capable of explicitly characterising the reaction mechanism and the energetic determinants governing hydrolysis within different assembly states.

Finally, our findings highlight a previously underappreciated role of client ligands as active participants in chaperone-mediated folding. In the case of GR, steroid binding not only stabilises specific assembly states but also modulates long-range allosteric communication throughout the Hsp90 machinery. This establishes a direct mechanistic connection between cellular metabolite availability and chaperone function. Variations in steroid concentration could therefore influence the progression of GR through the maturation cycle, stabilising specific assembly states and biasing receptor fate decisions. Experimental observations showing steroid-dependent changes in GR abundance, localisation and transcriptional activity are consistent with this interpretation [196–199].

More broadly, our results support a model in which metabolites and small-molecule ligands function as structural encoders of chaperone regulation, coupling chemical signals to the assembly and remodelling of protein-folding machines. Within this framework, cochaperones emerge as dynamic checkpoints that decode these signals and translate them into specific biological outcomes. Beyond providing new insights into GR regulation, these findings highlight the possibility of selectively targeting allosteric communication pathways within chaperone assemblies to manipulate client fate, opening new opportunities for therapeutic intervention in proteostasis-related disorders.

3.5 Conclusions

The biological function of multicomponent protein machineries emerges from a delicate balance of enzymatic activities, recognition events and the assembly of finely tuned, context-dependent complexes. This interplay favours the adaptation of key molecular machineries to changing cellular conditions.

Our integrative simulations reveal how these principles operate in the Hsp90–GR system at different stages of substrate maturation. We show that nucleotides and cochaperone identity actively model the dynamic states and the allosteric landscape of the chaperone cycle. Importantly, ligand binding in the client on their path to full maturation, exemplified in this work by steroids, emerges for the first time as an important modulator of complex stability and progression. At the same time, the closely related immunophilins FKBP51 and FKBP52 rewire interaction networks to bias GR fate toward cytosolic release or nuclear translocation. Within this model, asymmetric dynamics and sequential ATPase regulation further expand the repertoire of dynamic states accessible to Hsp90, enhancing its ability to adapt to the client, while co-chaperone regulation imparts selectivity and directionality to client remodelling by differentially interfering with the distinct allosteric pathways. The integration of chemical and structural information suggests a new model in which complex chaperone machineries are regulated by the cross-talk among their different components. Ligands (nucleotides or client-substrates) encode chemical information that translates into the modulation of the stability and functionally oriented dynamics of client-maturation complexes, while cochaperone recruitment defines selectivity in substrate recognition and directionality in its processing. Although future experiments are required to precisely define the progression of conformational states and the roles of the metabolites shown here, our findings provide a new hypothesis for how the chaperone machinery controls client fate.

Chapter 4

State-Dependent Effects of Hsp90β Phosphorylation

4.1 Introduction

As discussed in the previous chapter, correct protein folding is essential for maintaining cellular proteostasis under physiological conditions [13]. In vivo, the folding process is finely regulated by molecular chaperones, which interact with newly synthesised proteins as well as with proteins that have lost their native conformation following cellular stress [53].
Among these chaperones, Heat Shock Proteins (HSPs) constitute a highly conserved family found throughout living organisms. They play a fundamental role in cellular homeostasis, particularly in response to stress conditions such as elevated temperatures, oxidative damage, and exposure to toxins. Hsp90 is one of the major molecular chaperones and facilitates the correct folding of client proteins by transiently enclosing them within its client-binding cavity, also referred to as the lumen. The Hsp90 isoforms Hsp90α and Hsp90β are localised in the cytoplasm, whereas other members of the family, such as TRAP1, reside in the mitochondria [53, 54].
Both Hsp90 cytosolic isoforms interact with a broad range of client proteins, including kinases, hormone receptors, and transcription factors. To perform its functions efficiently, Hsp90 cooperates with several auxiliary proteins, known as cochaperones, which regulate both its ATP-driven conformational cycle and its ability to bind and release client proteins. One such cochaperone is Cdc37, which assists Hsp90β in the folding, stabilisation, and regulation of protein kinase clients. Of the two cytoplasmic isoforms, Hsp90β is constitutively expressed in eukaryotic cells under normal physiological conditions and plays a key role in the regulation of numerous signalling pathways, including those involving oncogenic proteins [200, 201].
Hsp90β functions as a homodimer and possesses a highly conserved architecture comprising three major domains, each associated with distinct structural and functional roles (see previous chapter). In addition to these structured domains, several loop regions contribute significantly to the protein's activity. These loops are highly dynamic and flexible elements that mediate molecular recognition and facilitate interactions with client proteins and cochaperones [202]. Among them, a particularly important loop is located within the middle domain. This region is often only partially resolved in crystallographic structures owing to its intrinsic flexibility [53]. Nevertheless, it plays a crucial role in the recruitment and scaffolding of client proteins and cochaperones. Consequently, structural alterations affecting this loop, including post-translational modifications, can significantly influence the functional behaviour of Hsp90β.
The functional cycle of Hsp90 is driven by ATP binding and hydrolysis, which induce a series of conformational changes that enable client-protein binding and maturation. The cycle begins in a fully open conformation following the release of ADP at the end of the previous cycle. In this state, the two protomers are spatially separated. ATP binding to the N-terminal domains promotes a progressive conformational rearrangement towards a semi-closed state. Subsequent structural

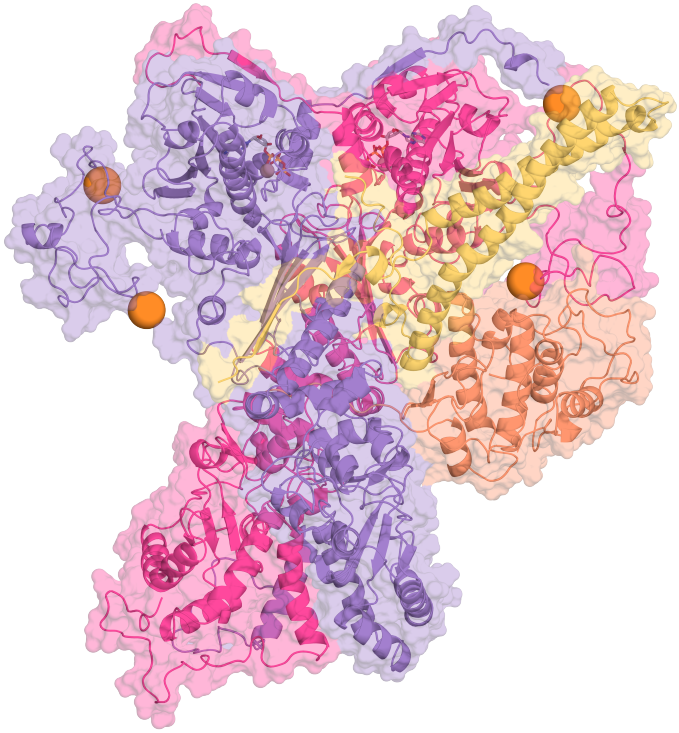

Figure 4.1: Hsp90β phosphorylated complex. *Structure of the complex between Hsp90 (monomer A in violet, monomer B in pink), cochaperone Cdc37 (in yellow) and CDK4 kinase (in light orange), with phosphorylation sites highlighted in orange.*

changes bring the protomers into closer contact, leading to the fully closed conformation. Hydrolysis of the first ATP molecule destabilises the corresponding protomer and facilitates hydrolysis of the second ATP molecule in the partner protomer. Following ATP hydrolysis, both nucleotides are released as ADP, allowing Hsp90β to return to its open conformation [55, 64].
Throughout this cycle, client proteins interact primarily with the middle domain of Hsp90β, occupying a binding cavity formed at the interface between the two protomers. The conformational transitions associated with the ATPase cycle enable client proteins to achieve their native folded state or to assemble into larger functional complexes.
Under pathological conditions, the activity of Hsp90β may become dysregulated. Prolonged cellular stress and oncogenic transformation can promote the formation of epichaperomes, highly stable multimeric assemblies composed of Hsp90β and its cochaperones. Unlike the transient interactions observed during normal chaperone activity, epichaperomes form persistent networks that support cell survival by stabilising aberrant signalling pathways and suppressing apoptotic responses [203, 204].
The formation of epichaperomes is frequently accompanied by altered expression levels of Hsp90β and changes in its ATPase activity. These alterations enhance the stabilisation of mutant, overexpressed, or misfolded proteins that would otherwise be degraded, thereby promoting the survival and proliferation of pathological cells. As a consequence, tumour cells become increasingly dependent on Hsp90β-mediated stabilisation of oncogenic signalling proteins.
The ATPase cycle itself may also be altered by oncogenic signalling, cellular stress, or post-translational modifications. Such perturbations can shift the equilibrium between open, semi-closed, and closed conformational states, thereby affecting interactions with both client proteins and cochaperones. For example, changes in ATP hydrolysis kinetics may increase the stability of oncogenic client proteins such as CDK4. CDK4 (Cyclin-Dependent Kinase 4) is a key regulator of cell-cycle progression and promotes the transition from the G1 phase to the S phase, during which DNA replication occurs. Under physiological conditions, Hsp90β contributes to the stabilisation of the

active conformation of CDK4. In tumour cells, however, dysregulated Hsp90β activity may prolong CDK4 stabilisation, thereby supporting uncontrolled cellular proliferation. Experimental evidence also suggests that Hsp90β can remain longer in intermediate conformational states, facilitating the recruitment of cochaperones such as Cdc37 and further enhancing the stabilisation of oncogenic kinase clients.
Phosphorylated Hsp90β has been detected in numerous pathological contexts, including breast cancer, gliomas, neurodegenerative disorders, and viral infections. Phosphorylation is typically catalysed by kinases that transfer a phosphate group from ATP to a target residue, generating ADP as a by-product. The introduction of a phosphate group adds negative charge to the protein surface and can promote the formation of new electrostatic interactions and hydrogen-bonding networks. Consequently, phosphorylation can significantly alter protein structure, dynamics, and intermolecular interactions [204, 205].
Numerous phosphorylation sites have been identified in Hsp90β. In this chapter, particular attention is devoted to the phosphorylation of Ser226 and Ser255 in both protomers. These residues are located within the middle domain and have been associated with conditions of severe cellular stress and tumour-cell proliferation. Experimental studies suggest that phosphorylation at these positions induces conformational changes that modulate interactions with both client proteins and cochaperones, including an increased affinity for Cdc37 [75, 152].
The aim of this study is to investigate how phosphorylation of Ser226 and Ser255 affects the conformational dynamics and allosteric communication of Hsp90β during key stages of its ATPase cycle. To this end, three representative systems were analysed: the semi-closed state (SC), the closed state (CL), and the complex state containing Hsp90β, Cdc37, and CDK4 (CO) (**Figure 4.1**).

4.2 Modelling, preparation and simulations

To investigate the effects of Ser226 and Ser255 phosphorylation on the different conformational states of Hsp90β, three structural models representing key stages of the ATPase cycle were prepared and subsequently simulated. The Closed Wild-Type state was generated first and served as a template for modelling the remaining systems. The initial structure, kindly provided by Prof. David A. Agard (University of California, San Francisco), was resolved from residues 2–219 and 274–724.
The missing N-terminal residue was added using PyMOL. To reconstruct the unresolved loop in protomer A, the MODELLER package was employed [118]. Twenty-five candidate models were initially generated and evaluated using the DOPE-HR (DOPE High Resolution) statistical potential. Each model was subsequently refined through up to 300 steps of conjugate-gradient energy minimisation applied to progressively larger residue intervals surrounding the modelled region.
Refinement was followed by MODELLER's slow molecular-dynamics annealing protocol. During this procedure, each model was heated from 0 K to 1000 K over 1100 molecular-dynamics steps of 4 fs each and subsequently cooled to 300 K over 4800 additional steps. The complete annealing protocol was repeated for two cycles.
The resulting models (mod01–mod25) were evaluated using both the DOPE score and the molecular probability density function (molpdf) score. Both metrics identified model mod04 as the best-performing structure. Consequently, mod04 was selected for all subsequent analyses. The reconstructed loop from protomer A was then duplicated into protomer B, yielding the final Closed Wild-Type Hsp90β model.
The Semi-Closed state was modelled starting from PDB structure 7KW7 [70], which contains Hsp90α in complex with two Hsp70 molecules, the cochaperone Hop, and a glucocorticoid receptor client protein in a semi-closed conformation. All proteins other than Hsp90 were removed, and the Hsp90α sequence was converted to the Hsp90β isoform through residue mutations performed in PyMOL. Missing regions were reconstructed by superimposing the Closed-state model onto the Semi-Closed structure and transferring coordinates from the corresponding resolved regions. In particular, three loops were modelled for each protomer: the N-terminal loop (residues 1–15), the charged linker

(residues 219–274), and the C-terminal loop (residues 700–724).
The Complex state was generated from PDB structure 5FWK [76], which contains Hsp90β bound to the cochaperone Cdc37 and the client protein CDK4. Missing residues were reconstructed using the Closed-state model following the same procedure described above. In this case, no sequence modifications were required because the experimental structure already corresponded to the Hsp90β isoform.
The three structural models were subsequently processed using the Reduce utility included in *AmberTools* (version 2021), which was employed to add hydrogen atoms and optimise side-chain orientations. Protonation states were then assigned using PROPKA (version 3.1) [120, 121]. Identical protonation states were maintained for both Wild-Type and phosphorylated systems.
Prior to phosphorylation and solvation, all structures underwent a preliminary vacuum energy minimisation using the *ParmEd* utility. This minimisation step was performed to relieve local steric strain introduced during loop modelling, particularly in regions adjacent to the anchor points used to graft reconstructed loops. Such relaxation was especially important for the Semi-Closed and Complex models, which contained multiple reconstructed segments.
Following generation of the minimised Wild-Type structures, phosphorylation was introduced at Ser226 and Ser255 in both protomers using PyMOL. The modification was implemented by replacing the hydroxyl hydrogen atom of each serine residue with a phosphate group and renaming the residue as SEP.
Crystallographic water molecules were retained by transferring conserved water positions from high-resolution X-ray structures of the Hsp90β N-terminal domain (PDB code 6N8Y) and middle domain (PDB code 3PRY). These structures were superimposed onto the corresponding domains of each Hsp90β model, and overlapping crystallographic waters were incorporated into the final systems.
All six systems were solvated in truncated-octahedral simulation boxes filled with TIP3P water molecules. A minimum solute-to-box-edge distance of 11.5 Å was imposed in all directions. System neutrality was achieved through the addition of sodium counterions.
Molecular dynamics simulations were subsequently performed according to the protocol described in **Chapter 2**. Wild-Type systems were simulated with Amber ff14SB forcefield [103], while Phosphorylated ones with its natural extension with phosphorylated residues phosaa14SB forcefield [206]. A total of six systems were simulated: the Semi-Closed (SC), Closed (CL), and Complex (CO) states, each in both Wild-Type and phosphorylated forms. Each system was simulated for a cumulative trajectory length of 4 µs.

4.3 Analyses and results

4.3.1 Phosphorylation reshapes hydrogen bonds network

To investigate the local structural effects induced by phosphorylation, we first analysed the hydrogen-bonding patterns in the vicinity of Ser226 and Ser255. Specifically, we monitored all potential hydrogen bonds formed between the oxygen atoms of the target serine residues and the nitrogen atoms of positively charged side chains, namely lysine and arginine residues.
Hydrogen-bond analysis was performed using the hbond module implemented in the CPPTRAJ package [174]. For the Wild-Type systems, interactions between serine hydroxyl oxygens and basic side chains were monitored according to the scheme:

$$\left[\mathrm{O}_{\gamma}^{\mathrm{SER}}\right]_{\mathrm{acc}} - \left[\mathrm{N}_{\zeta}^{\mathrm{LYS}}, \mathrm{N}_{\varepsilon}^{\mathrm{ARG}}, \mathrm{N}_{\eta 1}^{\mathrm{ARG}}, \mathrm{N}_{\eta 2}^{\mathrm{ARG}}\right]_{\mathrm{don}} \tag{4.1}$$

For the phosphorylated systems, all oxygen atoms belonging to the phosphate group were considered

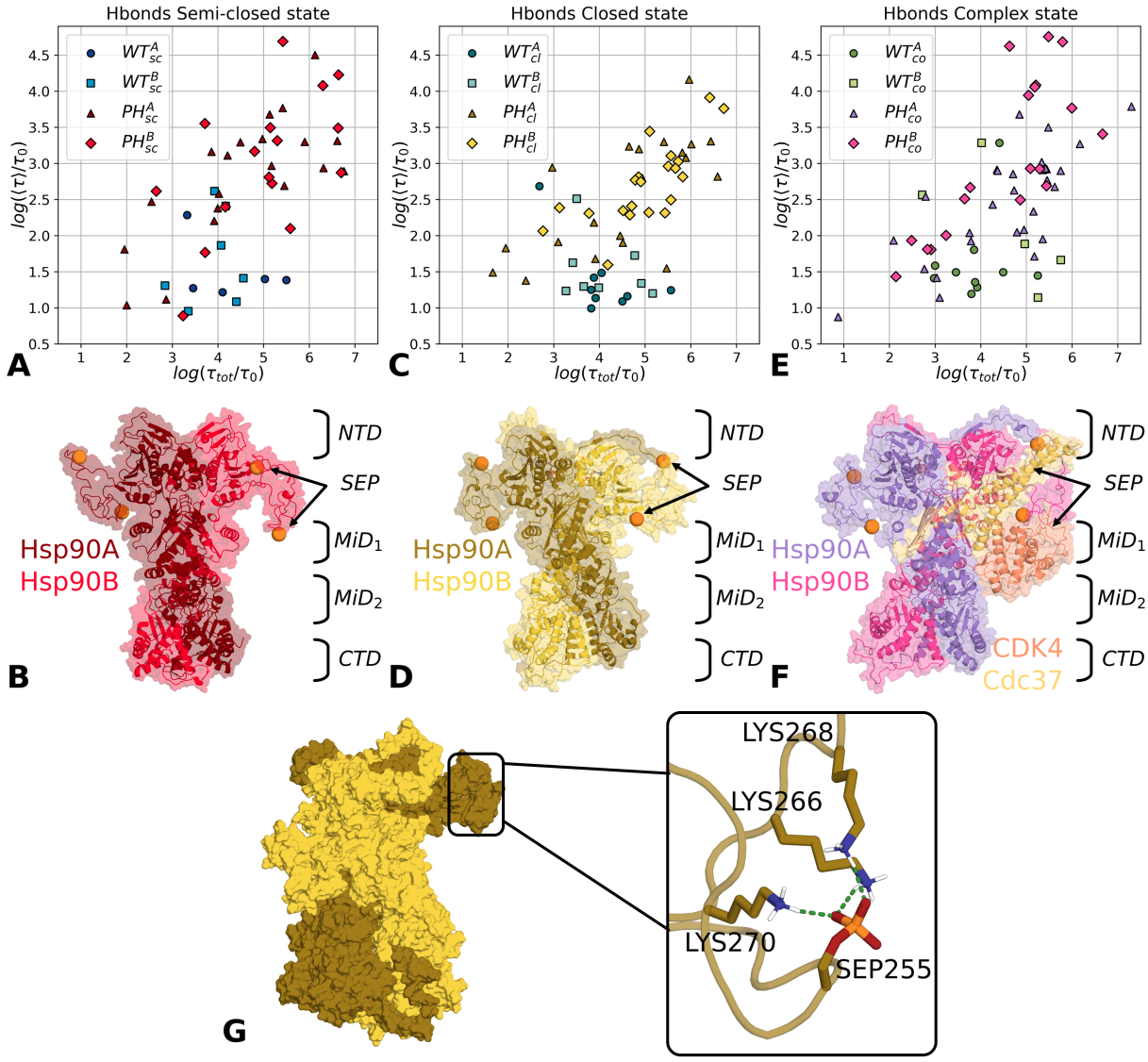

Figure 4.2: Hbond analysis on Hsp90β. *On the x-axis a measure related to the total lifetime, while on the y-axis a measure of the average lifetime of each hydrogen bond:* ***(A),(C),(E)*** *Hydrogen bonds in the Semi-Closed, Closed and Complex state respectively.* ***(B),(D),(F)*** *Structures of Semi-Closed, Closed and Complex state, with phosphorylation sites evidenced with orange spheres.* ***(G)*** *Frame with SEP226A performing three Hbonds*

as potential acceptors:

$$\left[\mathrm{O}_{\gamma}^{\mathrm{SEP}}, \mathrm{O}_{\delta 1}^{\mathrm{SEP}}, \mathrm{O}_{\delta 2}^{\mathrm{SEP}}, \mathrm{O}_{\delta 3}^{\mathrm{SEP}}\right]_{\mathrm{acc}} - \left[\mathrm{N}_{\zeta}^{\mathrm{LYS}}, \mathrm{N}_{\varepsilon}^{\mathrm{ARG}}, \mathrm{N}_{\eta 1}^{\mathrm{ARG}}, \mathrm{N}_{\eta 2}^{\mathrm{ARG}}\right]_{\mathrm{don}} \tag{4.2}$$

Because phosphorylation introduces an additional negative charge on the serine side chain, an increase in the frequency and persistence of hydrogen-bond formation with positively charged residues is expected.

Hydrogen bonds were monitored throughout the entire trajectories. Bond formation was defined using a geometric criterion requiring a donor–acceptor distance not exceeding 4 Å and a donor–hydrogen–acceptor angle of at least 135°:

$$\mathrm{dist}(D - A) \leq 4\,\text{Å}; \mathrm{angle}(D - \mathrm{H}_D - A) \geq 135^{\circ} \tag{4.3}$$

For residue pairs with multiple possible donor–acceptor atom combinations, a hydrogen bond was considered present if at least one atom pair satisfied these criteria. To quantify the stability of the detected interactions, two descriptors were calculated. The first was the total hydrogen-bond lifetime, τ_{TOT}, defined as the cumulative time during which a given hydrogen bond was present throughout the trajectory. The second was the average lifetime, $\langle\tau\rangle$, corresponding to the average duration of individual hydrogen-bonding events.

To avoid artificially fragmenting long-lived interactions, an intermittency criterion was employed in the lifetime analysis. Owing to thermal fluctuations, hydrogen bonds may undergo short-lived disruptions despite remaining effectively associated from a physico-chemical perspective. Therefore, a hydrogen bond was considered continuously formed if interruptions between two bonding events lasted fewer than 20 frames (which correspond to 1 ns). This procedure allowed neighbouring bonding intervals separated by very short disruptions to be merged into a single interaction event; indeed, strict geometric criteria do not take into account thermal fluctuation that could artificially reduce the length of an otherwise persistent interaction. Such 20-frame threshold was chosen as a compromise between treating short interruptions as transient fluctuations and preserving the distinction between genuinely separate hydrogen-bonding events. Moreover, increasing the threshold further would risk merging distinct hydrogen-bonding events and artificially increasing their apparent persistence.

For each detected hydrogen bond i, the quantities τ_{TOT}^{i} and $\langle\tau\rangle^{i}$ were calculated. Only interactions satisfying the following criteria were retained:

$$\tau_{TOT}^{i} \geq 2.5\,ns \wedge \langle\tau\rangle^{i} \geq 2.5\,ns \tag{4.4}$$

The selected interactions were subsequently represented in **Figure 4.2** by plotting $\log \frac{\langle\tau\rangle^{i}}{\tau_0}$ as a function of $\log \frac{\tau_{TOT}^{i}}{\tau_0}$, where $\tau_0 = 2.5$ ns. This threshold was chosen to exclude transient contacts while retaining interactions persistent on the nanosecond timescale.

Interestingly, phosphorylation consistently leads to the formation of hydrogen bonds that are less frequently observed in the corresponding Wild-Type systems, confirming that the introduction of the dianionic phosphate group promotes the establishment of novel interactions. Furthermore, a subset of hydrogen bonds is shared among all three phosphorylated systems, most notably the SEP255A–R196A interaction, suggesting that these contacts are particularly favourable and persist independently of the overall conformational state.

Nevertheless, each system is characterised by specific dominant interactions. In the Closed state, the SEP226B–R221B hydrogen bond is present for more than one quarter of the total simulation time, highlighting its remarkable stability and suggesting a potential role as a persistent intra-protein bridge. In the phosphorylated trimeric complex, several hydrogen bonds involve residues belonging to either Cdc37 or CDK4. Among these, the SEP255B–$\mathrm{R288}_{\mathrm{CDK4}}$ interaction emerges as one of the most persistent contacts observed across all simulated systems.

Next, we computed probability-density-normalised $C_\alpha - C_\alpha$ distance histograms for the residue pairs

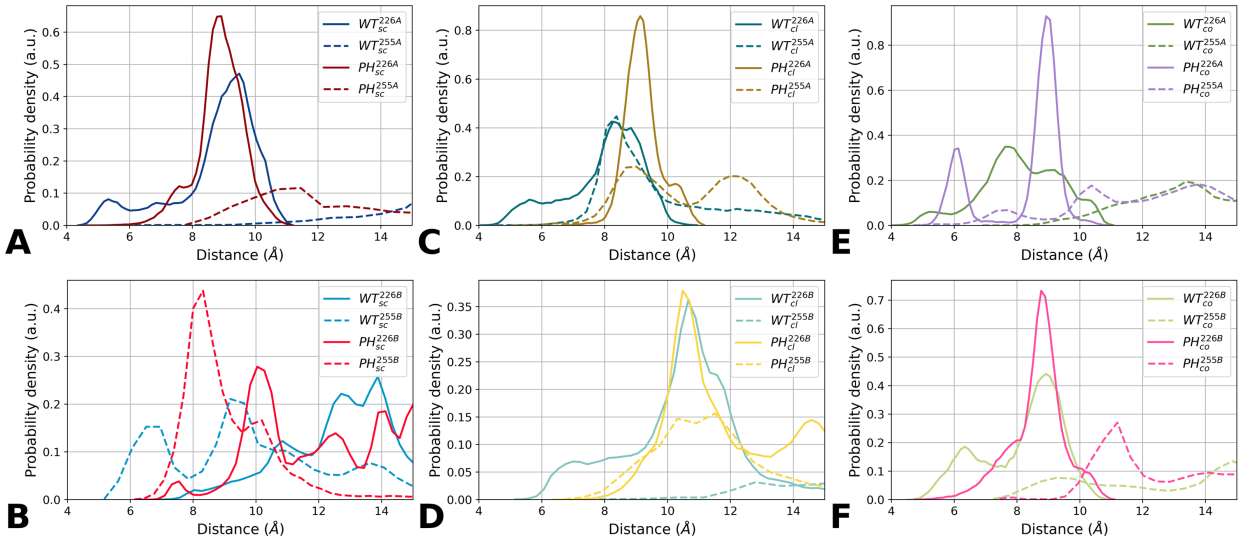

Figure 4.3: Distance distribution between C_α atom of serines and C_α of the Lys/Arg. *The longest lifetime in the phosphorylated states compared to the same WT distribution is plottedin dotted or solid line distinguish between the two serines in each monomer loop:* ***(A),(C),(E)*** *Distance distribution values for monomer A loop serine residues in Semi-Closed, Closed and Complex state in WT or phosphorylated form, respectively.* ***(B),(D),(F)*** *Distance distribution values for monomer B loop serine residues in Semi-closed, Closed and Complex states in WT or phosphorylated form, respectively.*

corresponding to the hydrogen bond exhibiting the longest τ_{TOT} for each phosphorylated serine residue and compared them with the corresponding residue pairs in the WT systems (**Figure 4.3**). In these plots, solid lines denote interactions involving Ser226, whereas dashed lines denote those involving Ser255. Different colours represent distinct monomers in the various conformational states. In all three Hsp90β states, phosphorylation sharpens the peaks of the $C_\alpha - C_\alpha$ distance distributions and shifts them towards shorter distances relative to the WT systems, indicating both a reduction in the average separation and a decrease in distance fluctuations. This behaviour suggests that the phosphate group attracts positively charged side chains, promoting the formation of a tighter and less mobile coordination shell. As a consequence, the sampling of more extended loop conformations is significantly reduced.

The results also indicate that the effects of phosphorylation are not restricted to the side chains directly involved in hydrogen bonding. Indeed, the distance distributions between backbone C_α atoms are also altered, suggesting that the local interactions induced by phosphorylation propagate to backbone dynamics and ultimately influence the conformational behaviour of the loop.

Furthermore, we computed radial distribution functions, $g(r)$ describing the probability density of $\mathrm{N}_\zeta^{\mathrm{LYS}}, \mathrm{C}_\zeta^{\mathrm{ARG}}$ atoms as a function of their distance r from each phosphoserine $\mathrm{O}_\gamma^{\mathrm{SER}}$. The cumulative coordination number, $N(r)$, defined as the running integral of $g(r)$, provides the average number of basic atoms located within a radius r. Each panel in **Figure 4.4** compares the corresponding distributions obtained for the phosphorylated and WT systems.

Across all six cases (two monomers in each of the three conformational states), the phosphorylated systems exhibit taller first-shell peaks in the $g(r)$ profiles, shifted towards shorter distances relative to the WT counterparts. In contrast, the WT distributions are broader and less pronounced. Consistently, the cumulative coordination number $N(r)$ increases more steeply in the phosphorylated systems, indicating a greater density of basic atoms in the immediate vicinity of the modified residues. These results further support the hypothesis that phosphorylation promotes the formation of a compact, positively enriched coordination environment that is absent in the WT systems. Overall, the data suggest that the phosphate groups act as electrostatic "molecular clamps", stabilising local

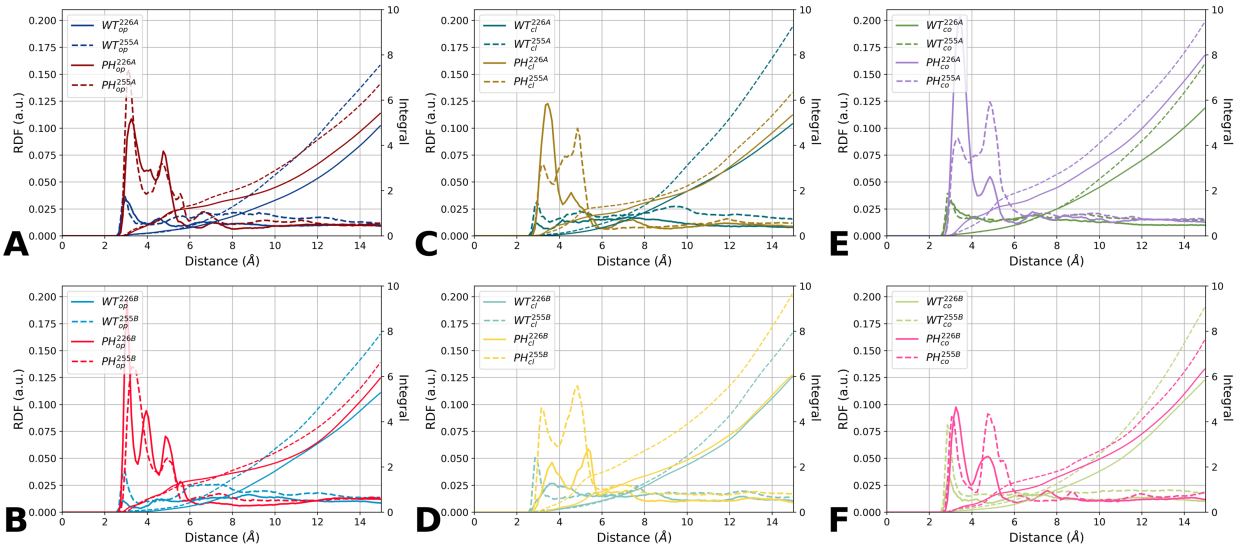

Figure 4.4: Radial distribution function. *Positive residues representative atoms* (N_{ζ}^{LYS}, C_{ζ}^{ARG}) *versus* O_{γ}^{SER} *in both phosphorylated and Wild Type systems are plotted; bold lines represent the rdf value, while thin lines represent the integrated value of rdf at increasing distances, dotted or solid line distinguish between the two serines in each monomer loop:* ***(A),(C),(E)*** *RDF and integral values for monomer A loop serine residues in Semi-Closed, Closed and Complex states in WT or phosphorylated form, respectively.* ***(B),(D),(F)*** *RDF and integral values for monomer B loop serine residues in Semi-Closed, Closed, Complex state in WT or phosphorylated form, respectively.*

interactions and restricting conformational freedom around the modified sites.

A possible consequence of the sequestration of positively charged residues around the phosphorylated serines is an increased exposure of negatively charged and polar groups to the solvent. To investigate this effect, sodium counterions were used as positively charged probes, and the average number of Na^+ ions within shells of increasing radius centred on the two charged loops was calculated (**Figure 4.5**).

The net charge difference between the WT and phosphorylated systems corresponds to four additional negative charges per monomer in the latter. However, the distribution of Na^+ ions around the loops varies substantially across the different conformational states and between the two monomers.

In the Semi-Closed state, the WT monomers display very similar ion distributions. In contrast, the phosphorylated system exhibits a systematic difference of approximately one unit charge between the two monomers at all shell radii, with monomer A being surrounded by a greater number of ions than monomer B. Relative to the WT system, phosphorylation results in an increase of approximately 2–3 associated sodium ions.

A similar trend is observed in the Closed state. The two WT monomers display comparable ion distributions, whereas the phosphorylated monomers exhibit a more differentiated behaviour.

The Complex state presents a markedly different scenario. In both the WT and phosphorylated systems, the two monomers display distinct ion distributions, differing by approximately two net charges. In the WT complex, monomer B is surrounded by a greater number of ions than monomer A. Upon phosphorylation, the opposite trend is observed, with the ion population becoming enriched around monomer A.

This behaviour may reflect differences in solvent accessibility associated with client binding. In the initial structure, monomer B is primarily involved in interactions with the client protein and is therefore expected to sample a smaller accessible volume within the lumen than monomer A, which remains comparatively free to explore the cavity. Consequently, changes in the local ion distribution may indirectly report on alterations in loop accessibility and protein–protein interactions.

Overall, these observations are consistent with phosphorylation promoting a stronger recruitment

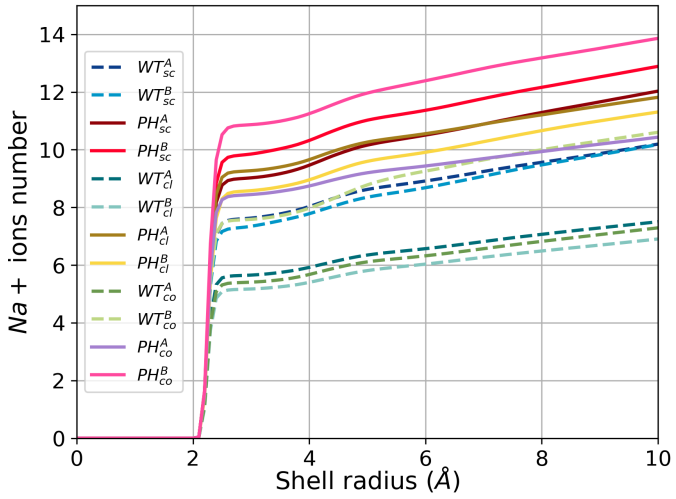

Figure 4.5: Sodium ions distribution. *Number of averaged sodium ions found in a given shell radius from the charged loop.*

of positively charged species around the modified loops and may indicate altered interactions with client proteins.

4.3.2 Phosphorylation affects coordination in Hsp90

To determine whether the observed modifications in contact networks affect the internal dynamics of the Semi-Closed, Closed, and Complex states of Hsp90β, and whether local perturbations in the loop regions propagate to distal regions of the protein, we computed pairwise distance-fluctuation (DF, see **Chapter 2**) matrices for all six systems.
To provide a compact representation of the dynamic differences induced by phosphorylation, we analysed the matrices:

$$\Delta DF = DF_{\mathrm{PHOS}} - DF_{\mathrm{WT}} \tag{4.5}$$

where positive values (red) indicate greater coordination (lower fluctuations) in the WT system relative to the phosphorylated one, whereas negative values (black) indicate the opposite behaviour (**Figure 4.7**).
The results reveal a complex, state-dependent allosteric response to phosphorylation. Within the N-terminal domains (NTDs), only limited changes in inter-NTD coordination are observed across the three states. A notable exception is the active-site region of protomer B in the Closed and Complex systems, where phosphorylation increases long-range coordination. This observation suggests that modifications at Ser226 and Ser255 may influence ATP processing and, consequently, the functional cycle of Hsp90β.
Analysis of MiD/CTD–NTD coupling reveals a reduction in coordination between the middle and C-terminal domains and the NTDs in both the Semi-Closed and Closed states. These results suggest that phosphorylation-induced perturbations in the loop regions partially decouple the middle domain from the nucleotide-binding machinery. In contrast, the Complex state exhibits the opposite behaviour, with a widespread increase in coordination between the NTDs and the MiD/CTD regions. A similar state dependence is observed for intra- and inter-protomer MiD/CTD interactions. The Semi-Closed and Closed systems display a modest reduction in coordination, whereas the Complex state exhibits increased internal coordination, consistent with stabilisation of the assembled

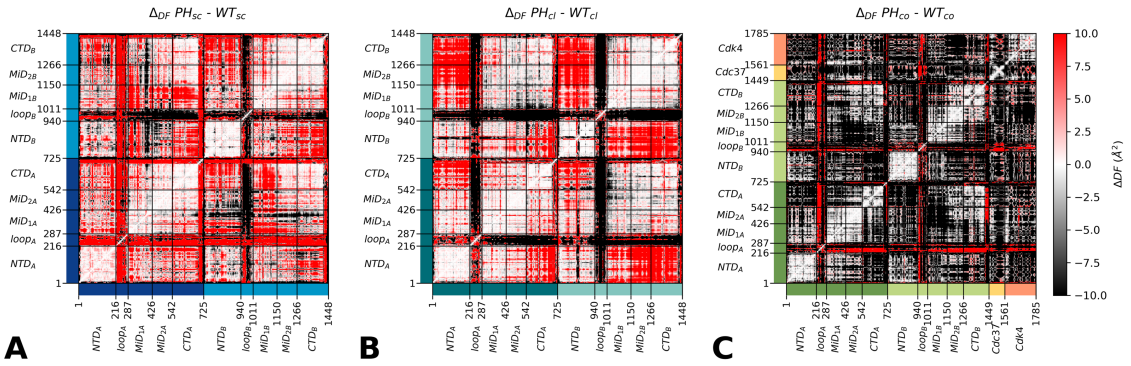

Figure 4.6:* ΔDF *matrices. *Matrices in different states:* ***(A)*** *Semi-Closed,* ***(B)*** *Closed and* ***(C)*** *Complex states.*

Hsp90β–Cdc37–CDK4 complex.

The loop regions show the most pronounced state-specific behaviour. In the Semi-Closed state, loop A loses coordination while loop B becomes more coordinated. In the Closed state, loop A gains coordination with the structural environment surrounding loop B while losing coordination with other regions of the protein, whereas loop B displays an overall increase in coordination. In the Complex state, loop A partially loses coordination with much of the assembly, while loop B specifically loses coordination with the cochaperone and client protein.

Analysis of the Hsp90β–Cdc37/CDK4 interface further indicates that phosphorylation enhances coordination between Hsp90β and both binding partners. This effect is particularly evident in the matrix region describing interactions between the loop regions and the client–cochaperone complex. In particular, phosphorylation of loop A markedly increases coordination with CDK4, providing a potential structural explanation for the stabilisation of the kinase-bound assembly.

To further investigate this hypothesis and characterise the conformational behaviour of the charged loops, we employed the grid command implemented in *CPPTRAJ* [174]. To analyse loop conformations, all trajectory frames were first aligned to the initial frame using only the backbone atoms belonging to the secondary-structure elements of Hsp90β. An average structure was then calculated from the aligned trajectory and used as the reference for a second alignment procedure, again performed on the backbone atoms of the secondary-structure elements. This protocol yielded trajectories in which the resolved domains of Hsp90β remained spatially fixed. A three-dimensional cubic grid with a spacing of 2.5 Å was subsequently generated using the grid command. For each voxel, the number of frames containing at least one atom belonging to the loop regions (residues 216–278) was recorded. The occupied volume was then visualised by plotting all voxels visited in at least 25% of the selected frames. The resulting occupancy volumes are shown in **Figure 4.7**.

In the Semi-Closed system, the loops occupy broadly similar regions of space in both the WT and phosphorylated forms. In protomer A, however, the phosphorylated loop samples a wider conformational space and occasionally extends towards the lumen, whereas the WT loop remains confined to a more compact and globular ensemble. In protomer B, the opposite trend is observed: the phosphorylated loop adopts a more compact conformation while still exploring regions located further from its average position compared with the WT loop.

In the Closed system, the WT loop of protomer A preferentially samples conformations extending towards the Hsp90 lumen, whereas the phosphorylated loop remains more compact. In protomer B, the overall behaviour is similar in both systems, although the phosphorylated loop occupies a slightly larger volume in front of the NTD.

In the Complex system, the phosphorylated loop of protomer A explores a broader conformational space and displays more frequent contacts with the lumen region, while the WT loop remains comparatively compact. The behaviour of protomer B is again similar in the two systems; however, the phosphorylated loop samples a more restricted conformational volume and exhibits a reduced

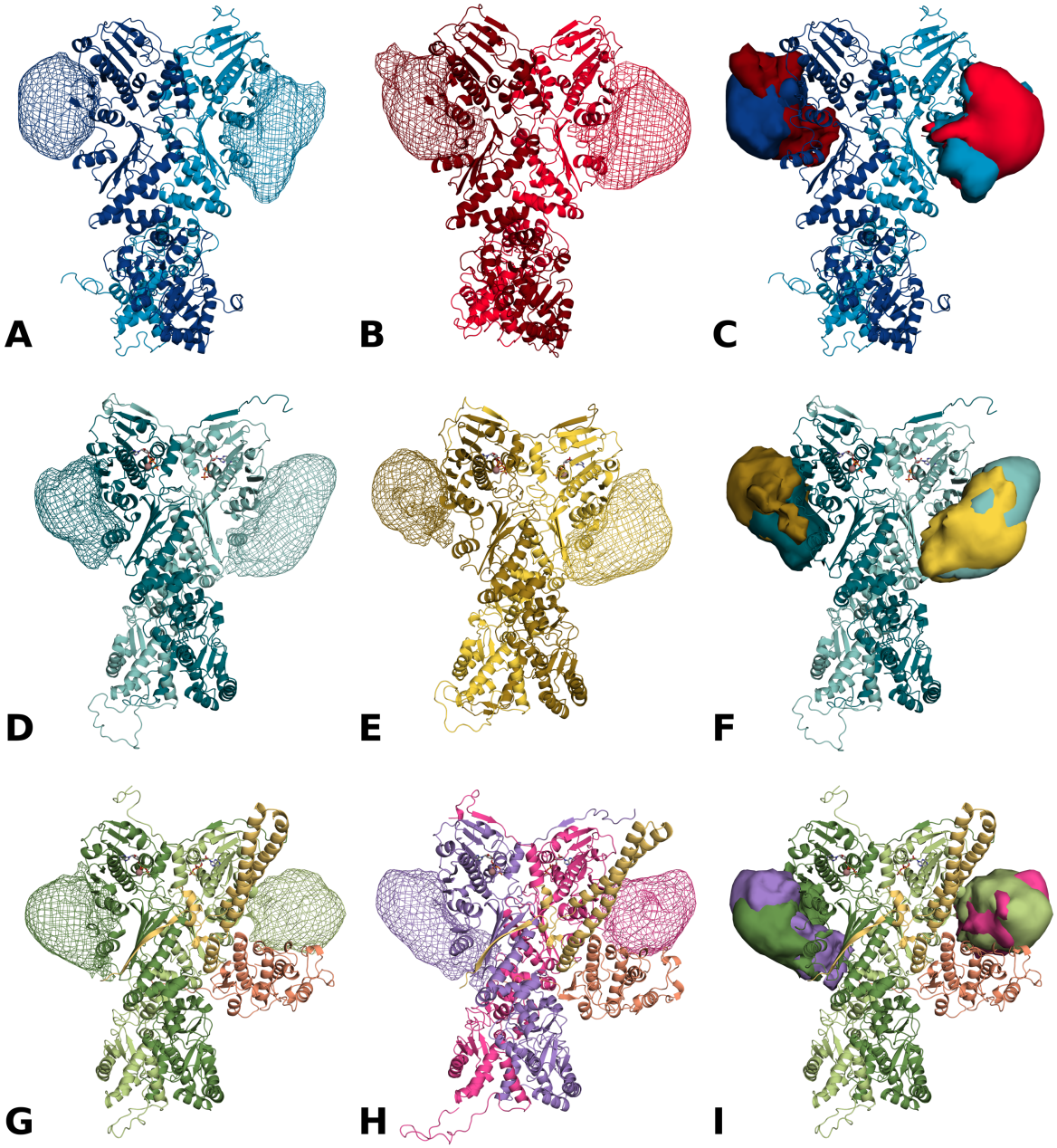

Figure 4.7: Charged loops exploration. ***(A),(D),(G)*** *Charged loops occupied volume in the Semi-Closed, Closed and Complex WT states, respectively.* ***(B),(E),(H)*** *Charged loops occupied volume in the Semi-Closed, Closed, Complex phosphorylated states, respectively.* ***(C),(F),(I)*** *Comparison between charged loops occupied volume in the Semi-Closed, Closed, Complex states respectively, between WT and phosphorylated state, with phosphorylated structure superposed onto the WT one and not displayed for better clarity.*

interaction surface with the CDK4 client. The differences in volume reported in **Figure 4.7** highlight a global tendency, as these differences are consistently observed across the different replicas and are therefore not attributable to a single outlier replica. The intrinsic stochasticity of the loop dynamics can nevertheless result in differences in the transient conformations reached by individual replicas. However, the main objective of this study was to identify global differences in the average conformational space explored by the loop upon phosphorylation. Therefore, differences in individual transient states were not further investigated.

4.3.3 Collective motions analysis

To gain further insight into the global behaviour of Hsp90β in the Wild-Type and phosphorylated systems, we applied a dimensionality-reduction approach based on the Tensor of Gyration (TOG) method, introduced in **Chapter 2**. In this framework, key protein domains are approximated as ellipsoids. Residues belonging to each domain are selected according to their DF score and subsequently mapped onto seven representative points: the centre of mass of the domain and two points along each semi-axis of the corresponding ellipsoid. This reduced representation facilitates the identification of large-scale collective motions through principal component analysis (PCA), as well as through the calculation of Superposition and Alignment maps (see **Chapter 2**), displayed in **Figure 4.8**. For this study, Hsp90 domains were subdivided in the same domains used in **Chapter 3**, while Cdc37 and Cdk4 were subdivided in one domain each (residues 20 to 108 for Cdc37 and residues 95 to 287 for Cdk4).

For the Semi-Closed state, the Superposition map indicates that the N-terminal domains (NTDs) are the regions most strongly affected by phosphorylation. Even after optimal translational alignment, the distance distributions between the NTDs and the remaining domains of the same protomer remain poorly super-imposable, suggesting that phosphorylation alters the conformational ensembles sampled during the dynamics. The Alignment map provides a more detailed picture, confirming that the NTDs undergo subtle but distinct motions. In protomer A, the NTD tends to remain closer to the middle and C-terminal domains, whereas in protomer B it displays a tendency to move away from them. Although these displacements are relatively small, typically on the order of a few tenths of an Ångström, the most pronounced effect of phosphorylation is observed in the inter-protomer motions. The Alignment map suggests an overall closure of the dimeric assembly, with the exception of NTDA, which tends to move away from NTDB. Indeed, most pairwise distance distributions require a positive translational shift to recover behaviour comparable to that observed in the WT system. Despite this overall trend, the magnitude of the displacement remains limited, with typical variations of approximately 2 Å between reduced points.

In the Closed state, the Superposition map similarly indicates that the conformational ensembles sampled by the WT and phosphorylated systems remain only partially comparable, even after alignment. The Alignment map reveals a behaviour broadly consistent with that observed in the Semi-Closed state: protomer A tends to undergo a slight closing motion, whereas protomer B exhibits a modest opening tendency following phosphorylation. Once again, the most evident effect is the increased proximity between the MiD domains of protomer A and protomer B, while NTDA moves further away from the remainder of the assembly.

The results obtained for the Complex state are consistent with those observed in the Closed state. The Superposition map highlights differences in conformational behaviour involving several Hsp90β domains, as well as regions belonging to the cochaperone and client proteins. The Alignment map again reveals a closing motion involving the MiD domains of the two protomers following phosphorylation. In contrast, the CTDB domain displays an overall displacement away from the remaining domains. Regarding the interaction partners, Cdc37 remains closer to the NTDs in the phosphorylated system, while CDK4 preferentially approaches the MiD and CTD regions. These observations are consistent with enhanced interactions between the client protein and phosphorylated Hsp90β.

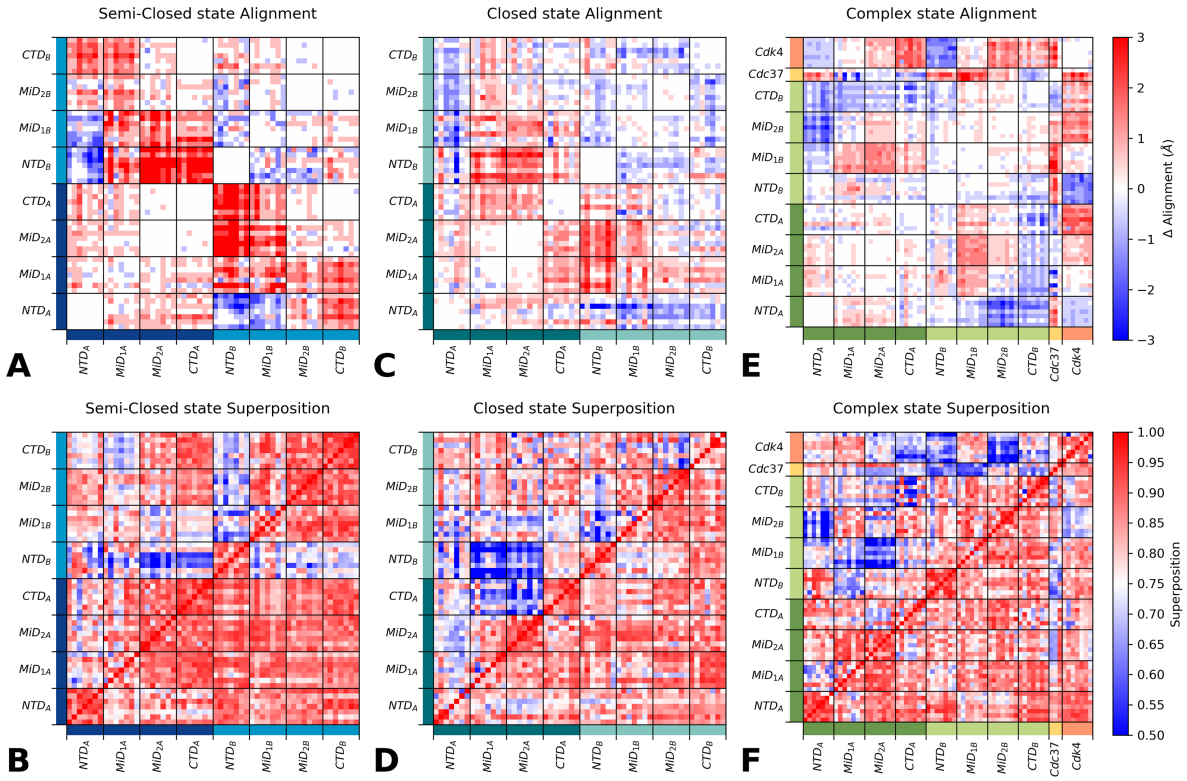

Figure 4.8: Alignment maps and Superposition maps. *Different TOG maps in states:* ***(A),(B)*** *Semi-Closed,* ***(C),(D)*** *Closed and* ***(E),(F)*** *Complex states*

Having identified the main structural differences between the phosphorylated and Wild-Type systems, we next performed principal component analysis (PCA) on the reduced TOG trajectories. To facilitate structural interpretation, the first principal component was projected onto the reduced-point representation of each system **Figure 4.9A-F**.

In the Semi-Closed state, the most evident difference is observed in the motion of NTDA. In the WT system, the dominant motion is directed outwards, whereas in the phosphorylated system the displacement vectors are more dispersed and do not define a single preferential direction. A second notable difference involves the MiD1B and MiD2B domains. In the WT system, these domains exhibit an outward-directed motion, while no comparable large-scale displacement is observed in the phosphorylated state.

The Closed state displays markedly different dominant motions in the two systems. In the WT system, the first principal component corresponds primarily to a partial closing motion of the CTDs, which move towards one another. In contrast, the phosphorylated system exhibits more heterogeneous motions in the CTD region, while the dominant displacement is associated with NTDA, which tends to move towards the lumen.

For the Complex state, PCA was performed both on the complete Hsp90β–Cdc37–CDK4 assembly and on the Hsp90β–CDK4 subsystem. This additional analysis was necessary because Cdc37 contains a double-helical domain that undergoes a large-amplitude swinging motion, accounting for a substantial fraction of the total variance. Consequently, PCA performed on the complete complex is dominated by fluctuations associated with this region and provides limited information on the dynamics of Hsp90β itself.

When the analysis is focused on Hsp90β, the first principal component corresponds to a global twisting motion around the lumen. This motion is observed in both systems but appears to be more pronounced in the phosphorylated state.

Despite the limited visual similarity between the dominant principal motions of the WT and

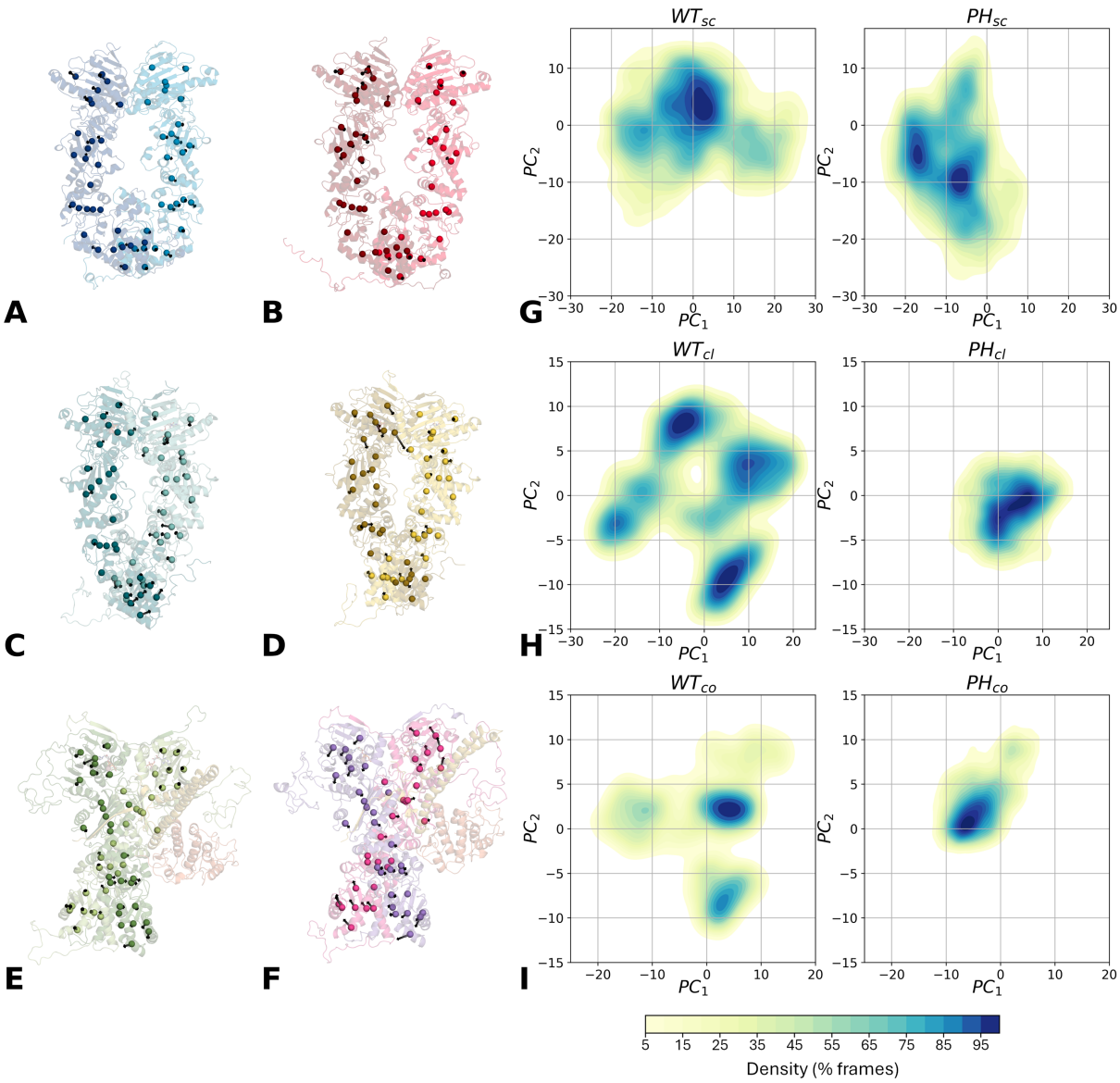

Figure 4.9: First principal component of TOG trajectories. ***(A),(C),(E)*** *WT Semi-Closed, Closed, Complex states respectively.* ***(B),(D),(F)*** *Phosphorylated Semi-Closed, Closed, Complex states respectivel.,* ***(G),(H),(I)*** *KDE plots comparison of the first two principal components between phosphorylated system and WT one in Semi-Closed, Closed, Complex states respectively.*

phosphorylated systems, we projected the trajectories of the phosphorylated systems onto the first two principal components obtained from the corresponding WT simulations **Figure 4.9G-I**. In all cases, the phosphorylated systems explored a smaller region of the principal-component space, further supporting the hypothesis that phosphorylation alters the ability of Hsp90β to sample its conformational landscape.

To establish a more direct connection between the principal motions and physically meaningful structural descriptors, we sought to identify collective variables capable of capturing the dominant conformational changes of Hsp90β. Two variables were selected. The first was the angle defined by the vectors connecting the centres of mass of the NTDs and CTDs, which provides a measure of the relative wrapping or twisting of the two protomers (**Figure 4.10A**). The second was a reduced lumen volume, calculated through triangulation of representative points facing the lumen cavity (**Figure 4.10B**).

This combination of collective variables highlights the effects of phosphorylation on the conformational cycle of Hsp90β more clearly than the PCA projections alone. In the Semi-Closed state, the phosphorylated system is characterised by a lower average lumen volume and a more compact NTD arrangement. Larger values of the angular variable correspond to increased twisting of the NTDs relative to the CTDs (**Figure 4.10C**). In the Closed state, the differences between the two systems become less pronounced, although the phosphorylated system displays a slightly more open lumen while maintaining a similar NTD–CTD angle distribution (**Figure 4.10D**).

The Complex state provides the clearest distinction between the two systems. The phosphorylated complex is characterised by a single basin corresponding to low lumen volumes, whereas the WT system samples a broader conformational region. In particular, the WT ensemble accesses partially open lumen conformations together with larger angular displacements between the NTDs and CTDs (**Figure 4.10E**). Overall, these results indicate that phosphorylation promotes a more compact conformational ensemble and restricts the large-scale motions accessible to the Hsp90β complex.

4.4 Conclusions

In this study, molecular dynamics simulations were employed to investigate the allosteric effects induced by phosphorylation of four serine residues in Hsp90β. Experimental evidence suggests that this post-translational modification may promote the development of certain carcinomas, potentially through enhanced interactions with cochaperones and client proteins. Our analyses provide an atomistic description of how phosphorylation alters the structural and dynamical properties of Hsp90β and its associated complexes [152, 203, 207].

A central finding of this work is the ability of phosphorylation to reshape the local interaction network surrounding the modified residues. The introduction of dianionic phosphate groups promotes the formation of persistent electrostatic interactions with nearby Arg and Lys residues, generating hydrogen-bonding patterns that are either absent or significantly less populated in the WT systems. In several cases, a single phosphoserine simultaneously engages multiple interaction partners (**Figure 4.2G**), highlighting its role as a multivalent electrostatic hub. Collectively, these interactions stabilise the charged loop region and reduce its conformational freedom.

The consequences of these local perturbations extend well beyond the phosphorylation sites. Loop-volume analyses reveal a more compact conformational ensemble in the phosphorylated systems, while TOG-based dimensionality reduction and collective-variable projections indicate a tendency towards reduced flexibility and more compact lumen conformations. These observations suggest that phosphorylation shifts the conformational equilibrium of Hsp90β towards more constrained structural states. Through allosteric coupling, these effects propagate to distal regions of the protein, altering inter-domain communication and modifying the collective motions that underlie Hsp90β function.

The impact of phosphorylation is particularly evident in the Hsp90β–Cdc37–CDK4 complex. Transient hydrogen bonds between the phosphorylated loops and residues belonging to the cochaperone

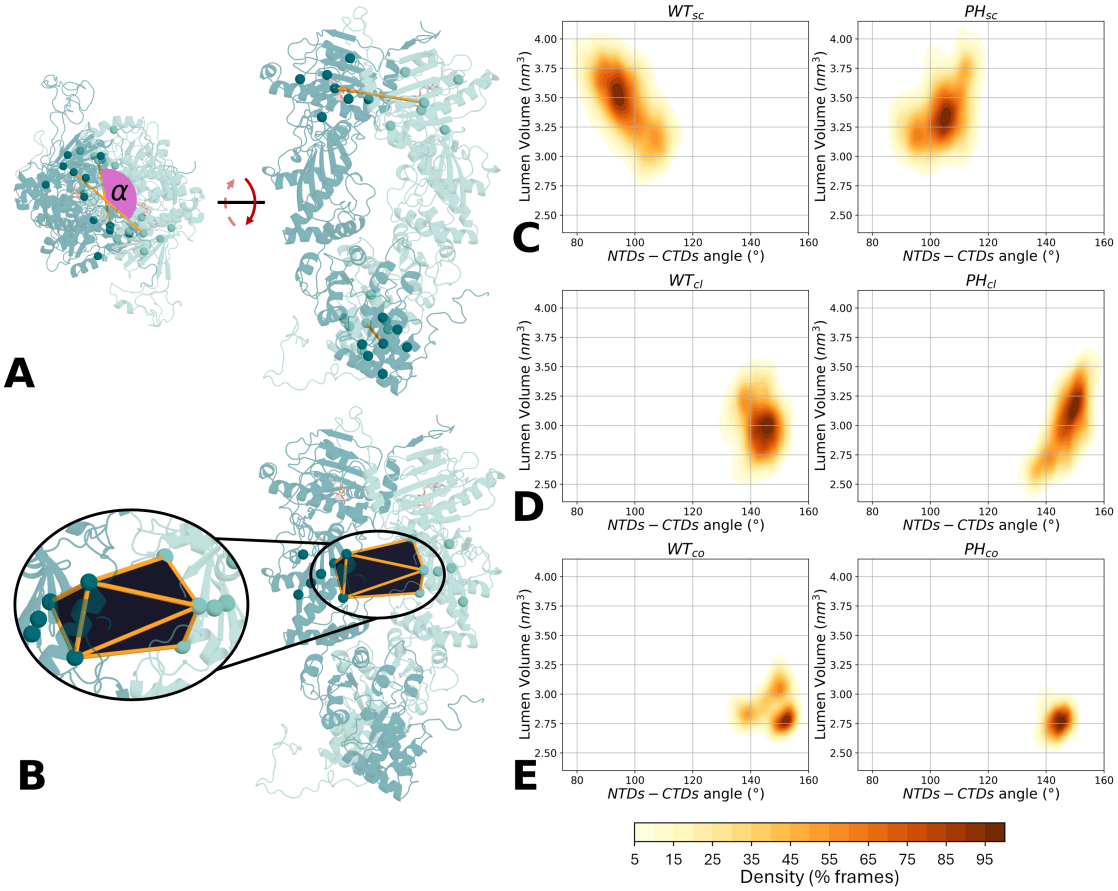

Figure 4.10: Collective variables capturing twisting-closure motion of Hsp90. ***(A)*** *First CV is the angle between NTDs and CTDs connecting vectors (twisting).* ***(B)*** *Second CV is the volume of Hsp90 lumen (triangulation of TOG points facing lumen).* ***(C-E)*** *KDE comparison plot of distribution of CVs in the Semi-Closed, Closed and Complex state.*

or client protein, together with the enhanced coordination observed across the assembly, suggest that phosphorylation favours the formation and persistence of intermolecular contacts. Such behaviour may contribute to a stabilisation of the complex and to an altered balance between assembly, client processing, and complex dissociation.

Overall, these results demonstrate how a local post-translational modification can propagate across multiple structural scales, from the formation of specific electrostatic interactions to global changes in protein dynamics and complex organisation. More broadly, this work highlights the ability of phosphorylation to modulate the functional landscape of molecular chaperones through coupled local and allosteric effects. While the present study is primarily mechanistic in nature, it suggests that selectively targeting specific conformational or interaction states—rather than globally inhibiting ATPase activity—may represent a promising strategy for modulating Hsp90-associated pathways [208, 209].

Chapter 5

Allosteric regulation of Eg-5 motor domain

5.1 Introduction

Cancer is one of the most prevalent diseases worldwide, and its incidence is projected to increase substantially over the next 25 years. Chemotherapy remains a cornerstone of cancer treatment; however, the clinical use of chemotherapeutic agents is often limited by their intrinsic toxicity, which is frequently linked to their mechanism of action. Consequently, adverse effects restrict the doses that can be safely administered. For example, the antimitotic agents currently used in clinical practice, including taxanes, vinca alkaloids and epothilones, target human tubulin and disrupt its normal dynamics. As tubulin is an essential component of the mitotic spindle, its inhibition results in mitotic arrest and subsequent cell death. However, tubulin is also required for the normal function of healthy cells, which largely explains the neurotoxicity associated with antimitotic agents and the dose limitations imposed by this adverse effect [80, 211, 212].

Mitotic kinesin-5, a motor protein required for the formation of the bipolar mitotic spindle and the correct alignment of chromosomes at the metaphase plate, has emerged as an attractive target for the development of novel antimitotic agents [213]. Its involvement in tumour initiation and progression is well established, and overexpression of kinesin-5 has been reported in a variety of cancer cell types [83, 214–218].

Structurally, kinesin-5 is a homotetramer arranged as two antiparallel dimers, giving rise to a characteristic dumbbell-shaped architecture (**Figure 5.1A**). Each end of the tetramer binds to a microtubule originating from opposite spindle poles, thereby cross-linking the microtubules. Each monomer consists of an N-terminal motor domain, a neck linker, a coiled-coil stalk and a tail domain [78, 219]. The motor domain is responsible for ATP hydrolysis and microtubule binding, whereas the neck linker coordinates the movement of the motor domains within each dimer. The coiled-coil stalk mediates the assembly of the four monomers, initially through the association of two monomers into a dimer, followed by the interaction of two dimers to form the tetramer. Furthermore, the stalk domain is thought to interact with the opposing dimer, reducing its movement and thereby enhancing the coordination of the entire complex [79]. Through ATP hydrolysis, kinesin-5, a plus-end-directed motor protein, generates the forces required for bipolar spindle assembly, cell elongation and the correct positioning of chromosomes within the mitotic spindle. Inhibition of kinesin-5 causes premature mitotic exit, ultimately leading to cell death through apoptosis or the formation of aberrant daughter cells (**Figure 5.1B**). In addition, kinesin-5 has been shown to exert a stabilising effect on growing microtubules, which is essential for spindle assembly [82].

The motor domain of kinesin-5 (**Figure 5.1C**) shares several structural features with the motor domains of other kinesin superfamily members, including a central core composed of eight β-sheets surrounded by six α-helices [219]. The catalytic region contains a phosphate-binding loop (P-loop) and two switch regions (SWI and SWII). Key residues within these domains, namely Ser233, Arg234

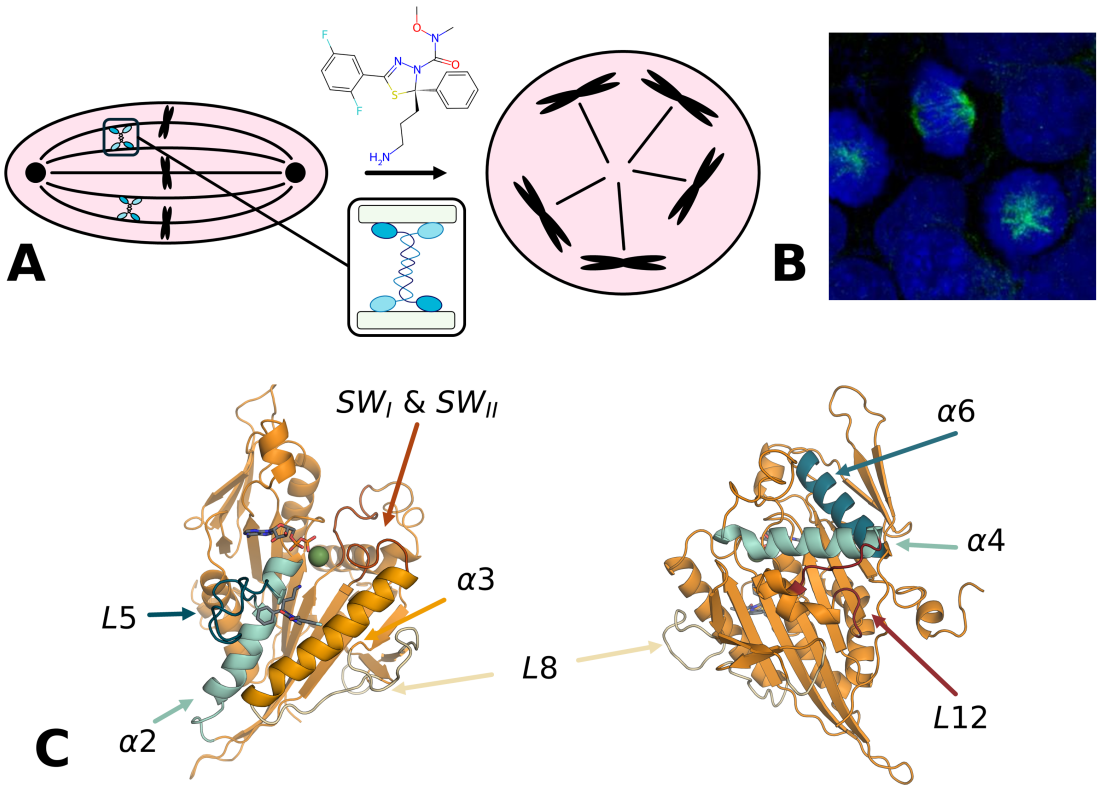

Figure 5.1: Kinesin-5 is necessary for mitotic spindle development. ***(A)*** *Schematic representation of filanesib inhibition of kinesin-5 in a mitotic cell, producing a monopolar spindle.* ***(B)*** *Fluorescence microscopy image of a normal bipolar mitotic spindle (middle) and two monopolar spindles due to kinesin-5 inhibition [210].* ***(C)*** *Kinesin-5 motor domain. Key regions (including L5) are labeled and highlighted in different color; remaining regions are rendered in dark orange.* ***(B)*** *panel was reproduced from reference [210] with permission of the Royal Society of Chemistry.*

and Glu264, the latter two forming a salt bridge, are sensitive to the presence of the ATP γ-phosphate and are directly involved in ATP hydrolysis across the kinesin family [81]. Similar to their counterparts in G proteins and myosins, the switch regions transmit the effects of nucleotide binding and hydrolysis to other functionally important regions through allosteric pathways. These regions include helix α4, which, together with loops L8 and L12, forms part of the microtubule-binding interface, and helix α6, which is directly connected to the neck linker.

Like other kinesins, kinesin-5 contains loop L5 (residues Glu116–Gly134), which protrudes from helix α2. Notably, L5 is the longest loop of this type within the kinesin superfamily [78–80]. It has been proposed that this and other subtle structural differences fine-tune the kinetic properties of individual kinesins, thereby adapting them to their specific cellular functions [220, 221]. For example, kinesin-1 is highly specialised for vesicle transport, a process that requires rapid movement but relatively low force generation. In contrast, kinesin-5 functions in mitotic spindle assembly, a process in which speed is less critical than the ability to withstand substantial mechanical forces generated within overlapping microtubules. Under extreme conditions, these forces may even trap kinesin-5 in a so-called "handbrake state" [222–224]. This functional specialisation is further illustrated by studies on a chimeric kinesin-5 from Xenopus laevis, in which the native motor domains were replaced with those of kinesin-1. Although the resulting protein retained the ability to move along antiparallel microtubules, it was unable to support proper bipolar spindle assembly [225]. An even more striking example of structural diversification within the kinesin superfamily is provided by kinesin-3 (KIF14), which possesses an N-terminal extension of approximately 350 amino acids—almost equivalent in size to the motor domain itself. This region is believed to confer enhanced processivity and improved performance in crowded intracellular environments [226].

Loop L5 plays a crucial role in kinesin-5 function, as demonstrated by mutagenesis studies and by the existence of a distinct class of kinesin-5 inhibitors that bind to a pocket formed by L5 together with helices α2 and α3. Interestingly, this binding pocket is observed only in crystallographic structures of kinesin-5 complexed with these inhibitors and is absent from representative ATP- and ADP-bound structures. It is therefore generally considered to be an induced pocket that forms upon inhibitor binding. The involvement of L5 in ATP hydrolysis is well established, as deletion mutants lacking this loop exhibit markedly reduced ATPase activity. Furthermore, L5 has been shown to contribute to ADP release and to prevent premature binding of the motor domain to microtubules prior to nucleotide release [227–230].

Despite their allosteric mode of action, L5-binding inhibitors do not appear to induce major structural changes in the crystallographic conformation of kinesin-5. This observation is somewhat counter-intuitive, as allosteric inhibition is often associated with substantial conformational rearrangements. An additional paradox arises from the observation that certain resistance-conferring mutations, namely Asp130Val, Ala133Asp and the corresponding double mutant, increase the enthalpic contribution to inhibitor binding while decreasing the overall binding free energy owing to a larger entropic penalty [231]. As noted previously, L5 is present in all kinesins; however, inhibitors developed against human kinesin-5 are highly selective and generally show no activity towards other kinesins, including kinesin-5 homologues from different species. Remarkably, chimeric non-human kinesin-5 proteins in which the native L5 sequence is replaced by the human counterpart become susceptible to inhibition by human kinesin-5 inhibitors [232]. This finding suggests that the allosteric mechanisms regulated by L5 are evolutionarily conserved, irrespective of loop length. Nevertheless, the precise molecular mechanism through which L5 exerts its regulatory role remains poorly understood.

The catalytic cycle of the kinesin-5 motor domain is intimately linked to its biological function and can be divided into three principal nucleotide states: ATP-bound, ADP-bound and apo (nucleotide-free). The ATP-bound and apo states exhibit high affinity for microtubules, whereas the ADP-bound state displays low affinity. The sequential transition between these states underlies the processive behaviour of the motor protein. Conceptually, kinesin movement can be likened to a "walking" mechanism (**Figure 5.2**), in which one motor head remains attached to the microtubule while the trailing head is translocated forward and subsequently re-establishes a strong microtubule interaction.

Once the leading head is firmly attached, the rear head can detach and advance, thereby continuing the cycle.

Initially, kinesin approaches the microtubule with both motor domains in the ADP-bound state [233] (**Figure 5.2A**). The leading motor domain then releases ADP and binds tightly to the microtubule (**Figure 5.2B**). ATP binding induces a conformational change that is transmitted through the motor domain to the neck linker, which is believed to promote forward displacement of the trailing ADP-bound head [81, 234] (**Figure 5.2C**). More recent studies on kinesin-1 have suggested that a Brownian ratchet mechanism provides the best description of motor stepping, with the neck linker acting primarily as a facilitating element rather than the sole driving force [235]. A similar mechanism has also been proposed for kinesin-5 [236]. Following translocation, the newly leading ADP-bound motor domain releases ADP [237] and binds strongly to the microtubule through conversion to the high-affinity apo state [220]. This transition permits the ATP-bound trailing head to hydrolyse ATP and detach from the microtubule. The apo state additionally induces a conformational change in tubulin that further enhances binding affinity [238, 239] (**Figure 5.2D**). Binding of a new ATP molecule to the leading head then initiates a new catalytic cycle (**Figure 5.2E-G**).

In processive transport kinesins, these conformational transitions are translated into directed motility. In contrast, kinesin-5 functions primarily as a force-generating and force-resisting motor. The mechanical force generated during nucleotide cycling is transmitted through the microtubules to which the motor is bound. Because the two ends of the kinesin-5 tetramer interact with antiparallel microtubules, this force promotes microtubule sliding and spindle pole separation. The alternating stepping and braking activities of kinesin-5 are therefore essential for proper spindle assembly [225].

To date, two distinct classes of kinesin-5 inhibitors have been identified. Both act through allosteric mechanisms but interfere with the catalytic cycle in fundamentally different ways. The largest and best-characterised class binds to the pocket formed by L5, $\alpha 2$ and $\alpha 3$ and is believed to stabilise an ADP-like state characterised by reduced ADP release and diminished microtubule affinity, as demonstrated by crystallographic and kinetic studies [241, 242]. Kinetic analyses further indicate that these inhibitors do not interfere with ATP binding, i.e., the transition from the apo to the ATP-bound state. Consequently, inhibitor binding must occur either in the ATP-bound or ADP-bound conformations. This conclusion is supported by calorimetric studies performed by Seth and co-workers using ispinesib, which demonstrated that inhibitor binding does not occur in the apo state [243]. Moreover, kinetic studies showing reduced ADP release in the presence of inhibitors [227, 241], together with the extensive collection of crystallographic structures of inhibitor-bound kinesin-5, strongly suggest that these compounds preferentially stabilise an ADP-like motor-domain conformation.

Several representatives of this inhibitor class, including ispinesib, AZD4877 [244], SB743921 [245], litronesib, and filanesib [246] have progressed to Phase I and Phase II clinical trials. These compounds have shown encouraging activity, particularly when administered in combination with other therapeutic agents such as proteasome inhibitors and dexamethasone. However, none has yet achieved regulatory approval. Among these compounds, filanesib is particularly attractive for mechanistic studies owing to its superior conformational stability and considerably longer half-life ($>$90 h compared with $<$30 h for related inhibitors) [219].

When combined with appropriate analytical approaches, molecular dynamics (MD) simulations constitute a powerful tool for investigating the allosteric mechanisms that regulate kinesin-5 function throughout the ATPase cycle and in the presence of allosteric inhibitors. Although several MD studies of the kinesin-5 motor domain have been reported [247–257], only a small proportion have explicitly focused on the allosteric processes underlying inhibition by L5-binding ligands.

In the present study, we employ extensive MD simulations to characterise and compare the allosteric communication networks operating within the kinesin-5 motor domain at key stages of its catalytic cycle. Particular emphasis is placed on the role of L5 and on the nucleotide-free apo state, which has received comparatively limited attention. Furthermore, we investigate the allosteric mechanism by which filanesib perturbs the functional cycle of kinesin-5, providing new insights into its mode

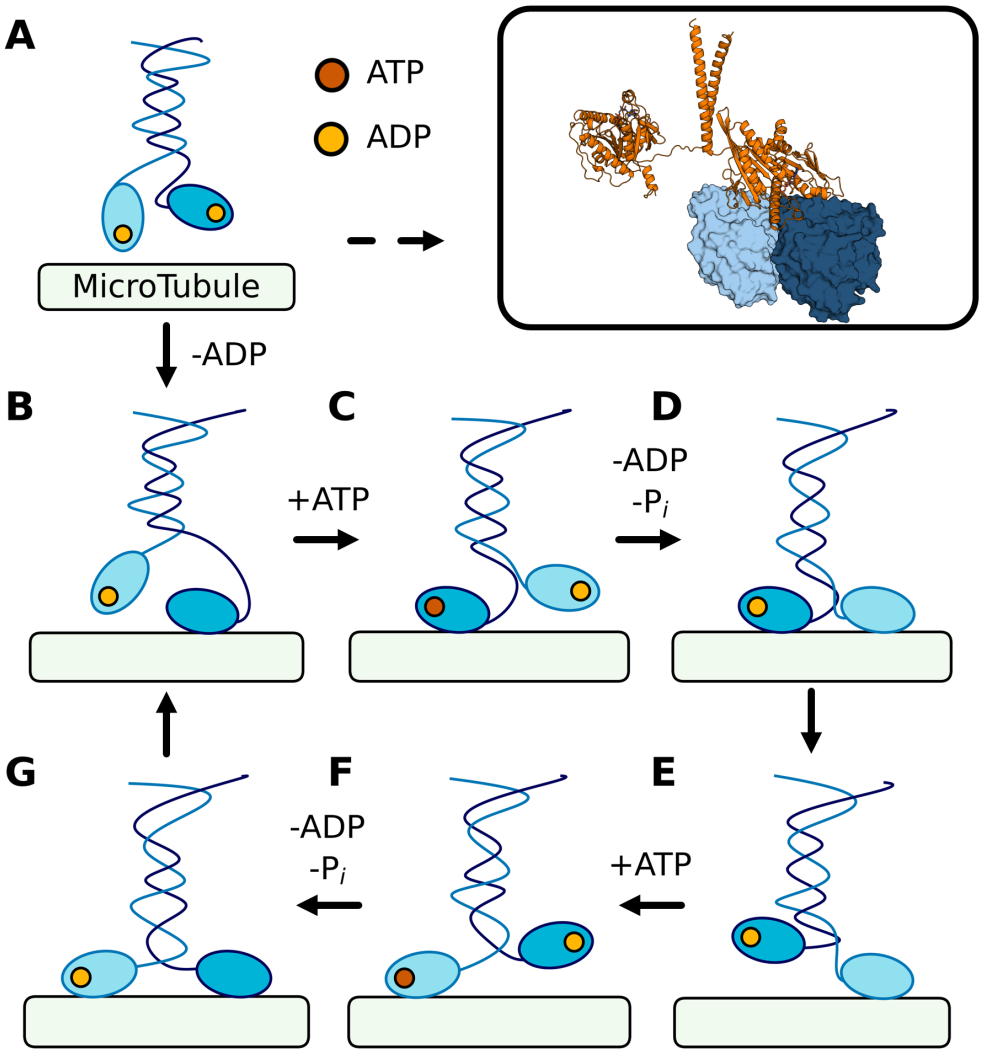

Figure 5.2: Kinesin-5 ATP cycle *Schematic representation of the catalytic cycle of a kinesin-5 dimer. In the separated panel, a model of a kinesin-5 dimer (both protomers in orange) with one of the motor domains bound to a ab-tubulin dimer (cyan and blue), using the Cryo-EM structure of kinesin-5 with AMPNP (PDB ID: 6TA4) as a template: from the model, it can be seen that the binding interface of kinesin-5 with the tubulin dimer is composed of L2, L8, SWII, α4, L12 and α6 [240]. Panels are described in the main text.*

of action. To the best of our knowledge, a systematic comparison of the ATP-bound, ADP-bound, apo and inhibitor-bound states using microsecond-scale MD simulations has not previously been reported.

5.2 Molecular dynamics and protein preparation

Molecular dynamics (MD) simulations were performed on four representative states of the kinesin-5 motor domain catalytic cycle. These comprised the ATP-bound state (ATP), the ADP-bound state (ADP), the nucleotide-free apo state (apo), and the ADP-bound state in complex with filanesib (ADPfil), which was included as an inhibited state deviating from the physiological catalytic cycle (**Figure 5.2**).

For each state, six independent MD replicas were generated and simulated for 1 µs, yielding a total simulation time of 24 µs across all systems. The resulting trajectories were subsequently combined and subjected to a series of analyses aimed at characterising the conformational dynamics of the motor domain and identifying patterns of allosteric communication associated with each catalytic state.

The ADP and ADPfil systems were prepared from their respective high-resolution crystal structures (PDB IDs: 1II6 [258] and 6HKY [259], respectively). The crystal structure 6HKY contains three monomers displaying only minor structural differences (RMSD relative to chain A: 0.5 Å for chain B and 3.3 Å for chain C). Chain A was selected for subsequent studies, as it contained the largest number of resolved residues.

The ATP-bound form was generated from the crystal structure of kinesin-5 in complex with the non-hydrolysable ATP analogue AMPPNP (PDB ID: 3HQD [260]) by replacing the imido nitrogen atom of the β–γ phosphate bridge with an oxygen atom.

At the time of this study, no experimental structure of the human apo form had been reported. Consequently, the apo model was generated from the human ATP-bound structure by removing the nucleotide and incorporating structural information from the apo form of the OSM-3 kinesin from Caenorhabditis elegans (PDB ID: 7A40). The OSM-3 kinesin exhibits high structural similarity to human kinesin-5, both in the ATP-binding region [261] and across the motor domain as a whole. Specifically, the backbone RMSD between the apo and AMPPNP-bound forms of OSM-3 (PDB ID: 7A5E) is only 1.4 Å, whereas the AMPPNP-bound OSM-3 structure differs from AMPPNP-bound human kinesin-5 by 2.4 Å. On this basis, we inferred that the apo form of human kinesin-5 would closely resemble the apo conformation of OSM-3.

A further justification for using the ATP-bound structure as the starting point for apo modelling was the observation that both the ATP-bound and apo states exhibit high affinity for microtubules and are therefore expected to adopt broadly similar conformations. This hypothesis is supported by previous studies showing that kinesin-5 inhibitors binding at the $\alpha4/\alpha6$ site, which stabilise microtubule association, induce conformations that more closely resemble the ATP-bound state than the ADP-bound state [240].

Despite the sequence differences between the two kinesins, the overall structures of apo OSM-3 and AMPPNP-bound human kinesin-5 are highly similar. The most notable differences are found in loop L5, which is substantially shorter in the C. elegans protein, and in helix $\alpha3$, which exhibits a slight tilt relative to its human counterpart. As the C-terminal region of helix $\alpha3$ is directly connected to the SWI domain, residues 192–205 of OSM-3, corresponding to the C-terminal portion of $\alpha3$ and the SWI region, were modelled into the homologous segment of human kinesin-5 (residues 221–234). Notably, the OSM-3 SWI region contains the conserved catalytically relevant sequence SSRSHS, corresponding to residues 232–237 in human kinesin-5. The remaining residues within the transplanted segment were subsequently mutated in silico to match the human sequence while preserving the original side-chain orientations. All other regions of the apo model, including loop L5, were inherited directly from the ATP-bound structure.

Missing residues were reconstructed in all systems using Modeller v10.5 [118], employing the

full-length protein sequence obtained from UniProt (accession P52732). This procedure included reconstruction of residues 1–16 in all systems except the ADP-bound structure, for which residues 1–15 were modelled. The resulting systems were solvated in cubic TIP3P water boxes extending 12 Å beyond the protein in all directions. Sodium and chloride ions were subsequently added using LEaP [66] to neutralise the systems and achieve a final ionic strength of 0.1 M.
The structure and binding pose of filanesib were extracted directly from the crystal structure 6HKY. The ligand was initially processed in PyMOL to ensure correct assignment of bond orders and saturation states. The aliphatic amine group was modelled in its protonated form, based on its close chemical analogy to a lysine side chain. Geometry optimisation was subsequently performed using Gaussian16 [125] at the B3LYP/6-31+G(d,p) level of theory [126, 262], followed by a single-point energy calculation at the HF/6-31+G(d,p) level. The resulting wavefunction was used to derive atomic partial charges according to the Merz–Kollman–Singh scheme [262]. Ligand parameters were then generated using Antechamber with the General Amber Force Field (GAFF) [128].
Protein residues were described using the ff14SB force field [103], while water molecules were represented using the TIP3P model [263]. Sodium and chloride ions were parameterised using the Joung and Cheatham ion parameters [122].
Molecular dynamics simulations were carried out using the protocol described in **Chapter 2**. Likewise, allosteric analyses were performed using the Shortest Path Map (SPM) and Distance Fluctuation (DF) approaches, as detailed in **Chapter 2**.
To characterise and monitor the conformational diversity of loop L5, a separate RMSD-based clustering analysis was performed using cpptraj [174]. Initially, side chains were removed from all trajectory frames, and the resulting backbone structures were aligned to the backbone of the first frame of each trajectory. Subsequently, a second alignment was carried out using only residues that were consistently classified as stable secondary-structure elements across the ATP, ADP, apo and ADPfil systems, from which an average structure was generated.
The trajectories were then fitted to a subset of helix $\alpha 2$ (residues 111–116 and 135–140) of the average structure in order to minimise the influence of global protein motions on the analysis of L5. Following this alignment, only the backbone atoms of residues 111–140, encompassing helix $\alpha 2$ and loop L5, were retained for further analysis.
Hierarchical agglomerative clustering based on pairwise RMSD values was subsequently performed using a clustering threshold of $\epsilon = 5$ Å. The complete procedure was applied independently to the apo, ATP-, ADP- and ADPfil-bound systems, yielding distinct conformational ensembles for each state.

5.3 Results

5.3.1 General results for Distance Fluctuation (DF) analysis

To facilitate comparison between the different catalytic states, changes in allosteric coordination were analysed by calculating differences in DF values between consecutive stages of the catalytic cycle (**Figure 5.3**). Specifically, the DF matrix of each state was subtracted from that of the preceding state. Thus, the effects of ATP hydrolysis were obtained by subtracting the ADP DF matrix from the ATP DF matrix, whereas the effects associated with ADP release were determined by comparing the apo and ADP states. As filanesib binds to the ADP-bound form, the effect of inhibitor binding was evaluated by subtracting the ADP DF matrix from that of the ADPfil system. The resulting difference matrices, hereafter denoted ΔDF, provide a quantitative representation of changes in allosteric coordination. Positive ΔDF values (red) indicate a reduction in residue coordination and, consequently, a weakening of allosteric coupling. Conversely, negative ΔDF values (black) correspond to increased coordination following a catalytic or binding event. Intermediate values (white) indicate little or no change in coordination between the two states. The original DF

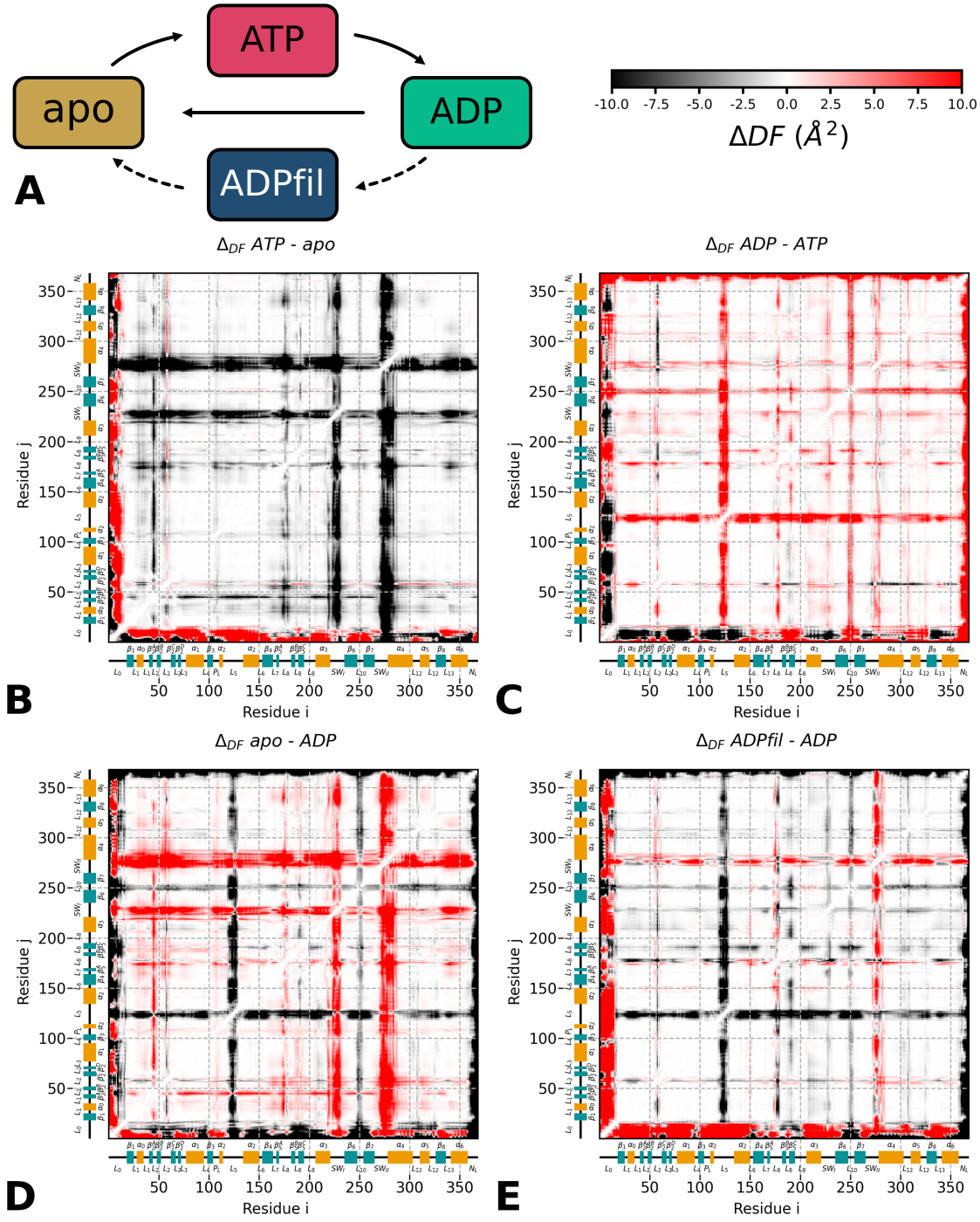

Figure 5.3: Allosteric modification in Eg5 cycle. ***(A)*** *Schematic representation of the catalytic cycle for the kinesin-5 motor domain.* ***(B-E)*** *ΔDF matrices for the catalytic steps of kinesin-5. The ΔDF values were calculated as DF scores of the end point minus DF scores of the starting point, following the order of the catalytic cycle. For each state, DF values are obtained as an average of the 6 independent MD replicas combined. All the axes contain a representation of the secondary structure of the motor domain of kinesin-5.*

matrices used to generate the ΔDF plots are provided in the **Appendix B** (see **Figure B.1**).
The SWI and SWII regions display behaviour consistent with their established role in nucleotide sensing and ATP hydrolysis [81, 249, 255]. Both domains contain residues directly involved in catalysis and are therefore expected to exhibit greater coordination in the ATP-bound state than in the ADP-bound state. Accordingly, ATP binding results in increased coordination throughout both switch regions (**Figure 5.3B**). ATP hydrolysis and subsequent release of the γ-phosphate lead to a moderate reduction in their coordination with the remainder of the motor domain. The most pronounced effect is observed upon complete nucleotide release, which causes the largest loss of coordination for both SWI and SWII. In contrast, filanesib binding selectively decreases the coordination of SWII while producing a moderate increase in the coordination of SWI. Inspection of the absolute DF matrices (**Figure B.1**) further reveals that L5 exhibits its lowest overall coordination in the ADP-bound state.
Analysis of the ΔDF matrices highlights a clear modulation of L5 coordination throughout the catalytic cycle. The allosteric coordination of this loop increases upon ADP release, coinciding with the transition to a high microtubule-affinity state, and remains elevated following ATP binding. These observations suggest a possible role for L5 in the regulation of microtubule association. Filanesib binding also increases the apparent coordination of L5; however, this effect is likely attributable, at least in part, to the local stabilisation typically induced by ligand binding. The potential contribution of L5 to kinesin-5 processivity is explored in greater detail in subsequent sections.
Marked and consistent ΔDF variations are also observed in the N-terminal segment (residues 1–16) throughout the catalytic cycle. Despite forming a stable α-helix in all simulations, this region exhibits low overall coordination in every state (**Figure B.1**). Previous studies have suggested that the N-termini of kinesin motor domains behave as intrinsically disordered regions (IDRs) involved in regulating processivity [226, 264], and examples of such behaviour have been reported for several non-human kinesin-5 proteins [265–267]. The persistence of a helical conformation in the present simulations is not necessarily inconsistent with this interpretation, as conventional force fields are known to favour more compact and ordered conformations of intrinsically disordered regions [268, 269].
Another region displaying substantial state-dependent behaviour is helix α0 (residues 30–36), whose coordination appears highly sensitive to nucleotide occupancy. Upon ATP binding, α0 exhibits reduced ΔDF values relative to loops L8 and helices α4 and α6, indicating stronger coordination with these regions. Conversely, ADP release produces a widespread increase in ΔDF values between α0 and the remainder of the protein (**Figure 5.3B,D**). Given its proximity to both the nucleotide-binding pocket and helix α6, which is directly connected to the neck linker, α0 may play an important role in coupling nucleotide-dependent conformational changes to motor-domain processivity.
The final region displaying substantial differences between catalytic states is loop L8 (residues 166–206), which is enriched in polar and charged residues, including serine, lysine, arginine, glutamate and aspartate. ATP binding produces a marked increase in the coordination of L8, whereas ATP hydrolysis results in a partial loss of coordination. The largest increase in ΔDF values is observed following ADP release (**Figure 5.3D**). In addition, filanesib binding decreases the ΔDF values of the C-terminal segment of L8 (residues 186–193).
These observations are particularly noteworthy in light of the computational study by Guo and co-workers, who demonstrated that salt bridges contribute more strongly than hydrogen bonds to kinesin-5–microtubule interactions and that these electrostatic contacts occur predominantly with β-tubulin [270]. The same study identified Glu166, Arg181 and Lys197 as major contributors to these interactions, all of which are located within L8. Furthermore, the authors proposed that kinesin-5 directionality may arise from preferential interactions between positively charged regions of the motor domain and the negatively charged surface of β-tubulin.
Taken together, these findings suggest that L8 may play an important role during the initial stages of microtubule recognition. Although kinesin-5 stepping is likely driven by a combination of Brownian

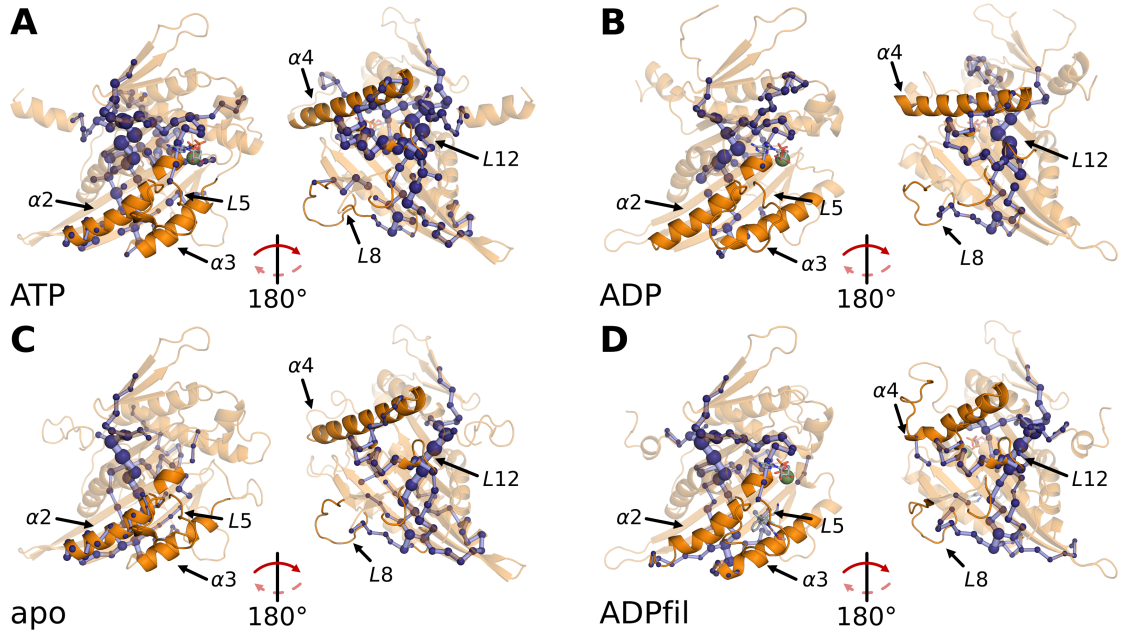

Figure 5.4: Shortest path maps (SPMs). *Results for ATP* ***(A)****, ADP* ***(B)****, apo* ***(C)*** *and ADPfil* ***(D)*** *forms. The nodes are represented by dark blue spheres, while paths are represented in a light blue color.*

diffusion of the detached motor head and neck-linker docking following nucleotide binding, an initial microtubule-recognition step mediated by L8 may also contribute to the process. Owing to its high charge density, L8 would be particularly well suited to establish long-range electrostatic contacts with the microtubule surface before the formation of more specific protein–protein interactions.
This hypothesis may account for the higher ΔDF values observed for L8 in the apo state relative to the ATP-bound state, despite both conformations exhibiting strong microtubule affinity. Following initial microtubule recognition, L8 could contribute to the correct orientation of the motor domain and facilitate the establishment of subsequent contacts. This interpretation is consistent with the low ΔDF values observed in the central region of L8, which indicate strong coordination with the remainder of the protein. Notably, the C-terminal segment of L8 is allosterically connected to helix $\alpha 3$, which forms part of the filanesib-binding site. As discussed above, this region exhibits reduced ΔDF values in the ADPfil system, suggesting local rigidification upon inhibitor binding. Such an effect could impair the conformational adaptability required for efficient microtubule association. This interpretation is supported by recent work from Alexandar and Ulaganathan, who reported that microtubule binding enhances both the flexibility and coordinated motions of the ATP-bound kinesin-5 motor domain [257].

5.3.2 General results for *SPM*

The DF analysis does not clearly distinguish the involvement of the motor-domain core, composed of eight β-sheets, across the different catalytic states. In all cases, the β-sheet core displays uniformly low $|\Delta DF|$ values, indicative of strong and persistent internal allosteric coupling. In this context, the SPM analysis provides valuable complementary information. While confirming the allosteric relevance of the β-sheet core, SPM additionally reveals the communication pathways connecting this structural scaffold with the surrounding functional domains (**Figure 5.4**).
In all catalytic states, the SPM network exhibits a prominent central pathway that traverses the β-sheet core diagonally, extending from $\beta 2$ to $\beta 4$. Notably, these β-sheets are directly connected to loops L2 and L8, respectively, both of which contribute to the microtubule-binding interface.

The branching pattern of this central pathway varies markedly with nucleotide occupancy. The ATP-bound state exhibits the most extensive and highly interconnected network, followed by the apo state, whereas the ADP and ADPfil states display substantially less branched communication pathways. In the ATP-bound form, a distinct allosteric pathway links the SWI and SWII regions to the β-sheet core. This pathway is absent from all other catalytic states and is consistent with the direct involvement of both switch regions in sensing the γ-phosphate of ATP. The presence of this communication route is therefore likely to have a significant influence on the allosteric behaviour of the motor domain.

The ATP-bound state also displays a highly connected branch linking the β-sheet core to loop L12 and helix α4. Although a similar pathway is present in the apo and ADP states, it is less extensive and exhibits lower significance. By contrast, this branch is completely absent in the ADPfil state. Furthermore, both ATP and apo display an additional extension of the β-sheet core network encompassing the entire β4 strand, which directly interacts with L8. This feature is not observed in either ADP or ADPfil.

Loop L12 and helix α4 form part of the microtubule-binding interface but differ functionally from L2 and L8. Whereas L2 and L8 primarily interact with β-tubulin through electrostatic interactions [270], L12 and α4 establish contacts predominantly with α-tubulin via hydrogen bonding. These weaker and more directional interactions are compatible with the dynamic attachment and detachment cycles required for kinesin motility.

An intriguing observation is that the allosteric connection between the β-sheet core and the α4/L12 region is considerably stronger in the ATP-bound state than in the apo state, despite both conformations exhibiting high affinity for microtubules. One possible explanation is that L2 and L8 mediate the initial high-affinity association of the motor domain with the microtubule through long-range electrostatic interactions, whereas ATP binding promotes additional structural rearrangements that facilitate the formation of more specific hydrogen-bonding interactions involving L12 and α4. Given the strong dependence of hydrogen bonds on both distance and orientation, such interactions would be expected to form only after initial motor-domain docking has occurred. Under this interpretation, ATP binding would promote a subsequent structural refinement of the kinesin–microtubule interface, effectively producing an ATP-dependent fitting process.

This hypothesis is supported by experimental observations in other kinesins demonstrating that the angle between the β-sheet core and the microtubule axis varies according to nucleotide occupancy, with apo and AMPPNP-bound forms adopting distinct orientations through a twisting motion of the motor domain [234]. Such nucleotide-dependent rotation may contribute to forward stepping by complementing the docking of the neck linker and facilitating displacement of the trailing motor head.

A further distinction among catalytic states involves loop L8. As discussed in the DF analysis, filanesib binding reduces the DF values of the C-terminal region of L8, indicating increased allosteric coordination. The SPM results reveal an allosteric pathway connecting the N-terminal region of helix α3 directly to L8 in all catalytic states. However, the extent and significance of this pathway vary considerably. The strongest and most extensive connection is observed in ADPfil, followed by ADP, whereas ATP and apo display substantially shorter communication pathways.

These observations support the notion that conformational flexibility of L8 may be important for efficient microtubule binding, whereas increased allosteric coupling and local rigidification correlate with lower microtubule affinity. Helix α3 occupies a strategically important position within this communication network, linking both the SWI region and L8. It therefore represents a plausible mediator through which nucleotide-dependent conformational changes are transmitted to the microtubule-binding interface. Nevertheless, the observation that apo and ADP both lack the γ-phosphate yet display distinct L8 behaviours suggests that nucleotide occupancy alone cannot fully account for the observed differences. The interaction between L8 and β4, identified above as a major component of the allosteric network, is likely to contribute significantly to this regulation.

Finally, a communication pathway connecting L5 to the P-loop through helix α2 is observed in

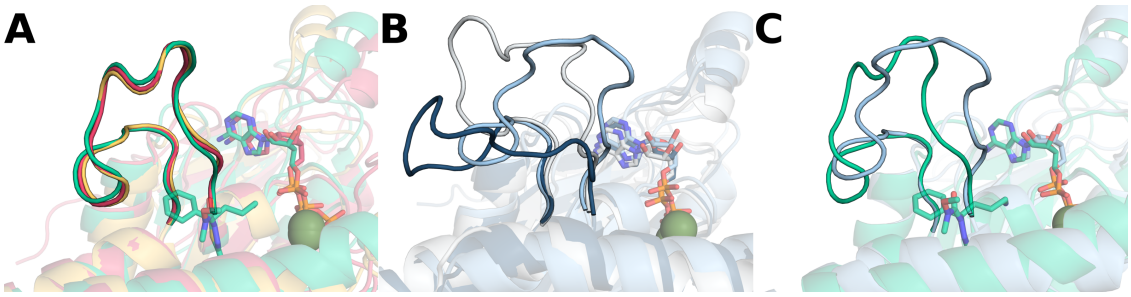

Figure 5.5: Clustering results of L5 conformations in all simulated states. ***(A)*** *Superposition of L5 in ATP (red), ADPfil (green) and apo (yellow).* ***(B)*** *Most representative conformations of L5 in ADP. The dark blue conformation is the most populated one (63%), followed by the cyan (24%) and the white one (13%).* ***(C)*** *Superposition of L5 conformation in ADPfil (green) and the most similar conformation in ADP (cyan).*

the ADP-, ADPfil- and ATP-bound states. This pathway is particularly pronounced in ATP and ADPfil, suggesting a persistent allosteric coupling between the inhibitor-binding region and the nucleotide-binding pocket.
Overall, the SPM analysis indicates that long-range allosteric communication within kinesin-5 is primarily mediated through the β-sheet core, which acts as a central communication hub linking the nucleotide-binding site to the microtubule-binding interface. In addition, L8 exhibits stronger allosteric coupling in the low microtubule-affinity states, supporting its potential role as a key regulator of microtubule recognition and binding.

5.3.3 Formation of the L5 binding pocket

A distinctive feature of the kinesin-5 L5 binding pocket is its apparent absence in ligand-free, ATP-bound and ADP-bound crystallographic structures. In contrast, other allosteric pockets within kinesin-5, including the site formed by helices α4 and α6, remain at least partially detectable in the absence of bound ligands. As discussed previously, kinetic studies indicate that L5 inhibitors bind preferentially to either the ATP- or ADP-bound forms of the motor domain. Consequently, the L5 pocket is generally regarded as an induced binding site that forms upon ligand binding [271].
The present DF analysis supports this interpretation. Relative to the other catalytic states, the transition to the ADP-bound form is accompanied by increased ΔDF values involving residues of L5 and the remainder of the motor domain, indicating a reduction in allosteric coordination and an increase in conformational flexibility. Such behaviour is consistent with the enhanced structural plasticity expected for a region required to adopt alternative conformations during formation of a cryptic binding site. In the ADPfil system, ΔDF values decrease to levels comparable to those observed in the apo and ATP-bound states. However, this reduction is most likely a consequence of filanesib binding itself, as ligands frequently stabilise neighbouring residues and promote more coordinated motions within the surrounding protein environment.
These variations in DF values suggest that conformational transitions of L5 throughout the catalytic cycle may have important functional consequences. This interpretation is supported by previous experimental studies. Fluorescence measurements performed by Cochran and Gilbert led to the proposal of an “open–closed” model for L5 [272], whereas Muretta and co-workers suggested the existence of three distinct L5 conformational states associated with different stages of the catalytic cycle. Notably, their study identified the ADP-associated conformation as the most mobile of the three states [228].
To further investigate this hypothesis, an RMSD-based clustering analysis of L5 conformations was performed across all simulated catalytic states (**Figure 5.5**). The results revealed that the dominant

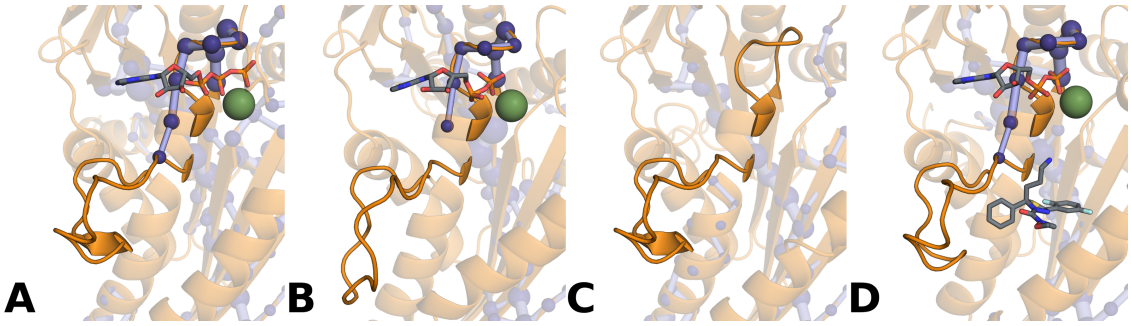

Figure 5.6: SPM shows allosteric pathway. *Connection between L5 (Glu116-Gly134) and P-loop (Gly105-Thr112) in ATP* ***(A)****, ADP* ***(B)****, apo* ***(C)*** *and ADPfil* ***(D)****.*

conformations sampled in the ATP-bound and apo states were essentially identical. Remarkably, the same conformation was also found to be the most populated state in the ADPfil simulations. By contrast, L5 displayed substantially greater conformational heterogeneity in the ADP-bound state, populating three major clusters. Only one of these clusters resembled the dominant conformation observed in the inhibitor-bound system (**Figure 5.5C**).
The observed behaviour suggests a functional analogy between L5 and the kinesin neck linker, which alternates between docked and undocked conformations during the catalytic cycle. In this context, L5 may similarly undergo transitions between a more ordered, "docked" conformation associated with the ATP-bound and apo states, both characterised by high microtubule affinity, and a more flexible, "undocked" conformation associated with the ADP-bound state, which displays lower microtubule affinity. Within this framework, filanesib appears to stabilise a conformation of L5 that closely resembles the docked state observed in the ATP-bound and apo forms (**Figure 5.5A**).
Interestingly, the clustering analysis further shows that ADP-state conformations compatible with filanesib binding originate from relatively sparsely populated clusters (**Figure 5.5B**; cyan and white conformations). This observation is consistent with the notion that the L5 pocket behaves as a cryptic binding site, sampling binding-competent conformations only transiently in the absence of ligand. Filanesib binding would therefore selectively stabilise one of these pre-existing conformational states, shifting the equilibrium towards a pocket-open conformation capable of accommodating the inhibitor.

5.3.4 Loop 5 is allosterically connected to the ATP binding site

A particularly intriguing feature of L5-targeting inhibitors is that, although they stabilise L5 in a conformation resembling those observed in the apo and ATP-bound states, they simultaneously maintain the kinesin motor domain in an ADP-like functional state rather than inducing extensive structural rearrangements throughout the protein. Experimental studies have demonstrated that these inhibitors slow ADP release, thereby reducing the affinity of kinesin-5 for microtubules [241]. Furthermore, previous molecular dynamics investigations have shown that L5-binding inhibitors increase the binding affinity of ADP for the nucleotide-binding pocket [248, 255].
Behnke-Parks and co-workers proposed that the allosteric effects of L5 inhibitors are transmitted to the SWI region through helix $\alpha 3$ [273]. However, the present analyses do not provide direct evidence supporting such a communication pathway. Instead, the *SPM* analysis reveals a prominent allosteric connection linking L5 directly to the nucleotide-binding site through the P-loop (**Figure 5.6**). This pathway is clearly observed in both the ATP-bound and ADPfil states, whereas a weaker connection is detected in the ADP-bound form. By contrast, no equivalent pathway is observed in the apo state.
In both ATP and ADPfil simulations, perturbations originating from Gly115 within L5 propagated through the allosteric network to reach the P-loop, with the strongest communication observed in the ATP-bound state. This observation is particularly noteworthy given the functional importance of neighbouring residues within L5. For example, mutagenesis studies have identified Glu116 as a critical

determinant of kinesin-5 activity [274], and structural studies have shown that this residue forms an ionic interaction with the primary amine of filanesib and numerous other kinesin-5 inhibitors [253]. Consequently, Glu116 represents a key element of the allosteric network linking the inhibitor-binding site to the catalytic machinery and should be considered an important target in the rational design of future inhibitors.
Additional evidence supporting this communication network can be found within the nucleotide-binding site itself. McGrath and co-workers demonstrated that Thr107 within the P-loop plays a critical role in ATP hydrolysis through its interaction with Glu270 of the SWI region [249]. Glu270 additionally forms the anionic component of the conserved SWI–SWII salt bridge (Arg240–Glu270), which is essential for coordinating the catalytic cycle. The existence of an allosteric pathway linking L5 to the P-loop therefore provides a plausible mechanistic explanation for how perturbations at the inhibitor-binding pocket can influence nucleotide binding and release without requiring large-scale conformational changes throughout the motor domain.
The DF and SPM analyses further indicate that L5 motions are coupled to those of the SWI region in the ADP-, ATP- and ADPfil-bound states. As discussed in the Introduction, the substantial structural variability of L5 across the kinesin superfamily has long suggested a role in fine-tuning the kinetic properties of individual motor proteins. The present results provide strong support for this hypothesis. Specifically, L5 appears to occupy a strategically important position within the kinesin-5 allosteric network, simultaneously communicating with the nucleotide-binding pocket and with elements of the microtubule-binding interface. Through these connections, L5 is well positioned to coordinate the coupling between nucleotide state and microtubule affinity that underlies the mechanochemical cycle of kinesin-5.

5.3.5 Trp127 is crucial for the functional interaction between L5 and helix α3

Previous studies have highlighted the importance of Trp127 within L5, noting that this residue is displaced upon inhibitor binding and, in several inhibitor-bound structures, contributes directly to ligand stabilisation through π-π stacking interactions [243, 248]. Inspection of the crystal structure of kinesin-5 bound to the ATP analogue AMPPNP (PDB ID: 3HQD [260]) further reveals a potential π-π interaction between Trp127 and Tyr211 located on helix α3. Notably, this interaction is absent in the ADP-bound structure.
Additional evidence for the functional significance of Trp127 has been obtained from mutagenesis studies. Substitution of Trp127 by either alanine or cysteine results in a marked reduction in kinesin-5 catalytic activity, an effect attributed to impaired ADP release [228, 229]. Interestingly, partial restoration of activity was observed when the fluorescent probe monobromobimane (mBBr) was covalently attached to the cysteine mutant [228]. Given the comparable size and aromatic character of mBBr and the indole ring of tryptophan, this observation suggests that the probe may partially mimic the native interaction mediated by Trp127.
The possibility that Tyr211 functions as an aromatic anchoring point for Trp127 is particularly attractive from a mechanistic perspective. Tyr211 is located within the region of helix α3 directly connected to the SWI domain, raising the possibility that nucleotide-dependent conformational changes within the catalytic centre are transmitted to L5 through the α3 helix. Such a mechanism would provide a direct structural link between the nucleotide-binding site and the conformational dynamics of L5.
The present DF analysis is consistent with this interpretation. Detailed examination of the ATP- and apo-state trajectories reveals that all residues within L5 do not contribute equally to the loop's allosteric behaviour. Instead, Trp127 displays markedly lower DF values than neighbouring residues, indicating stronger coordination with the surrounding protein environment. DF values increase progressively on either side of Trp127, suggesting that this residue occupies a uniquely constrained position within the loop (**Figure 5.7**). Furthermore, no alternative interactions between L5 and helix α3 appear capable of providing a comparable degree of coupling. This observation strongly

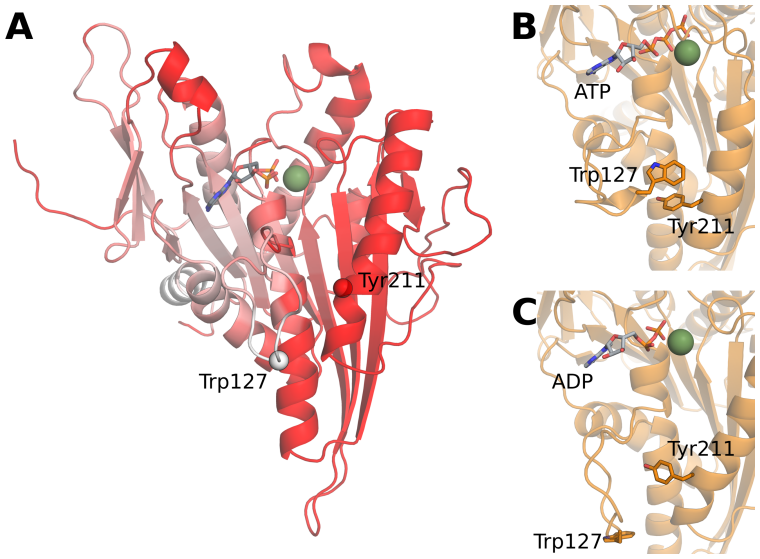

Figure 5.7: Trp-Tyr stacking. ***(A)*** *Projection of* $\Delta DF_{127,j}$ *scores (ADP - ATP) for Trp127. The red color follows the same metrics as the matrix in* ***Figure 5.3C*** *and indicates a decrease of coordination of Trp127 with the rest of the protein.* ***(B-C)*** *Crystal structures of ATP* ***(B)*** *and ADP* ***(C)****. Trp127 and Tyr211 are shown as sticks. In ATP, both residues have an orientation that suggests a* π*-*π *interaction, while in ADP this is not observed [210, 256].*

supports the hypothesis that Trp127 serves as the principal anchoring residue connecting L5 to helix $\alpha 3$.

Even the adjacent residue Glu128, which could potentially interact electrostatically with Lys207 on helix $\alpha 3$, exhibits substantially higher DF values than Trp127. Quantitative analyses of the putative Trp127–Tyr211 π-π interaction and the Glu128–Lys207 ionic interaction are provided in the Supporting Information (**Figure B.2** and **Figure B.3**, respectively). Together, these data indicate that aromatic interactions involving Trp127 are likely to contribute more significantly to L5 stabilisation than neighbouring electrostatic contacts.

The proposed role of Trp127 is further supported by evolutionary evidence. Similar L5–$\alpha 3$ interactions appear to be conserved across divergent kinesin-5 homologues. For example, in the kinesin-5 motor domain of Plasmodium falciparum (PDB ID: 7NB8 [275]), Tyr244 and Tyr245 occupy positions analogous to Tyr211 in human kinesin-5 and are therefore well positioned to establish comparable aromatic interactions. An alternative but functionally analogous arrangement is observed in the kinesin-5 Cut7 from Schizosaccharomyces pombe (PDB ID: 6S8M [267]), where a salt bridge between Asp208 within L5 and Arg298 on helix $\alpha 3$ appears to stabilise L5 in a docked conformation.

These observations suggest that the existence of a stabilising interaction between L5 and helix $\alpha 3$ is more strongly conserved than its precise chemical nature. While human kinesin-5 appears to employ an aromatic Trp127–Tyr211 interaction, other species may achieve an equivalent structural outcome through alternative aromatic or electrostatic contacts. Consequently, disruption of these conserved L5–$\alpha 3$ interactions may represent a viable strategy for targeting kinesin-5 homologues in species for which no L5-directed inhibitors have yet been identified.

More broadly, multiple sequence alignment of kinesin-5 homologues from selected species (**Figure B.4**) reveals substantial conservation of a tryptophan residue at positions analogous to Trp127, together with an aromatic residue, typically tyrosine or phenylalanine, at positions corresponding to Tyr211. This conservation further supports the functional importance of the L5–$\alpha 3$ interaction and its role in regulating the kinesin-5 catalytic cycle.

5.3.6 Allosteric inhibitors lock L5 and increase the ADP binding affinity

As discussed previously, filanesib and other L5-targeting inhibitors are known to stabilise kinesin-5 in a low microtubule-affinity state while simultaneously slowing ADP release from the nucleotide-binding pocket. Experimental studies have consistently shown that these inhibitors trap the motor domain in an ADP-like functional state despite binding at a site remote from the catalytic centre [241].
At first sight, one possible explanation would be that inhibitor binding prevents the interaction between Trp127 and Tyr211, thereby locking L5 in an inactive conformation and inhibiting the progression of the catalytic cycle. However, the present results do not support such a mechanism. Quantitative analysis of the Trp127–Tyr211 interaction (**Figure B.2**) indicates that filanesib actually favours π-π stacking between these residues. Moreover, the clustering analysis demonstrates that filanesib stabilises L5 in the same predominant conformation observed in the ATP-bound and apo states. Although transient disruption of the Trp127–Tyr211 interaction can still occur in the ADPfil simulations, as previously reported [259], these events represent only a minor fraction of the sampled conformational ensemble.
Additional evidence arguing against disruption of the Trp127–Tyr211 interaction as the primary inhibitory mechanism comes from mutagenesis studies. Tcherniuk and co-workers reported that the Trp127Ala mutation confers only moderate resistance to S-trityl-L-cysteine (STLC), likely through loss of aromatic interactions with the inhibitor itself, while having little effect on the activity of monastrol [276]. If inhibition relied primarily on interference with the Trp127–Tyr211 interaction, mutations of Trp127 would be expected to significantly alter the efficacy of multiple L5 inhibitors rather than only selected compounds.
Previous molecular dynamics studies have shown that L5 inhibitors increase the binding affinity of ADP for the nucleotide-binding pocket [248, 255]. In the absence of inhibitor, ATP binds more strongly than ADP, a behaviour that is consistent with the physiological catalytic cycle, where ATP must be readily captured by the apo motor domain and ADP must subsequently be released following hydrolysis [255]. Given the allosteric connection identified between L5 and the nucleotide-binding site, we initially considered whether L5 might directly interact with the bound nucleotide. A similar hypothesis was previously proposed by Harrington and co-workers, who suggested that the conformational flexibility of L5 could occasionally bring the loop into direct contact with the nucleotide-binding pocket; but, to achieve those conformations, they had to run simulations at high temperature (600 K) [277].
Inspection of the ATP-bound structure revealed a potential interaction between Glu118 and the ribose moiety of the nucleotide, raising the possibility that this residue contributes directly to nucleotide stabilisation. However, quantitative analysis of hydrogen-bond occupancy (**Figure B.5**) demonstrated that the Glu118–nucleotide interaction is actually less prevalent in the ATP, apo and ADPfil states, where L5 adopts a docked conformation. Instead, this interaction is most frequently observed in the ADP state. These observations therefore exclude a stabilising role of Glu118 in nucleotide binding.
Closer inspection of the ADPfil simulations reveals the origin of this effect. The protonated amine of filanesib forms a persistent salt bridge with Glu118, effectively sequestering the residue and preventing its interaction with ADP. This behaviour is clearly reflected in the radial distribution function shown in **Figure B.5E**. Independent support for this interpretation is provided by the nucleotide-density analysis (**Figure B.6**). Whereas ADP in the uninhibited state occupies a more solvent-exposed position within the binding pocket, the nucleotide in ADPfil shifts towards a conformation closely resembling that observed in ATP. The improved overlap between ATP and ADPfil nucleotide densities further supports this conclusion. Interestingly, multiple sequence alignment of kinesin-5 homologues (**Figure B.4**) reveals partial conservation of either glutamate or aspartate at the position corresponding to Glu118, suggesting functional importance for this residue.
Taken together, these findings suggest that Glu118 contributes to nucleotide release rather than nucleotide retention. By facilitating partial displacement of ADP towards the solvent, Glu118 may promote nucleotide dissociation and thus ensure efficient progression through the catalytic cycle.

Sequestration of Glu118 by filanesib would consequently slow ADP release and increase apparent nucleotide affinity. This interpretation is fully consistent with the kinetic experiments of Waitzman and co-workers, who demonstrated that L5 accelerates ADP release during microtubule binding and that both L5 deletion mutants and STLC treatment reduce the rate of ADP dissociation [227]. Similarly, ispinesib has been shown experimentally to slow ADP release even under conditions where nucleotide exchange already occurs slowly in the absence of microtubules [241].
Although impaired ADP release alone would be sufficient to slow mitotic progression, it does not explain the characteristic monopolar spindle phenotype produced by kinesin-5 inhibitors [242]. This phenotype arises from collapse of the highly tensioned mitotic spindle and therefore requires a direct effect on kinesin–microtubule interactions. In this regard, Chen and co-workers demonstrated that L5 inhibitors induce spindle collapse by trapping kinesin-5 in a low microtubule-affinity state, causing motor domains to dissociate from the spindle microtubules [278].
The present allosteric analyses provide a plausible explanation for this behaviour. Both DF and SPM results indicate that filanesib preserves and even enhances the allosteric coupling between L8 and the rest of the motor domain, a feature otherwise associated with low microtubule-affinity states. Initially, this effect might be attributed to direct perturbation of helix $\alpha 3$ by the difluorophenyl group of filanesib. However, no SPM pathway connecting the inhibitor-binding pocket to L8 through $\alpha 3$ was detected. Furthermore, such a mechanism would fail to explain why structurally distinct L5 inhibitors, including monastrol, produce similar cellular phenotypes despite lacking interactions within this region.
Instead, the results point towards an alternative mechanism involving the $\alpha 2$–$\beta 4$–L8 communication network. The C-terminal segment of helix $\alpha 2$ participates in an allosteric pathway connected to $\beta 4$, which in turn forms part of the L8 communication network. This pathway is observed in ATP, apo and ADPfil but is largely absent in the ADP state. The similarity between ATP and ADPfil is particularly evident in the supplementary ΔDF matrix comparing these states (**Figure B.7**). Consequently, the enhanced coupling observed in ADPfil may arise from the stabilisation of an ATP-like L5 conformation in the presence of a bound ADP molecule. In effect, filanesib appears to decouple L5 conformation from nucleotide state, generating a hybrid allosteric configuration in which L5 adopts an ATP-like arrangement while the catalytic centre remains trapped in an ADP-like state. This combination would simultaneously impair nucleotide release and reduce microtubule affinity.
Overall, the present results support a mechanistic model in which filanesib acts through two complementary effects. First, binding of the inhibitor sequesters Glu118 and stabilises ADP within the nucleotide-binding pocket, thereby slowing nucleotide release. Second, inhibitor-induced locking of L5 into an ATP-like conformation alters long-range allosteric communication with the microtubule-binding interface, particularly through the $\alpha 2$–$\beta 4$–L8 pathway, ultimately reducing microtubule affinity. Together, these effects provide a coherent molecular explanation for both the inhibition of kinesin-5 ATPase activity and the characteristic monopolar spindle phenotype observed experimentally.
Finally, although a systematic investigation of resistance mutations was beyond the scope of the present work, the combined DF and SPM analyses presented here provide a framework for interpreting the effects of distal mutations capable of compensating for the allosteric perturbations introduced by filanesib. Such analyses may also facilitate the identification of previously unrecognised sites with the potential to generate resistance through long-range allosteric effects, particularly at positions distant from the L5 binding pocket itself.

5.4 Conclusions

The mechanochemical activity of kinesin-5 depends on the coordinated progression of the ATPase cycle within its motor domains. Changes in nucleotide occupancy reorganise the allosteric communication pathways that connect the catalytic centre, the microtubule-binding interface and the

regulatory elements of the motor domain. In this work, microsecond-scale atomistic molecular dynamics simulations were employed to characterise, for the first time in a systematic manner, the allosteric behaviour of four key functional states of the kinesin-5 motor domain: ATP-bound (ATP), ADP-bound (ADP), apo, and the ADP-bound form in complex with the allosteric inhibitor filanesib (ADPfil).

The combined DF and SPM analyses demonstrate that the switch regions, SWI and SWII, are strongly coordinated with the remainder of the motor domain in the ATP-bound state and progressively lose coordination as the catalytic cycle advances. The analyses further identify loop L8 as a key allosteric component whose communication with the rest of the protein is strongly dependent on nucleotide state. In ATP, this communication is mediated primarily through the β4–L8 interface, whereas filanesib binding preserves a pattern of enhanced L8 coupling characteristic of low microtubule-affinity states.

A major outcome of this study is the detailed characterisation of the conformational behaviour of loop L5 throughout the catalytic cycle. The simulations reveal that L5 alternates between two principal conformational states: a docked state, characterised by strong allosteric coupling to the motor domain and predominating in the ATP and apo forms, and an undocked state, characterised by greater flexibility and predominantly sampled in the ADP form. This behaviour is closely associated with the interaction between Trp127 and Tyr211, which acts as a key structural anchor connecting L5 to helix α3. The clustering analysis further demonstrates that filanesib stabilises L5 in a conformation closely resembling the docked ATP/apo state despite the motor domain remaining functionally trapped in an ADP-bound state. These observations support the notion that the L5 pocket behaves as a cryptic allosteric site whose binding-competent conformations are selectively stabilised by inhibitor binding.

The present work also provides new insights into the role of L5 in regulating nucleotide release. The results suggest that Glu118 contributes to ADP dissociation by interacting with the nucleotide in the ADP state and promoting its displacement towards the solvent-exposed region of the binding pocket. Filanesib disrupts this process by sequestering Glu118 through a persistent ionic interaction, thereby preventing its participation in nucleotide release. Together with the enhanced allosteric communication observed between L5 and the P-loop, this mechanism provides a structural explanation for the experimentally observed increase in ADP affinity and reduced nucleotide release induced by L5 inhibitors.

Perhaps the most significant finding of this study is that filanesib does not simply stabilise an ADP-like state. Instead, the inhibitor generates a hybrid allosteric state in which different regions of the motor domain resemble distinct stages of the catalytic cycle. While the nucleotide-binding pocket remains functionally trapped in an ADP-like configuration, L5 adopts an ATP/apo-like conformation and the microtubule-binding interface retains features associated with low-affinity states. This decoupling of normally coordinated allosteric processes provides a coherent mechanistic explanation for both the inhibition of nucleotide exchange and the loss of microtubule affinity characteristic of kinesin-5 inhibitors.

Overall, the results presented in this chapter establish L5 as a central regulator of kinesin-5 allostery, linking the nucleotide-binding site to the microtubule-binding interface through an extensive communication network. By combining long-timescale molecular dynamics simulations with complementary allosteric analyses, this work provides a detailed molecular framework for understanding kinesin-5 inhibition and offers a basis for the future design of allosteric modulators targeting kinesin motor proteins.

Chapter 6

Protein Conformational Motion: FAN Method

6.1 Introduction

Proteins are the foundational building blocks of living organisms. They oversee a wide range of functions via diverse highly regulated and finely-tuned mechanisms, many of which are still elusive at atomistic resolution [1, 2]. In general, the variable combinations of the twenty aminoacids in proteins' polypeptidic chains determine the possible 3D arrangements, i.e. the tertiary structures, that are responsible for biological activities. However, recent experimental and computational studies have pushed forward the ensemble vision of proteins and their functions. In this model, each protein can be viewed as an ensemble of conformations populating a variably broad basin on the free energy landscape [19, 27]. Functionally distinct states can be different basins representing different global conformations, separated by free energy barriers, or different dynamic states determined by ligand binding, aminoacid modifications, or covalent changes.
The evolution from static protein models towards a more dynamic picture of structure-dynamics-activity relationships view implies considering the internal, potentially coordinated/ordered, motions of constitutive residues explicitly into account [134, 279]. This would be particularly important in the full understanding of allostery [18], the prime mechanism of biological functional regulation, whereby a modification at one point of the protein reverberates in a variation at a distal site, modifying affinity towards a primary binding event, enzymatic kinetics, conformational change mechanisms, or the selection of surfaces for engaging a binding partner in protein-protein interaction (PPI) complex. Many models have been proposed to rationalize this mechanism. They entail sequence-based approaches (ranging from the analysis of sequence motifs to coevolutionary patterns [15]), structure-based approaches (including the analysis of properties such as mobility, electrostatic profiles, solvent accessibility), and AI approaches (where datasets from experiments and computation are analysed to identify patterns connecting different protein sites).
Results in recent years have made it evident that the functional dynamics depend upon the coordinated spatial fluctuations of the residues that make up the structure [15]. In this framework, several methods were developed and applied to extract functionally oriented dynamic properties from, e.g., MD simulations [34]. Some key examples include the shortest path method (SPM), the characterization of the key contributors to principal components of motion, and the evolution of signals from one site to a distal one via non-equilibrium simulations [135].
To analyse internal protein dynamics and extract information on how perturbations at specific points diffuse their effect throughout the whole 3D structure, we previously introduced distance fluctuation (DF) analysis [133]. Briefly, from MD trajectories of a certain protein (in different conditions), the matrix of distance fluctuations can be computed, in which each element of the matrix corresponds to the fluctuation in the distance between pairs of residues. This parameter is invariant under translations and rotations of the molecules and does not depend on the choice of a particular protein

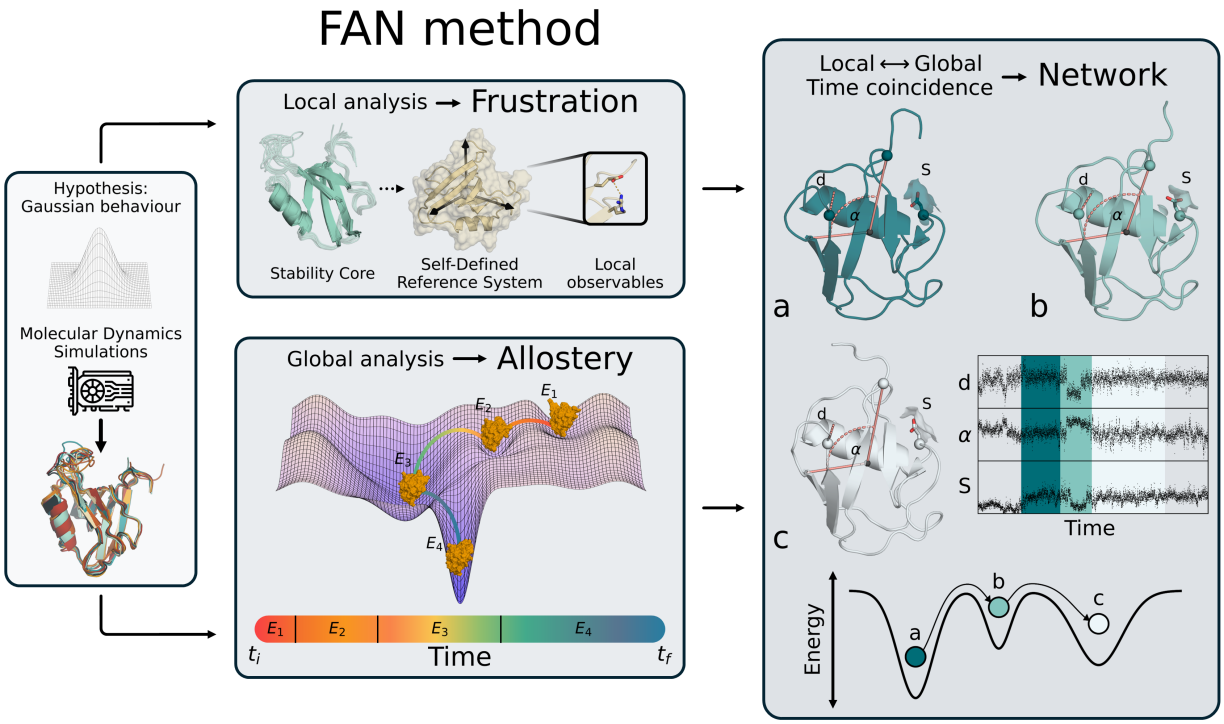

Figure 6.1: Schematic map of FAN analysis method. *Starting from the available MD simulation data and from the hypothesis that variables describing protein motion are gaussianly distributed, two complementary analysis are developed. The first is related to Frustration, the algorithm identifies the regions of the protein Stability Core, i.e. the residues which remain coupled during the simulation, and use this reference system to study those residues which display a local, multimodal, conformational distribution. The second analysis performs an RMSD-based procedure to identify, during the simulation, the Ensemble conformation achieved by the protein during its motion, i.e. capturing the basins characterizing an Ensemble view of Allostery. These two complementary analysis are employed by the final analysis which uncovers the hidden Network inside proteins, namely searching for time coincidences between local and global variables. The dynamical, intertwined, motion inside protein can thus be decoupled, highlighting the connections that activate protein conformational motion.*

reference structure. The DF matrix can be used to assess the intrinsic flexibility of proteins and how it changes upon ligand binding, mutations, PTMs. Importantly, the DF profile of specific (groups of) residues may change as a function of perturbations, allowing to highlight allosteric changes. One potential weakness of the approach is that the results do not provide an automatic and quantitative definition of which aminoacids may be more relevant in determining allosteric mechanisms (see **Chapter 2** for a detailed explanation).

It is well known that even small mutations in protein sequence can lead to enormous changes in the equilibrium system. One of our aims in this work is to develop a simplified view of the intricate interaction network without losing the single amino-acid perspective. However, finding the right key to understanding the problem by considering every single residue could be practically and theoretically impossible. We then tried to develop a smart way to tackle the problem from the collective perspective. Starting from these considerations, here we introduce a novel algorithm to extract the information on the protein motions that can be key determinants of conformational changes. The advantage of the method is that, consistent with the original implementation, it does not require the choice of a reference structure and provides key mechanistic information in a self-consistent framework. The method is validated on several, diverse test systems.

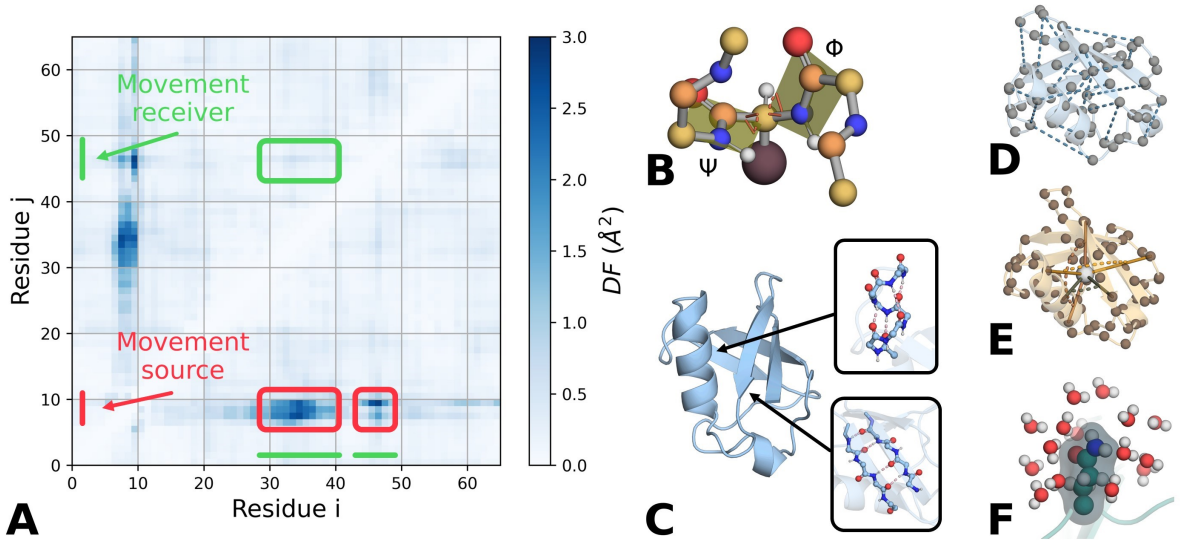

Figure 6.2: Basic descriptors for protein motion. ***(A)*** *Qualitative illustration of the inspiration of the model from a DF matrix: region around residues 5-10 can be seen as a movement source since it shows averagely high values of distance fluctuation with respect to other residues, while regions 22-38 and 45-50 can be seen as movement receivers because they keep fixed during the dynamics with low DF value and they both "see" the motion of the movement source;* ***(B-F)*** *Graphical representation of the chosen descriptors:* ***(B)*** *Dihedral angles ψ, ϕ of the backbone.* ***(C)*** *Secondary Structures calculated by DSSP.* ***(D)*** *Pairwise distance between all residue pairs.* ***(E)*** *Pairwise angle between residues and Centre of Geometry of C_α atoms.* ***(F)*** *Solvent Accessible Surface Area of each residue.*

6.1.1 Method Overview

FAN (Frustration Allostery Network) method was built to catch and determine the maximum amount of information from molecular dynamics simulations data. The algorithm is based on different, consecutive computational tools that try to connect the concepts of local motion and global ensembles to build a unified framework that can characterize and identify allosteric mechanisms in Molecular Dynamics simulations. Starting from the hypothesis that, at equilibrium, all observables probability density functions behave as gaussian functions, we developed a simple definition of Frustration based on the *detectability* concept. We then found regions of protein stably folded (i. e. *bones*, see below) and successively the stability core of a protein. This stability core was then used as a self-defined collective reference system which, together with local motion descriptors, was then able to describe movements with high frustration that could be defined as the ones that can capture the essence of residue motion. We further developed a definition of Statistical Equilibrium Ensembles based on pairwise RMSD and *detectability*. This definition enabled us to divide the trajectory into Ensemble time windows, that identify intervals where the protein does not considerably change conformation. Finally, comparing the time development of local observable distributions with the evolution of the total simulation, we could determine whether a given variable is connected to the Ensemble changes and how much this connection is relevant. Possible couplings between connected variables were further evaluated, establishing connected allosteric endpoints. A simple scheme describing the steps of the algorithm is presented in **Figure 6.1**.

6.2 Theoretical and Methodological Framework

The study and the analysis of protein functional motions is a well-known problem in computational biochemical physics [131]. The search for optimal descriptors with respect to which build and

visualize Free Energy profiles has been a subject of intense investigation and affected biological simulations since their birth [6]. Because of the intrinsic complexity related to protein systems, a potentially infinite list of descriptors could be defined, where every variable gives a different view of the system. The general goal is to have a "good" description of the dynamics providing the maximum amount of information possible. Here, we set out to use general and unbiased principles applying them to standard simulation outputs.

The general framework for allostery can be traced back to the Monod-Wyman-Changeux model [18], whereby the response of proteins to signals translates into allosteric pathways or collective allosteric conformational changes that enable biological functions [20].

From the structural and dynamical point of view, a general concept successfully applied many times is the one of Frustration [37–39]. Frustration reflects the fact that the ruggedness of the energy landscape for specific protein substructures may oppose its stabilization at equilibrium [37, 280]. It has in fact been established that frustration tends to localize around those protein regions which are fluctuating more and that need an external factor (i.e. a ligand) to be stabilized [30]. Frustration can thus be established through an analysis of equilibrium trajectories, exploring the capability of a given residue to populate more than one equilibrium state [281, 282].

Finally, another important concept in the definition of collective functional protein motions is the one of Networks [41, 279]. Proteins, in their native states, can be viewed as tightly packed 3D objects in which residues exchange information through a complex Network, which may ultimately underlie the response of the system to a perturbation.

In this framework, we ask whether it is possible to develop a general model to automatically detect the salient traits of internal protein dynamics that define allosteric regulation and functional dynamics (or conformational changes). Specifically, we aim to combine the ability of providing a general view on protein dynamics and allosteric relationships given by DF analysis with the capability of highlighting relevant pathways for the transmission of signals obtainable by Network analysis of protein structure and dynamics.

To proceed along this avenue, we first set out to use a descriptor of protein dynamics that is independent from a reference structure and invariant to rotations and translations. Distance-Fluctuation matrices report on the coordination propensities of pairs of aminoacids: one can associate small DF values to small deviations between the residues involved in the pair, indicating that they are moving in a coordinated way [133]. Large DF values report on pairs where one or both residues sample a diversity of relative positions/states. In this context, one can calculate the average DF value for all the pairs in which one specific aminoacid is involved (this would amount to calculating the average value over a column/row of the DF matrix). In our model, (groups of) residues characterized by high average values of such overall DF parameter could be defined as *motion sources*, i.e. regions with a tendency to fluctuate more and with a dispersed equilibrium position. Instead, residues characterized by high fluctuations with respect to the *movement sources* but low average DF values with respect to the rest of the sequence can be seen as *motion receivers*, namely regions that tend to occupy a fixed position with respect to the rest of the protein. Such residues can be used as probes to characterize motions of *motion sources*. Based on these definitions, one then needs first to uniquely identify *motion receivers* (see **Figure 6.2**).

Next, we analysed a set of structural/dynamic descriptors that could be exploited to ideally identify and track motions within the system under exam, with the aim of quantitatively discriminating between *motion sources* and *receivers*. The first two descriptors are the DSSP-based definition of residue secondary structure and backbone dihedral angles, that recapitulate topological stability and conformational properties of the chain. Next, to identify significant conformational changes that characterize specific parts of the protein, we defined two additional descriptors: the $C_\alpha - C_\alpha$ distance and the $C_\alpha - COG - C_\alpha$ angle, where the COG is the centre of geometry of all C_α (this can be considered a fixed point of the system, because in equilibrium simulation of small protein, local chain movements do not affect the global mean position). The angle was chosen to detect small rotational motions that could not be detected using only the $C_\alpha - C_\alpha$ distance. Such distances

and angles were first calculated for all possible residue-residue pairs, and then further analyses will reduce the number of inspected values to one distance/angle per residue (see the next paragraphs). The final parameter we included in our analysis is the residue-based Solvent Accessible Surface Area, to monitor possible side chains conformational changes that could hint binding-prone conformation with other entities (proteins, ligands, ...).
The starting point for the development of the model is the statistical analysis of the distributions of the different descriptors, that can be divided into two categories: discrete (DSSP) and continuous (all the others).

6.2.1 Statistical analysis of descriptors

We assume that, at equilibrium, all the non-discrete descriptors are distributed according to continuous Gaussian distributions around one (or more) equilibrium position. Moreover, we expect that stable regions display a single gaussian distribution (for example a dihedral angle inside a helix), while less stable (i.e. frustrated) regions can exhibit multiple gaussian distributions sampling different equilibrium values.
We then developed a strategy to define the most likely, minimal number of gaussian peaks describing a given probability distribution. After running a number of preliminary statistical tests, the one that is found to give the best results in terms of reproducibility, stability and easy automation is the integral of the absolute difference between the distribution of the values of a certain observable sampled during the computer simulation, $s(x)$, and its gaussianly fitted probability distribution $f(x)$ (both normalised). This difference is called ΔC (**Equation 6.1**).

$$\Delta C = \int |f(x) - s(x)| dx \tag{6.1}$$

Then, every probability distribution analysed was recursively fitted with an increasing number of gaussians until the ΔC was under a previously defined threshold, while the maximum fitted number of gaussians was set to 5. The threshold was defined with some tests on gaussian randomly generated data, searching for an optimal convergence between the known parameters used for generating the data and the fit results.
Once the best fitting probability distribution has been found, we need a "detectability" definition. Indeed, even though a probability distribution could be well described by a double gaussian, it may not return a clear distinction of the two states that underpin such distribution, implying that such states remain "not detectable". This may reflect the fact that a protein energy landscape typically displays a rugged profile with many energetically close microstates. In this context, we introduced the detectability parameter g, which describes the separation between two gaussians. If

$$|\mu_1 - \mu_2| > g(\sigma_1 + \sigma_2) \tag{6.2}$$

is verified, where $\mu_{1,2}$ are the gaussian means and $\sigma_{1,2}$ are the dispersions, then the two gaussians are considered to be separated and the two equilibrium positions detectable (as default value, $g = 1$ was set).
This type of analysis reports on the degree of frustration that would characterize a certain protein region; multiple peaks in the distributions can be interpreted as the tendency of a specific subregion to populate multiple dynamic states. Identifying multimodal distributions can thus be a viable proxy to determine which residues are the most prone to undergo conformational changes, possibly related to biochemical tasks.
To avoid overfitting, we tuned ΔC using a value that was a good compromise between simplicity (high ΔC means that the fitting procedure is less restrictive and thus a lower number of gaussians can fit a given probability distribution) and complexity (low ΔC means that the distribution needs to be fitted with a higher number of gaussians). In the case in which the system is described by more than 2 gaussians, a vast range of possible detectable states arises. To analyse and filter all these

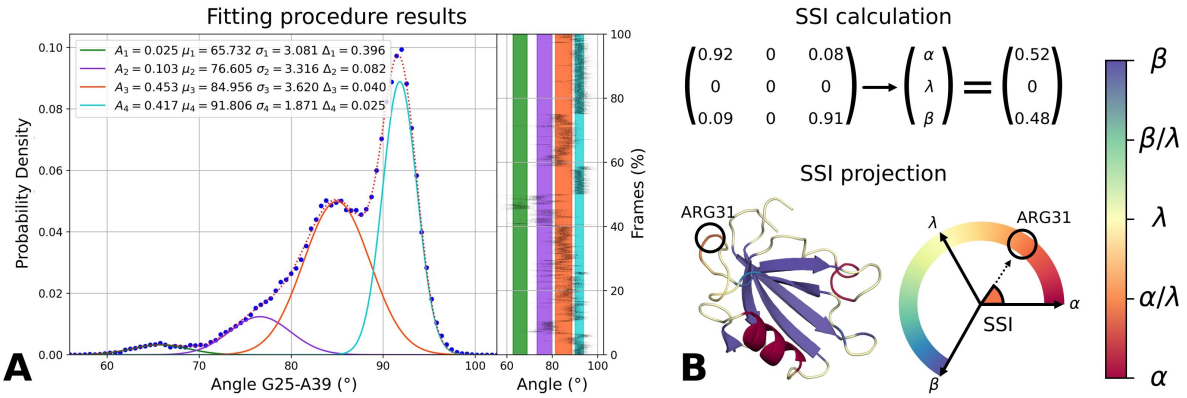

Figure 6.3: Examples of the two different approaches for continuous and discrete variables. ***(A)*** *Multiple gaussian fit of normalized probability distribution of distance between a residue in the stability core and residue 32 of Second PDZ domain system, on the top left the gaussian parameters of the fit are showed, together with ΔC_i difference between fit and simulated distribution, with i representing the number of the iteration; on the right of panel* ***A*** *the values of the distance are plotted as a function of the simulation time, together with some coloured bands in the regions $\mu_i \pm g\sigma_i$.* ***(B)*** *Definition of Secondary Structure Index and values plotted on the structure, as an example we plotted the Markov Matrix of residue 31 that displayed a mixed α/λ state; in the 3D structure of the protein, only residues classified as bones (see below) were represented with the standard helix/arrow representation.*

possibilities we define the number of Detected Peaks DP that is the number of disjoint intervals coming from all the possible intervals $U_i = [\mu_i - g\sigma_i; \mu_i + g\sigma_i]$. For example, if the correct number of gaussians describing a probability distribution is 4 and we have that:

$$(U_1 \cap U_2 \neq \varnothing) \wedge (U_3 \cap U_4 \neq \varnothing) \wedge (U_1 \cap U_2) \cap (U_3 \cap U_4) = \varnothing \tag{6.3}$$

DP=2 since we have 2 distinct "detectable" intervals.
To exploit the variations in the discrete variables in DSSP, 3 secondary structure states are defined for every residue: α (helix state), β (sheet state), λ (loop state). Then, a 3×3 Markov matrix was built per residue, with the probabilities of undergoing all possible transitions after a timestep τ (fixed to 50 ps, the trajectory output time). The Master Equation is then solved, finding the equilibrium probability of being in a certain equilibrium state.
To extract a single parameter value and identify the equilibrium distribution, we develop the Secondary Structure Index: we associated a unitary vector in the 2D plane to each secondary structure state. The α state is associated with a vector having an angle of 0° with the x axis, the λ state is associated with a vector with an angle of 120° with the x axis, and β state is associated with a vector having an angle of 240° with the x axis (see **Figure 6.3B** for a clearer view). Every state-vector is then multiplied by the equilibrium probability calculated before. The Secondary Structure Index is calculated as the angle between the sum of the 3 reweighted state-vectors and the x axis.
In this framework, angles close to an equilibrium value indicate a high probability of finding a given residue in that secondary structure, while angles far from equilibrium values indicate secondary structure instability.
In our analysis we are mostly interested in residues that populate α or β states, and we consider as stable residues with a Secondary Structure Index in the interval [−2.7°:2.7°] for α state and [237.3°:242.7°] for β state (those values correspond to the Secondary Structure Index of residues with equilibrium probability of 0.95 of being in α/β state and 0.05 of being in λ state).

6.2.2 Minimal Frustration Defines an Optimal Network

The introduction of the gaussian model for different variables permits to explore and detect different protein motions. To interpret the results coming from the fitting procedure, we defined 3 parameters: the Frustration F_i, defined as $F_i = DP_i$, where DP_i is the number of detected peaks for the variable i. It is used to identify which residues tend to display different equilibrium positions during a simulation. We next introduce the Splitting parameter D_i, reporting on the separation of the Gaussian peaks for a certain variable. This parameter is defined as a measure of the occupancy of an interval by gaussian peaks. The interval is in turn defined as the difference between:

$$L_l = \mu_{\min} - g\sigma_{\mu_{\min}}; L_u = \mu_{\max} + g\sigma_{\mu_{\max}} \tag{6.4}$$

where $L_{l,u}$ stands for lower/upper Limit, $\mu_{\min,\max}$ and $\sigma_{\mu_{\min,\max}}$ are the mean and variance of the smallest/biggest peak. Then, we calculated the fraction D_i of the interval $[L_l; L_u]$ occupied by the sub-intervals $\mu_i \pm g\sigma_{\mu_i}$ (considering possible intersections). D_i close to one means that the interval is almost completely occupied by gaussian, while D_i close to zero means that the different gaussians do not fill entirely the interval, so they are well separated. The last parameter is related to the fluctuation of an observable: we analyse its dispersion to determine how much a variable fluctuates with respect to a reference value. In a gaussian model, the sigma represents the conformational space the system explores around equilibrium position. Thus, σ is a measure of available conformations, and we can relate larger σ to higher space exploration. The conformational entropy is defined as:

$$S_c^i = \sum_{j=1}^{n} \ln \frac{\sigma_{i,j}}{\sigma_0} \tag{6.5}$$

where $\sigma_{i,j}$ represents the variability of observable i (with relative probability A_j from the fitting procedure reported in the previous paragraph) and σ_0 is a "standard" equilibrium dispersion (in the case of pairwise distance between C_α it is the average σ between $C_\alpha^i - C_\alpha^{i+1}$, for dihedrals it is the dispersion of the dihedral inside an helix, for C_α angles it is the fluctuation of the angle $C_\alpha^i - COG - C_\alpha^{i+1}$ and for SASA it was chosen the lowest fluctuation measured during the analysis on test systems). With this set of parameters, we can start to identify the sets of residues to be used as internal reference that define the relative motions in the proteins.

Since secondary structures are typically among the most stable points in a protein, they could be considered as good candidates for the identification of stable regions. Combining the results from Secondary Structure Index and Frustration of dihedral angles, we define protein "Bones" as protein substructures of at least 2 residues that have a single detectable gaussian in dihedral angles and a stable secondary structure in α or β state.

This definition is made more stringent by including the condition that distance and angle descriptors of every possible residue-residue pair in each bone have $F_i = 1$. This amounts to stating that bone substructures are internally rigid and do not undergo conformational transitions (self-consistence criterion). This condition is necessary to avoid cases that arise, e.g., from bending motions in long helices, where some residues at one end of the helix are effectively displaying a multiple gaussian behaviour with respect to one another. In these cases, the bone region is reduced to the biggest self-consistent ensemble within the previously defined bone.

The further combination of self-consistent subregions of Bones define the ensemble of Motion Receivers, defined as Stability Core. This core can be used to probe the motions of other residues in the protein; the advantage of this approach is that it naturally provides a consistent internal reference. In this framework, the analysis is simplified by the fact that the motion of every residue can be described in terms of translation and rotation with respect to two residues of the Stability Core (one for translation and the other one for rotation). The best probes are then given by the residues inside the core with the lowest D and the highest value of F, respectively. This allows to capture the various equilibrium positions for the moving residue outside the core.

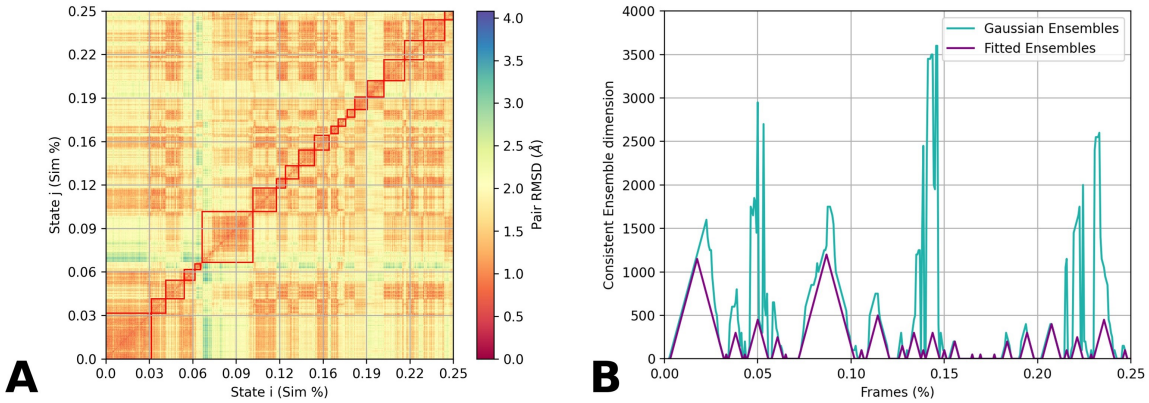

Figure 6.4: Examples of Pairwise RMSD Ensembles. ***(A)*** *Pairwise RMSD values of a system in a Replica and red squares represent the Ensembles found during the fitting procedures.* ***(B)*** *The y-value of the cyan line represents the theoretical dimension of a RMSD square centred on the x-value, while the purple one is the best approximation on the assumption that a RMSD Ensemble must extend with the same time length forward and backward in time from its centre.*

Finally, the last descriptor we consider is the Solvent Accessible Surface Area for the various residues. The Gaussian fitting method was applied also on the SASA calculated per residue, enhancing different possible orientations with respect to the outer environment and to the protein itself. Thus, non monogaussian behaviour can hint the possibility of entropic Protein-Protein (Ligand) binding. An exposure threshold was defined to simplify the analysis: only residues that had an exposed area greater than half a water molecule (0.123 nm^2) for at least half of the simulation time underwent to the fitting analysis. This observable was chosen to measure side-chain conformational space and to connect it with inner protein motion. In fact, even if we could, in principle, be able to define protein motions with C_α motions only, they do not actively participate in biochemical protein tasks. Thus, it was necessary to study possible connections between inner protein motions and changes in side-chain orientation and solvent exposure to define whether some residue could actively play a role during collective protein motions.
Overall, the procedure we have developed permits to identify in a natural and direct way the different ensembles of equilibrium positions ideally from simple MD simulations. The next step is to establish whether identified motions reverberate throughout the whole structure, affecting the entire ensemble of residues, or if their effect is simply confined to local conformational changes.

6.2.3 Determination of time evolution of the collective ensemble

To progress on this pathway, we wanted to determine time windows where distinct global conformational ensembles are populated. To initially define them, we make use of the pairwise RMSD matrix obtainable from MD simulations. Our working hypothesis here is that as different conformational ensembles are populated, the averaged pairwise RMSD between the two ensembles must be different (see **Figure 6.4A**). To quantify this difference, we define the conformational Ensemble as a group of consecutive frames where the pairwise RMSD probability distribution has $F = 1$. This assumption can be qualitatively verified in common diagonal square-like patterns (**Figure 6.4A**).
To practically identify these ensembles, we defined a minimum time-length T_0 (necessary to have a statistically consistent number of frames) and a time-step Δt. Starting from the frame $\frac{T_0}{2}$ until the frame $N - \frac{T_0}{2}$, the fitted distribution of pairwise RMSD was calculated at time t_i in the interval $[t_i - t_s; t_i + t_s]$, where $t_s = T_0/2 + j \cdot \Delta t$, where j is a progressive integer. The iterative fitting procedure on the distribution was done at every frame of the trajectory, increasing the dimension of the ensemble step after step until two distinct gaussians were found. What we obtained was the

time development of the size of detectable ensembles (see **Figure 6.4B**): the time is plotted on the x-axis and on the y-axis the number of frames needed to find another ensemble.
The final graph could look, depending on the system, very rough. Thus, we approximated it to a triangle-like function where the x coordinate of the vertices of triangles represents the centroids of ensembles, and the y coordinate represents how long backward and forward in time the ensembles extend. This way of defining ensembles gives many advantages: the distribution of the ensembles depends on the dynamics of the system, e.g. it depends on the characteristic values of RMSD spanned during the simulations and not on a pre-defined threshold; the dimension of the ensembles is defined progressively; steep transitions between states can easily be detected thanks to the highly different pairwise RMSD values; there is no need to define a priori the type or the length of the transition state, since it is automatically detected. An example of equilibrium states projected on pairwise RMSD matrix can be seen in **Figure 6.4**.
At the end of the Ensemble analysis, we obtained a sequence of time intervals where we know that the protein remains in the same equilibrium conformation. Moreover, we also obtained equilibrium gaussian distributions of single variable descriptors. Thus, we could combine local motion to ensemble one, to assess the impact of a single residue to the protein motion.

6.2.4 Finding the Allosteric Impact

According to an ensemble view of the allostery, where protein is seen as a collection of different states, we thought to describe as allosteric a region in the protein where local motions is connected to Ensemble one. We then defined the Allosteric Impact Vector (AIV) of a given observable ν as vector containing different information about the allosteric impact of the variable. In the case in which, for example, $\nu = \psi$ the AIV would be a collection of 3 vectors of dimension $N - 1$, with N number of residues. Since the relation between a local variable and the global one could be difficult to establish, in this work we propose (and measure) 3 different quantities answering 3 different questions:

1. Inside a global equilibrium time window, does a local variable display only one equilibrium position?

2. Does a change in equilibrium position of a local variable coincide with an ensemble change?

3. If local equilibrium position change coincides with an ensemble change, how different are the ensembles near the sudden switch and what is the impact of local variable change on the global ensemble difference?

To answer these questions, we defined a way to find changes in equilibrium positions in local observables. For each variable we know its equilibrium interval values (i.e. the disjoint gaussian regions $I^a, \ldots, I^n$). With this result, we can assign each point of a given variable $\nu(t)$ to the corresponding equilibrium region, finding time intervals where $\nu(t)$ belongs continuously to the same equilibrium position. To do that, we employed different criterion to obtain the most likewise assignment for each point. An average value $\bar{\nu}(t)$ is assigned to each point $\nu(t)$ calculating its mean value in the interval $[\nu(t - T_0/2), \nu(t + T_0/2)]$, with T_0 set to 1 ns. If $\bar{\nu}(t)$ falls inside one of the $I^a, \ldots, I^n$, then that point is given a coarse representation $q(\nu)$ that assigns the right interval. However, averages of points near a change in equilibrium position, could not fall inside one of $I^a, \ldots, I^n$. Then the vector of global probabilities P_{th} was calculated as the equilibrium probability of finding a point in one of the intervals:

$$P_{\text{th}} = \begin{pmatrix} \int_{I^a_{\min}}^{I^a_{\max}} \sum_{i \in I^a} A_i e^{-\frac{(x-\mu_i)^2}{\sigma_i^2}} dx \\ \vdots \\ \int_{I^n_{\min}}^{I^n_{\max}} \sum_{i \in I^n} A_i e^{-\frac{(x-\mu_i)^2}{\sigma_i^2}} dx \end{pmatrix} \tag{6.6}$$

with $I^a_{\text{min,max}}$ being the borders of interval a and A_i, μ_i, σ_i are the fitted gaussian parameters. Then, the vector of local probabilities can be calculated around $\nu(t)$ in $[\nu(t - T_0/2), \nu(t + T_0/2)]$:

$$P^v_{\text{sim}}(t) = \begin{pmatrix} \frac{1}{T_0} \sum_{i \in E} \chi^{I^a}(v(t)) \\ \vdots \\ \frac{1}{T_0} \sum_{i \in E} \chi^{I^n}(v(t)) \end{pmatrix} \tag{6.7}$$

where E is the interval around $\nu(t)$ and $\chi^{I^a}(i)$ is the characteristic function of the interval I^a (1 if i belongs to the interval and 0 otherwise). Thus, for every interval E^j we can calculate the quantity $P_r^{i,E^j} = P_{\text{sim}}^{i,E^j} / P_{\text{th}}^i$, that is the ratio between the probability of finding the variable inside equilibrium position i at time interval E. We then stated that if:

$$P_r^{i,E} \geq h \wedge P_r^{k \neq i,E} < h \tag{6.8}$$

stands, inside interval E we can assume that only a single equilibrium population is present, with h being the Allosteric threshold, that is the probability threshold over which a local frustrated variable can be assigned to a single equilibrium position.

This condition was necessary to ensure that just one among all the possible theoretical peaks was effectively occurring more frequently than the other ones. Analysing P_r, we can assess whether a single variable can be considered predominant inside a particular ensemble while, if none of the previous conditions was satisfied, no equilibrium position was assigned at point $\nu(t)$.

What we obtained then was the coarse approximation $q(\nu)$ of the variable ν during the trajectory. This procedure was done at every instant of the trajectory, assigning an equilibrium position according to equilibrium probabilities with no bias on changing in the surrounding environment. To assess the impact of the variable on the total Ensemble, we calculated the approximated value $q_E(\nu)$ of the variable inside time intervals defined during the Ensemble analysis. In this way, we could measure the impact of changes in local equilibrium position to global ensembles. The approximation procedure then is just a trick to simplify the analysis and enhance differences and correlations.

To answer question 1) we then measured the superposition between $q(\nu)$ and $q_E(\nu)$: the more this superposition is equal to 1, the more local and global equilibrium position coincide.

Another interesting insight that we can get from $q(\nu)$ is about time intervals where $q(\nu)$ remained stable. Indeed, from $q(\nu)$ we could get the time instants of "jumps" between local equilibrium positions, while from the previous analysis, we obtained time instants of "jumps" between global ensembles. We then answered question 2) by measuring the probability that a jump in local equilibrium position at time t_l^j (time of local jump) happens near an ensemble jump in global equilibrium position $t_E^j \pm t_d$ (time of ensemble jump $\pm$ time delay) and normalizing the number of time coincidences on the total number of t_l^j.

To answer the third question, we calculated the impact of a local variable change to the global structure by summing the average pairwise RMSD between consecutive states near single variables jump. The AIVs could then be built by using these 3 scores based on: superposition between local

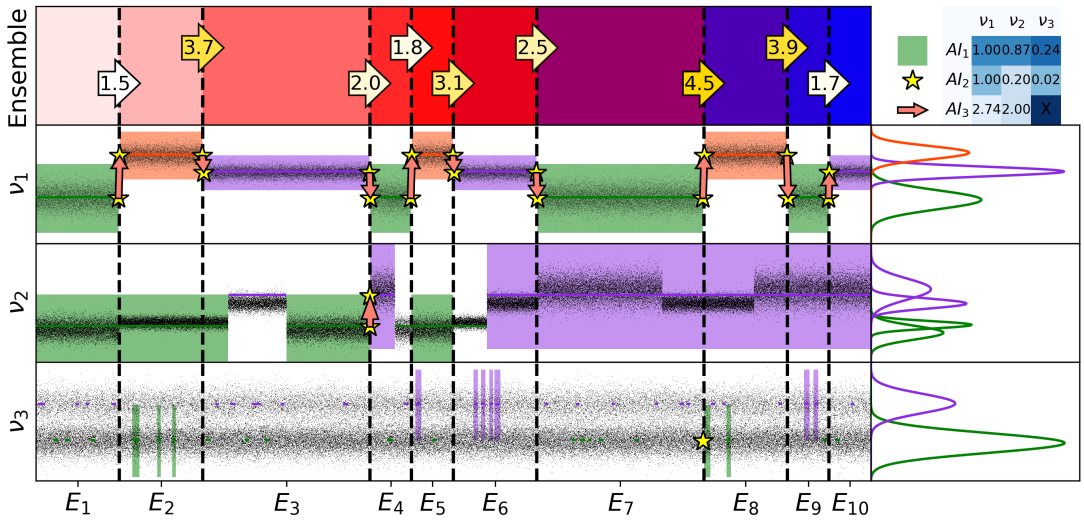

Figure 6.5: Pictorial illustration of Allosteric Impact. *On the top we have the time progression of the Ensemble with colours representing the absolute simulation time from 0 (light red) to 1 (blue), the arrows represent the averaged pairwise RMSD between the two ensembles coloured from white (low averaged difference between the ensemble) to dark yellow (high averaged difference). On the top right there is the Allosteric Impact Matrix for variables $\nu_{1,2,3}$; these data were randomly generated just to pictorially explain the AIV definition. The first line of AIV represents the superposition between quantized local variables defined around small time intervals and coarse grained local variables defined between Ensemble time borders; the second line represents the frequency of events that happened near an Ensemble change; the third line represents the averaged difference between Ensembles detected by a given variable. These three aspects are represented by superposition between large rectangle and small line (1), stars (2), arrows (3). Thus, ν_1 is perfectly coherent with the Ensemble and has a perfect superposition, its changes happen near the Ensemble one and has also an optimal averaged time coincidence and its impact on the global changes is given by the total average. Instead ν_2 has a good global superposition but a bad time coincidence, meaning that probably this variable is marginal in the Ensemble changes, but it could happen that a local change is reflected in the Ensemble. Finally, ν_3 badly represents the global changes, mainly because local changes are too fast or completely random and despite frustrated, the variable is not connected to the global motion.*

and global view, coincidence between local and global changes, impact of local variable on ensemble differences.

Thus, this analysis is intrinsically connected to an ensemble view of the Allostery: when the conformation changes, it is possible to say that the protein finds itself in a different structural ensemble with (possibly) different functions. Determine which variables (and relative motions) have a major impact on the intricate collective mechanism could help in explaining why some residues are more relevant than others for protein dynamics and how they impact on the collective motion. Once the components of AIV were calculated, we projected these values on the AIV_1,AIV_2 plane. We are then interested in those points that display higher values for both vectors, with possible high values on AIV_3. We decided to restrict our analysis on those points in the area $[Q^{0.5}(AIV_1), Q^{0.5}(AIV_2)]$, where $Q^{0.5}$ is the quantile representing the values higher than the median.

6.2.5 Signalling pathways

After retaining only those points that contributed the most to global allosteric movement, we tried to identify some possible endpoints of "signalling pathways" looking for time coincidences between

local changes. The signal from a residue I to a residue J was measured similarly to the coincidence between local changes and global ones, comparing the superposition between local variation on I and J and vice versa.
However, variables displaying large number of switching events could give rise to random coincidences leading to an error in allosteric signalling quantification. Thus, four thresholds were employed to filter out random coincidences, retaining just significant events. These were related to: minimum number of coincidences observed, minimum efficiency between signal that were sent from residue I and received from residue J, maximum delay between coincident events, number of replicas in which the coincidence was observed (respectively $\varepsilon_n, \varepsilon_f, \varepsilon_d, \varepsilon_R$).
These parameters were tuned using as test system a collection of 500 random variables, each displaying a random number of gaussian peaks (between 1 and 5), random equilibrium values, random variance and population percentage. Such signals were then mixed, generating a random number of swaps between equilibrium population. Tuning the parameters enabled the discrimination of random coincidence events, retaining only significant pairs characterized by one (or more) pair of connected variables highlighting the possibility of allosteric connection between distal sites of the protein.

6.2.6 Molecular Dynamics Simulations

Six systems were modelled for this work to inspect the ability of our method to characterize allosteric communication inside proteins and also to study how different biochemical events can alter these processes. The three main systems are Ubiquitin protein [283, 284], Second PDZ domain [285] and Checkpoint Kinase [286], while biochemical events modelled are: tail deletion (removing four C-terminal disordered residues in Ubiquitin), peptide binding (simulating Second PDZ domain in its bound and unbound form), ligand binding (simulating Checkpoint Kinase in its bound and unbound form). The systems were chosen both for their biological interest and for their well-known results and benchmarks in the Molecular Dynamics community. The results will be presented following the methodological approach, starting from Frustration measurements and Network analysis and arriving to Allosteric Impact results.
Systems were modelled starting from the following PDB codes: 1UBQ (both ubiquitin forms), 3LNX and 3LNY (apo and peptide bound forms of second PDZ domain), 1IA8 and 2E9N (apo and ligand bound form of Checkpoint kinase 4). Modelling and simulations followed the procedure described in **Chapter 2**.

6.3 Results

6.3.1 Ubiquitin

Ubiquitin protein is one of the most studied systems in molecular dynamics simulation [284, 287–289]. This is due to its small number of aminoacids together with its well folded core and widespread presence in many biological organisms. Many studies highlighted that, at equilibrium conditions, Ubiquitin can display small conformational changes occurring at the $ms - \mu s$ timescales. The objective here is to verify where FAN method can capture, measure and identify those movements, and possible modifications of collective motions after deletion of four C-terminal residues.
The SSI analysis in **Figure 6.7A,B** highlights the presence of some unstably folded residue, in particular 3 amino-acids around residue 40 exhibited the tendency to fold as helix, even if the folded form was not achieved for the entire simulation (red circles is upper left panels).
From the analysis of Splitting, Frustration and Conformational entropy, the small loop region around residue 10 seems to be the most important in terms of conformational exploration (see **Figure 6.6A,B**). The main difference between the two systems is located in the region of the cut

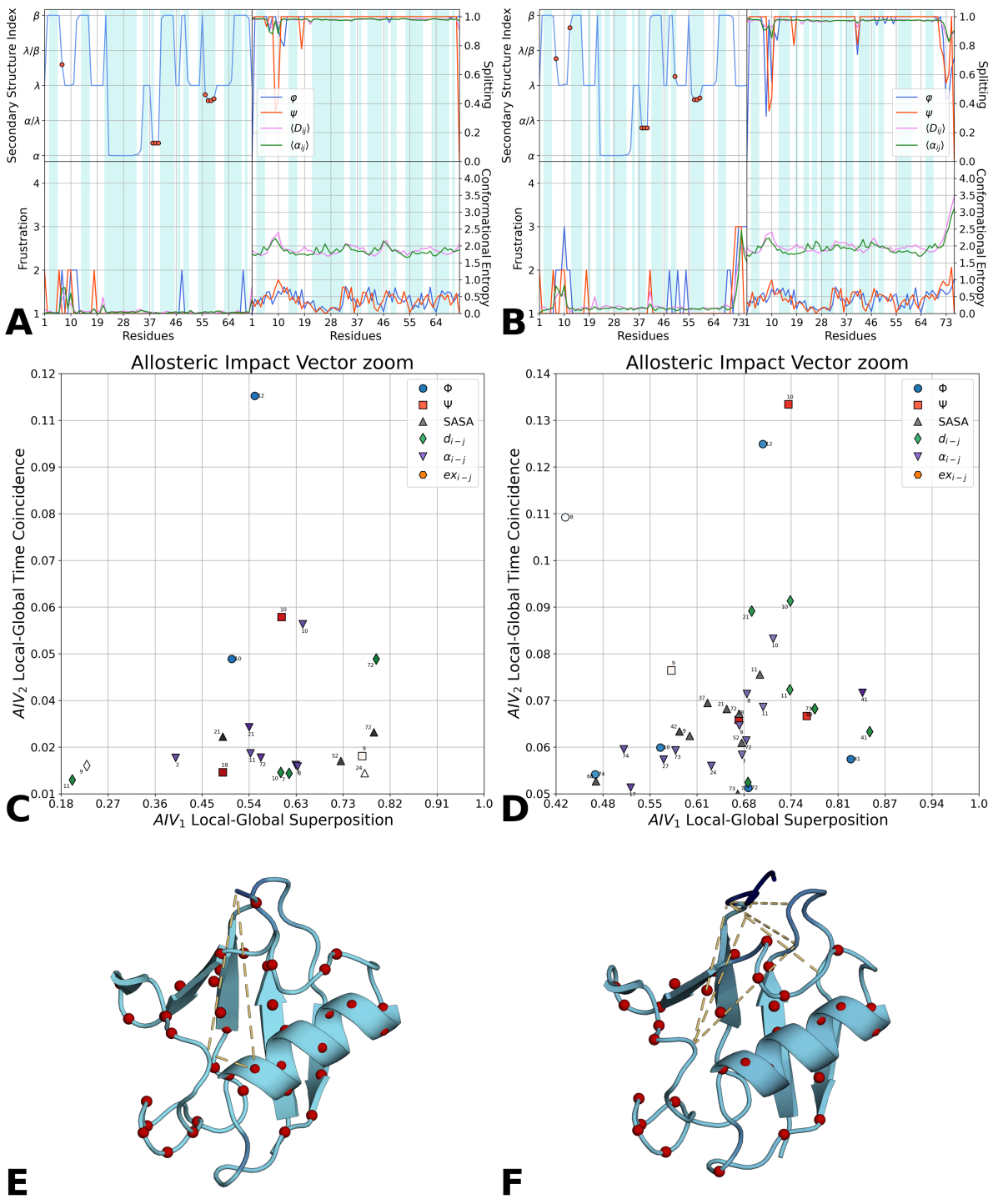

Figure 6.6: FAN results for Ubiquitin systems. ***(A,B)*** *Plot of SSI, Splitting, Frustration and Conformational Entropy calculated for each residue and observable in truncated ubiquitin and in full length ubiquitin respectively; in the case of Distance and Angle the plotted result is the averaged one, because we wanted to highlight regions that averagely displayed high fluctuation.* ***(C,D)*** *Regions with points that have higher Allosteric Impact Vector values for both systems, on the x axis there is the first component of AIV while the second is on the y axis.* ***(E,F)*** *Ubiquitin structures coloured by averaged distance Frustration value, red spheres represent points in the stability core, while yellow dotted lines represent allosteric coincidences found during the analysis.*

loop, which is exhibiting a larger frustration and entropy. Other minor differences arise in the dihedral values of Splitting and Frustration values of some residues which are likely to depend on the different phase space exploration during dynamics.
As expected from a stably folded protein like ubiquitin, many residues displayed a Splitting value equal to one, meaning that dihedral angles and residue-residue distances/angles tend to explore just one equilibrium position. Instead, the region around the first loop (residues 5-12) clearly displays a tendency to explore more equilibrium positions, especially for ψ angle, highlighting possible conformational changes, in both forms of the protein. This agrees with the Frustration analysis which shows that even if different equilibria are achieved, no more than 2 distinct gaussian peaks are observed. Finally, Conformational entropy profiles show that, apart from the C-terminal tail, both the complete protein and the cut one have the same behaviour. This leads to the conclusion that the terminal loop does not play a major global role in the definition of conformational sampling.
Starting from this information, the Allosteric Impact Vector was calculated after the global Ensemble evolution, showing interesting results. Both systems display a similar population of points, but the complete ubiquitin, reasonably, has more points contributing to possible allosteric events, being the longer tail more prone to undergo structural changes that trigger the ensemble ones. To inspect the most interesting part of the plane, we selected the values exceeding the 0.5 quantile for both AIV_1 and AIV_2 (**Figure 6.6C,D**). In the plot, also the residues responsible for allosteric events were labelled, showing that the first loop is the part of the protein mostly responsible for allosteric conformational changes.
Finally, the two structures coloured by distance frustration were plotted (**Figure 6.6E,F**, light blue means frustration close to one, while darker blue means frustration close to, or higher than two) together with stability core residues (red spheres). Moreover, coincidences were searched between these allosterically relevant residues and plotted connecting with yellow dashed line residues displaying high coincidences. Interestingly, in both systems, the C-terminal tail is playing a signalling role, being connected to distal residues. The complete ubiquitin structure indeed has a greater communication network, involving many different residues across the entire protein.
An example of such signalling network is reported in **Figure 6.7**, where a triangulation of allosteric events is observable by means of coincidence analysis. From this analysis, the main variables involved in this complex movement are angles/distances from the core of residue 7 and 72, together with solvent accessible surface area of residue 52. This could suggest a strong allosteric coupling inside Ubiquitin, with a complex network that transmits important changes from one site to another. This analysis also reduces the number of frustrated variables which display more conformations to the ones that are actually connected, discriminating between random events and strongly coupled ones. Nothing can be said about time connection or cause-effect phenomena; even if it is, in principle, possible to determine the start of a different Ensemble, the intrinsic noisy nature of signal does not enable a robust way to determine time correlation, but just to highlight hidden connection between different regions of the protein.

6.3.2 Second PDZ Domain

Second PDZ Domain is another well-studied system for allostery, signalling and intra-protein communications [285, 290–293]. This is due both to its small structure that enables longer simulation times and to its binding mode with the substrate, which can be considered a case of allostery without conformational change, since the active site where the substrate binds does not change shape upon binding [23, 25]. Second PDZ Domain was simulated starting from pdb codes 3LNX and 3LNY, where the protein is present, respectively, in its apo and peptide-bound form.
Analysis of SSI, Splitting, Frustration and Conformational Entropy revealed that the main source of conformational differences is located inside the region 23-34, where the loop is exhibiting different equilibrium positions in both apo and peptide bound form (**Figure 6.8A,B**). Moreover, the region between residues 67-78 exhibits larger value of frustration in the apo form with respect to the bound

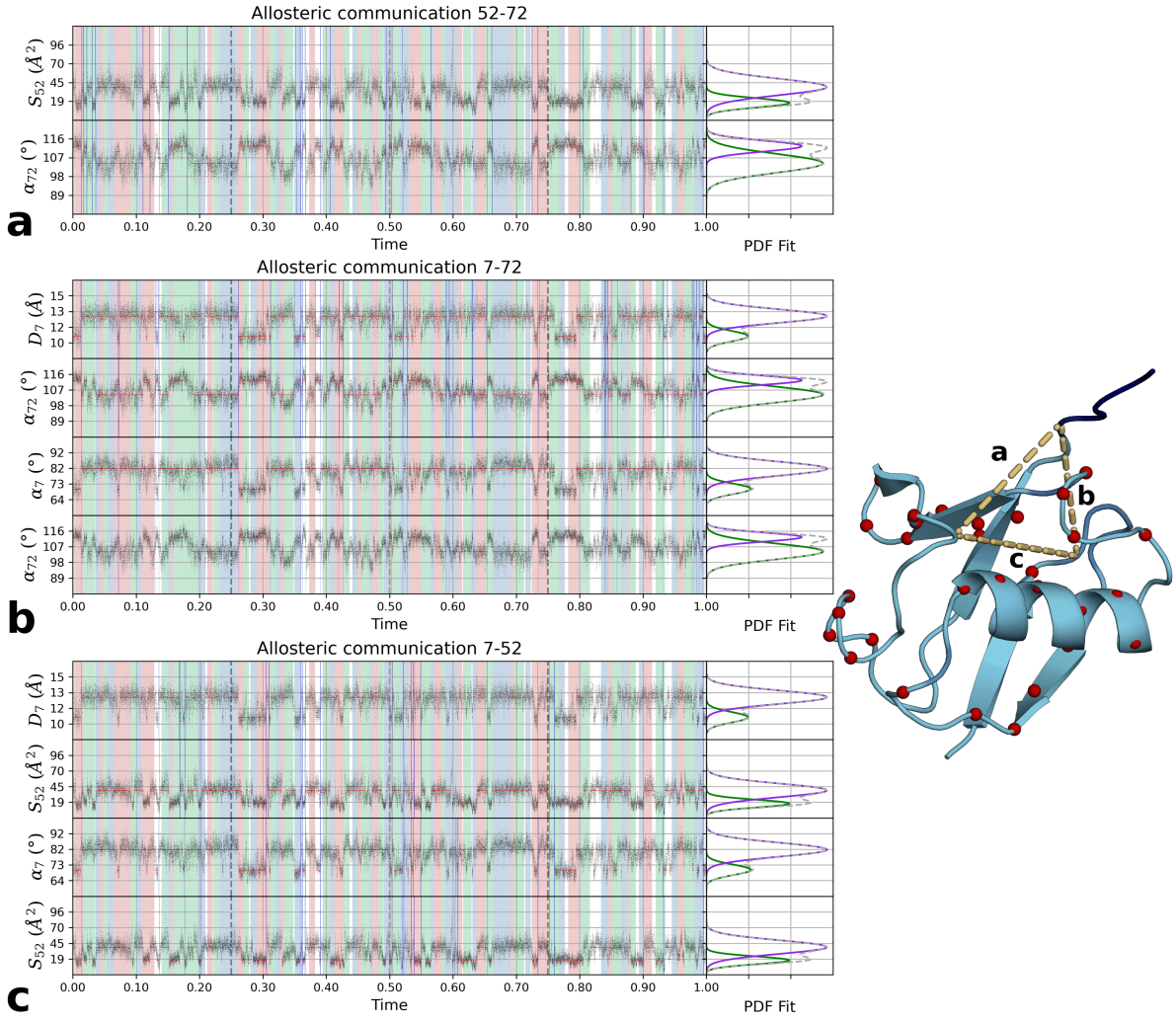

Figure 6.7: Example of allosteric communication inside Ubiquitin. *3 yellow dashed lines connect residues 7,52 and 72; labels on dotted lines coincide with plots highlighting the coincidences Ensemble changes between different variables of the relevant residues. Time intervals where the same global ensemble is observed are shown with the same colour, so a change in colour means change in Ensemble, but same/different colour does not necessarily mean same/different Ensemble. Coincidences between variables are plotted with vertical blue lines.*

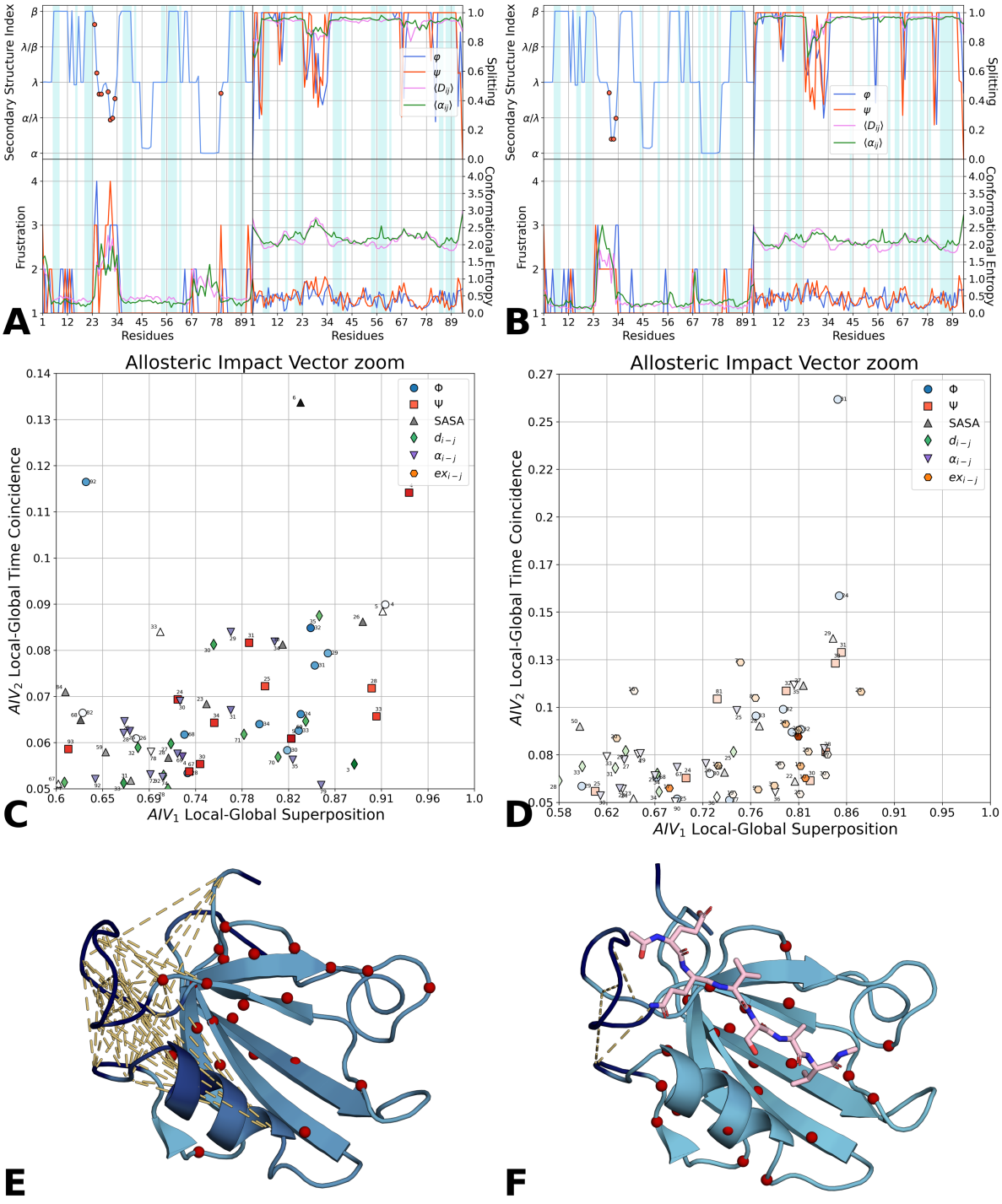

Figure 6.8: FAN results for Second PDZ domain systems. ***(A,B)*** *Plot of SSI, Splitting, Frustration and Conformational Entropy calculated for each residue and observable in apo and peptide-bound form respectively; in the case of Distance and Angle the plotted result is the averaged one, because we wanted to highlight regions that averagely displayed high fluctuation.* ***(C),(D)*** *Regions with points that have higher Allosteric Impact Vector values for both systems, on the x axis there is the first component of AIV while the second is on the y axis.* ***(E),(F)*** *Second PDZ domain structures coloured by averaged distance Frustration value, red spheres represent points in the stability core, while yellow dotted lines represent allosteric coincidences found during the analysis.*

form. This region corresponds to a stably folded helix that is involved in the binding region of the peptide, that even if always stably folded during the simulation (SSI index in the helix region), is not involved in the stability core in the case of apo structure. Entropy plots of the two states are comparable, apart from a slightly higher distance/angle entropy in the region of the first loop (residues 23-34) in the apo structure.
Allosteric impact plots (**Figure 6.8C,D**) highlight that the variables corresponding to residues in the loop-helix region are the ones most involved in coincidences between a local and global changes. In the peptide bound plot, the presence of variables connected with peptide conformations are visible.
In **Figure 6.8E,F** results are plotted onto the structure, showing a substantial modification of the behaviour of second PDZ domain in presence or absence of peptide. Notably, the frustration profile has a radical change when the peptide is present, globally stabilizing the protein and reducing the number of available conformations. In the apo form, instead, frustration is localized around helix and loop, pointing out that, even if stably folded, the helix is performing a global motion which detaches it from the stability core. Indeed, another important modification is related to the stability core, which is different in the two systems. Peptide bound structure has the aforementioned helix included in the core, highlighting that it actively participates in binding, and proving that our method is capable of capturing the tight dynamical difference between the two systems.
The analysis of time coincidences further proves what observed before, in fact the number of coincidences drastically decreases when the peptide is bound. Indeed, this is consistent with what previously stated, since a reduced frustration means reduced number of equilibrium position and reduced number of conformational changes in time, which are thus less probable. Instead, the apo form shows an extremely high number of coincidences in the region of the binding pocket. This possibly highlights that a concerted motion is necessary to explore different conformations, and that helix movements affect loop ensembles. Peptide binding stops these signalling events, reducing the number of available conformations.

6.3.3 Checkpoint kinase

As third test systems the Checkpoint kinase 1 was chosen for its known allosteric modulation and was simulated in the apo and inhibitor-bound form. Structurally, Chk1 consists of an N-terminal catalytic kinase domain and a C-terminal regulatory domain connected by a flexible linker region. The kinase domain contains the ATP-binding pocket and substrate recognition motifs characteristic of the protein kinase family. The allosteric regulation of Chk1 is of particular biochemical interest. Rather than being controlled solely through ATP binding at the active site, Chk1 activity is modulated by conformational communication between its regulatory and catalytic domains [286, 294, 295].
In **Figure 6.9** results from different index of FAN method are reported. The main difference arising in terms is located in the N-terminal domain. In fact, while SSI index looks similar in both cases, Splitting, Frustration and Entropy indices are particularly different between the two systems in this region (residues 1-40). The difference is visible mostly in distance and angle values (green and pink lines, respectively), indicating a possible rotating movement generated in the β-sheet region of N-terminal domain. Higher frustration and entropy are seemingly generated by the introduction of the inhibitor in the binding site region, that affects this domain, while the other seems to remain unperturbed.
The projection of Frustration on the kinase structure highlights that the β-sheet region is the one which is mostly affected, and that the binding also restricts the stability core to be located on the lower, C-terminal region. While in the apo structure the stability core was dispersed on the entire protein, in the inhibitor-bound case many residues appear cooperatively destabilized. This is supported by the fact that the N-terminal domain maintains a stable fold, but all the residues are excluded from the major, stable core, indicating a possible coherent movement.
Allosteric communication appears to be largely modified: inhibitor stabilizes the pocket, diminishing the frustration arising from loops in the region connecting the two domains. Instead, a larger number

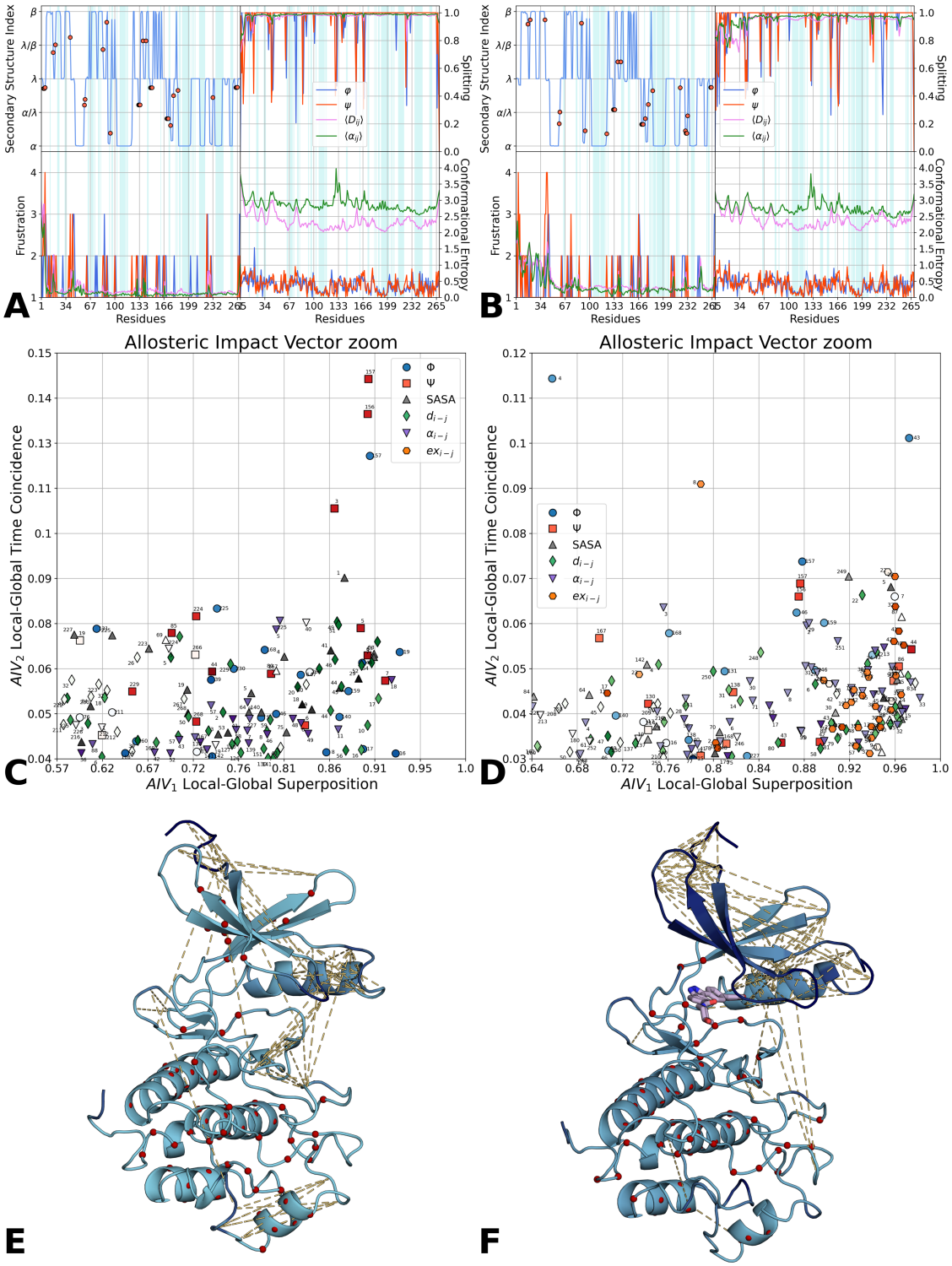

Figure 6.9: FAN results for Checkpoint Kinase Chk1 systems. ***(A,B)*** *Plot of SSI, Splitting, Frustration and Conformational Entropy calculated for each residue and observable in apo and inhibitor-bound form respectively; in the case of Distance and Angle the plotted result is the averaged one, because we wanted to highlight regions that averagely displayed high fluctuation.* ***(C,D)*** *Regions with points that have higher Allosteric Impact Vector values for both systems, on the x axis there is the first component of AIV while the second is on the y axis.* ***(E,F)*** *Checkpoint kinase Chk1 structures coloured by averaged distance Frustration value, red spheres represent points in the stability core, while yellow dotted lines represent allosteric coincidences found during the analysis.*

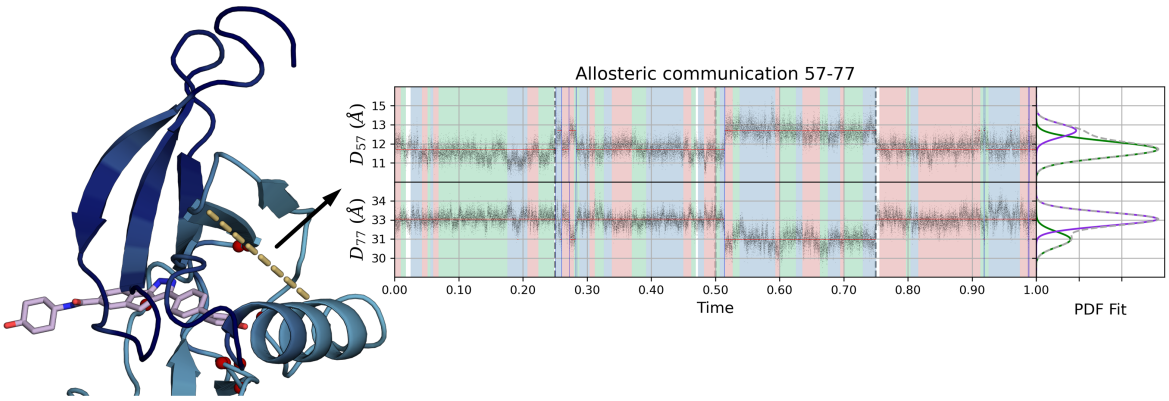

Figure 6.10: Example of allosteric connection between residues in Checkpoint kinase 1. *Inspection of the characterized variables suggests that both residues are performing the same motion with respect to C-terminal domain, highlighting how the ligand decouples the two domains.*

of coincident motions are happening in the β-sheet region. This may be due to a local, coherent roto-translational motion of the whole N-terminal domain, whose main motion around C-terminal domain is, on average, the same for many residues. Thus, being the domain motion the same for all its components, clearly many coincidence events will take place. An example is the allosteric connection found between residues 55 and 77 (see **Figure 6.10**). Here, both allosteric variables represent the distance of the two residues from two points, located in the stability core, which best describe their movement. Thus, in this case, the coincidence between these variables should be considered as a coherent motion of the two points with respect to the stability core.
Then, in this case, the insertion of the ligand provokes a conformational rearrangement that affects the N-terminal domain, generating a slight motion that disentangles the two domains which are instead more compact in the apo case. Our method proved to detect such motion and to identify which residues could be ones that could best describe different motions in terms of a physico-chemical collective variable.

6.4 Discussion

The complexity and the variety of protein conformational motion have always fascinated scientists, that, using both experimental and computational methods, tried to shed light into the inner mechanisms that govern such phenomena. The development of theoretical principles like allostery, or the adaptation of frustration in complex systems to the protein dynamics helped in the explanation of molecular motions. Here we provided a new computational method, named FAN, which can combine the local analysis of many, reference-independent quantities with the ensemble characterization of protein conformations. Inserting those key concepts in a framework where every quantity can be described with a gaussian probability, we managed to demonstrate that a fine characterization of conformational ensembles for proteins is indeed possible, identifying those residues which take part in the stabilization of the protein, acting like a strong scaffold, namely the stability core. Interestingly, such stability core can be modified upon the action of external elements, i.e. substrates, that stabilize one frustrated part of the protein, promoting it inside the common scaffold.
The analysis of pairwise RMSD values, demonstrated that ensemble motions can be coupled to local ones, highlighting the presence of subtle mechanisms governing the switching between two (or more) conformational ensembles. Indeed, such ensembles proved to be effectively describing the nature of possible protein function, pointing out that frustrated regions govern the tight balance between different conformations, enabling proteins to exploit their functional role. Finally, time coincidences proved to be the right key to identify and understand allosteric mechanisms at the base of signalling

processes inside proteins, bridging local conformational changes to global ones.
Our next goal is to expand this method to the analysis of larger proteins, including multi-domain ones, and to tune it to capture the nature of conformational motion also in larger or coarse-grained system, connecting systems with different scales.

Chapter 7

Conclusion and Perspectives

This work has demonstrated the extensive application of Molecular Dynamics (MD) simulations to the investigation of complex biomolecular systems, with the aim of elucidating the molecular mechanisms governing protein–protein interactions and allosteric regulation. By examining a diverse range of biochemical processes, from the complete maturation cycle of the glucocorticoid receptor (GR) to more localized allosteric events, including ligand binding and post-translational modifications (PTMs), this thesis has shown that MD simulations constitute a powerful framework for probing protein dynamics and revealing the mechanistic basis of biological function. The integration of state-of-the-art analytical approaches, including Distance Fluctuations (DF) and the Shortest Path Map (SPM), enabled the identification of previously unrecognized features of the systems under investigation. In addition, novel computational methodologies were developed to address limitations of existing approaches, including TOG for the characterisation of large-scale motions in globular proteins, the internal D-NEMD method for the analysis of subtle responses in non-equilibrium molecular dynamics simulations, and the FAN method for the identification of allosteric communication pathways from MD trajectories.

The methods employed in this thesis provide complementary perspectives on allosteric communication, rather than constituting alternative approaches to the same aspect of protein dynamics. Each method captures a different component of the overall mechanism underlying protein function, and its suitability therefore depends on the biological system and the specific question being addressed. For large, multidomain proteins and macromolecular complexes, the TOG method can be particularly advantageous because it provides a description of global molecular motions through collective coordinates, facilitating the identification of large-scale conformational changes that may be difficult to characterise using residue-based approaches alone. SPM, in contrast, is particularly suited to identifying equilibrium communication pathways between functionally relevant regions. However, because SPM relies on the structural representation and alignment of the conformational ensemble, its interpretation can become more challenging in systems containing highly mobile or structurally heterogeneous regions. In such cases, the FAN method can provide complementary information by identifying stable regions of the protein and characterising distinct equilibrium states of more mobile regions. Combined with SPM, this information can help distinguish persistent structural communication pathways from changes associated with conformational heterogeneity, thereby providing a more complete description of the equilibrium allosteric mechanism.

D-NEMD addresses a different aspect of allostery by explicitly probing the response of a system to an external perturbation. It is therefore particularly informative when the objective is to investigate how a perturbation propagates through the protein and how the system departs from and subsequently evolves towards its equilibrium behaviour. Its main limitation is the substantially higher computational cost associated with performing multiple non-equilibrium simulations, which can restrict its application to very large systems or extensive conformational sampling. Nevertheless, D-NEMD provides information that cannot be obtained from equilibrium analyses alone. In particular, combining SPM with D-NEMD can bridge the structural and dynamical descriptions of allostery:

SPM can identify putative communication pathways present in the equilibrium ensemble, whereas D-NEMD can test how perturbations propagate along, or potentially deviate from, these pathways. The combination can therefore distinguish structural connectivity from an actual dynamic response to perturbation, providing a more comprehensive view of the mechanisms underlying allosteric communication than either approach considered independently. Overall, TOG, SPM, FAN and D-NEMD should therefore be regarded as complementary tools whose relative utility depends on the size, conformational heterogeneity and dynamical regime of the system, as well as on whether the primary objective is to characterise equilibrium structure and communication or the response to an external perturbation.

The investigation of the glucocorticoid receptor maturation cycle provided new insights into the structural and dynamical principles underlying the function of the Hsp90 chaperone machinery. Through dynamical nonequilibrium molecular dynamics (D-NEMD) simulations, it was possible to characterise, at an unprecedented level of detail, the asymmetric structural response of the Hsp90 catalytic pocket following ATP hydrolysis. Remarkably, this asymmetric behaviour was preserved in both complexes containing different immunophilins, suggesting the existence of a robust and finely regulated allosteric mechanism that operates independently of the specific molecular context, but with different extents depending on the global allosteric modulation. These findings contribute to a deeper understanding of how Hsp90 coordinates its ATPase cycle with cochaperone binding and client recognition, highlighting the central role of long-range communication within this highly dynamic molecular machine.

The computational investigation of phosphorylation at two serine residues within the Hsp90 charged loop further demonstrated how relatively small chemical modifications can exert significant long-range effects on protein dynamics. Analysis of collective motions, particularly through principal component analysis, revealed a measurable stiffening of the chaperone that emerged even within the timescales accessible to conventional MD simulations. These results emphasise the sensitivity of protein dynamics to post-translational modifications and reinforce the concept that subtle local perturbations can propagate through extensive allosteric networks to modulate protein function. The study of the catalytic cycle of human Kinesin-5 (Eg5) illustrated the ability of MD simulations to accurately capture the conformational landscape of functionally important structural elements, such as loop 5. The resulting characterisation of its conformational dynamics provided a novel mechanistic interpretation of inhibitor binding, complementing and extending available experimental evidence. This work further demonstrates how computational approaches can bridge the gap between static structural information and the dynamic processes that ultimately determine protein activity and pharmacological response.

A further contribution of this thesis is the development of a novel computational framework for the study of allostery, designed to reconcile the description of signalling events with the ensemble-based view of allosteric regulation. By enabling the identification of subtle intra-protein communication pathways, this methodology reveals how distant regions of a protein can interact through finely balanced modulations of structural fluctuations and energetic frustration. Beyond providing a more comprehensive description of allosteric communication, this framework offers a versatile tool for investigating the molecular consequences of subtle environmental changes, ligand binding events, or other biochemical modifications, thereby expanding the repertoire of computational strategies available for studying protein regulation.

Throughout this thesis, analyses were in several cases performed on concatenated trajectories from multiple independent replicas of the same system. A limitation of this approach is that it does not explicitly characterise the variability between individual replicas, and therefore does not directly assess the reproducibility of the observed behaviours across independent simulations. A replica-wise analysis could provide additional information on the variance and consistency of the properties investigated. The primary aim of this thesis was, however, to characterise ensemble-level properties that could be compared with experimental observations and, where possible, to identify the microscopic mechanisms underlying protein motion. The analyses were therefore primarily designed to

characterise the behaviour of the overall simulated ensemble rather than to systematically investigate differences between individual replicas. A dedicated analysis of replica-to-replica variability would constitute a valuable extension of the present work, but was beyond the scope of this thesis.
Overall, this work demonstrates that Molecular Dynamics simulations can serve as a powerful molecular "microscope", capable not only of reproducing experimentally observed behaviour but also of providing mechanistic and predictive insight into the dynamical principles governing biomolecular function. By combining advanced simulation techniques with the development of novel analytical methodologies, this thesis contributes to a more detailed understanding of how local molecular perturbations propagate through complex biological systems to regulate protein activity. More broadly, the approaches presented here establish a flexible computational framework that can be readily extended to the investigation of other biomolecular systems, ultimately contributing to the rational interpretation of biological mechanisms and supporting future developments in structural biology, biophysics and structure-based drug discovery.

Appendix A

Supplementary files for Chapter 3

In this Section, supplemental files for **Chapter 3** are reported.

Table A.1: *A summary of the TOG domain division (final selection) for each protein belonging to the GR activation cycle multiprotein assemblies.*

Proteins	Domains	Selections
Hsp90A	NTD_A	[M30; D66], [K74; G114], [G132; A166], [G168; T195]
	MiD_{1A}	[W297], [R299; E332], [Q334; D350], [K358; N397], [R400; L403], [Q405; E422]
	MiD_{2A}	[S442; E537]
	CTD_A	[E547; L549], [L551], [E553; S602], [K632; L671]
Hsp90B	NTD_B	[M30; P67], [E75; I110], [G114; E120], [D127; A166], [G168; T195]
	MiD_{1B}	[W297; P301], [D303; E332], [Q334; F349], [K357; E432]
	MiD_{2B}	[S442; E537]
	CTD_B	[E547; L549], [L551; P552], [D554; S602], [K632; L672]
GR	LBD	[R558; E631], [L636; S765]
Hsp70C	NBD	[A4; R76], [W90; R187], [G191; G224], [T226; S254], [N256; I379]
	SBD	[D395; T502]
Hsp70S	NBD	[A4; I379]
Hop	Hop_1	[D217; E329]
	Hop_2	[L359; Q469]
	Hop_3	[D484], [K486; L539]
p23	Core	[A4; D102]
	α-tail	[M117; G131]
FKBP51	FK_1	[G32; V37], [K52; D68], [H71; R73], [E75; F79], [I87; K99], [E101; K108], [A126; L134]
	FK_2	[G147; N163], [G165; R175], [D182; G220], [L241; E251]
	TPR_1	[T261; L346]
	TPR_2	[L362; K397]
FKBP52	FK_1	[G50; S69], [L71; E110], [N125; K138]
	FK_2	[I150; A226], [A241; K254]
	TPR_1	[E264; E303], [F306; D349]
	TPR_2	[G355; I400]

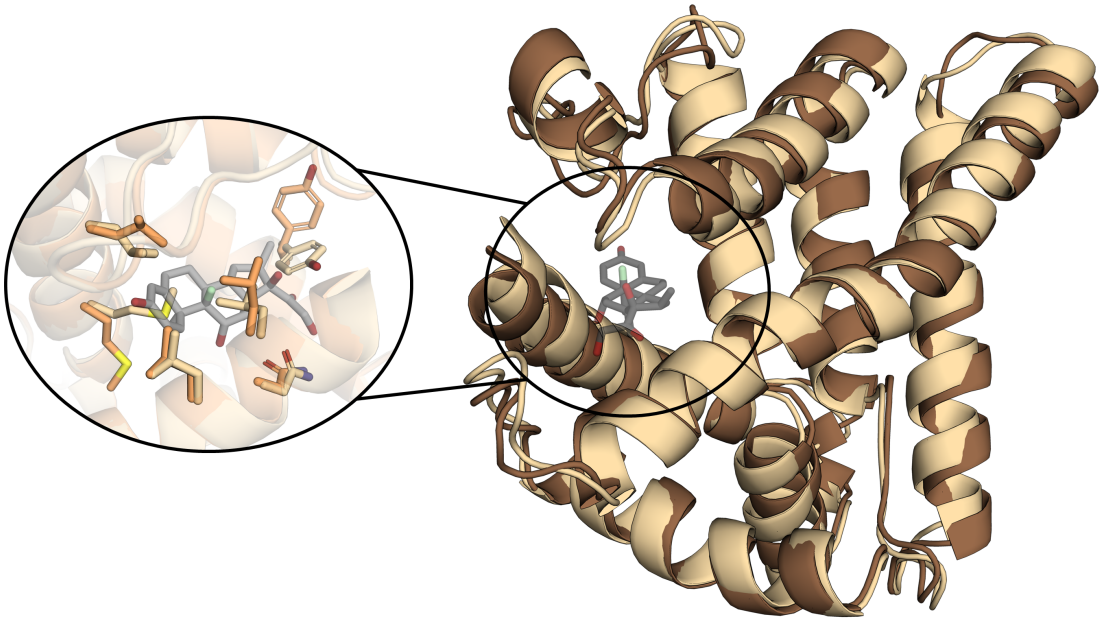

Figure A.1: *Synthetic agonist DEX (dexamethasone) docked into the GR active site. On the right, we superimpose the GR core backbone from frames isolated from MD simulations of the Loading-DEX complex (light yellow) and Maturation+DEX complex sharing the lowest RMSD pair. On the left, a zoomed view of the DEX docking pose (grey sticks) in the GR active site: light yellow sticks represent the residue side chains before trimming; while dark yellow sticks represent their new post-docking positions, highlighting the active site remodeling.*

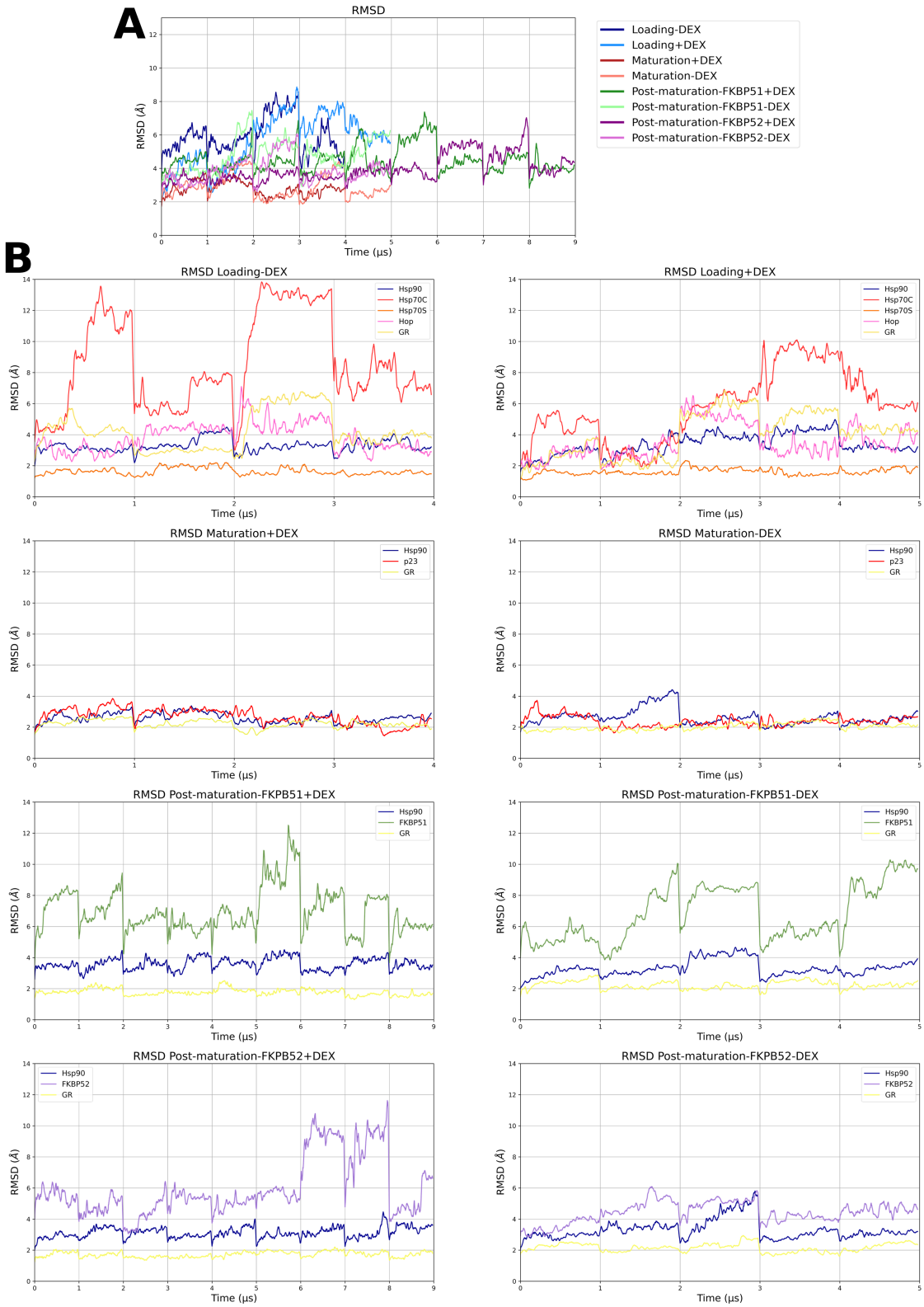

Figure A.2: *RMSD plots of* ***(A)*** *the C_αs of the secondary structures along each simulation and* ***(B)*** *the C_αs of each single protein component.*

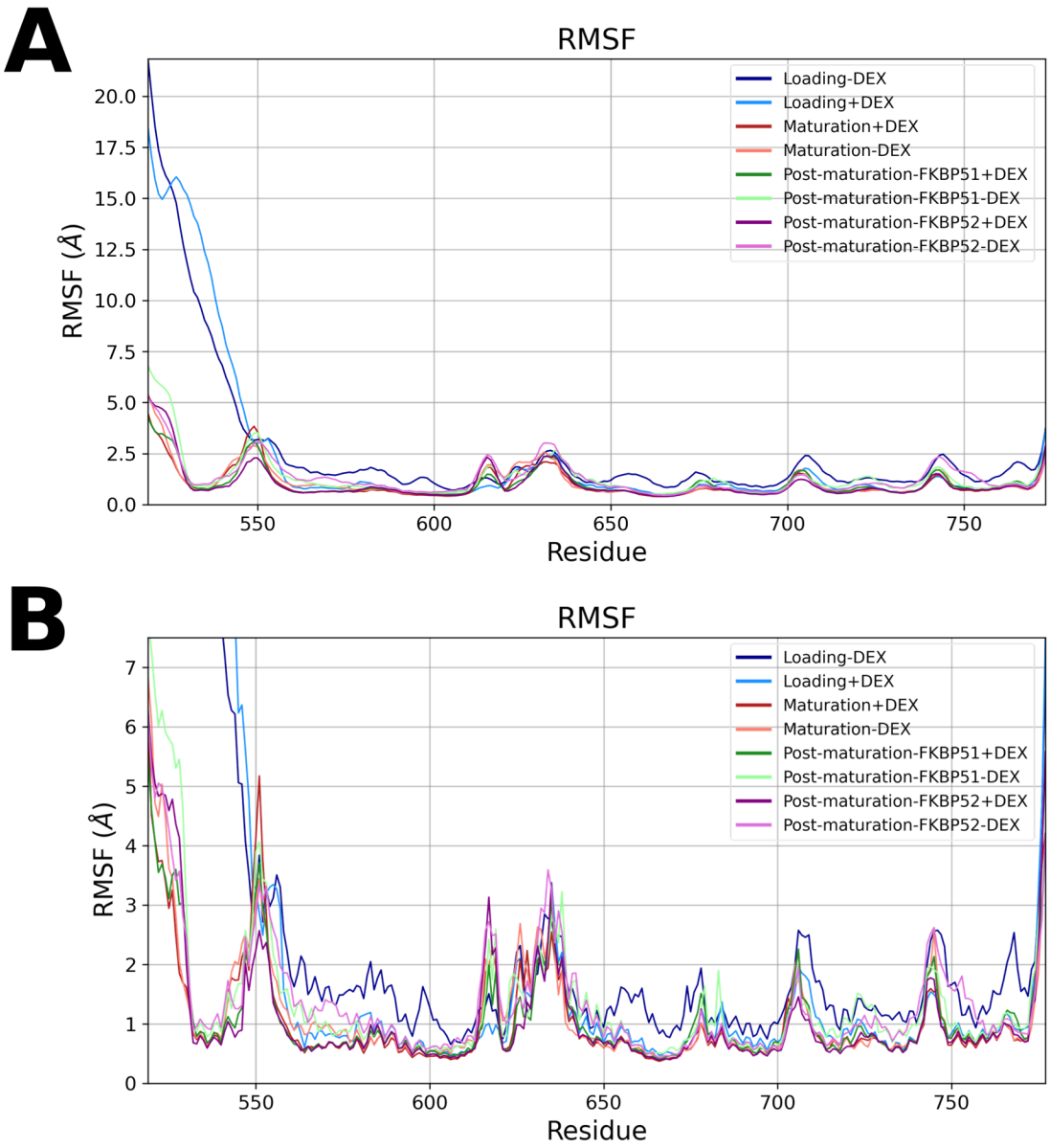

Figure A.3: *GR RMSF plot for all equilibrium simulations (equilibrium systems are in darker colors; perturbed systems in lighter colors). For better visualization, the plot lines are raw data convoluted over blocks of 5 residues.* ***(A)*** *Raw RMSF result.* ***(B)*** *A zoomed view in the 0-7.25 Å RMSF range.*

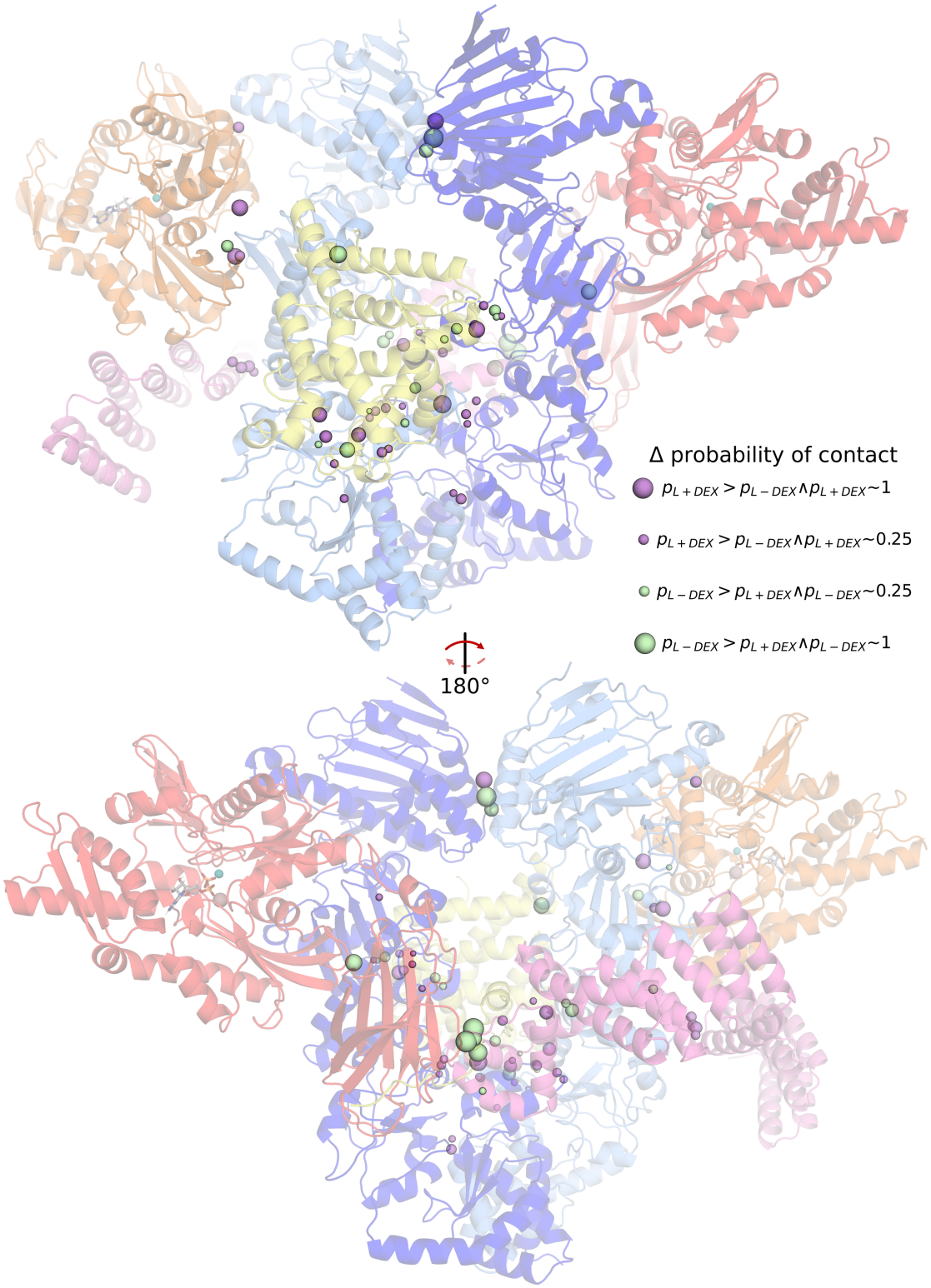

Figure A.4: *Projection of differences in probability of contact onto the Loading complex structure (the front view on the top; the rear view on the bottom). Each sphere indicates the presence of a statistically significant modified interaction between each inter-protein residue pair contained in the unperturbed and perturbed contact matrices. Spheres mark the midpoint between the two C_α atoms involved in each interaction.*

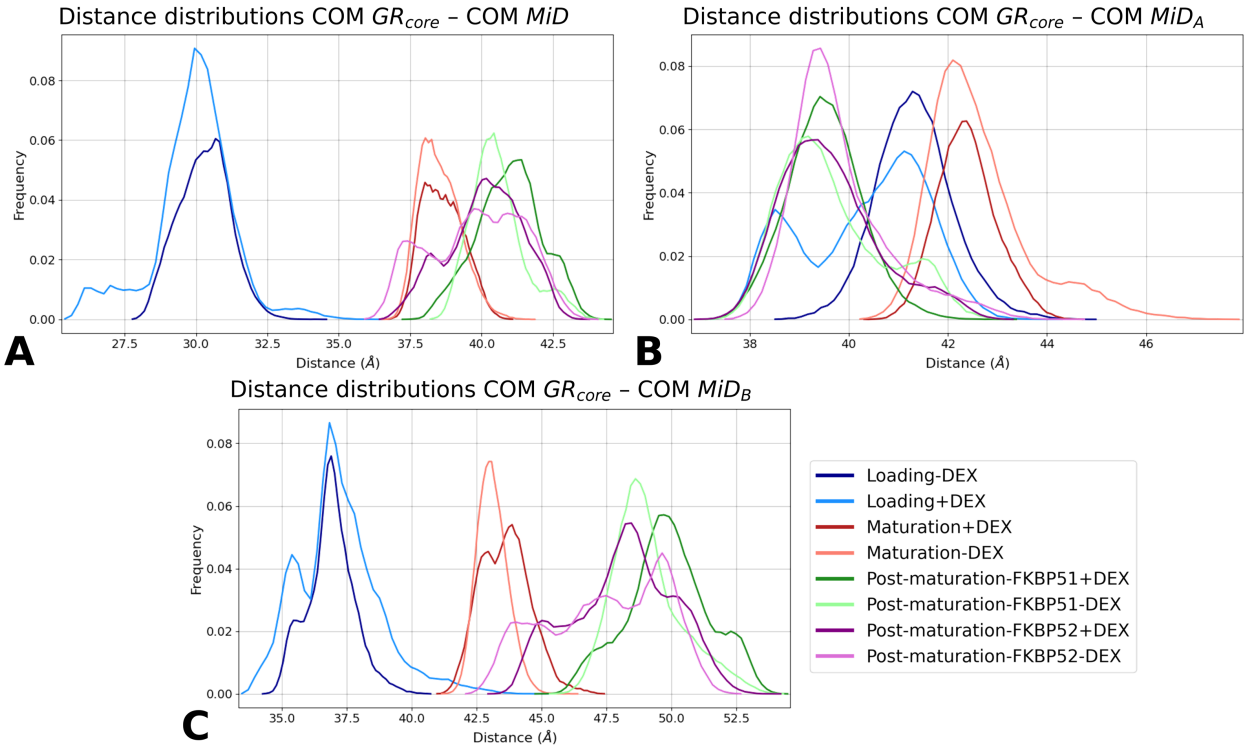

Figure A.5: *Distance distributions between the center of mass of the GR core (residues 561-777) and:* ***(A)*** *the center of mass of the entire MiD (residues 282-546; both protomers);* ***(B)*** *the center of mass of* MiD_A*; (C) the center of mass of* MiD_B*.*

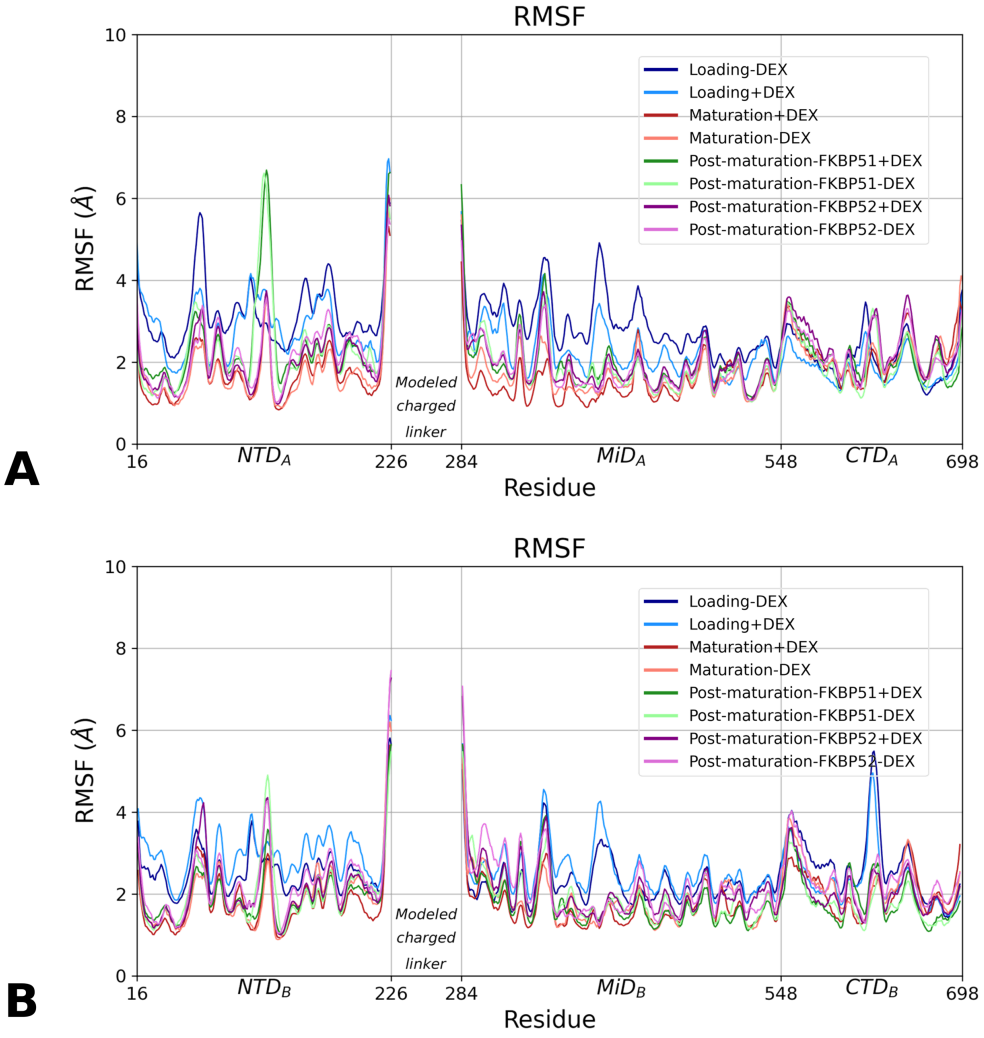

Figure A.6: *The RMSF of Hsp90 in equilibrium simulations of all complexes. Darker colors represent the WT simulations and lighter colors the perturbed ones.* ***(A)*** *Results for protomer A.* ***(B)*** *Results for protomer B.*

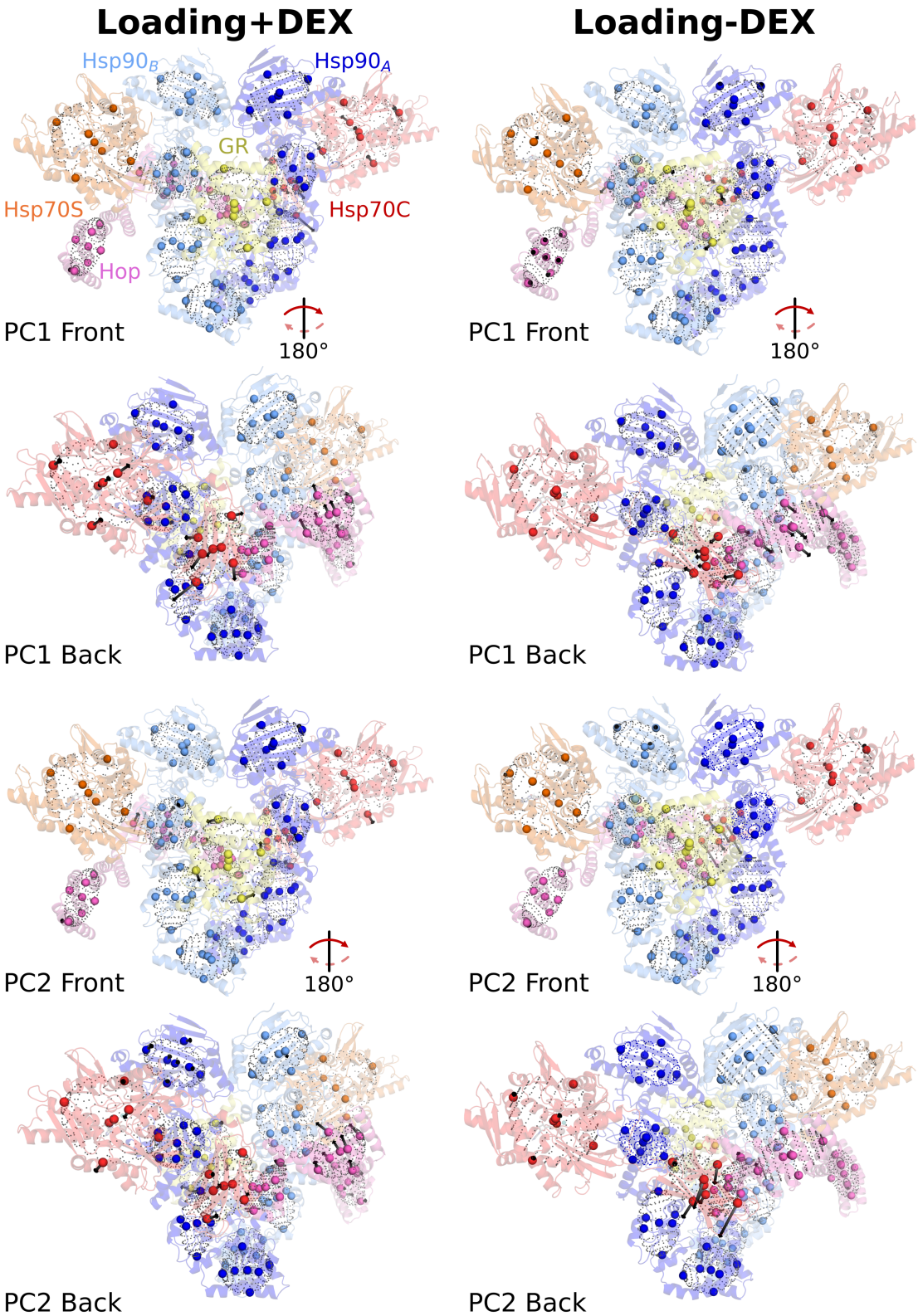

Figure A.7: *Projection of the first two PCA eigenvectors (vectors as arrows) onto structures of the Loading complex, from unperturbed (Loading-DEX; left) and perturbed simulations (Loading+DEX; right). Spheres represent the TOG-derived reduced points, while dots outline the shape of the identified ellipsoids that include each cluster of spheres. For each PC, the front and rear views are provided.*

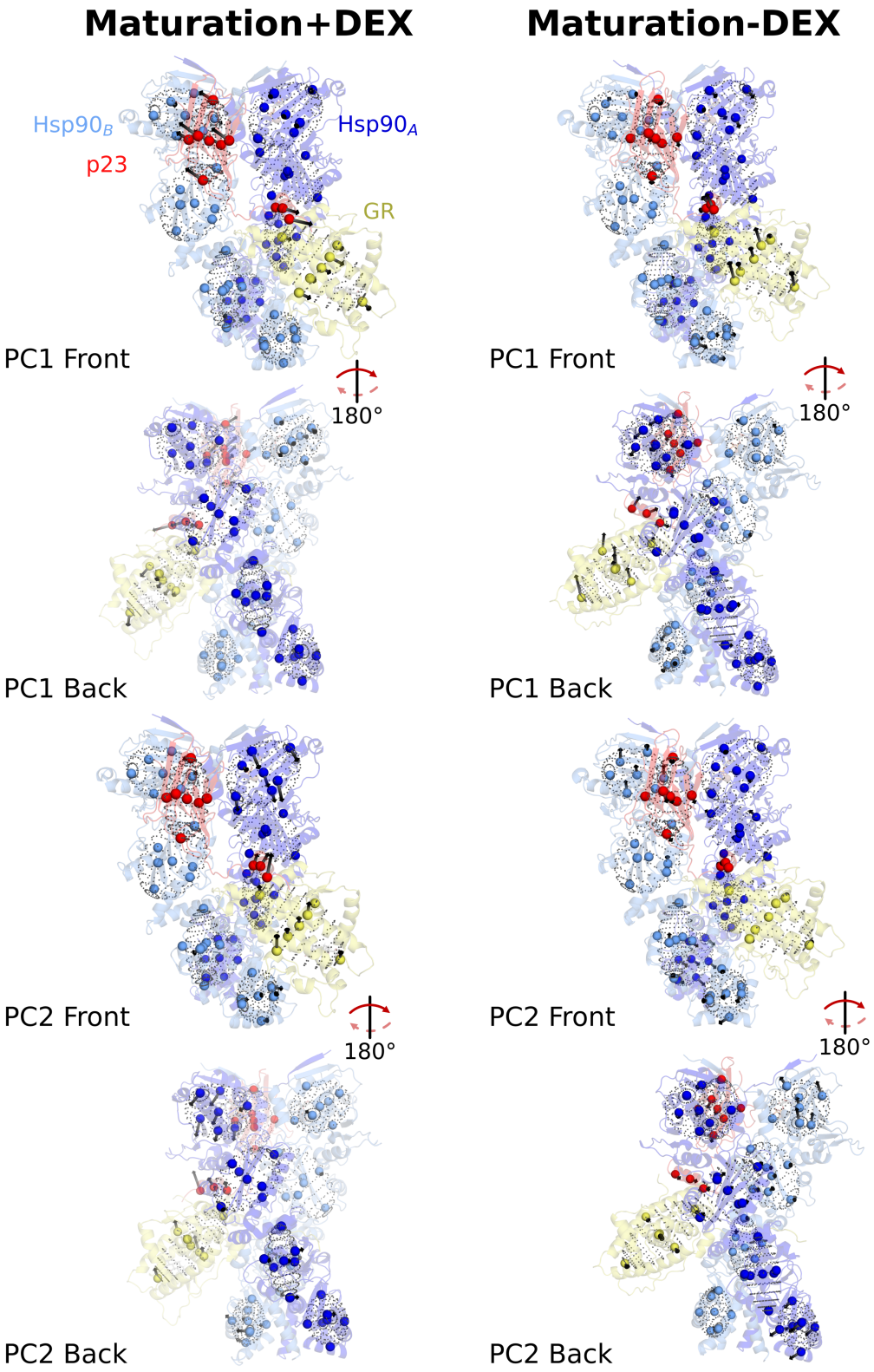

Figure A.8: *Projection of the first two PCA eigenvectors (vectors as arrows) onto structures of the Maturation complex, from unperturbed (Maturation+DEX; left) and perturbed simulations (Maturation-DEX; right). Spheres represent the TOG-derived reduced points, while dots outline the shape of the identified ellipsoids that include each cluster of spheres. For each PC, the front and rear views are provided.*

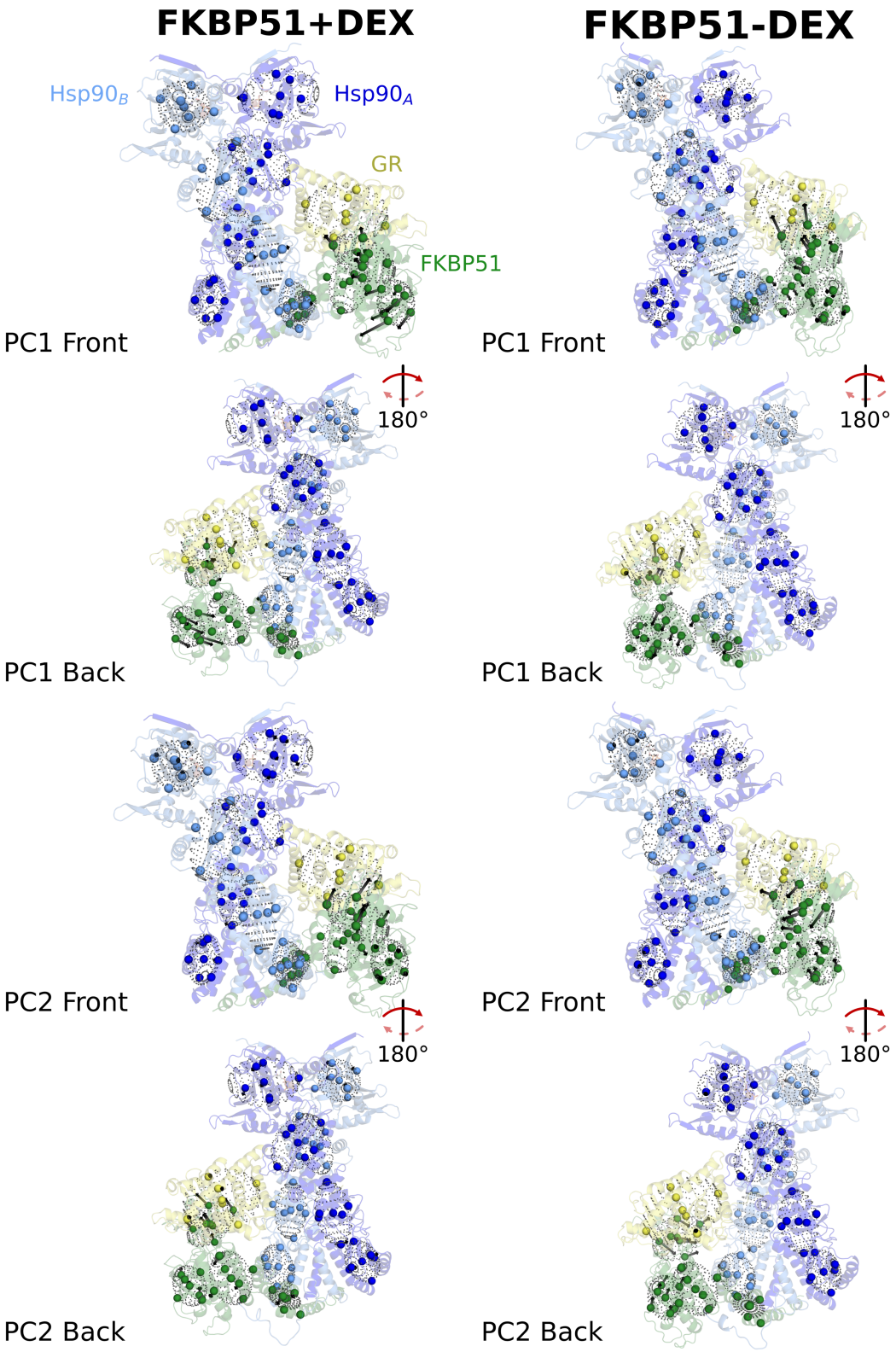

Figure A.9: *Projection of the first two PCA eigenvectors (vectors as arrows) onto structures of the FKBP51 complex, from unperturbed (FKBP51+DEX; left) and perturbed simulations (FKBP51-DEX; right). Spheres represent the TOG-derived reduced points, while dots outline the shape of the identified ellipsoids that include each cluster of spheres. For each PC, the front and rear views are provided.*

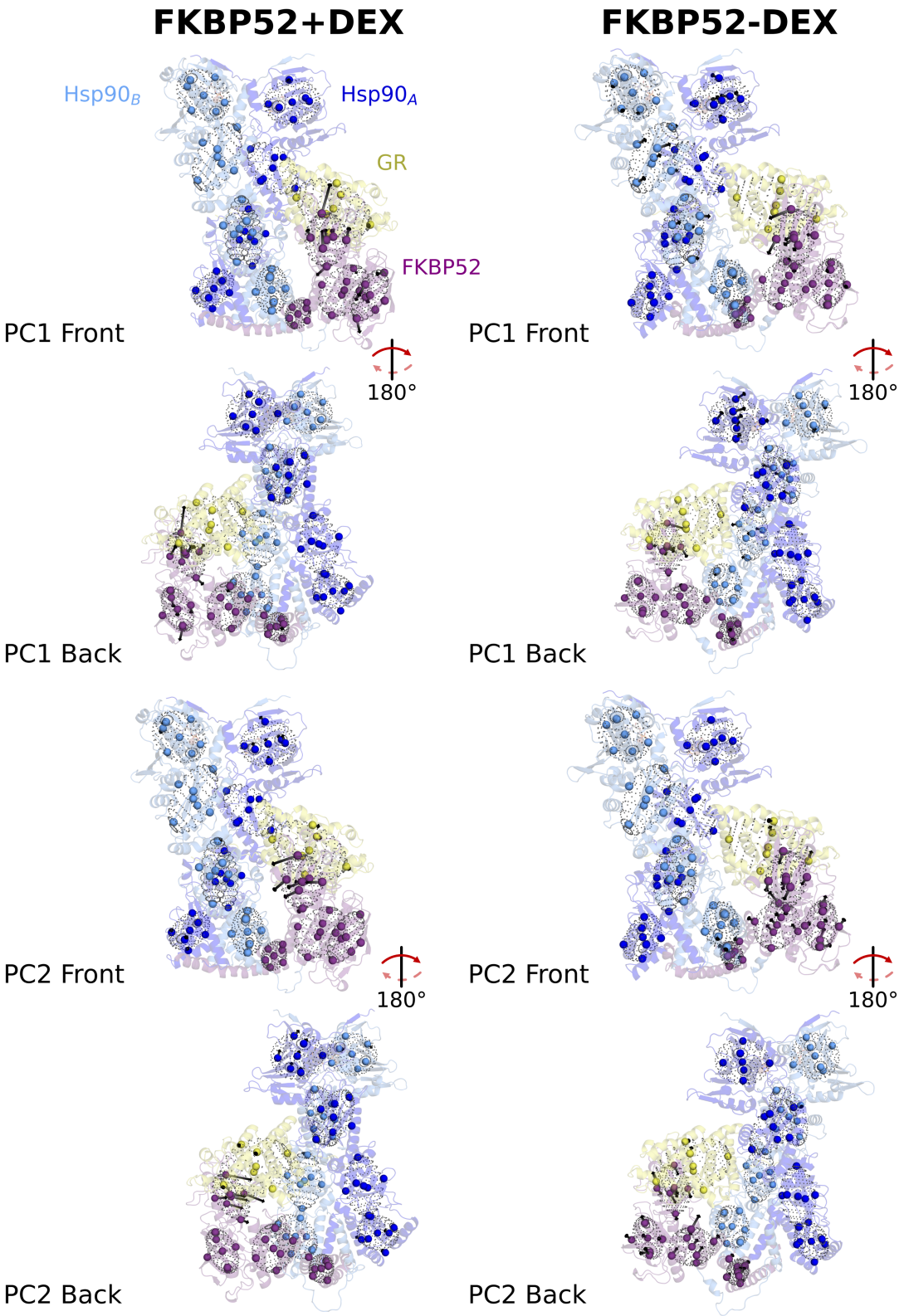

Figure A.10: *Projection of the first two PCA eigenvectors (vectors as arrows) onto structures of the FKBP52 complex, from unperturbed (FKBP52+DEX; left) and perturbed simulations (FKBP52-DEX; right). Spheres represent the TOG-derived reduced points, while dots outline the shape of the identified ellipsoids that include each cluster of spheres. For each PC, the front and rear views are provided.*

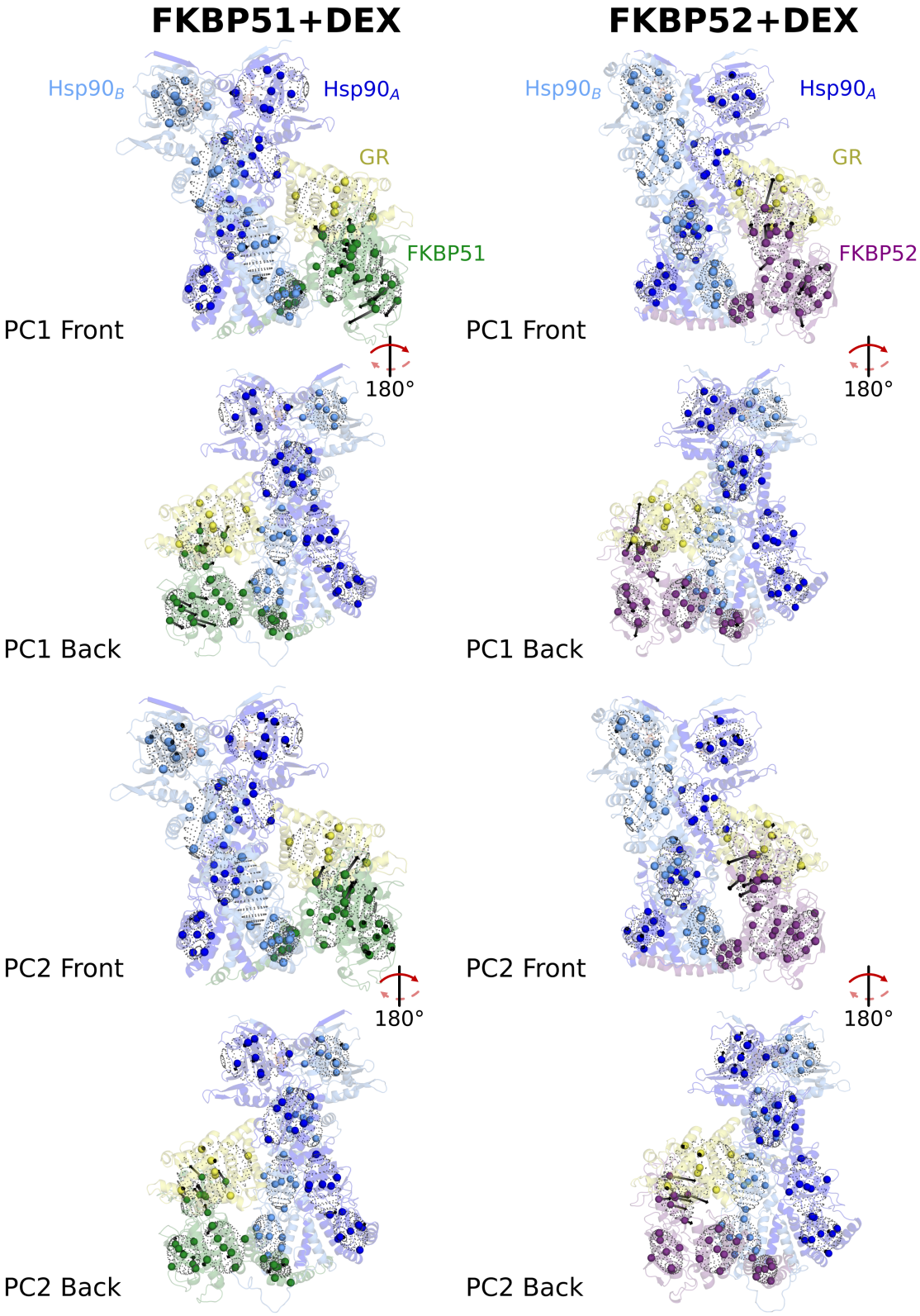

Figure A.11: *Projection of the first two PCA eigenvectors (vectors as arrows) onto the relative structures of the immunophilin complex, from FKBP51+DEX (left) and FKBP52+DEX (right) simulations. Spheres represent the TOG-derived reduced points, while dots outline the shape of the identified ellipsoids that include each cluster of spheres. For each PC, the front and rear views are provided.*

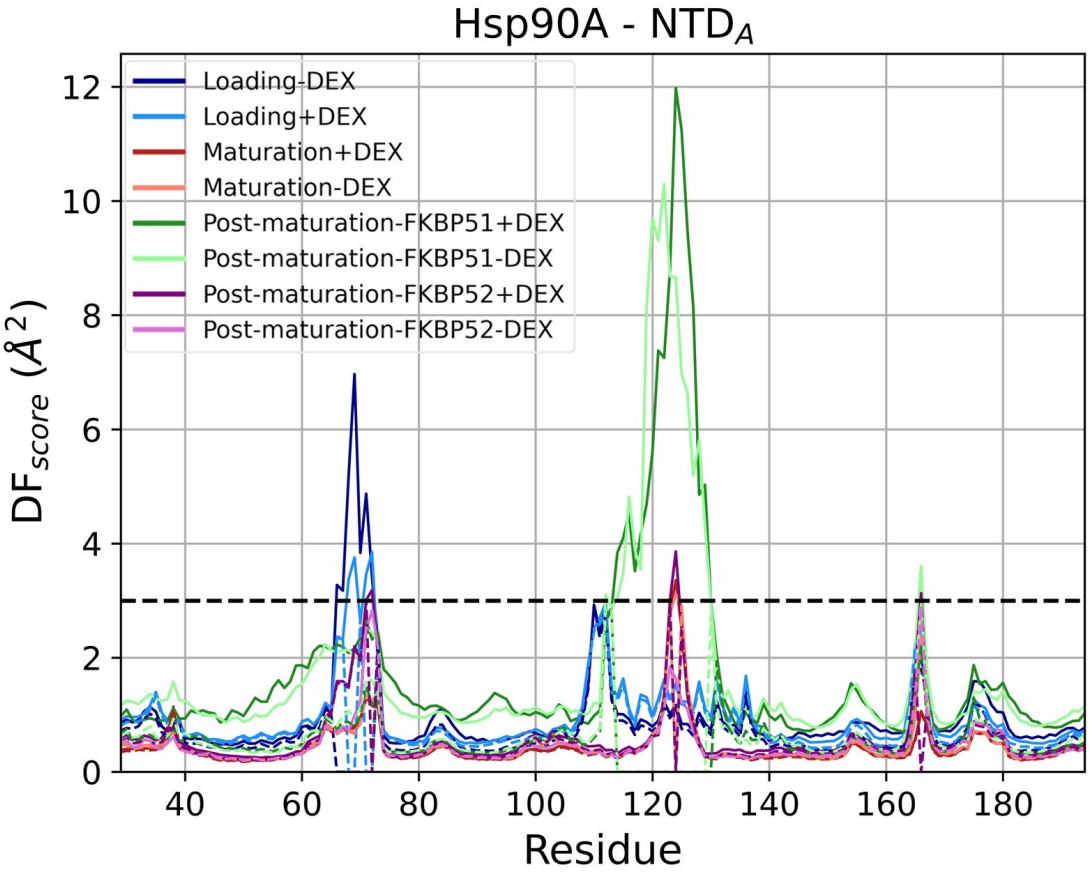

Figure A.12: *The DF_{score} plot of the NTDA for all the simulated systems. The original DF_{score} score is represented by coloured solid lines. The black dotted line represents the threshold to iteratively discard highly flexible amino acids. Coloured dotted lines represent the DF score after the iterative selection termination.*

Appendix B

Supplementary files for Chapter 5

In this Section, supplemental files for **Chapter 5** are reported.

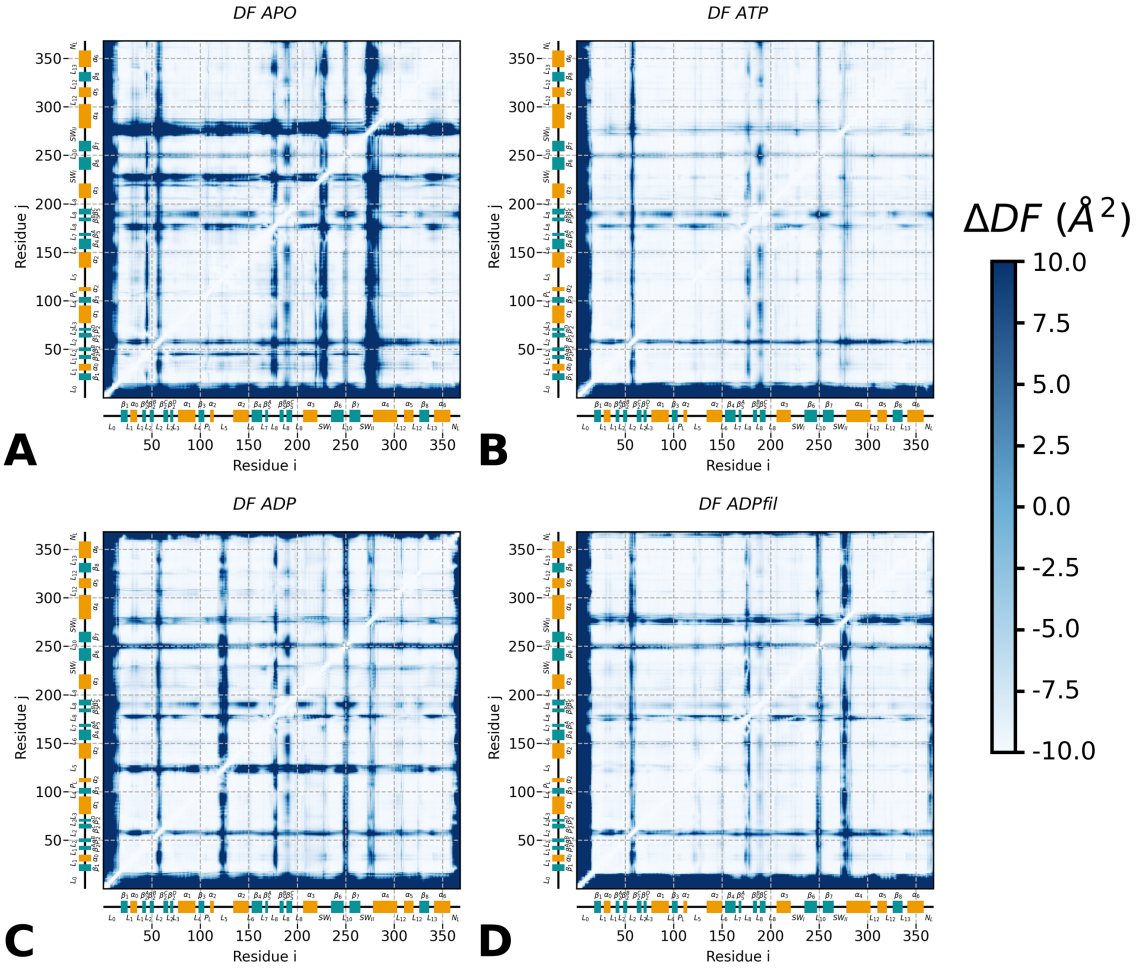

Figure B.1: *DF results* ***(A)*** *ATP,* ***(B)*** *ADP,* ***(C)*** *apo and* ***(D)*** *ADPfil. A threshold of 10* Å^2 *is applied, so that DF values higher than 10* Å^2 *are automatically rescaled to 10* Å^2*. Intensity of the blue color is correlated to the DF values.*

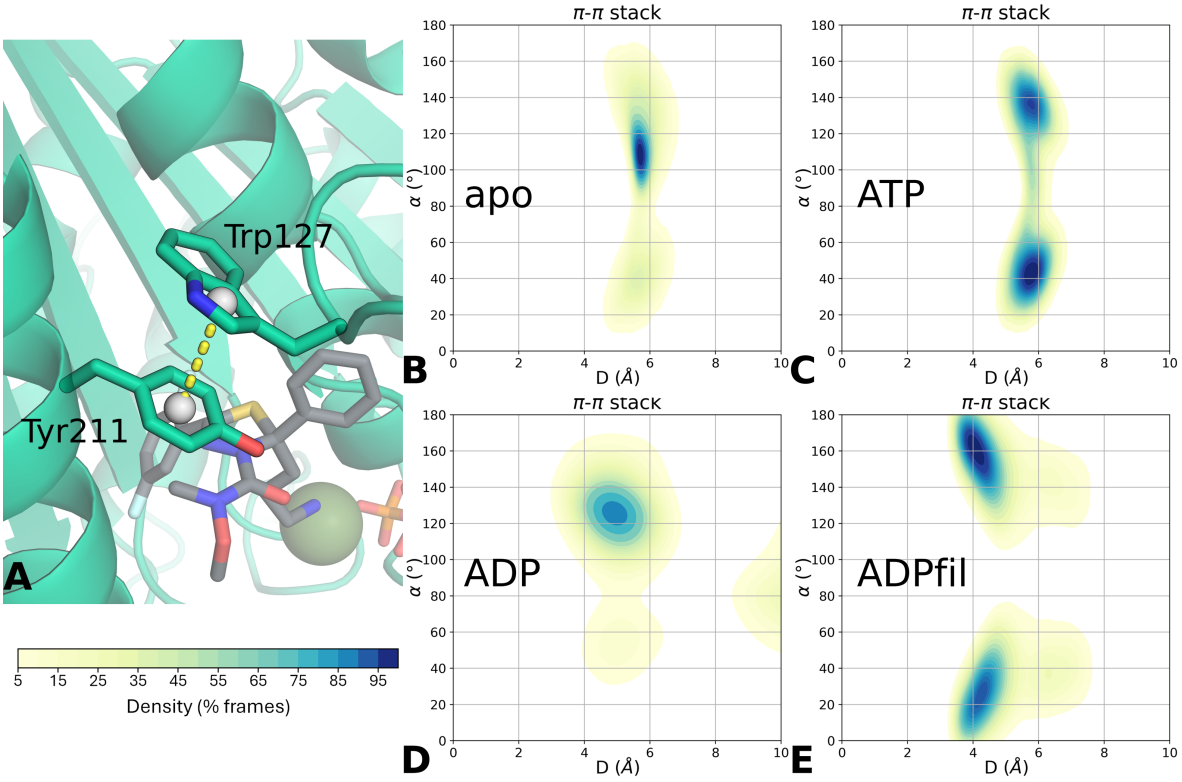

Figure B.2: *Tyr211-Trp127 interaction.* ***(A)*** *Static structure taken from ADPfil simulation highlighting the distance between center of mass of the two rings.* ***(B-D)*** *Kernel Density Estimate (KDE) plot of the distance between center of mass of the two rings and the angle between the two planes. Only a minor percentage of frames in ADP could be involved in $\pi - \pi$ interaction, while in ATP and ADPfil a stronger interaction is present.*

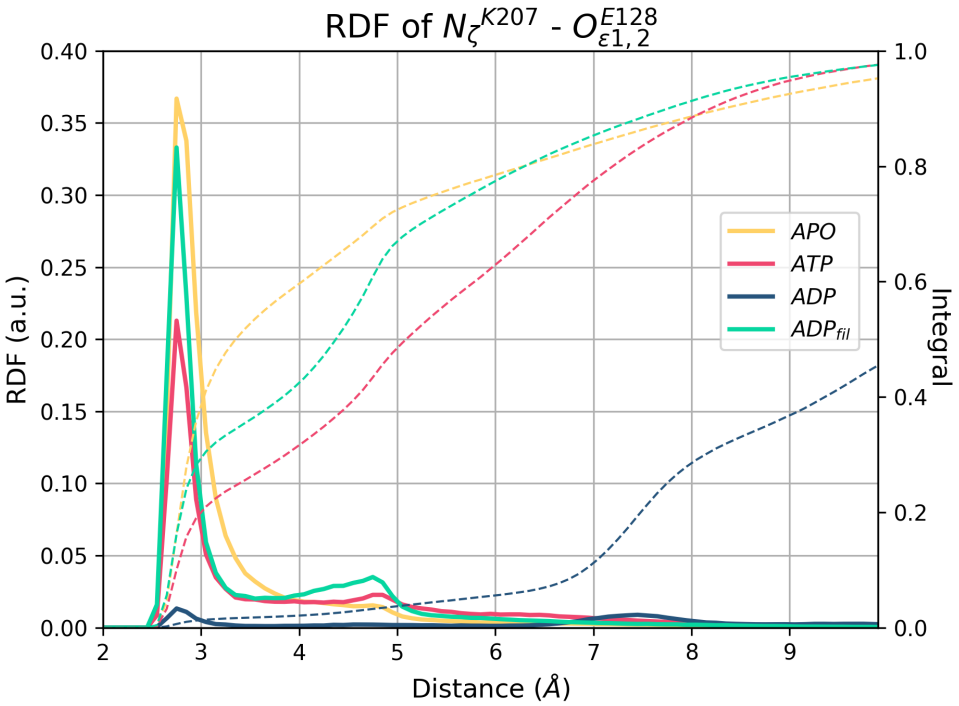

Figure B.3: *Radial distribution function of Glu128 oxygens ($O_{\epsilon 1}, O_{\epsilon 2}$) from Lys207 nitrogen ($O_\zeta$). The RDF values are plotted as solid lines, while dashed lines represent RDF integrals for each system. The peak visible between 2.5 and 3.5 Å highlights the interaction between the two sidechains, which is present only in apo, ATP and ADPfil.*

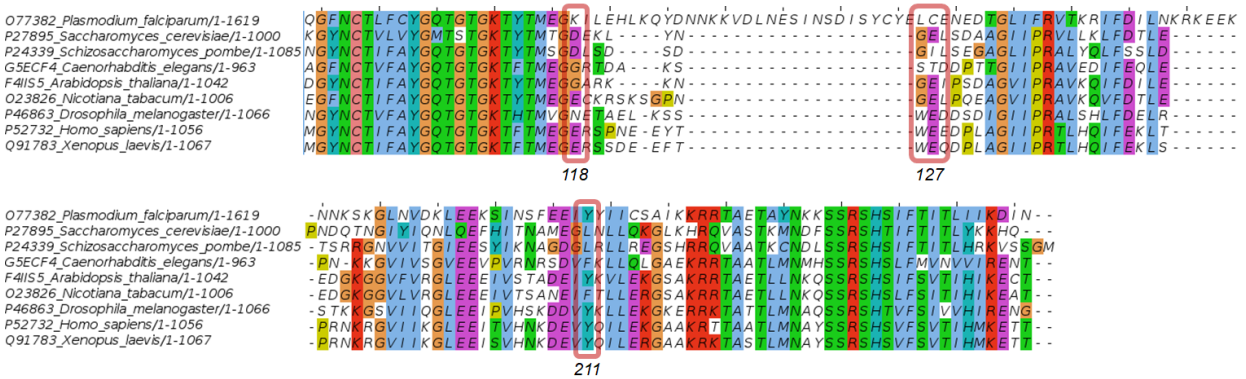

Figure B.4: *Multiple Sequence Alignment of selected kinesin-5 homologues, on the left we list the Uniprot accession codes of each protein, together with species names and protein length. Sequences were aligned with the Clustal Omega server [296] and results were visualized through Jalview [297].*

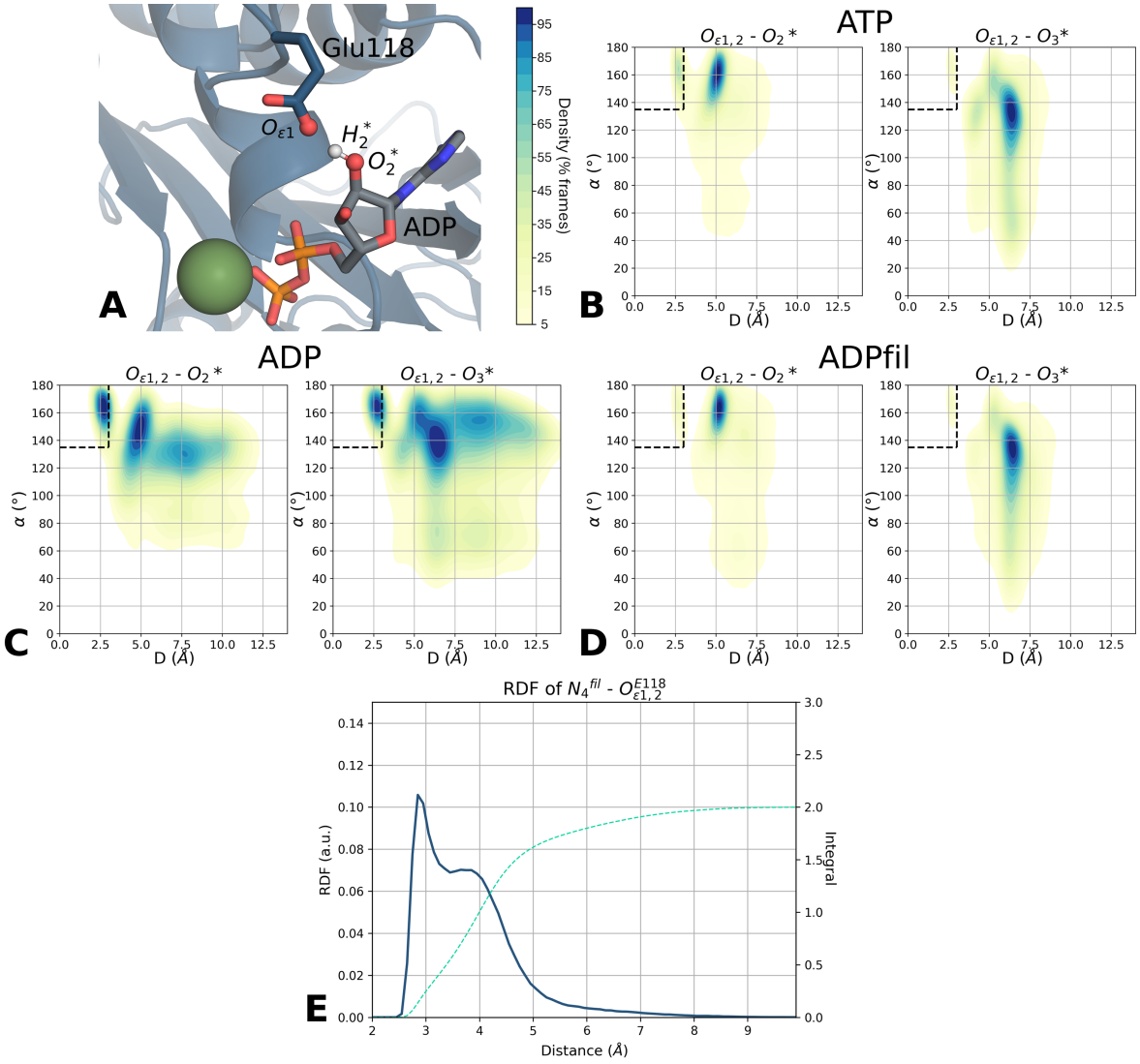

Figure B.5: *Glu118 interaction with the nucleotide in every state and with filanesib in ADPfil. Glu118 interacts more strongly with the nucleotide in ADP.* ***(A)*** *Static structure taken from ADP simulation highlighting the atoms involved in hydrogen bond interactions.* ***(B-D)*** *Kernel Density Estimate (KDE) plot of the distance between one oxygen of Glu118 and one of the two oxygens of ADP versus the O-H$\cdots$O angle in all the states of the cycle. Only a minor percentage of frames in ATP and ADPfil is involved in hydrogen bonds, while in ADP a stronger interaction is present.* ***(E)*** *The radial distribution function between* N_4 *in the* NH^{3+} *group of filanesib and Glu118 oxygens shows, with an intergral of almost 2 within the first 8 Å, that Glu118 is sequestered fairly stably by filanesib.*

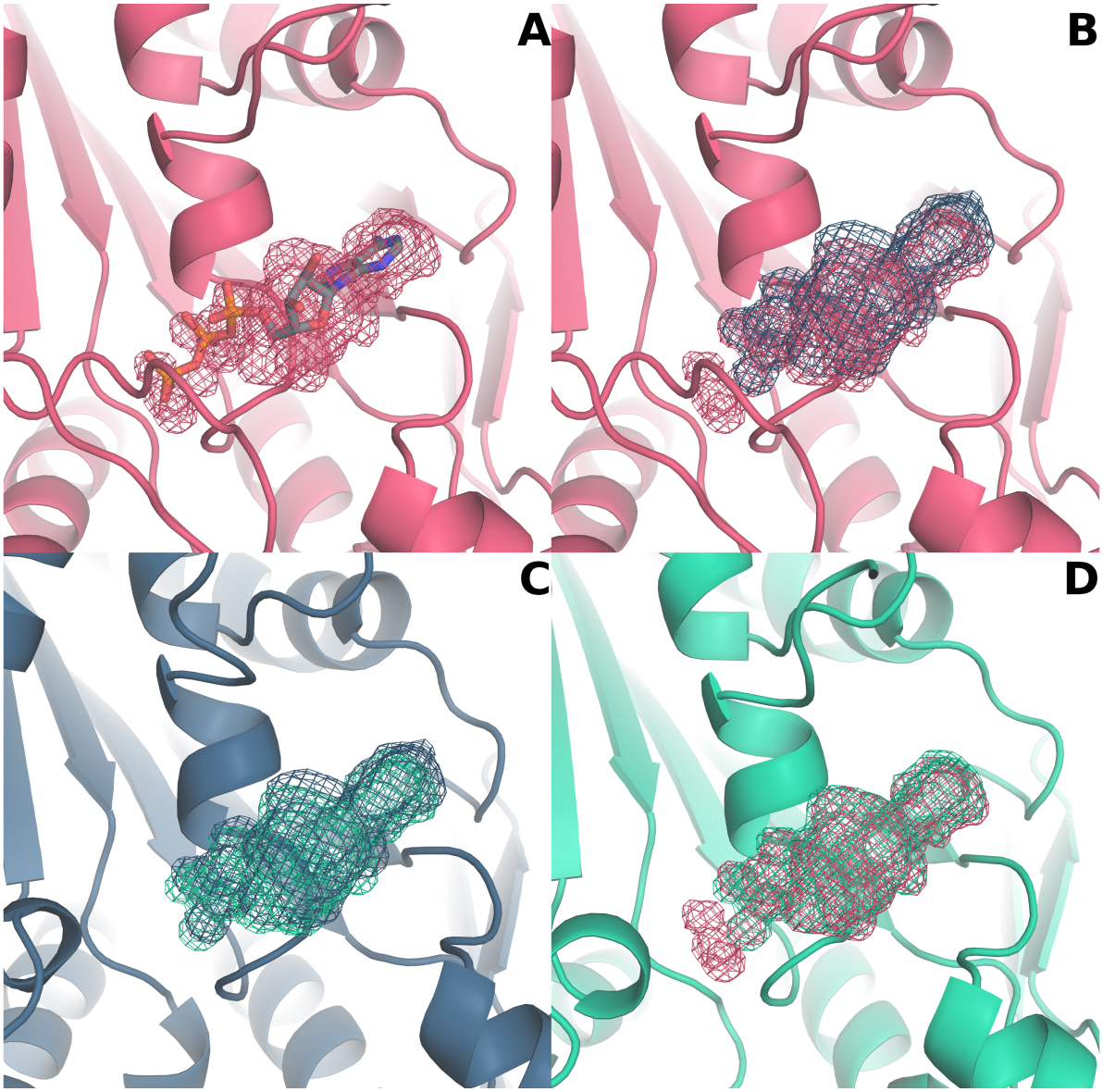

Figure B.6: *Results of the grid command on the nucleotide in every state.* ***(A)*** *3D density of ATP inside the pocket during simulations of ATP superimposed onto a frame of the MD.* ***(B)*** *Superposition between the density of ATP in ATP (red) and ADP in ADP (blue) onto the ATP structure; the ADP density seems to be partially shifting towards the upper right part of the pocket.* ***(C)*** *Superposition between the density of ADP in ADP (blue) and ADP in ADPfil (green) on ADP; density of ADP in ADPfil seems to partially reacquire similarity with ATP conformations.* ***(D)*** *Superposition between the density of ADP in ADPfil state (green) and ATP in ATP (red) on the ATP structure; density of ADP in ADPfil state seems to match almost entirely the one of ATP in the pocket, supporting the hypothesis that filanesib blocks the ADP state in an ATP-like conformation.*

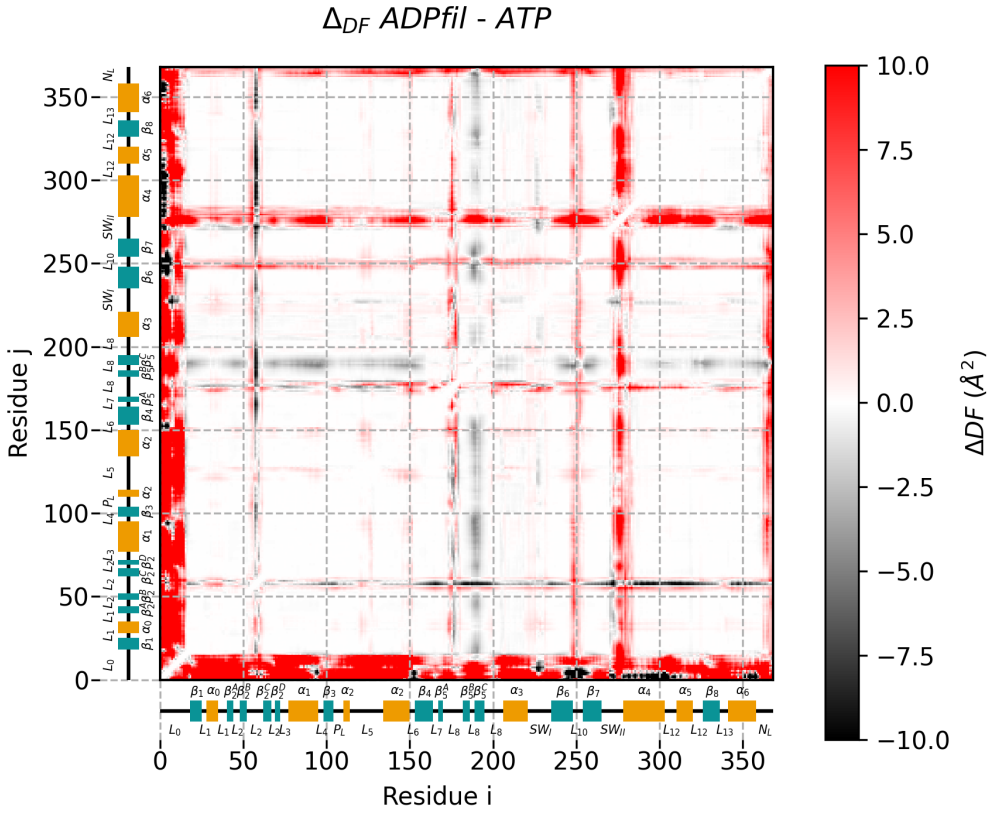

Figure B.7: *ΔDF matrix between ADPfil and ATP. With respect to the ADP-ATP matrix, the fluctuations on L5 are reduced, painting a picture of the ADPfil state as an "ATP-like ADP state", with locked L5. Additionally, it can be seen that the final portion of L8 presents a higher correlation with respect to ATP.*

Bibliography

1. Finkelstejn, A. V. & Pticyn, O. B. *Protein physics: a course of lectures* (Academic Press, Amsterdam Boston, 2002).
2. Kessel, A. & Ben-Tal, N. *Introduction to Proteins: Structure, Function, and Motion* 2nd ed. (CRC Press, Boca Raton, FL, 2018).
3. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181,** 223–230 (1973).
4. Zwanzig, R., Szabo, A. & Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 20–22 (1992).
5. Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **338,** 1042–1046 (2012).
6. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 6679–6685 (2005).
7. Dinner, A. R., Šali, A., Smith, L. J., Dobson, C. M. & Karplus, M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences* **25,** 331–339 (2000).
8. Brockwell, D. J. & Radford, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Current Opinion in Structural Biology* **17,** 30–37 (2007).
9. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5,** 789–796 (2009).
10. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353,** aac4354 (2016).
11. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *International Journal of Analytical Chemistry* **2016,** 1–6 (2016).
12. Chiti, F. & Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **75,** 333–366 (2006).
13. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting Proteostasis for Disease Intervention. *Science* **319,** 916–919 (2008).
14. Nussinov, R., Tsai, C.-J. & Jang, H. Protein ensembles link genotype to phenotype. *PLoS Comput Biol* **15,** e1006648 (2019).
15. Marchetti, F., Capelli, R., Rizzato, F., Laio, A. & Colombo, G. The Subtle Trade-Off between Evolutionary and Energetic Constraints in Protein–Protein Interactions. *J. Phys. Chem. Lett.* **10,** 1489–1497 (2019).
16. Morra, G. & Colombo, G. Relationship between energy distribution and fold stability: Insights from molecular dynamics simulations of native and mutant proteins. *Proteins* **72,** 660–672 (2008).
17. Tzeng, S.-R. & Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Current Opinion in Structural Biology* **21,** 62–67 (2011).

18. Monod, J., Wyman, J. & Changeux, J.-P. On the Nature of Allosteric Transitions: A Plausible Model. *Journal of Molecular Biology* **12,** 88–118 (1965).

19. Hilser, V. J. An Ensemble View of Allostery. *Science* **327,** 653–654 (2010).

20. Koshland Daniel E., J., Némethy, G. & Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **5,** 365–385 (1966).

21. Tsai, C.-J. & Nussinov, R. A Unified View of "How Allostery Works". *PLoS Comput Biol* **10** (ed Livesay, D. R.) e1003394 (2014).

22. Wodak, S. J. *et al.* Allostery in Its Many Disguises: From Theory to Applications. *Structure* **27,** 566–578 (2019).

23. Cooper, A. & Dryden, D. T. F. Allostery without conformational change: A plausible model. *Eur Biophys J* **11,** 103–109 (1984).

24. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of *all* dynamic proteins? *Proteins* **57,** 433–443 (2004).

25. Tsai, C.-J., Del Sol, A. & Nussinov, R. Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. *Journal of Molecular Biology* **378,** 1–11 (2008).

26. Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. Dynamically driven protein allostery. *Nat Struct Mol Biol* **13,** 831–838 (2006).

27. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508,** 331–339 (2014).

28. Nussinov, R. Introduction to Protein Ensembles and Allostery. *Chem. Rev.* **116,** 6263–6266 (2016).

29. Fenwick, R. B., Esteban-Martín, S. & Salvatella, X. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur Biophys J* **40,** 1339–1355 (2011).

30. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448,** 325–329 (2007).

31. Papaleo, E. *et al.* The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **116,** 6391–6423 (2016).

32. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 18249–18254 (2009).

33. Shaw, D. E. *et al.* Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **330,** 341–346 (2010).

34. Colombo, G. Computing allostery: from the understanding of biomolecular regulation and the discovery of cryptic sites to molecular design. *Current Opinion in Structural Biology* **83,** 102702 (2023).

35. Guo, J. & Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **116,** 6503–6515 (2016).

36. Hacisuleyman, A. & Erman, B. Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin. *PLoS Comput Biol* **13** (ed Briggs, J. M.) e1005319 (2017).

37. Ferreiro, D. U., Komives, E. A. & Wolynes, P. G. Frustration in biomolecules. *Quart. Rev. Biophys.* **47,** 285–363 (2014).

38. Li, J. *et al.* Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. *eLife* **6,** e30688 (2017).

39. White, J. T., Li, J., Grasso, E., Wrabl, J. O. & Hilser, V. J. Ensemble allosteric model: energetic frustration within the intrinsically disordered glucocorticoid receptor. *Phil. Trans. R. Soc. B* **373,** 20170175 (2018).
40. Oliveira, A. S. F., Ciccotti, G., Haider, S. & Mulholland, A. J. Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur. Phys. J. B* **94,** 144 (2021).
41. Post, M., Lickert, B., Diez, G., Wolf, S. & Stock, G. Cooperative Protein Allosteric Transition Mediated by a Fluctuating Transmission Network. *Journal of Molecular Biology* **434,** 167679 (2022).
42. Ali, A. A. A. I., Gulzar, A., Wolf, S. & Stock, G. Nonequilibrium Modeling of the Elementary Step in PDZ3 Allosteric Communication. *J. Phys. Chem. Lett.* **13,** 9862–9868 (2022).
43. Berezovsky, I. N. & Nussinov, R. Multiscale Allostery: Basic Mechanisms and Versatility in Diagnostics and Drug Design. *Journal of Molecular Biology* **434,** 167751 (2022).
44. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267,** 585–590 (1977).
45. Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117,** 5179–5197 (1995).
46. MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
47. Seifert, U. Stochastic thermodynamics: Principles and perspectives. *Eur. Phys. J. B* **64,** 423–431. arXiv: `0710.1187[cond-mat.stat-mech]` (2008).
48. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314,** 141–151 (1999).
49. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics* **120,** 11919–11929 (2004).
50. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).
51. Noé, F., Tkatchenko, A., Müller, K.-R. & Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **71,** 361–390 (2020).
52. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50,** D439–D444 (D1 2022).
53. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat Rev Mol Cell Biol* **18,** 345–360 (2017).
54. Buchner, J. Hsp90 & Co. – a holding for folding. *Trends in Biochemical Sciences* **24,** 136–141 (1999).
55. Biebl, M. M. & Buchner, J. Structure, Function, and Regulation of the Hsp90 Machinery. *Cold Spring Harb Perspect Biol* **11,** a034017 (2019).
56. Felts, S. J. *et al.* The hsp90-related Protein TRAP1 Is a Mitochondrial Protein with Distinct Functional Properties. *Journal of Biological Chemistry* **275,** 3305–3312 (2000).
57. Röhl, A., Rohrberg, J. & Buchner, J. The chaperone Hsp90: changing partners for demanding clients. *Trends in Biochemical Sciences* **38,** 253–262 (2013).
58. Prodromou, C. & Bjorklund, D. M. Advances towards Understanding the Mechanism of Action of the Hsp90 Complex. *Biomolecules* **12,** 600 (2022).

59. Hainzl, O., Lapina, M. C., Buchner, J. & Richter, K. The Charged Linker Region Is an Important Regulator of Hsp90 Function. *Journal of Biological Chemistry* **284,** 22559–22567 (2009).

60. Li, J., Soroka, J. & Buchner, J. The Hsp90 chaperone machinery: Conformational dynamics and regulation by co-chaperones. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823,** 624–635 (2012).

61. Hiller, S. Chaperone-Bound Clients: The Importance of Being Dynamic. *Trends in Biochemical Sciences* **44,** 517–527 (2019).

62. Mayer, M. P. & Le Breton, L. Hsp90: Breaking the Symmetry. *Molecular Cell* **58,** 8–20 (2015).

63. Goloubinoff, P., Sassi, A. S., Fauvet, B., Barducci, A. & De Los Rios, P. Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins. *Nat Chem Biol* **14,** 388–395 (2018).

64. Hessling, M., Richter, K. & Buchner, J. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat Struct Mol Biol* **16,** 287–293 (2009).

65. Cunningham, C. N., Krukenberg, K. A. & Agard, D. A. Intra- and Intermonomer Interactions Are Required to Synergistically Facilitate ATP Hydrolysis in Hsp90. *Journal of Biological Chemistry* **283,** 21170–21178 (2008).

66. Karagöz, G. E. *et al.* Hsp90-Tau Complex Reveals Molecular Basis for Specificity in Chaperone Action. *Cell* **156,** 963–974 (2014).

67. Ferraro, M. *et al.* Allosteric Modulators of HSP90 and HSP70: Dynamics Meets Function through Structure-Based Drug Design. *J. Med. Chem.* **62,** 60–87 (2019).

68. Timmermans, S., Souffriau, J. & Libert, C. A General Introduction to Glucocorticoid Biology. *Front. Immunol.* **10,** 1545 (2019).

69. Kadmiel, M. & Cidlowski, J. A. Glucocorticoid receptor signaling in health and disease. *Trends in Pharmacological Sciences* **34,** 518–530 (2013).

70. Wang, R. Y.-R. *et al.* Structure of Hsp90–Hsp70–Hop–GR reveals the Hsp90 client-loading mechanism. *Nature* **601,** 460–464 (2022).

71. Noddings, C. M., Wang, R. Y.-R., Johnson, J. L. & Agard, D. A. Structure of Hsp90–p23–GR reveals the Hsp90 client-remodelling mechanism. *Nature* **601,** 465–469 (2022).

72. Noddings, C. M., Johnson, J. L. & Agard, D. A. Cryo-EM reveals how Hsp90 and FKBP immunophilins co-regulate the glucocorticoid receptor. *Nat Struct Mol Biol* **30,** 1867–1877 (2023).

73. Galigniana, M. D., Radanyi, C., Renoir, J.-M., Housley, P. R. & Pratt, W. B. Evidence That the Peptidylprolyl Isomerase Domain of the hsp90-binding Immunophilin FKBP52 Is Involved in Both Dynein Interaction and Glucocorticoid Receptor Movement to the Nucleus. *Journal of Biological Chemistry* **276,** 14884–14889 (2001).

74. Tatro, E. T., Everall, I. P., Kaul, M. & Achim, C. L. Modulation of glucocorticoid receptor nuclear translocation in neurons by immunophilins FKBP51 and FKBP52: Implications for major depressive disorder. *Brain Research* **1286,** 1–12 (2009).

75. Backe, S. J., Sager, R. A., Woodford, M. R., Makedon, A. M. & Mollapour, M. Post-translational modifications of Hsp90 and translating the chaperone code. *Journal of Biological Chemistry* **295,** 11099–11117 (2020).

76. Verba, K. A. *et al.* Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352,** 1542–1547 (2016).

77. Verba, K. A. & Agard, D. A. How Hsp90 and Cdc37 Lubricate Kinase Molecular Switches. *Trends in Biochemical Sciences* **42,** 799–811 (2017).

78. Mann, B. J. & Wadsworth, P. Kinesin-5 Regulation and Function in Mitosis. *Trends in Cell Biology* **29,** 66–79 (2019).

79. Bodrug, T. *et al.* The kinesin-5 tail domain directly modulates the mechanochemical cycle of the motor domain for anti-parallel microtubule sliding. *eLife* **9,** e51131 (2020).

80. Chandrasekaran, G., Tátrai, P. & Gergely, F. Hitting the brakes: targeting microtubule motors in cancer. *Br J Cancer* **113,** 693–698 (2015).

81. Cross, R. A. Review: Mechanochemistry of the kinesin-1 ATPase. *Biopolymers* **105,** 476–482 (2016).

82. Chen, Y. & Hancock, W. O. Kinesin-5 is a microtubule polymerase. *Nat Commun* **6,** 8160 (2015).

83. Ding, S. *et al.* Overexpression of Eg5 predicts unfavorable prognosis in non-muscle invasive bladder urothelial carcinoma. *Int J of Urology* **18,** 432–438 (2011).

84. Daan Frenkel, B. S. *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, Amsterdam Boston, 2002).

85. M. P. Allen, D. J. T. *Computer Simulation of Liquids* (Clarendon Press - Oxford, Amsterdam Boston, 1987).

86. Huang, K. *Statistical Mechanics* 2nd (John Wiley & Sons, New York, 1987).

87. Van Gunsteren, W. F. & Berendsen, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. *Molecular Simulation* **1,** 173–185 (1988).

88. Cuendet, M. A. & Van Gunsteren, W. F. On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of Chemical Physics* **127,** 184102 (2007).

89. Lebowitz, J. L., Percus, J. K. & Verlet, L. Ensemble Dependence of Fluctuations with Application to Machine Computations. *Phys. Rev.* **153,** 250–254 (1967).

90. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52,** 255–268 (1984).

91. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81,** 511–519 (1984).

92. Hoover, W. G. Constant-pressure equations of motion. *Phys. Rev. A* **34,** 2499–2500 (1986).

93. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81,** 3684–3690 (1984).

94. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31,** 1695–1697 (1985).

95. Grest, G. S. & Kremer, K. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A* **33,** 3628–3631 (1986).

96. Bussi, G. & Parrinello, M. Accurate sampling using Langevin dynamics. *Phys. Rev. E* **75,** 056707 (2007).

97. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **17,** 1302–1322 (1978).

98. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126,** 014101 (2007).

99. Tsai, D. H. The virial theorem and stress calculation in molecular dynamics. *The Journal of Chemical Physics* **70,** 1375–1382 (1979).

100. Bussi, G., Zykova-Timan, T. & Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics* **130,** 074101 (2009).

101. Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **52,** 24–34 (1983).

102. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23,** 327–341 (1977).

103. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11,** 3696–3713 (2015).

104. Tian, C. *et al.* ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **16,** 528–552 (2020).

105. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369,** 253–287 (1921).

106. Deserno, M. & Holm, C. How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. *The Journal of Chemical Physics* **109,** 7678–7693 (1998).

107. Petersen, H. G. Accuracy and efficiency of the particle mesh Ewald method. *The Journal of Chemical Physics* **103,** 3668–3679 (1995).

108. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics* **98,** 10089–10092 (1993).

109. De Leeuw, S. W., Perram, J. W. & Smith, E. R. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. A.* **373,** 27–56 (1980).

110. De Leeuw, S. W., Perram, J. W. & Smith, E. R. Simulation of electrostatic systems in periodic boundary conditions. II. Equivalence of boundary conditions. *Proc. R. Soc. A.* **373,** 57–66 (1980).

111. De Leeuw, S. W., Perram, J. W. & Smith, E. R. Simulation of electrostatic systems in periodic boundary conditions. III. Further theory and applications. *Proc. R. Soc. A.* **388,** 177–193 (1983).

112. Ciccotti, G. & Ferrario, M. Non-equilibrium by molecular dynamics: a dynamical approach. *Molecular Simulation* **42,** 1385–1400 (2016).

113. Oliveira, A. S. F. *et al.* Identification of the Initial Steps in Signal Transduction in the $\alpha4\beta2$ Nicotinic Receptor: Insights from Equilibrium and Nonequilibrium Simulations. *Structure* **27,** 1171–1183.e3 (2019).

114. Oliveira, A. S. F. *et al.* A General Mechanism for Signal Propagation in the Nicotinic Acetylcholine Receptor Family. *J. Am. Chem. Soc.* **141,** 19953–19958 (2019).

115. Sofia F. Oliveira, A. *et al.* The fatty acid site is coupled to functional motifs in the SARS-CoV-2 spike protein and modulates spike allosteric behaviour. *Comput Struct Biotechnol J* **20,** 139–147 (2022).

116. Castelli, M. *et al.* Decrypting Allostery in Membrane-Bound K-Ras4B Using Complementary *In Silico* Approaches Based on Unbiased Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **146,** 901–919 (2024).

117. Castelli, M. *et al.* Molecular mechanisms of chaperone-directed protein folding: Insights from atomistic simulations. *Protein Science* **33,** e4880 (2024).

118. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234,** 779–815 (1993).

119. Pearlman, D. A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91,** 1–41 (1995).

120. Case, D. A. *et al.* AmberTools. *J. Chem. Inf. Model.* **63,** 6183–6191 (2023).

121. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).

122. Joung, I. S. & Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **112,** 9020–9041 (2008).

123. Allnér, O., Nilsson, L. & Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **8,** 1493–1502 (2012).

124. Meagher, K. L., Redman, L. T. & Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* **24,** 1016–1025 (2003).

125. Frisch, M. J. *et al. Gaussian 16 Revision C.01* Gaussian Inc. Wallingford CT. 2016.

126. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37,** 785–789 (1988).

127. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97,** 10269–10280 (1993).

128. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25,** 1157–1174 (2004).

129. Loncharich, R. J., Brooks, B. R. & Pastor, R. W. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N -acetylalanyl- N'-methylamide. *Biopolymers* **32,** 523–535 (1992).

130. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* **13,** 952–962 (1992).

131. Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **17,** 412–425 (1993).

132. Palma, J. & Pierdominici-Sottile, G. On the Uses of PCA to Characterise Molecular Dynamics Simulations of Biological Macromolecules: Basics and Tips for an Effective Use. *ChemPhysChem* **24,** e202200491 (2023).

133. Morra, G., Potestio, R., Micheletti, C. & Colombo, G. Corresponding Functional Dynamics across the Hsp90 Chaperone Family: Insights from a Multiscale Analysis of MD Simulations. *PLoS Comput Biol* **8** (ed Livesay, D. R.) e1002433 (2012).

134. Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design. *WIREs Comput Mol Sci* **11,** e1502 (2021).

135. Casadevall, G., Casadevall, J., Duran, C. & Osuna, S. The shortest path method (SPM) webserver for computational enzyme design. *Protein Engineering, Design and Selection* **37** (ed Eva, P.) gzae005 (2024).

136. Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik* **1,** 269–271 (1959).

137. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109,** 1528–1532 (2015).

138. Bunne, C. *et al.* How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell* **187,** 7045–7063 (2024).

139. Goldberg, A. P. *et al.* Emerging whole-cell modeling principles and methods. *Current Opinion in Biotechnology* **51,** 97–102 (2018).

140. Sali, A. From integrative structural biology to cell biology. *Journal of Biological Chemistry* **296,** 100743 (2021).

141. Rodina, A. *et al.* Systems-level analyses of protein-protein interaction network dysfunctions via epichaperomics identify cancer-specific mechanisms of stress adaptation. *Nat Commun* **14,** 3742 (2023).

142. Gershenson, A. & Gierasch, L. M. Protein folding in the cell: challenges and progress. *Current Opinion in Structural Biology* **21,** 32–41 (2011).

143. Mecha, M. F., Hutchinson, R. B., Lee, J. H. & Cavagnero, S. Protein folding in vitro and in the cell: From a solitary journey to a team effort. *Biophysical Chemistry* **287,** 106821 (2022).

144. Lim, C. M. & Vendruscolo, M. *Proteostasis Signatures in Human Diseases* 2024.

145. Miles, J., Scherz-Shouval, R. & Van Oosten-Hawle, P. Expanding the Organismal Proteostasis Network: Linking Systemic Stress Signaling with the Innate Immune Response. *Trends in Biochemical Sciences* **44,** 927–942 (2019).

146. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **82,** 323–355 (2013).

147. Ciechanover, A. & Kwon, Y. T. Protein Quality Control by Molecular Chaperones in Neurodegeneration. *Front. Neurosci.* **11** (2017).

148. Bhattacharya, K. & Picard, D. The Hsp70–Hsp90 go-between Hop/Stip1/Sti1 is a proteostatic switch and may be a drug target in cancer and neurodegeneration. *Cell. Mol. Life Sci.* **78,** 7257–7273 (2021).

149. Brehme, M. *et al.* A Chaperome Subnetwork Safeguards Proteostasis in Aging and Neurodegenerative Disease. *Cell Reports* **9,** 1135–1150 (2014).

150. Hartl, F. U. Molecular chaperones in cellular protein folding. **381** (1996).

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

152. Chiosis, G., Digwal, C. S., Trepel, J. B. & Neckers, L. Structural and functional complexity of HSP90 in cellular homeostasis and disease. *Nat Rev Mol Cell Biol* **24,** 797–815 (2023).

153. Silbermann, L.-M., Vermeer, B., Schmid, S. & Tych, K. The known unknowns of the Hsp90 chaperone. *eLife* **13,** e102666 (2024).

154. Kurop, M. K., Huyen, C. M., Kelly, J. H. & Blagg, B. S. The heat shock response and small molecule regulators. *European Journal of Medicinal Chemistry* **226,** 113846 (2021).

155. Truman, A. W., Bourboulia, D. & Mollapour, M. Decrypting the chaperone code. *Journal of Biological Chemistry* **296,** 100293 (2021).

156. Ali, M. M. U. *et al.* Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* **440,** 1013–1017 (2006).

157. Pearl, L. H. & Prodromou, C. Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. *Annu. Rev. Biochem.* **75,** 271–294 (2006).

158. Shiau, A. K., Harris, S. F., Southworth, D. R. & Agard, D. A. Structural Analysis of E. coli hsp90 Reveals Dramatic Nucleotide-Dependent Conformational Rearrangements. *Cell* **127,** 329–340 (2006).

159. Lavery, L. A. *et al.* Structural Asymmetry in the Closed State of Mitochondrial Hsp90 (TRAP1) Supports a Two-Step ATP Hydrolysis Mechanism. *Molecular Cell* **53,** 330–343 (2014).

160. Dollins, D. E., Warren, J. J., Immormino, R. M. & Gewirth, D. T. Structures of GRP94-Nucleotide Complexes Reveal Mechanistic Differences between the hsp90 Chaperones. *Molecular Cell* **28,** 41–56 (2007).

161. Sahasrabudhe, P., Rohrberg, J., Biebl, M. M., Rutz, D. A. & Buchner, J. The Plasticity of the Hsp90 Co-chaperone System. *Molecular Cell* **67,** 947–961.e5 (2017).

162. Sabbagh, J. J. *et al.* Targeting the FKBP51/GR/Hsp90 Complex to Identify Functionally Relevant Treatments for Depression and PTSD. *ACS Chem. Biol.* **13,** 2288–2299 (2018).

163. Li, J., Richter, K., Reinstein, J. & Buchner, J. Integration of the accelerator Aha1 in the Hsp90 co-chaperone cycle. *Nat Struct Mol Biol* **20,** 326–331 (2013).

164. Lorenz, O. R. *et al.* Modulation of the Hsp90 Chaperone Cycle by a Stringent Client Protein. *Molecular Cell* **53,** 941–953 (2014).

165. Dahiya, V. *et al.* The switch from client holding to folding in the Hsp70/Hsp90 chaperone machineries is regulated by a direct interplay between co-chaperones. *Molecular Cell* **82,** 1543–1556.e6 (2022).

166. Kirschke, E., Goswami, D., Southworth, D., Griffin, P. R. & Agard, D. A. Glucocorticoid Receptor Function Regulated by Coordinated Action of the Hsp90 and Hsp70 Chaperone Cycles. *Cell* **157,** 1685–1697 (2014).

167. Lopez, A. *et al.* Client binding shifts the populations of dynamic Hsp90 conformations through an allosteric network. *Sci. Adv.* **7,** eabl7295 (2021).

168. Balega, B. *et al.* Dynamical nonequilibrium molecular dynamics simulations reveal allosteric networks, signal transduction mechanisms, and sites associated with drug resistance in biomolecular systems. *Molecular Physics* **123,** e2428350 (2025).

169. Meli, M., Morra, G. & Colombo, G. Simple Model of Protein Energetics To Identify Ab Initio Folding Transitions from All-Atom MD Simulations of Proteins. *J. Chem. Theory Comput.* **16,** 5960–5971 (2020).

170. Serapian, S. A. *et al.* The Answer Lies in the Energy: How Simple Atomistic Molecular Dynamics Simulations May Hold the Key to Epitope Prediction on the Fully Glycosylated SARS-CoV-2 Spike Protein. *J. Phys. Chem. Lett.* **11,** 8084–8093 (2020).

171. Castelli, M. *et al.* How aberrant N-glycosylation can alter protein functionality and ligand binding: An atomistic view. *Structure* **31,** 987–1004.e8 (2023).

172. Johnston, R. C. *et al.* Epik: p K_{a} and Protonation State Prediction through Machine Learning. *J. Chem. Theory Comput.* **19,** 2380–2388 (2023).

173. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **49,** 534–553 (2006).

174. Roe, D. R. & Cheatham, T. E. Parallelization of CPPTRAJ enables large scale analysis of molecular dynamics trajectory data. *J Comput Chem* **39,** 2110–2117 (2018).

175. Serapian, S. A., Moroni, E., Ferraro, M. & Colombo, G. Atomistic Simulations of the Mechanisms of the Poorly Catalytic Mitochondrial Chaperone Trap1: Insights into the Effects of Structural Asymmetry on Reactivity. *ACS Catal.* **11,** 8605–8620 (2021).

176. Kashefolgheta, S. & Vila Verde, A. Developing force fields when experimental data is sparse: AMBER/GAFF-compatible parameters for inorganic and alkyl oxoanions. *Phys. Chem. Chem. Phys.* **19,** 20593–20607 (2017).

177. Baischew, A., Engel, S., Taubert, M. C., Geiger, T. M. & Hausch, F. Large-scale, in-cell photocrosslinking at single-residue resolution reveals the molecular basis for glucocorticoid receptor regulation by immunophilins. *Nat Struct Mol Biol* **30,** 1857–1866 (2023).

178. Hua, G., Paulen, L. & Chambon, P. GR SUMOylation and formation of an SUMO-SMRT/NCoR1-HDAC3 repressing complex is mandatory for GC-induced IR nGRE-mediated transrepression. *Proc. Natl. Acad. Sci. U.S.A.* **113** (2016).

179. Oakley, R. H. & Cidlowski, J. A. Cellular Processing of the Glucocorticoid Receptor Gene and Protein: New Mechanisms for Generating Tissue-specific Actions of Glucocorticoids. *Journal of Biological Chemistry* **286,** 3177–3184 (2011).

180. Backe, S. J., Heritz, J. A. & Mollapour, M. Hsp70 and Hsp90 post-translational modifications and translating the chaperone code. *Cell Stress and Chaperones* **30,** 100118 (2025).

181. Rehn, A. *et al.* A methylated lysine is a switch point for conformational communication in the chaperone Hsp90. *Nat Commun* **11,** 1219 (2020).

182. Nicolaides, N. C. *et al.* Functional characterization of the hGRαT556I causing Chrousos syndrome. *Eur J Clin Investigation* **46,** 42–49 (2016).

183. Kino, T. & Chrousos, G. P. Pathologic Human GR Mutant Has a Transdominant Negative Effect on the Wild-Type GR by Inhibiting Its Translocation into the Nucleus: Importance of the Ligand-Binding Domain for Intracellular GR Trafficking. *The Journal of Clinical Endocrinology & Metabolism* **86,** 5600–5612 (2001).

184. Charmandari, E. *et al.* A Novel Point Mutation in Helix 11 of the Ligand-Binding Domain of the Human Glucocorticoid Receptor Gene Causing Generalized Glucocorticoid Resistance. *The Journal of Clinical Endocrinology & Metabolism* **92,** 3986–3990 (2007).

185. Charmandari, E. *et al.* A Novel Point Mutation in the Ligand-Binding Domain (LBD) of the Human Glucocorticoid Receptor (hGR) Causing Generalized Glucocorticoid Resistance: The Importance of the C Terminus of hGR LBD in Conferring Transactivational Activity. *The Journal of Clinical Endocrinology & Metabolism* **90,** 3696–3705 (2005).

186. Gestwicki, J. E. Multi-protein complexes as drug targets. *Cell Chemical Biology* **29,** 713–715 (2022).

187. Johnson, O. T. & Gestwicki, J. E. Multivalent protein–protein interactions are pivotal regulators of eukaryotic Hsp70 complexes. *Cell Stress and Chaperones* **27,** 397–415 (2022).

188. Laine, E. & Carbone, A. Protein social behavior makes a stronger signal for partner identification than surface geometry. *Proteins* **85,** 137–154 (2017).

189. Nussinov, R., Tsai, C.-J. & Jang, H. Signaling in the crowded cell. *Current Opinion in Structural Biology* **71,** 43–50 (2021).

190. Schmid, S. & Hugel, T. Controlling protein function by fine-tuning conformational flexibility. *eLife* **9,** e57180 (2020).

191. Sohmen, B. *et al.* The Onset of Molecule-Spanning Dynamics in Heat Shock Protein Hsp90. *Advanced Science* **10,** 2304262 (2023).

192. Ferguson, A. *et al.* Functional Dynamics within the Human Ribosome Regulate the Rate of Active Protein Synthesis. *Molecular Cell* **60,** 475–486 (2015).

193. Tanaka, K. The proteasome: Overview of structure and functions. *Proc. Jpn. Acad., Ser. B* **85,** 12–36 (2009).

194. Pokorná, P., Aupič, J., Fica, S. M. & Magistrato, A. Decoding Spliceosome Dynamics through Computation and Experiment. *Chem. Rev.* **125,** 9807–9833 (2025).

195. Chen, Z. *et al.* De novo protein identification in mammalian sperm using in situ cryoelectron tomography and AlphaFold2 docking. *Cell* **186,** 5041–5053.e19 (2023).

196. Zhou, L., Roth, M., Papakonstantinou, E., Tamm, M. & Stolz, D. Expression of glucocorticoid receptor and HDACs in airway smooth muscle cells is associated with response to steroids in COPD. *Respir Res* **25,** 227 (2024).

197. Shimojo, M. *et al.* Differences in down-regulation of glucocorticoid receptor mRNA by cortisol, prednisolone and dexamethasone in HeLa cells. *Endocr. J.* **42,** 629–636 (1995).

198. Vicent, G. P., Pecci, A., Ghini, A., Piwien-Pilipuk, G. & Galigniana, M. D. Differences in Nuclear Retention Characteristics of Agonist-Activated Glucocorticoid Receptor May Determine Specific Responses. *Experimental Cell Research* **276,** 142–154 (2002).

199. Hadley, K. E., Louw, A. & Hapgood, J. P. Differential nuclear localisation and promoter occupancy play a role in glucocorticoid receptor ligand-specific transcriptional responses. *Steroids* **76,** 1176–1184 (2011).

200. Morra, G., Verkhivker, G. & Colombo, G. Modeling Signal Propagation Mechanisms and Ligand-Based Conformational Dynamics of the Hsp90 Molecular Chaperone Full-Length Dimer. *PLoS Comput Biol* **5** (ed Jacobson, M. P.) e1000323 (2009).

201. Radli, M. & Rüdiger, S. G. Dancing with the Diva: Hsp90–Client Interactions. *Journal of Molecular Biology* **430,** 3029–3040 (2018).

202. Johnson, J. L. Evolution and function of diverse Hsp90 homologs and cochaperone proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823,** 607–613 (2012).

203. Rodina, A. *et al.* The epichaperome is an integrated chaperome network that facilitates tumour survival. *Nature* **538,** 397–401 (2016).

204. Chu, F., Sharma, S., Ginsberg, S. D. & Chiosis, G. PTMs as molecular encoders: reprogramming chaperones into epichaperomes for network control in disease. *Trends in Biochemical Sciences* **50,** 892–905 (2025).

205. Roychowdhury, T. *et al.* Phosphorylation-driven epichaperome assembly is a regulator of cellular adaptability and proliferation. *Nat Commun* **15,** 8912 (2024).

206. Raguette, L. E. *et al.* phosaa14SB and phosaa19SB: Updated Amber Force Field Parameters for Phosphorylated Amino Acids. *J. Chem. Theory Comput.* **20,** 7199–7209 (2024).

207. Pillarsetty, N. *et al.* Paradigms for Precision Medicine in Epichaperome Cancer Therapy. *Cancer Cell* **36,** 559–573.e7 (2019).

208. Longhi, E. *et al.* Hijacking client instability to selectively disrupt HSP90-driven glucocorticoid receptor activation. *Pharmacological Research* **228,** 108205 (2026).

209. Torielli, L. *et al.* Design of multi-target peptide modulators for protein chaperone networks. *Structure* **33,** 1944–1957.e6 (2025).

210. González-Hernández, E. *et al.* Dihydropyrimidine-2-thiones as Eg5 inhibitors and L-type calcium channel blockers: potential antitumour dual agents. *Med. Chem. Commun.* **10,** 1589–1598 (2019).

211. Triarico, S. *et al.* Vincristine-Induced Peripheral Neuropathy (VIPN) in Pediatric Tumors: Mechanisms, Risk Factors, Strategies of Prevention and Treatment. *IJMS* **22,** 4112 (2021).

212. Da Costa, R. *et al.* Taxane-induced neurotoxicity: Pathophysiology and therapeutic perspectives. *British J Pharmacology* **177,** 3127–3146 (2020).

213. Pérez-Melero, C. KSP inhibitors as antimitotic agents. *Current Topics in Medicinal Chemistry* **14,** 2286–2311 (2014).

214. Zhou, Y. *et al.* KIF11 is upregulated in colorectal cancer and silencing of it impairs tumor growth and sensitizes colorectal cancer cells to oxaliplatin via p53/GSK3β signaling. *J. Cancer* **12,** 3741–3753 (2021).

215. Gu, X. *et al.* KIF11 manipulates SREBP2-dependent mevalonate cross talk to promote tumor progression in pancreatic ductal adenocarcinoma. *Cancer Medicine* **11,** 3282–3295 (2022).

216. Hu, Z.-D. *et al.* KIF11 Promotes Proliferation of Hepatocellular Carcinoma among Patients with Liver Cancers. *BioMed Research International* **2021** (ed Maida, M.) 2676745 (2021).

217. Liu, L. *et al.* Overexpression of Eg5 correlates with high grade astrocytic neoplasm. *J Neurooncol* **126,** 77–80 (2016).

218. Li, Z., Yu, B., Qi, F. & Li, F. KIF11 Serves as an Independent Prognostic Factor and Therapeutic Target for Patients With Lung Adenocarcinoma. *Front. Oncol.* **11,** 670218 (2021).

219. Garcia-Saez, I. & Skoufias, D. A. Eg5 targeting agents: From new anti-mitotic based inhibitor discovery to cancer therapy and resistance. *Biochemical Pharmacology* **184,** 114364 (2021).

220. Cochran, J. C., Krzysiak, T. C. & Gilbert, S. P. Pathway of ATP Hydrolysis by Monomeric Kinesin Eg5. *Biochemistry* **45,** 12334–12344 (2006).

221. Mickolajczyk, K. J. & Hancock, W. O. Kinesin Processivity Is Determined by a Kinetic Race from a Vulnerable One-Head-Bound State. *Biophysical Journal* **112,** 2615–2623 (2017).

222. Tao, L. *et al.* A Homotetrameric Kinesin-5, KLP61F, Bundles Microtubules and Antagonizes Ncd in Motility Assays. *Current Biology* **16,** 2293–2302 (2006).

223. Valentine, M. T., Fordyce, P. M., Krzysiak, T. C., Gilbert, S. P. & Block, S. M. Individual dimers of the mitotic kinesin motor Eg5 step processively and support substantial loads in vitro. *Nat Cell Biol* **8,** 470–476 (2006).

224. Shimamoto, Y., Forth, S. & Kapoor, T. M. Measuring Pushing and Braking Forces Generated by Ensembles of Kinesin-5 Crosslinking Two Microtubules. *Developmental Cell* **34,** 669–681 (2015).

225. Cahu, J. & Surrey, T. Motile microtubule crosslinkers require distinct dynamic properties for correct functioning during spindle organization in *Xenopus* egg extract. *Journal of Cell Science* **122,** 1295–1300 (2009).

226. Zhernov, I., Diez, S., Braun, M. & Lansky, Z. Intrinsically Disordered Domain of Kinesin-3 Kif14 Enables Unique Functional Diversity. *Current Biology* **30,** 3342–3351.e5 (2020).

227. Waitzman, J. S. *et al.* The Loop 5 Element Structurally and Kinetically Coordinates Dimers of the Human Kinesin-5, Eg5. *Biophysical Journal* **101,** 2760–2769 (2011).

228. Muretta, J. M. *et al.* Loop L5 Assumes Three Distinct Orientations during the ATPase Cycle of the Mitotic Kinesin Eg5. *Journal of Biological Chemistry* **288,** 34839–34849 (2013).

229. Brier, S., Lemaire, D., DeBonis, S., Forest, E. & Kozielski, F. Molecular Dissection of the Inhibitor Binding Pocket of Mitotic Kinesin Eg5 Reveals Mutants that Confer Resistance to Antimitotic Agents. *Journal of Molecular Biology* **360,** 360–376 (2006).

230. Moores, C. A. Kinesin-5 mitotic motors: Is loop5 the on/off switch? *Cell Cycle* **9,** 1286–1290 (2010).

231. Talapatra, S. K., Anthony, N. G., Mackay, S. P. & Kozielski, F. Mitotic Kinesin Eg5 Overcomes Inhibition to the Phase I/II Clinical Candidate SB743921 by an Allosteric Resistance Mechanism. *J. Med. Chem.* **56,** 6317–6329 (2013).

232. Liu, L., Parameswaran, S., Liu, J., Kim, S. & Wojcik, E. J. Loop 5-directed Compounds Inhibit Chimeric Kinesin-5 Motors. *Journal of Biological Chemistry* **286,** 6201–6210 (2011).

233. Brendza, K. M., Rose, D. J., Gilbert, S. P. & Saxton, W. M. Lethal Kinesin Mutations Reveal Amino Acids Important for ATPase Activation and Structural Coupling. *Journal of Biological Chemistry* **274,** 31506–31514 (1999).

234. Kikkawa, M. The role of microtubules in processive kinesin movement. *Trends in Cell Biology* **18,** 128–135 (2008).

235. Xie, P. An Overview of Contrasting Experimental Results on Dynamics of Kinesin-1 Molecular Motors: Insight into the Underlying Mechanism. *Biomolecules* **15,** 1453 (2025).

236. Rosenfeld, S. S., Xing, J., Jefferson, G. M. & King, P. H. Docking and Rolling, a Model of How the Mitotic Motor Eg5 Works. *Journal of Biological Chemistry* **280,** 35684–35695 (2005).

237. Cochran, J. C. *et al.* Mechanistic Analysis of the Mitotic Kinesin Eg5. *Journal of Biological Chemistry* **279,** 38861–38870 (2004).

238. Morikawa, M. *et al.* X-ray and Cryo-EM structures reveal mutual conformational changes of Kinesin and GTP-state microtubules upon binding. *EMBO J* **34,** 1270–1286 (2015).

239. Shi, X.-X. *et al.* Investigating role of conformational changes of microtubule in regulating its binding affinity to kinesin by all-atom molecular dynamics simulation. *Proteins* **86,** 1127–1139 (2018).

240. Peña, A. *et al.* Structure of Microtubule-Trapped Human Kinesin-5 and Its Mechanism of Inhibition Revealed Using Cryoelectron Microscopy. *Structure* **28,** 450–457.e5 (2020).

241. Lad, L. *et al.* Mechanism of Inhibition of Human KSP by *Ispinesib*. *Biochemistry* **47,** 3576–3585 (2008).

242. Crevel, I.-T., Alonso, M. & Cross, R. Monastrol stabilises an attached low-friction mode of Eg5. *Current Biology* **14,** R411–R412 (2004).

243. Sheth, P. R. *et al.* Thermodynamics of Nucleotide and Inhibitor Binding to Wild-Type and Ispinesib-Resistant Forms of Human Kinesin Spindle Protein. *Biochemistry* **48,** 11045–11055 (2009).

244. Jones, R. *et al.* Phase II study to assess the efficacy, safety and tolerability of the mitotic spindle kinesin inhibitor AZD4877 in patients with recurrent advanced urothelial cancer. *Invest New Drugs* **31,** 1001–1007 (2013).

245. O'Connor, O. A. *et al.* The addition of granulocyte-colony stimulating factor shifts the dose limiting toxicity and markedly increases the maximum tolerated dose and activity of the kinesin spindle protein inhibitor SB-743921 in patients with relapsed or refractory lymphoma: results of an international, multicenter phase I/II study. *Leukemia & Lymphoma* **56,** 2585–2591 (2015).

246. Ocio, E. M. *et al.* Filanesib in combination with pomalidomide and dexamethasone in refractory MM patients: safety and efficacy, and association with alpha 1-acid glycoprotein (AAG) levels. Phase Ib/II Pomdefil clinical trial conducted by the Spanish MM group. *Br J Haematol* **192,** 522–530 (2021).

247. Alegaon, S. G., Ranade, S. D., Gharge, S., Kavalapure, R. S. & Ulaganathan, V. Computational and experimental repositioning of quinoline analogues as KSP inhibitors: insights from free energy landscape and PCA analysis. *J Comput Aided Mol Des* **39,** 65 (2025).

248. Ogunwa, T. H., Laudadio, E., Galeazzi, R. & Miyanishi, T. Insights into the Molecular Mechanisms of Eg5 Inhibition by (+)-Morelloflavone. *Pharmaceuticals* **12,** 58 (2019).

249. McGrath, M. J., Kuo, I.-F. W., Hayashi, S. & Takada, S. Adenosine Triphosphate Hydrolysis Mechanism in Kinesin Studied by Combined Quantum-Mechanical/Molecular-Mechanical Metadynamics Simulations. *J. Am. Chem. Soc.* **135,** 8908–8919 (2013).

250. Sebastian, J., Raghav, D. & Rathinasamy, K. MD simulation-based screening approach identified tolvaptan as a potential inhibitor of Eg5. *Mol Divers* **27,** 1203–1221 (2023).

251. Elseginy, S. A. Identifying and characterising promising small molecule inhibitors of kinesin spindle protein using ligand-based virtual screening, molecular docking, molecular dynamics and MM-GBSA calculations. *J Comput Aided Mol Des* **38,** 16 (2024).

252. Elseginy, S. A. Identification, characterization and biological studies of kinesine spindle protein inhibitor, benzo[b][1,4]oxazine-6-carboxamide derivative, via 3D-pharmacophore virtual screening. *Journal of Molecular Structure* **1312,** 138508 (2024).

253. Jia, N. *et al.* Binding patterns of inhibitors to different pockets of kinesin Eg5. *Archives of Biochemistry and Biophysics* **756,** 109998 (2024).

254. Alexandar, S. P., Yennamalli, R. M. & Ulaganathan, V. Coarse grained modelling highlights the binding differences in the two different allosteric sites of the Human Kinesin EG5 and its implications in inhibitor design. *Computational Biology and Chemistry* **99,** 107708 (2022).

255. Zhang, W. Exploring the Intermediate States of ADP−ATP Exchange: A Simulation Study on Eg5. *J. Phys. Chem. B* **115,** 784–795 (2011).

256. Rahman, M. A. *et al.* Thermochemical, spectral and biological activity prediction of some methylpyridine derivatives: a computational approach. *Discov. Chem.* **2,** 173 (2025).

257. Alexandar, S. P. & Ulaganathan, V. Investigation of the role of microtubule on the activity of the mitotic kinesin EG5 using multiscale modelling: unravelling molecular mechanisms. *J Comput Aided Mol Des* **39,** 92 (2025).

258. Turner, J. *et al.* Crystal Structure of the Mitotic Spindle Kinesin Eg5 Reveals a Novel Conformation of the Neck-linker. *Journal of Biological Chemistry* **276,** 25496–25502 (2001).

259. Indorato, R.-L. *et al.* Is the Fate of Clinical Candidate Arry-520 Already Sealed? Predicting Resistance in Eg5–Inhibitor Complexes. *Molecular Cancer Therapeutics* **18,** 2394–2406 (2019).

260. Parke, C. L., Wojcik, E. J., Kim, S. & Worthylake, D. K. ATP Hydrolysis in Eg5 Kinesin Involves a Catalytic Two-water Mechanism. *Journal of Biological Chemistry* **285,** 5859–5867 (2010).

261. Varela, P. F. *et al.* Structural snapshots of the kinesin-2 OSM-3 along its nucleotide cycle: implications for the ATP hydrolysis mechanism. *FEBS Open Bio* **11,** 564–577 (2021).

262. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J Comput Chem* **5,** 129–145 (1984).

263. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79,** 926–935 (1983).

264. Seeger, M. A. & Rice, S. E. Intrinsic disorder in the kinesin superfamily. *Biophys Rev* **5,** 233–247 (2013).

265. Stock, M. F., Chu, J. & Hackney, D. D. The Kinesin Family Member BimC Contains a Second Microtubule Binding Region Attached to the N terminus of the Motor Domain. *Journal of Biological Chemistry* **278,** 52315–52322 (2003).

266. Edamatsu, M. Molecular properties of the N-terminal extension of the fission yeast kinesin-5, Cut7. *Genet. Mol. Res.* **15** (2016).

267. Von Loeffelholz, O. & Moores, C. A. Cryo-EM structure of the Ustilago maydis kinesin-5 motor domain bound to microtubules. *Journal of Structural Biology* **207,** 312–316 (2019).

268. Holehouse, A. S. & Kragelund, B. B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat Rev Mol Cell Biol* **25,** 187–211 (2024).

269. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **115** (2018).

270. Guo, W. *et al.* Using a comprehensive approach to investigate the interaction between Kinesin-5/Eg5 and the microtubule. *Comput Struct Biotechnol J* **20,** 4305–4314 (2022).

271. Brier, S., Lemaire, D., DeBonis, S., Forest, E. & Kozielski, F. Identification of the Protein Binding Region of S-Trityl-l-cysteine, a New Potent Inhibitor of the Mitotic Kinesin Eg5. *Biochemistry* **43,** 13072–13082 (2004).

272. Cochran, J. C. & Gilbert, S. P. ATPase Mechanism of Eg5 in the Absence of Microtubules: Insight into Microtubule Activation and Allosteric Inhibition by Monastrol. *Biochemistry* **44,** 16633–16648 (2005).

273. Behnke-Parks, W. M. *et al.* Loop L5 Acts as a Conformational Latch in the Mitotic Kinesin Eg5. *Journal of Biological Chemistry* **286,** 5242–5253 (2011).

274. Kim, E. D. *et al.* Allosteric Drug Discrimination Is Coupled to Mechanochemical Changes in the Kinesin-5 Motor Core. *Journal of Biological Chemistry* **285,** 18650–18661 (2010).

275. Cook, A. D. *et al.* Cryo-EM structure of a microtubule-bound parasite kinesin motor and implications for its mechanism and inhibition. *Journal of Biological Chemistry* **297,** 101063 (2021).

276. Tcherniuk, S., Van Lis, R., Kozielski, F. & Skoufias, D. A. Mutations in the human kinesin Eg5 that confer resistance to monastrol and S-trityl-l-cysteine in tumor derived cell lines. *Biochemical Pharmacology* **79,** 864–872 (2010).

277. Harrington, T. D. *et al.* Analysis of the interaction of the Eg5 Loop5 with the nucleotide site. *Journal of Theoretical Biology* **289,** 107–115 (2011).

278. Chen, G.-Y. *et al.* Eg5 Inhibitors Have Contrasting Effects on Microtubule Stability and Metaphase Spindle Integrity. *ACS Chem. Biol.* **12,** 1038–1046 (2017).

279. La Sala, G., Decherchi, S., De Vivo, M. & Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent. Sci.* **3,** 949–960 (2017).

280. Lubchenko, V. Competing interactions create functionality through frustration. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 10635–10636 (2008).

281. Gianni, S. *et al.* Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 14141–14146 (2014).

282. Tzul, F. O., Vasilchuk, D. & Makhatadze, G. I. Evidence for the principle of minimal frustration in the evolution of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **114** (2017).

283. Ganoth, A., Tsfadia, Y. & Wiener, R. Ubiquitin: Molecular modeling and simulations. *Journal of Molecular Graphics and Modelling* **46,** 29–40 (2013).

284. Komander, D. & Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **81,** 203–229 (2012).

285. Morra, G., Genoni, A. & Colombo, G. Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains. *J. Chem. Theory Comput.* **10,** 5677–5689 (2014).

286. Li, Y. *et al.* Computational simulation studies on the binding selectivity of Wee1 and Checkpoint kinase 1 by molecular dynamics simulation combined with free energy calculations. *Journal of Biomolecular Structure and Dynamics* **40,** 1172–1181 (2022).

287. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 5915–5920 (2013).

288. Fenwick, R. B. *et al.* Weak Long-Range Correlated Motions in a Surface Patch of Ubiquitin Involved in Molecular Recognition. *J. Am. Chem. Soc.* **133,** 10336–10339 (2011).

289. Kony, D. B., Hünenberger, P. H. & Van Gunsteren, W. F. Molecular dynamics simulations of the native and partially folded states of ubiquitin: Influence of methanol cosolvent, pH, and temperature on the protein structure and dynamics. *Protein Science* **16,** 1101–1118 (2007).

290. He, Y., Liwo, A., Weinstein, H. & Scheraga, H. A. PDZ Binding to the BAR Domain of PICK1 is Elucidated by Coarse-grained Molecular Dynamics. *Journal of Molecular Biology* **405,** 298–314 (2011).

291. Gianni, S. *et al.* A PDZ domain recapitulates a unifying mechanism for protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 128–133 (2007).

292. Stevens, A. O. & He, Y. Allosterism in the PDZ Family. *IJMS* **23,** 1454 (2022).

293. Kong, Y. & Karplus, M. Signaling pathways of PDZ2 domain: A molecular dynamics interaction correlation analysis. *Proteins* **74,** 145–154 (2009).

294. Al-Shar'i, N. & Musleh, S. S. CHK1 kinase inhibition: identification of allosteric hits using MD simulations, pharmacophore modeling, docking and MM-PBSA calculations. *Mol Divers* **26,** 903–921 (2022).

295. Mai, N. T. *et al.* Estimation of the ligand-binding free energy of checkpoint kinase 1 via non-equilibrium MD simulations. *Journal of Molecular Graphics and Modelling* **100,** 107648 (2020).

296. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7,** 539 (2011).

297. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).