DEVELOPMENT OF A HYBRID METHODOLOGY FOR INVESTIGATION AND MANIPULATION OF FUNCTIONAL MECHANISMS OF BIOLOGICAL MACROMOLECULES WITH A FOCUS ON NON-GLOBULAR PROTEINS

by

Burçin Acar

B.S., Molecular Biology and Genetics, Istanbul University, 2008

M.S., Biomolecular Sciences, University of Groningen, 2010

Submitted to the Institute of Biomedical Engineering

in partial fulfillment of the requirements

for the degree of

Doctor

of

Philosophy

Boğaziçi University

2021

ACKNOWLEDGEMENTS

Firstly, I would like to pronounce my gratitude to my thesis advisor, Prof. Türkan Haliloğlu for her guidance and encouragement on performing independent research, while providing me the opportunity to collaborate with great scientists. I am also deeply thankful to my co-advisor Ahmet Ademoğlu for giving me the chance to work in an interdisciplinary field as computational biology.

I am grateful to the members of my thesis committee Prof. Kutlu Ülgen, Assoc. Prof. Bora Garipcan, Prof. Özlem Keskin, Assistant Prof. Ezgi Karaca, who have provided valuable comments and discussions on my thesis.

It was a pleasure to collaborate with Assoc. Prof. Oded Lewinson, who enriched this thesis by providing the experimental results on BtuCD and sharing his biological knowledge. With Prof. Nir Ben-Tal, their valuable suggestions and discussions contributed a lot to this study. Additionally, I very much appreciate Assoc. Prof. Hamdi Torun for passing his technical and scientific knowledge on AFM.

I would like to thank to all my friends from Polymer Research Center and Boğaziçi for their support and fellowship, with special thanks to Burcu Aykaç Fas for a joyful teamwork that enabled us to develop a useful computational tool. Many thanks to my friends from Istanbul Lindy Hoppers for making my life colorful and enjoyable.

I would like to acknowledge the financial support from TUBITAK projects 101T088, 115M418 and 112T569, BAP projects 11160, 16542 and 8622 (COST), BETIL fund and NATO SPS projects 984622 and G5685.

Finally, I thank to my parents, Şennur and İrfan Acar; for always supporting me.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Burçin Acar, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

DEVELOPMENT OF A HYBRID METHODOLOGY FOR INVESTIGATION AND MANIPULATION OF FUNCTIONAL MECHANISMS OF BIOLOGICAL MACROMOLECULES WITH A FOCUS ON NON-GLOBULAR PROTEINS

Anisotropic Network Model (ANM) guided Langevin Dynamics (LD) method (ANM-LD) is an enhanced sampling algorithm, in-house developed to study conformational changes between functional protein structures, that are not possible by traditional techniques. In this thesis, the applicability of ANM-LD was validated on various non-globular systems, then assessed on vitamin B12 importer BtuCD by means of experimental observations and comparison of computational outcomes with maltose importer $MalFGK_2$ and lipid-linked oligosaccharide flippase PglK. ANM-LD succeeded to extract the mechanistic differences among these transporters while predicting fluctuations and allosteric couplings of BtuCD residues in agreement with previous experiments and observed FRET intensities. The dynamically key residues enabling the sampled transition were defined as functional residue networks and their estimated perturbation response were highly agreeable with the functional assays of the BtuCD mutants on these sites (25 out of 26 mutants functioned as predicted). Later ANM-LD algorithm was advanced to improve sampling, then tested on case systems c-Src kinase and BtuCD. In c-src kinase, these enhancements enabled to predict dynamically key sites that overlapped with known oncogenic mutation sites. In BtuCD, diversification of guiding modes resulted in alternative transition pathways. Consequently with its practicality and modularity; ANM-LD stands as an efficient tool to study protein dynamics and their working mechanisms and to extract allosteric communication networks, toward the aim of controlling protein function.

Keywords: Computational Sampling Methods, Protein Dynamics, Enhanced Sampling Methods, ANM-LD, Conformational Change, Elastic Network Model.

ÖZET

GLOBÜLER OLMAYAN PROTEİNLER MERKEZİNDE BİYOLOJİK MAKROMOLEKÜLLERİN İŞLEVLERİNİN ARAŞTIRILMASI VE YÖNLENDİRİLMESİ İÇİN MELEZ BİR YÖNTEMİN GELİŞTİRİLMESİ

Anizotropik ağ modeli (ANM) rehberli Langevin Dynamics (LD) simülasyonu (ANM-LD), geleneksel yöntemlerin yetersiz kaldığı protein konformasyon değişikliklerini incelemek amacıyla laboratuvarımızda geliştirilmiş, ileri bir hesaplamalı örnekleme yöntemidir. ANM-LD'nin performansı, öncelikle çeşitli küresel olmayan proteinlerde değerlendirilmiş, daha sonra B12 vitamini taşıyıcı proteini BtuCD'de elde edilen sonuçlar, deneysel gözlemler ve maltoz alıcı $MalFGK_2$ ve lipid bağlı oligosakkarit taşıyıcı PglK proteinleri için sağlanan hesaplamalı sonuçlar ile karşılaştırılarak detaylı irdelenmiştir. ANM-LD, BtuCD rezidü hareketlerini, gözlemlenen FRET yoğunluklarıyla uyumlu olarak tahmin ederken, bu proteinler arasındaki işlevsel mekanizma farklılıklarını da saptamayı başarmıştır. Yapılar arası geçişi mümkün kılan anahtar rezidüler, işlevsel rezidü ağları olarak tanımlanmış ve bu rezidüler için tahmin edilen sarsım tepkileri, BtuCD'nin bu bölgelerde oluşturulan mutantlarının işlev deneyleriyle oldukça uyumlu sonuç vermiştir (26 mutantın 25'i). Daha sonra simülasyon algoritması, örneklemenin iyileştirilmesi amacıyla, c-Src kinaz ve BtuCD sistemleri üzerinde denererek, çeşitli yönlerden geliştirilmiştir. Bu iyileştirmeler c-Src kinazda onkogenik bölgelerle örtüşen anahtar rezidülerin belirlenmesini sağlamıştır. BtuCD'de ise simülasyona rehberlik eden mod havuzunun çeşitlendirilmesi başka geçiş patikalarının bulunmasıyla sonuçlanmıştır. Özetle, ANM-LD; pratikliği ve esnekliği ile, proteinlerin dinamiklerini ve çalışma mekanizmalarınıincelemek, ve allosterik iletişim ağlarını çıkarmak ve protein işlevini tasarlamak hedeflerine yönelik etkili bir araç olarak durmaktadır.

Anahtar Sözcükler: Hesaplamalı Örnekleme Yöntemleri, Protein Dinamiği, Gelişmiş Örnekleme Yöntemleri, ANM-LD, Konformasyon Değişimi, Elastik Ağyapı Modeli.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

AD	Cumulative angular deviation score
c	Co-variance matrix of residue fluctuations
C	Dynamic cross correlation matrix of residue fluctuations
C_α	Alpha carbon atom
df	the difference in the atomic distances
DF	Deformation factor
D_{max}	Maximum size parameter from SAXS experiments
H	Hessian matrix
k_B	Boltzmann constant
m_i	Atomic mass
$modemax$	Number of available slow ANM modes for ANM-LD
N	Number of amino acid residues on protein
p	Probability
R_c	ANM cutoff distance
R_g	Radius of gyration
R_i	Positional vector of atom i
r_{MC}	Pseudo-random variable to be used in MC
R_{new}	Atomic positions of output conformation by ANM-LD
R_{old}	Atomic positions of starting structure
RF_i	Random force applied on atom i
ΔR	Positional fluctuation vector
ΔR_{ANM}	Positional changes caused by ANM perturbations
$\Delta R_{Eng.Min.}$	Positional changes caused by energy minimization
ΔR_{LD}	Positional changes caused by Langevin Dynamics
S	Conditional entropy
t	Time variable
T	Absolute temperature in degrees Kelvin
t_{rest}	Overlap threshold for restriction runs

TE_{ij}	Transfer entropy from residue i to j for a given time delay
tr	Trace of a square matrix
U_{best}	The most overlapping ANM vector with difference vector
U_k	Eigenvector for kth slow mode
V	Potential energy
Å	Angstrom
γ	Interaction force constant for ANM
Γ	Kirchhoff matrix
γ_{LD}	Damping coefficient for Langevin Dynamics
λ_k	Eigenvalue for kth slow mode
μ_{WT}	Mean of backbone rotation angles for wild type protein
τ	Time delay factor
ϕ	Backbone rotation angles
χ^2	Discrepancy value between scattering curves

LIST OF ABBREVIATIONS

3D	Three Dimensional
ABC	ATP-Binding Cassette
Aβ protein	amyloid beta protein
AdK	Adenylate Kinase
ADP	Adenosine Diphosphate
AFM	Atomic Force Microscopy
AMP	Adenosine Monophosphate
ANM	Anisotropic Network Model
ANM-LD	Anisotropic Network Model guided Langevin Dynamics
ATP	Adenosine Triphosphate
BtuCD	Vitamin B_{12} Transporter
coMD	Collective Molecular Dynamics
COSMIC	Catalogue of Somatic Mutations in Cancer
COST	European Cooperation in Science and Technology
cryo-EM	Cryogenic Electron Microscopy
c-Src	Cellular Src Kinase
ClyA	Cytolytic Pore Toxin Protein
DCCM	Dynamic Cross Correlation Map
DEER	Double Electron Electron Resonance
ENM	Elastic Network Model
EPR	Electron Paramagnetic Resonance
FimD	Fimbrial Usher Protein
FRET	Fluorescence Resonance Energy Transfer
GE	Generalized Ensemble
GNM	Gaussian Network Model
GPCR	G-Protein Coupled Receptor
IaMD	Integrated Accelerated Molecule Dynamics
IDP	Intrinsically Disordered Protein

IF	Inward Facing Conformation
KD	Kinase Domain
LD	Langevin Dynamics
$MalFGK_2$	Maltose Transporter Protein
MDeNM	Molecular Dynamics with Excited Normal Modes
MC	Monte Carlo
MD	Molecular Dynamics
MDeNM	Molecular Dynamics with Excited Normal Modes
MSF	Mean Square Fluctuations
MntABC	Manganese Transporter Protein
NBD	Nucleotide Binding Domain
NM	Normal Mode
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
OF	Outward Facing Conformation
PapC	P Fimbrial Usher Protein
PC	Principal Component
PCA	Principal Component Analysis
PDB	Protein Data Bank
PglK	Lipid Oligosaccharide Flippase Protein
PSP	Perturb-Scan-Pull
RE	Replica Exchange
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
SAXS	Small Angle X-ray Scattering
SBP	Substrate Binding Protein
TE	Transfer Entropy
TMD	Transmembrane Domain
WT	Wild Type

1. INTRODUCTION

Proteins are biological macromolecules through which cells execute various vital functions, such as catalyzing chemical reactions, building cellular architectures, transporting sustenance and waste materials; and many more. Like other molecules, they obey physical laws of thermodynamics and although they attain specific shapes due to atomic interactions, they are truly dynamic [1–4]. The resulting litheness makes proteins readily respond to environmental changes (e.g. pH, temperature), functional/modulation events (e.g. binding of ligands, biomolecular partners and substrates) and post-translational modifications. These modifications may affect the distribution of visited states in an ensemble of possible protein conformations (i.e. a shift in the ensemble of conformations), in which each conformation corresponds to a distinct free energy value [2]. The structural/conformational and dynamical changes related with these altering phenomena might occur with the changes in enthalpic interactions and/or in overall entropy of the biological system. Eventually, the resulting conformational plasticity and dynamics enable proteins to perform biological functions by providing the necessary mechanistic steps and are evolutionarily conserved for each protein [5–9]. All these factors make dynamics an indispensable component of the protein function and an interesting target for studying the functional mechanisms with the aim of understanding and control either through protein engineering or external modifiers serving to therapeutic (i.e. drug discovery) and biotechnological (i.e. product and process improvements) purposes [10–14].

Understanding the role of dynamics in the molecular mechanism of protein function requires the elucidation of several long-standing questions. How amino acid sequences through attained structures and dynamics create functional proteins [5, 15, 16]? How do these aspects interrelate throughout the evolution so that the proteins are fine-tuned to perform specific tasks yet diversify in function [17, 18]? How do proteins interconvert between different conformational states [19–21]? How do they respond to environmental changes [22, 23]? How is protein function regulated

on molecular level [24, 25]? How does allosteric signal arise and dissipate in terms of dynamic couplings between various sites of a protein [26, 27]? How do perturbations on protein molecules such as mutations modify function through the dynamic changes [28]? These questions can be diversified and have been targeted by various methods in protein science for a complete understanding of protein dynamics and related functional mechanisms and they are also at the center of this thesis.

1.1 Protein Dynamics Can Be Summarized in Terms of a Free Energy Landscape

In order to approach above mentioned, complex biological questions; protein dynamics was approximated as wandering of the molecule on a complex energy landscape on which every possible/accessible conformation of a protein is depicted with a corresponding free energy [29]. Proteins are hypothesized to sweep this energy landscape as they fluctuate. Various stable conformations of a protein belong to discrete energy wells on this landscape and are generally separated by high energy barriers [6] as exemplified on Figure 1.1. Herein, the landscape is defined with two axes of reaction coordinates or collective variables along which major conformational changes occur. The fluctuations of the protein in these energy wells define equilibrium dynamics [30]. Interchange between different stable conformations might occur through multiple transition pathways imposed by the topology of the landscape and correspond to transition dynamics of a protein [6, 31].

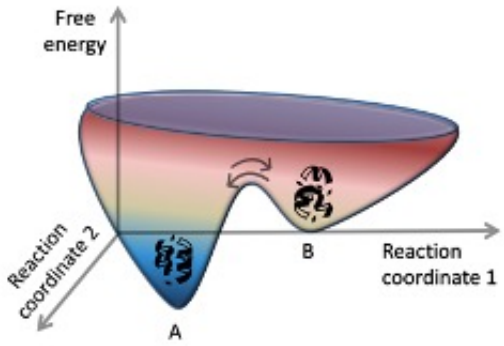

Figure 1.1 Schematic representation of a free energy landscape of a protein with two stable conformational states A and B corresponding to distinct minima.

The free energy landscape of a protein is also not resistant to change [32]. The physicochemical events such as ligand-binding [33], protein-solvent interactions with entropic outcomes [34], post-translational modifications, mutations and environmental changes such as temperature and pH are known to change the energy landscape hence lowering or raising the energy barriers between conformational states [32]. All these phenomena make studying protein dynamics challenging. Yet in order to understand the dynamic behavior of proteins in relation with their functional mechanisms, methods that can sample this energy landscape in full cover are needed. To this aim, both experimental and computational methods have been exploited widely as complementary tools.

1.2 Experimental Methods for the Exploration of Protein Dynamics

Experimental methods are essential for protein dynamics studies, as they provide the structures and measure the physical properties of biologically relevant states from an ensemble of visited conformations. X-ray crystallography experiments yield high resolution structures of biomolecules [35, 36] which slot other observed physical parameters into the larger picture of functional mechanism and be further utilized by molecular simulations of protein dynamics. However, the method requires the protein to be in a crystal form, a difficult and nonnative state to obtain, and then

to be bombarded by X-rays which in the end might cause radiation damage on the sample material. The first disadvantage is averted in another X-ray based method, small-angle X-ray scattering (SAXS) which examines the protein in solution instead of a crystal and interpreting the scattering curves, although the quality of the measurement outcome is limited to a small number of observables such as molecular size and shape [37]. Another powerful tool to study proteins in solution yet averting radiational damages is nuclear magnetic resonance (NMR) spectroscopy, which provides direct dynamic information with atomic resolution in addition to deciphering structures. NMR is particularly useful for proteins with size between 5 and 25 kDa [38]. For larger systems, obtaining high quality data is difficult due to crowdedness and overlap in resonance peaks of different nuclei which obstruct the determination of their spatial position. In this respect, cryo-electron microscopy (cryo-EM) is a more appropriate technique for large systems and multimolecular complexes as the surface traces of electron beams hit on the sample are easier to interpret as 3D structures [39]. However, the sample is prone to radiation damage due to electron bombardment and the resolution is generally in medium quality [40]. Atomic force microscopy (AFM) is another method which can be used for structure determination of large systems besides their mechanical features and flexibility, without the harm of radiation [41]. AFM can perform at room temperature in contrast with cryo-EM and without any need for a crystallized sample as opposed to X-ray crystallography. It does not require the use of any electromagnetic beam to scan protein surface but utilizes the touch of a mechanic probe on the sample. These features make AFM a valuable tool to study proteins under native-like conditions [42, 43]. In recent years, another method called Terahertz microscopy which uses lower frequency radiation to target the protein samples and measure resulting THz electric field to obtain long-range vibrations of proteins gained popularity [44, 45]. Other biochemical techniques such as Förster resonance energy transfer (FRET) and mass spectrometry (MS) are also widely used for obtaining observables such as inter-residual distances to support protein dynamics studies. All these experimental tools procure the physicochemical observations required to construe the role of dynamics in functional mechanisms of proteins.

1.3 Computational Methods Complement Protein Dynamics Research

Although experimental methods are highly informative on protein dynamics, they are expensive, time consuming and work with limitations imposed by the duration of the conformational transition of interest, physical properties of the system and environmental conditions. For example, solution NMR is not suitable for membrane-embedded systems [46] and fast conformational changes [47], while cryo-EM requires working at low temperatures. Therefore, in order to obtain a complete picture on functional mechanism of proteins, computational methods are exploited both to elucidate parts of protein dynamics omitted by experiments and to predict a potential outcome of an experiment before investing time and money.

To apply computational methods in the protein dynamics studies has various advantages. Firstly, these methods are cheaper and more time-efficient than the experimental methods while still being able to provide atomic-level description. In addition, a simulation renders not only stable conformations of a protein, but also the transient states, even the rare conformations [48]. It enables to detect transition pathways between different states of a protein, sometimes more than one pathway [49]. Estimating fluctuations for each residue, computational tools provide a means to deduce dynamic coupling between different sites of the protein. The deduced allosteric interactions between different sites of the protein help for the understanding of functional mechanism and its further control with potential drug molecules in case of disease-forming malfunction or low yield activity of the protein. In conclusion, computational methods are highly informative on protein dynamics; however, it is important to keep in mind that they are based on physicochemical assumptions and require experimental validation for evaluating the biological and physical relevance of deduced molecular mechanisms and dynamic behaviors.

As discussed earlier, free energy landscape of a protein is a useful scheme to feature its complex dynamic behavior. The efficiency of a computational method

therefore is based on its ability to sample this landscape as completely as possible in a reasonable time. Various methods were developed with this aim. Molecular Dynamics (MD) is a standardized method for conformational sampling, as it provides a time-evolution of the system by solving Newton's laws of motion [50]. It even brought a Nobel Prize to its developers, as it provides a way to create multiscale models for complex (bio) chemical systems including proteins [51]. Despite of the reliability and physical relevance of MD simulations, it has certain drawbacks. For modeling large systems and sampling long-time changes, MD simulations take a significant amount of time and computational power. Additionally, the MD sampling of proteins deviating from the globular behavior is challenging since previously performed simulations yielded unrealistic compactness for the sampled conformations [52, 53]. The optimizations in the interactions among solvent, water and protein atoms were shown to improve this [54, 55].

In order to avoid the computation time required for solving equation of motions as in MD, Monte Carlo (MC) based sampling techniques has been proposed, in which conformational iterations take through conformational space in tentative time units, as number of iterations. For this purpose, randomly chosen residues and/or atoms of the protein are perturbed and evaluated based on the energy difference between the initial and perturbed states (i.e. Metropolis criterion) repeatedly. Although MC-based methods are faster, conventionally they do not provide information about the evolution of the system in real time, when time factor of a certain dynamic behavior is needed [56]. However, they can still be useful to calculate probabilities and the free energies of the conformations in an ensemble of representative configurations obtained from the simulations [57], and to decipher order of events with a Markovian approach [58, 59].

Another approach to avoid time and size limitation problem of conformational sampling is using normal modes (NM)-based methods to dissect equilibrium dynamics of proteins into orthogonal modes and to predict thermal fluctuations of the system. Elastic network models (ENMs), as an NM-based approach, are widely used as coarse-grained models of protein structures to assume interaction of the neighboring nodes (mostly C_α atoms) to be defined by harmonic potentials and

calculate residual fluctuations based on these [60]. Two popular ENMs are: Gaussian Network Model (GNM) [61] and Anisotropic Network Model (ANM) [62], which assume isotropic and anisotropic fluctuations for the protein respectively. Although these methods are highly informative on the collective dynamics of the system at equilibrium, their contribution alone is limited in investigation of non-equilibrium dynamics such as large conformational changes or long-time dynamics. Morphing techniques such as normal mode interpolation that interpose with the two end-point conformations of a target protein was developed to compensate this shortfall. Although these methods provide fast computation, the energetics of the sampled conformations are difficult to be ensured [63, 64].

The application limitations of above-mentioned algorithms, for dynamic events such as rare and high-energy states, long-time events and large conformational changes, necessitate developing and improving sampling techniques to reach wider applicability. Thus, many coarse-grained methods as well as enhanced sampling methods were developed with the aim of a complete sampling of the conformational space and understanding the protein dynamics hence the functional mechanism. Some hybrid methods also combine experimental derivatives with computational sampling techniques in order to obtain biologically and physically relevant conformation trajectories [65, 66]. One of the main aims of these enhanced sampling methods is to sample protein dynamics more efficiently especially by overcoming the problem of trapping in one of the energy minima on the relevant free energy landscape [67]. Simulated annealing provides a solution at this point by using temperature change during the simulation; however it is hard to obtain accurate thermodynamic average parameters with this method [65, 68]. The tools which correspond to the time gap between computational methods and time-span of biological events are also needed to fully sample the protein dynamics. Current enhanced sampling methods developed toward this aim, can be classified in two basic categories: generalized ensemble (GE) algorithms and multiscale sampling [69]. GE algorithms solve trapping problem by performing a random walk in the potential energy space using non-Boltzmann weight factor for each conformational state. Known examples of this approach are multicanonical algorithm, umbrella sampling

and replica exchange (RE) methods [65]. On the other hand, multiscale sampling algorithms combine coarse-grained simulations with all-atom simulations [69–72] and are discussed further in the next section, as such algorithm is evaluated and advanced as part of this thesis.

1.4 Scope and Aim of This Work

Multiscale sampling methods are particularly useful for addressing transition dynamics of proteins between their various stable conformations. In literature, many methods exploit the fast steering capacity of normal modes to guide conventional numerical sampling methods such as MD and MC [73, 74]. While NMA based-ENM tools are used to predict collective mode shapes to direct the conformational perturbations, MD and MC methods are exploited to evaluate and minimize energetics of atomic interactions and local fluctuations of the perturbed structure. Collective molecular dynamics (coMD) is one of these approaches [75], which deforms the initial and target structures of a transition along the modes predicted by ANM to bring together a trajectory. Another method proposed a unidirectional approach in which the initial structure is perturbed in soft ANM modes toward a target conformation and energetically relaxed by MC [76]. Besides these methods, the MDeNM (molecular dynamics with excited normal modes) method performs multiple-replica short MD simulations in which motions are kinetically excited by adding additional atomic velocities along NM vectors [77, 78]. Another approach called iterANM-IaMD combines ANM, to construct an initial conformation space of protein by iterations in ANM modes, with Integrated accelerated Molecule Dynamics (IaMD) to explore the relevant free energy landscape [79, 80]. On the other hand, the perturb-scan-pull (PSP) method integrates elastic network-based perturbation response calculations with steered MD [81].

In our group, a similar multiscale sampling method has been developed called ANM-driven Langevin Dynamics (ANM-LD) [82–84]. ANM-LD utilizes the structure's ANM modes to guide conformational changes through cycles of perturbations toward

a target conformation, each perturbation followed by a short time LD simulation for the energy minimization. The ANM mode of perturbation is decided based on the overlap of corresponding mode vector with the difference vector between initial and target conformations. After perturbation along the selected mode vector, LD is then used to relax the conformation considering the stochastic solvent-protein interactions implicitly. The iteration of cycles is continued until root mean square distance (RMSD) of the created conformations to the final conformation reaches a plateau.

In this thesis, ANM-LD method was planned to be evaluated and advanced on various aspects and biological systems. In Chapter 2, non-globular proteins are summarized in general and case protein systems BtuCD and c-Src kinase, used in this study are discussed in more detail. The general ANM-LD protocol applied in this thesis is given in Chapter 3 along with the performed advancements on the algorithm (i.e. diversifying mode selection criteria, incorporating ligand molecules, developing all-atom ANM-LD simulations and integration with experimental variables) and methods for trajectory analysis.

In Chapter 4, the physical and biological relevance of obtained ANM-LD trajectories were addressed and aimed, for membrane proteins. Toward this aim, ANM-LD was applied to study transition dynamics, corresponding to functional cycles of conformational changes, of three ATP-binding cassette (ABC) transporters, namely vitamin B12 and maltose importers (i.e. BtuCD and $MalFGK_2$ respectively) and oligosaccharide flippase PglK. The conformational changes imposed by their different topologies enable a comparative approach to decipher dynamic and allosteric behavior of these transporters resulting with distinct functional mechanisms. Thus, directionality of allosteric signaling in conformational transitions of these systems and underlying causality which can be deciphered by transfer entropy calculations, was deduced. Furthermore, BtuCD was selected as a case membrane system to assess ANM-LD based deductions on dynamics thoroughly, by comparing computed residue pair distances with observed FRET intensities. By dissection of complex transition dynamics into elemental normal modes and prediction of hinge residues underlying these modes, key dynamic residue networks were detected. The dynamical effects of

the perturbations on these residues were predicted by a perturbation response analysis based on ANM-LD simulations. The functional role of these hinge residues was validated by single point mutation experiments at these positions and functional assays of the generated mutants in comparison with WT.

After ensuring the application potential of ANM-LD on membrane protein dynamics, various advances were implemented on the ANM-LD algorithm in order to improve the sampling capacity of the method (Chapter 5). Firstly, different ANM mode selection algorithms were implemented to ANM-LD protocol and resulting trajectories of a large globular protein, c-Src kinase were evaluated based on calculated energy landscape of the system with two known collective variables. Restriction of the ANM mode pool for perturbation was presented as a procedure to assess the roles of certain ANM modes on the transition dynamics. Secondly, ligand molecules were incorporated into the ANM-LD algorithm in order to consider role of ligand presence on the obtained trajectory and related dynamic behavior. Thirdly, in order to improve the resolution of the conformational trajectory, all-atom ANM-LD simulation algorithms were developed. Finally, the integration of ANM-LD algorithm with experimental restraints obtained from AFM and SAXS data is performed.

Finally, the thesis is concluded in Chapter 6 and prospective studies are discussed in Chapter 7, while the publications produced from this thesis are listed in Chapter 8.

2. NON−GLOBULAR PROTEINS AND CASE SYSTEMS FOR EVALUATION AND ADVANCEMENT OF ANM-LD SIMULATIONS

2.1 Non-globular Proteins as Challenging Systems to be Studied

As their name suggests, non-globular proteins are defined by what they are not, rather than what they are, since the features of these proteins deviate from the physicochemical characteristics of globular proteins. Globular proteins are obviously spherical and mostly soluble in water since they have a nonpolar core and a polar surface, while other proteins differ in shape and solubility and some even prefer lipid environment due to the differences in their amino acid composition [85]. High tertial structure complexity is displayed by globular proteins, as observed by various combinations of α and β elements with the participation of loop regions, while much simpler structural units might constitute other proteins [86]. Today there is vast knowledge available on globular proteins owing to their feasibility to be studied by experimental methods and the accumulation of the research done. On the other hand, non-globular proteins still need a clear definition, classification and understanding which were also stated as the aim of the COST (European Cooperation in Science and Technology) action in 2015 with code BM1405 to which part of this research had contributed. For today, the proteins which do not show the properties owned by the globular proteins, especially the spherical and compact tertiary structure and high solubility in water, are named as non-globular proteins in the literature [87–89].

The interest for non-globular proteins increased in recent years. Because after mining the putative proteins in the genome, many proteins which do not obey the globular characteristics were discovered and many found to be related with challenging diseases such as Alzheimer and cancer [90, 91]. Non-globular proteins

have been overlooked because of being hard to be experimentally scrutinized due to their features such as not forming proper crystals and precipitation in solution. But today there are more experimental and computational tools available along with vast amount of information on proteins to target these proteins. Finally, understanding the dynamics of non-globular proteins might also improve the current knowledge on the link between the protein dynamics and the function.

Non-globular proteins can be vaguely classified into three categories: intrinsically disordered proteins (IDPs), repeat proteins (RPs, encompassing also a large portion of the membrane proteins) and aggregative/fibrous proteins. IDPs are the proteins which do not adopt a defined 3D structure under physiological conditions, and they renovated the structure-function paradigm that assumes a defined structure for functioning of a protein [92, 93]. They mostly function in intracellular signaling in which they serve as promiscuous signal transmitters [94]. IDPs owe their functional characteristics to their different sequence profile [95] and as a result still rugged but relatively flat free energy landscape comparing with the globular proteins [96]. The difference in the energy landscape is caused by the lack of secondary structural restrictions [96]. Repeat proteins are low-complexity proteins that include repeating structural/functional motifs and differ from the globular proteins in their intramolecular interactions. In opposition to the globular proteins, the direct interactions between the residues distant in space are absent in repeat proteins. These biophysical characteristics of the repeat proteins cause them to have different energy landscapes, thermodynamic and kinetic features [90, 97]. Many transmembrane proteins are in this group (e.g. an usher protein PapC that functions in fimbria biogenesis [98, 99]) and the positioning of membrane proteins in a lipid environment makes the study of their dynamics challenging. Aggregative/fibrous proteins are proteins with high tendency to form highly organized structures with large number of proteins take place in the process. For fibrous proteins this organization is functional as in the case of collagen or cytolytic pore toxins such as ClyA [100, 101], while for aggregative proteins it causes dysfunction as happens with $\alpha-$synuclein in Alzheimer [102]. Fibrous proteins have simpler tertiary structure than globular proteins and hence with their highly organized quaternary structures

investigation of their dynamics might require different approaches and/or parameters [85].

In this study, repeat and fibrous proteins along with membrane proteins were studied to improve and broaden the applicability of ANM-LD simulations. Disordered proteins were excluded since they have relatively shallow energy landscape making barrier-jumping irrelevant. The repeat proteins FimD and PapC, fibrous protein ClyA and membrane proteins BtuCD, PglK and MntAB were simulated by ANM-LD simulations, evaluated based on final RMSDs to the target conformation. Yet membrane protein BtuCD was chosen as the case system to exemplify the application of ANM-LD to nonglobular systems in the scope of this thesis. Additionally, c-Src kinase was used as a model globular system to evaluate the proposed advances on the ANM-LD algorithm. These two systems are discussed in more detail below.

2.2 Case Systems

Two proteins were chosen as exemplary systems for evaluation and advancement of ANM-LD simulations. Firstly, BtuCD was chosen as the main representative system of membrane proteins to assess the application of regular ANM-LD simulations on deducing the dynamics of these systems. Then for the advancement of ANM-LD method, algorithmic implementations were first tested on a globular protein named c-Src kinase and then also applied to BtuCD to ensure their usage in membrane embedded systems. The structural and functional properties of these two proteins are discussed in detail below.

2.2.1 A Model Globular System: c-Src Kinase

c-Src kinase functions in cellular signaling by phosphorylating specific tyrosine residues on various cellular targets with vital functions in cell cycle progression, apoptosis and more [103]. It transfers the γ phosphate of ATP to the hydroxyl

oxygen in phenol ring of the relevant tyrosine, releasing an ADP molecule [104]. Due to its role in post-translational modification of various proteins, proper regulation of c-Src kinase activity is crucial for the organism [105]. Indeed, in many types of cancer, the role of overactive c-Src kinase is shown [106]. This reveals the need of understanding the inactivation and activation dynamics of c-Src kinase, defined by the conformational transitions between the inactive and active states of the enzyme, and why the active state is favored upon perturbations such as mutations.

There is a 23.3 Å root mean square deviation (RMSD) difference between the known inactive (2SRC [107]) and active (1Y57 [108]) conformations of c-Src kinase with the most complete structure. Inactive state corresponds to a closed, compact conformation of folded domains, while active state is present in open form (Figure 2.1). The kinase comprises of five domains: SH4, unique, SH3, SH2 and SH1 (kinase) domains. In the available X-ray crystal structures of c-Src kinase only SH3, SH2 and SH1 domains are present (residues 84-533), therefore they are used in this study, with given PDB IDs. Two major differences between these structures are the orientations of C-helix (residues 304-316) and Activation (A)-loop (residues 404-435), which are thought to occur via the hinge residues on the SH2-kinase linker. C-helix rotates inward to contribute to catalytic site formation and A-loop unfolds during the activation (Shukla et al., 2014).

The phosphorylation of c-Src kinase on two different tyrosine sites affects the protein dynamics and function (Shukla et al., 2014). Phosphorylated Tyr527 (chicken) in the C-terminus of the protein interacts with the SH2 domain, resulting in activity loss [109, 110]. Vice versa dephosphorylation of Tyr527, results in the activation of the kinase, since the basal activity of the apo protein is high [111, 112]. On the other hand, phosphorylation of Tyr416 in the activation loop (A-loop) is required for complete kinase activity [113].

Two collective variables are used in the literature to define the free energy landscape on which the conformational changes of c-Src kinase can be summarized. First variable is df, which is defined as the difference in the atomic distances between

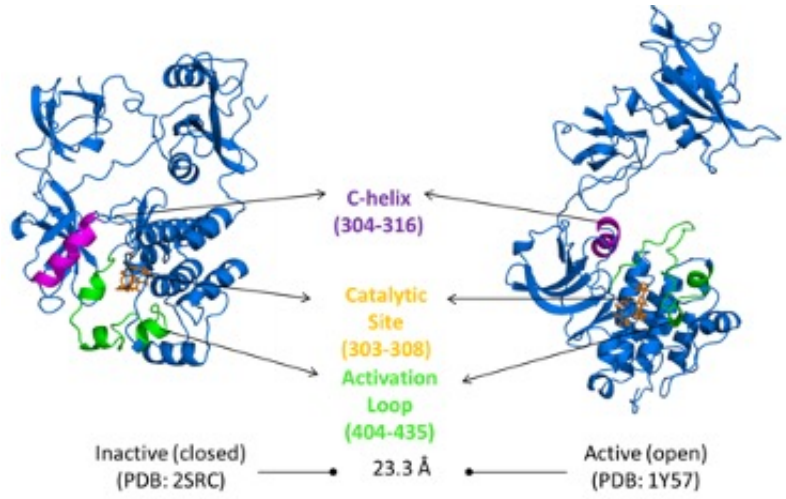

Figure 2.1 Inactive and active conformations of c-Src kinase with 23.3 Å RMSD in between.

E310(CD)-R409(CZ) and E310(CD)-K295(NZ) [114]. This difference corresponds to a shift in hydrogen bonding of E310 residue on C-helix upon activation (changing from R409 to K295) and inactivation (changing from K295 to R409). These residues are located in close proximity of the catalytic site and the change in their hydrogen bonding pattern affects the catalytic site dynamics. Second variable is the RMSD of A-loop, which unfolds through the activation and thus it is a useful variable to assess the visited conformations. In this study, these two variables are used to evaluate obtained ANM-LD transition pathways between inactive and active states.

Mutations on c-Src kinase might change its function, leading to malfunction in the extreme such as lost or constant activities. The mutations R95W, T96I, D117N, W260A and R318Q are known to cause increased activity while K295M and Y527F result in a kinase-defective molecule [115–117]. Additionally, there are 127 amino acid positions with missense single point mutations submitted in COSMIC database [118] related with cancer formation (Table 2.1). These mutation sites are assessed in the perspective of dynamic information obtained from ANM-LD simulations in order to contribute to understanding potential mechanisms for the aberrant functional outcome.

Table 2.1
List of missense single point mutations on c-Src kinase related with cancer formation that were submitted to COSMIC database.

Amino Acid Mutations
A88V, E93V, R95M, E97Q, T98A, S101F, R107Q, I110F, N113D, T114I, E115Q, W118R/C, W119C, H122D, S123L, L124I/P, G130C, P133S, A138T, P139L, S140P, D141N, Q144K, E147K, K152R, R160L/Q/W, E166G, P168L, V174M, E178K, T179S, A183T, S187P, D192N, N193K, K195N, H201L, K203R, R205H, D208N, G211S, F212L, R217C/H/S, V227L, T243S, V244L, T247M, K249N, P250L/Q, T252A, Q253E, A256T, K257N, A259D, P263H/T, S266L, R268Q/W, L273P, Q275E, T289N, T290A, G300D, P304S, A306V, Q309H, M314I/K, L317P, E320D, Q324R, V337I/L, T338A/M/R, E339K, M341I, S342C, G344E, L350I/V, L358M, R359G/Q/W, D365N, Q369E, M374V, V377M, V383I, R385G, R388C, V394M, N397K, L398M, V399L, D404H/N, E415Q, Y416H, A418V, G421C, A422V, P425S/T, W428C, A430T, E432K, L435R, Y436S, V445L, S447F, L451V, T453S, R460L/Q/W, V467A, R469C, D473N, R480W, P484L, P488L, S490F/T, H492Q, L494F, R500Q/W, P503L/S, E504D, E505K, R506Q/W, P507H, T508I, F509L, E510G/K/V, T521M, E524D/K and Q528K.

2.2.2 A Membrane Embedded System - BtuCD

BtuCD is another ABC-transporter that uses ATP hydrolysis to perform and regulate molecule (i.e. vitamin B12) uptake in *E. coli* while being located on the outer membrane. It comprises of two transmembrane BtuC forming the passage pathway at their interface and two cytosolic BtuD monomers (nucleotide binding domain-NBD) including the ATP-binding regions. In order to perform its gating function for vitamin B12 transport, BtuCD has to go through various conformational states coupling dynamics of transmembrane and nucleotide-binding domains allosterically. Starting from apo state (PDB code: 1L7V [119]), some of these states could be captured by X-ray crystallography. ATP-bound holo protein of outward-facing BtuCD (PDB code: 4R9U [120]) is able to bind to BtuF protein that carries the vitamin B12 on periplasmic region of BtuC dimer. Upon the binding of BtuF-vitamin B12 complex protein changes to inward-facing holo conformation (PDB code: 4FI3 [121]). Following the ATP-hydrolysis, the vitamin B12 is released to the cytoplasm and bound BtuF keeps the BtuCD protein in an asymmetrical inward

facing conformation (PDB code: 2QI9 [122]). The unbinding of BtuF results in the outward-facing apoprotein conformation that again might be converted to holo conformation by ATP binding in another cycle of conformational change. Two gate regions cytoplasmic gate I (residues 143, 144, 147, 148) and II (residues 83-85) on BtuC monomers (Figure 2.2) are functional in the regulation of vitamin B12 uptake during these conformational changes by their opening/closing mechanisms. Another key component of the uptake process is the ATP-binding sites I and II since ATP is required for the transport to take place. In this study apo and holo conformations with PDB codes 1L7V and 4R9U were used along with IF protein conformations corresponding to ATP BtuF-bound (4FI3) and BtuF-bound protein (2QI9) and the interchange between these conformations were studied by ANM-LD simulations

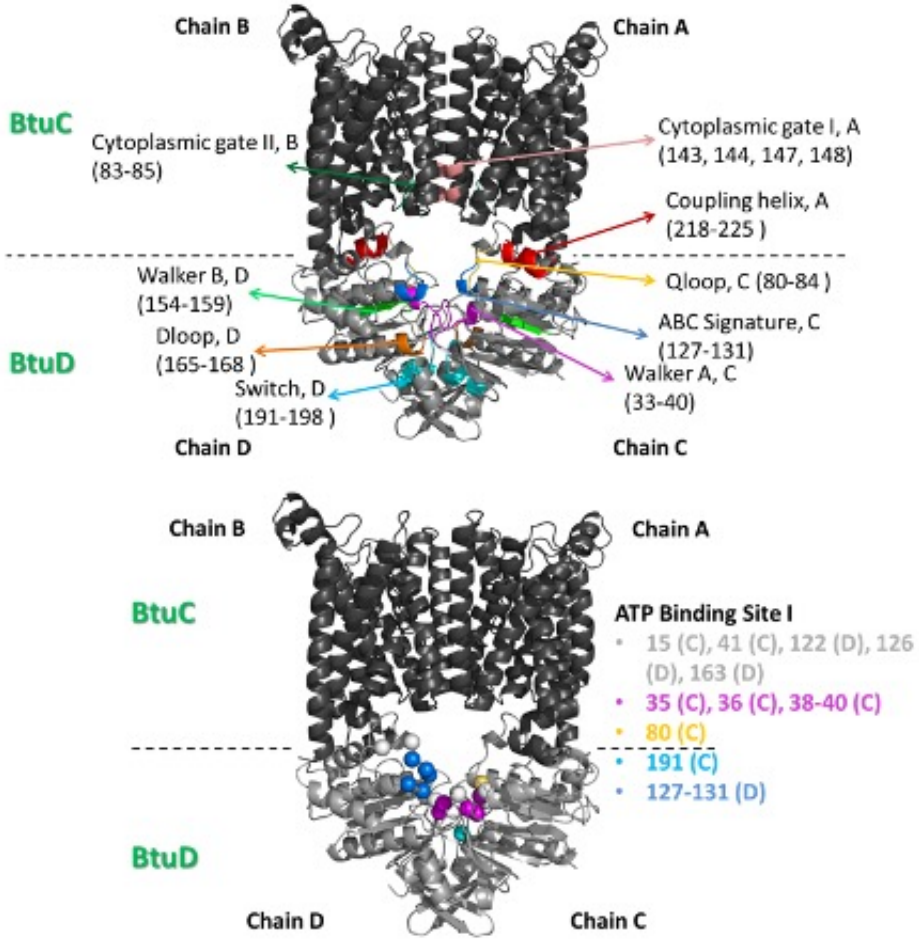

Figure 2.2 Functional regions of BtuCD depicted on its cartoon representation.

3. MATERIALS AND METHODS

In the present work, both computational and experimental methods are utilized due to their complementary aspects in studying protein dynamics and discussed below under successive sections. Following these, the hybrid approaches used to combine the computational methods with experimental restraints are described in the last section.

3.1 Computational Methods

Elastic network models (ENMs), with their computational efficiency and ability to explain internal equilibrium dynamics of proteins imposed by the owned contact topology [123–125] comprise the basis of the computational approach used in this study. Combined with Langevin Dynamics (LD) for the consideration and evaluation of interatomic interactions and related energies [126], ENM-based Anisotropic Network Model (ANM) was used to develop an enhanced sampling method called ANM-LD for sampling transition dynamics between known conformations of a protein [83, 84]. ANM-LD is the main computational method used in this study. Post-processing and analyses were performed on the obtained conformational transitions pathways by ANM-LD. Additionally, various advancing changes on ANM-LD algorithm were performed to elevate the sampling efficiency.

3.1.1 Simulated PDB Structures

Inactive and active conformations of c-Src kinase with PDB codes 2SRC [107] and 1Y57 [108] respectively were studied in the present work. For BtuCD, apo form (PDB ID 1L7V [119]), ATP-bound structure (PDB ID 4R9U [120]), BtuF-bound conformation (PDB ID 2QI9 [122]) and ATP & BtuF-bound configuration (PDB ID 4FI3 [121]) were used for studying its dynamics. For PglK, the apo (PDB ID 5C78 [127]) and ADP-bound (PDB ID 5C73 [127]) structures were simulated.

3.1.2 Elastic Network Models

Elastic network models (ENMs) are simplified representations of protein structures as coarse grained nodes, mostly C_α atoms, connected by harmonic springs if they are within a cutoff distance (R_c = 7-18 Å) of each other [128, 129]. These springs describe the interresidual interactions between neighboring amino acids. The harmonic definition results a quadratic potential energy function for calculating the intramolecular and intermolecular interactions of the system. The model captures the equilibrium fluctuations of the proteins around a minimum energy conformation as dictated by the contact topology and assesses the intrinsic dynamics of the system as a weighted linear combination of orthogonal modes by the use of normal mode analyses (NMA) [130]. NMA-based methods can be used to explore potential movements around an equilibrium conformation by mean of calculating the eigenvectors and eigenvalues associated to different normal modes [131]. Each resulting ENM mode is thus defined by its eigenvector and eigenvalue and represents a global collective, correlated and complex form of motion for the entire protein [132]. Any conformational change around equilibrium can be represented as an eigenvalue-weighted combination of normal modes [133].

Two major types of ENMs, Anisotropic [62] and Gaussian Network Models [61], ANM and GNM respectively, are used in this study and described further in following sections.

3.1.2.1 Anisotropic Network Model (ANM). In ANM, the anisotropic fluctuations of the system are predicted by consideration of the three-dimensional orientations of each interaction between the nodes (i.e. C_α atoms) within a predefined cutoff distance R_c [62] , which equals to 13 Å in this study. A Hessian matrix (H) corresponding to second derivative, the harmonic force constant of the intramolecular potential energy function V with respect to the global coordinates of the system is generated to extract orthogonal modes of motion [62, 134]. The potential function V is defined as follows in which ΔR corresponds to positional fluctuations and γ is the

interaction constant:

$$V = \frac{\gamma}{2} \cdot \Delta R^T \cdot H \cdot \Delta R \tag{3.1}$$

The pseudoinverse of the Hessian matrix (H^{-1}) renders 3N-6 intrinsic modes of motion for a system of N residues, according to the following equation of correlation between fluctuations of residue i and j [62].

$$< \Delta R_i \cdot \Delta R_j >= \frac{3k_B \cdot T}{\gamma} \cdot tr[H^{-1}{}_{ij}] = \frac{3k_B \cdot T}{\gamma} \cdot \Sigma_k tr[\lambda_k^{-1} \cdot U_k \cdot U_k^T]_{ij} \tag{3.2}$$

In Eq. 3.2, k_B is the Boltzmann constant and T is the absolute temperature in degrees Kelvin while λ_k is kth eigenvalue and U_k is kth eigenvector for kth mode of motion. Among the resulting normal modes, high frequency modes describe local motions while low frequency modes represent global conformational changes, due to the direct proportion between mode frequencies and the energy required for the corresponding movement [62, 135]. These low frequency modes, which are termed slow modes, were found to overlap with the observed functional protein motions and to be conserved during evolution as they represent the movements of lowest energetic cost [136, 137].

3.1.2.2 Gaussian Network Model (GNM). In GNM, the predicted equilibrium fluctuations of a protein are assumed to be Gaussian and the interactions between the nodes are coded as either present or absent without adding their orientational information [131]. Along this definition, the Hessian matrix (H) in ANM (Eq. 3.1) is replaced with the Kirchhoff matrix (Γ), and the correlation between ΔR_i and ΔR_j is decomposed into N-1 modes by the following equation.

$$< \Delta R_i \cdot \Delta R_j >= \frac{3k_B \cdot T}{\gamma} \cdot \Gamma^{-1} = \frac{3k_B \cdot T}{\gamma} \cdot \Sigma_k [\lambda_k^{-1} \cdot U_k \cdot U_k^T]_{ij} \tag{3.3}$$

Minima points of mean square fluctuations (MSFs) for a given GNM mode k are defined as the hinge residues enabling the relevant movement of mode k. GNM is more robust in hinge detection than ANM due to its simplistic approach [62, 138] and thus preferred in this study for this aim.

Additionally, GNM can be used to infer directionality and causality in allosteric communication between various regions of protein, based on Schreiber's transfer entropy concept and its GNM derivations [139, 140]. In the basic GNM transfer entropy model, all modes contribute to the calculation of allosteric transfer entropy; yet in the present work, instead of including all the modes, we used only the GNM modes relevant for a functional conformational transition (i.e. GNM modes corresponding to the ANM modes which are popularly used in ANM-LD simulations). This moderates the computational load while focusing on the directionality and causality of allosteric communication related with a specific conformational change [83].

Transfer entropy ($\mathrm{TE}_{ij}(\tau)$) provides an estimate of the direction of information flow between two residues i and j for a certain time delay τ between their fluctuations. This definition represents the degree to which the present movement of residue i decreases the amount of uncertainty for the future movement of residue j after τ time. If, $\mathrm{TE}_{ij}(\tau) > \mathrm{TE}_{ji}(\tau)$, then the dynamics of residue i affect the dynamics of residue j, meaning a causal directional relationship between dynamics of residues i and j. The transfer entropy $\mathrm{TE}_{ij}(\tau)$ from $\Delta R_i(t)$ of residue i at time t to that of $\Delta R_j(t+\tau)$ of residue j at time t+ τ was formulated [139] as:

$$TE_{ij}(\tau) = S(\Delta R_j(t+\tau)|\Delta R_i(t)) - S(\Delta R_j(t+\tau)|\Delta R_i(t), \Delta R_j(t)) \tag{3.4}$$

where the conditional entropies are:

$$S(\Delta R_j(t+\tau)|\Delta R_i(t)) = - < \ln p(\Delta R_i(0), \Delta R_j(\tau)) > + < \ln p(\Delta R_i(0)) > \quad (3.5)$$

$$S(\Delta R_j(t+\tau)|\Delta R_i(t), \Delta R_j(t)) = - < \ln p(\Delta R_i(0), \Delta R_j(\tau)) > + < \ln p(\Delta R_i(0), \Delta R_j(0), \Delta R_j(\tau)) > \quad (3.6)$$

The details of the formulations for time correlation of fluctuations above were given in the work of [139]. Using the equations in the latter reference, we present net entropy transfer $\mathrm{TE}_{ij}(\tau)$-$\mathrm{TE}_{ji}(\tau)$ values whose positive values correspond to i as entropy source while negative values as entropy acceptors.

3.1.3 Langevin Dynamics

Given an instantaneous conformation of a molecule and the initial velocities of its constituent atoms, molecular dynamic simulations numerically solve the classical equations of motion by modeling interatomic interactions using empirical potential energy functions [141–143]. As a result of this defined energy, the moving forces exerted on an atom by the rest of the system is calculated, rendering new positions and velocities at each time step. Due to the detailed atomistic nature of the information provided by MD simulations, the technique is considered gold standard in the computational studies of protein dynamics [74]. Starting with short-time dynamics, as the computational power increases more complex and long dynamic events such as ligand binding, protein activation and protein folding were targeted by MD [144]. Yet, there are still limitations to the application of MD on sampling long-time and large-scale dynamics of proteins due to the substantial computer power

required such as Anton which pushes the current limit of MD simulations to the millisecond range [145].

Another challenge to performing MD simulations is the consideration of solvent atoms explicitly in the equations of motion particularly for the large systems having high number of atoms alone [50]. One solution to overcome this handicap is to introduce a stochastic term to the equation of motion as a representative of random collisions between protein and solvent atoms [146].

$$m_i \cdot \frac{d^2 R_i}{dt^2} = -\frac{dV}{dR_i} - m_i \cdot \gamma_{LD} \cdot \frac{dR_i}{dt} + RF_i(t) \tag{3.7}$$

In LD simulations, the friction/damping coefficient γ_{LD} and random force RF_i whose mean is zero are added to the Newtonian equation of motion to represent solvent-induced random movements of the molecule which also enable energy barrier-crossing motions [146, 147]. In the end, this stochastic component provides an improved sampling of the conformational landscape [148].

In this study, LD, performed in Amber 11-Sander [149] biomolecular simulation package, was integrated into an enhanced simulation method called ANM-LD.

3.1.4 ANM-LD Simulations

To explore the conformational transition pathways connecting functional states of proteins, which is difficult for molecular dynamics, a hybrid molecular simulation method named ANM-LD was designed as an enhanced sampling technique in our laboratory. ANM-LD integrates a low-resolution C_α-based anisotropic network model (ANM) with stochastic all-atom implicit solvent Langevin dynamics (LD) simulations. ANM-LD utilizes the structure's intrinsic dynamics imposed by its contact topology and defined by ANM modes to guide conformational changes toward

a target conformation [83, 84].

The ANM-LD algorithm is an in-house MATLAB program (can be downloaded from **https://github.com/prclab/anmldmutant/**) which iterates cycles of ANM calculations [62] and consecutive call-outs of energy minimization and LD simulation steps in Amber 11-Sander [149] biomolecular simulation package toward a target conformation. An ANM-LD cycle consists of positional perturbations of atoms in selected ANM mode, energy minimization and short time LD simulations of the perturbed conformations. The cycles of ANM-LD simulations do not reflect the real time of the simulated transitions, yet provides information on the transition mechanism and the order of events that cannot be achieved with regular molecular dynamics (MD) simulations [83, 84]. Having accessible internal dynamics as the only bias toward the target conformation enables the determination of physically and biologically relevant, realistic transition pathways [83, 84]. Since utilization of ANM-LD requires that the initial and target conformations include the same number of atoms, the nucleotides and substrate binding proteins are not considered in the regular ANM-LD simulations.

With given initial structure, ANM-LD constructs the Hessian matrix (H) for each current initial structure by setting a distance threshold radius (R_c = 13 Å) to define the neighbor interactions between C_α atoms at each iteration cycle. Then the correlation between atomistic fluctuations is decomposed into 3N-6 ANM modes [62]. Among these, the ANM mode U_{best} that overlaps most with the difference vector between the structurally aligned initial conformation of the current cycle and the target conformation is determined. By the application of a perturbation in the value of a deformation factor (DF = 0.2-1 Å) along U_{best} on the current structure, an orthogonal transformation of the system is done with positional changes ΔR_{ANM} $(DF \cdot U_{best})$. The energy of the perturbed conformation is minimized for 500 steps (steepest descent for the first 50 steps and then conjugate gradient for the rest; using the Amber 11-Sander biomolecular simulation program), resulting in positional changes $\Delta R_{Eng.Min.}$. This is followed by 100 steps of LD simulation with a time step of 0.2 fs (20 fs = 100 steps · 0.2 fs) at temperature T = 310 K with a damping

constant of $\gamma_{LD} = 5\ ps^{-1}$, leading to further structural relaxation with conformational changes of ΔR_{LD}. The resulting conformation (intermediate) undergoes another cycle of ANM-LD, and more iteration follow until the system gets close enough to the goal conformation or when the generated changes become too small. A simplified equation of positional sampling in each ANM-LD cycle is given below as:

$$\Delta R_{new} = \Delta R_{old} + \Delta R_{ANM} + \Delta R_{Eng.Min.} + \Delta R_{LD} \tag{3.8}$$

in which R_{old} represents the atomic positions of structure entering into the ANM-LD cycle and R_{new} corresponds to the positions of output conformation. The ANM-LD method is summarized on Figure 3.1 in general.

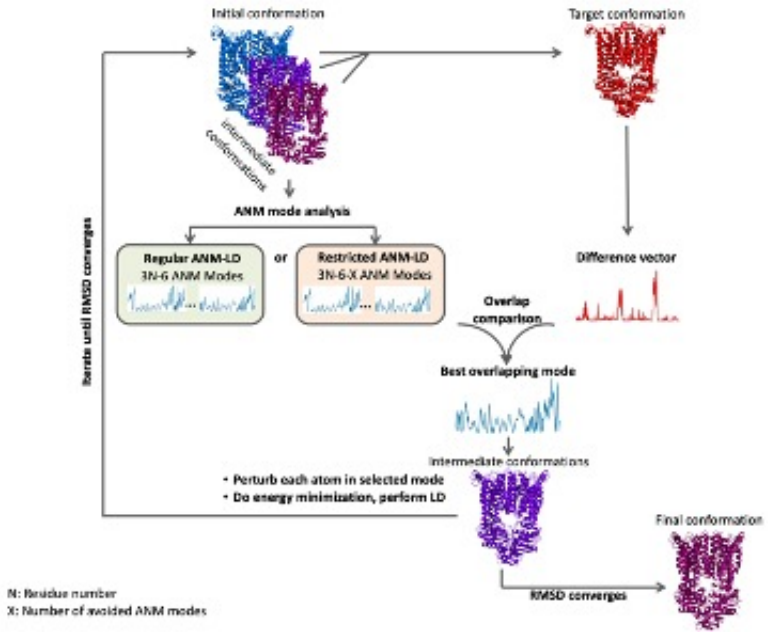

Figure 3.1 Schematic representation of ANM-LD algorithm.

Before starting the ANM-LD simulations, missing residues of initial and target structures are modeled using ModLoop [150] and energetically minimized [149]. After extracting the ANM modes, either all or a set of ANM modes can be used to compare with the difference vector and choose the best overlapping mode to perturb the

structure. The number of slow modes comprising this set of ANM modes is defined by the parameter named modemax. First 30 slow modes are preferably used in the simulations in order to promote the use of global modes in the transition, except for the cases in which more local changes must occur for a transition to be performed. In these cases, modemax=100 or higher values can be used. Once the ANM mode is selected to perturb the structure, another parameter called deformation factor (DF) is called to define the amplitude of the perturbation occurring on the whole protein in terms of RMSD. Based on the size of the conformational difference between initial and target structures DF value is set between 0.1 and 1 Å in proportion to the size. Another parameter γ_{LD} defines the collision frequency (in ps^{-1}) and effective viscosity of the solvent surrounding the protein. For membrane proteins in lipid environments range of 1-5 ps^{-1} are used. Larger values up to the collision frequency of liquid water (50 ps^{-1}) are suitable for cytosolic systems, yet smaller values should be preferred for the sampling quality and stability of integration [151, 152].

3.1.4.1 Evaluation of ANM-LD. ANM-LD simulations were performed for a diversified set of proteins including globular and non-globular proteins. These include Adenylate Kinase (AdK), membrane proteins such as FimD (repeat protein) and ClyA (fibrous protein) (unpublished data), and the ABC transporters of the present study. Starting with RMSDs in range of 3-24 Å between the initial and target conformations depending on the system studied, the target conformation was converged to within an RMSD of 1.4-3.1 Å. Last but not least, the transition trajectory between the open and closed states of Adenylate Kinase (AdK), as determined by ANM-LD, included the known structures of homologous proteins, even though these were not used to guide the simulation (unpublished data). This observation strongly supports the applicability of ANM-LD in identifying conformational transition trajectories that are biologically and physically relevant.

3.1.5 Assessment of ANM-LD Outputs

Two main outputs of an ANM-LD simulation are the conformational trajectory of the sampled transition and the selected ANM mode pool which are used in guiding the simulation. Various dynamic parameters can be extracted from the output trajectory such as covariance-based variables; backbone atomic positional fluctuations (as depicted by mean square fluctuations- MSF) and dynamical cross-correlations between residue pairs. On the other hand, the ANM modes which lead the certain conformational transitions provide additional normal mode based dynamic information. For instance, the network of hinge residues enabling these modes of motion appear as dynamic allosteric network of residues with potential roles in functional mechanism [83]. Both covariance-based and normal mode-based approaches to analyze ANM-LD outputs are discussed below.

3.1.5.1 Covariance-based approaches. MSFs and cross-correlations are two concepts deduced from covariance matrix (c) of residual position vectors (R_i and R_j) obtained from ANM-LD trajectories. The diagonal of the covariance matrix gives the MSFs of the protein, while normalization of the covariance matrix results in the cross-correlation matrix (C) of the studied molecule.

$$c_{ij} =< (R_i - < R_i >)(R_j - < R_j >) >=< R_i \cdot R_j > - < R_i > \cdot < R_j > \tag{3.9}$$

$$C_{ij} = \frac{c_{ij}}{\sqrt{c_{ii} \cdot c_{jj}}} = \frac{< R_i \cdot R_j > - < R_i > \cdot < R_j >}{\sqrt{(< R_i^2 > - < R_i >^2) \cdot (< R_j^2 > - < R_j >^2)}} \tag{3.10}$$

MSF is a measure of how much a residue fluctuates from its mean position during the simulation. In this regard, it provides additional information on mobility

of residues to the difference vector between two stable conformations. For example, if a functional gate is closed in both initial and target conformations, it might still move and visit open states during the simulation before it again arrives the closed state at the target conformation.

The cross-correlation maps disclose the dynamic coupling between any residue pair on the protein. The cross-correlation values range between -1 and 1. Positively correlated residues travel in parallel vectors in space and time, while negatively correlated residues perform anti-parallel movements to each other. The values close to zero, either correspond to persistent orthogonal movements or uncorrelated behavior of two residues. These correlations assume synchronous movements of residue pairs; however, a delay might appear between coupled movements of residue pairs. So, time delay τ can be inserted into the cross-correlation equation as follows:

$$C_{ij}(\tau) = \frac{< R_i(t) \cdot R_j(t+\tau) > - < R_i(t) > \cdot < R_j(t+\tau) >}{\sqrt{(< R_i(t)^2 > - < R_i(t) >^2) \cdot (< R_j(t+\tau)^2 > - < R_j(t+\tau) >^2)}} \quad (3.11)$$

In Eq. 3.11, t represents an instantaneous pseudo-time variable and τ is an additive time delay factor in terms of ANM-LD cycles. When τ=0, the equation is reduced to the well-known time-independent (equal time) correlations ranging from -1 to 1. For τ values larger than 0, the same equation gives time-delayed correlations between R_i at time t, and R_j at later time t+τ. The case of i=j yields the auto-correlations of residues at both equal and delayed times depending on the τ value.

There are two main parameters in definition of the correlation behavior; the time window of the conformational transition (part of the trajectory from which the conformations and positional vectors are extracted) and the time delay factor τ. These correlations thus reflect the cooperativity of residues within a specific time window of the transition pathway, as well as with a certain time delay factor applied in the case of time-delayed behavior. In time-delayed correlations, τ is the time (in

ANM-LD steps) chosen to reflect order of events. Nevertheless, the calculations were repeated for different time windows and time delay values. Time-delayed cross-correlations reveal the co-movements of lead (preceding) and lag (delayed) residues or group of residues. For any two residues i and j, if $C_{ij}\ (\tau) > C_{ji}\ (\tau)$ for a given time delay τ, the fluctuations of residue i precede the fluctuations of residue j. Since the conformational change of i precedes that of j, residue i is likely the driving residue and residue j is the driven one. When examining the time-delayed cross-correlations between functional domains/residues one can infer causality relations, also supported with GNM transfer entropy calculations described above.

3.1.5.2 Normal mode-based approaches. In ANM-LD algorithm, certain ANM modes are repeatedly selected to guide the conformational changes toward the target conformation due to their directing role toward the target. These recurrent modes disclose the movements required for the conformational change relevant for the functional mechanism [83, 84]. In general, determining these ANM modes enables the dissection of complex transition dynamics into a set of simpler modes of motion. These modes arise from the contact topology and allosteric communication of their underlying hinge residues and resulting mechanistic role of these residues causing coupling and decoupling between different regions of the protein [61, 62].

Hinge residues which generate a certain motion of an ANM mode, can be deduced by finding the most similar GNM mode based on the resemblance between their mode shapes (i.e. overlapping minima points). GNM is more robust in hinge detection than ANM [62, 138] and the minima points of a GNM mode shape correspond to the hinge residues. These hinge residues are used to predict allosteric network of residues since they cooperate to form a dynamic behavior [83]. Of the hinge residues identified by GNM analysis, the ones which are evolutionarily conserved are assumed to form a functional allosteric network. To scale the conservation of the hinge residues independent ConSurf [153] calculations are conducted with studied proteins. Homologues with 35% to 95% sequence identity are collected using HMMER [154] with one iteration and an E-value of 0.0001 against the

clean UniProt database. 300 unique sequences are retrieved, and multiply aligned using MAFFT-L-INS-i. The rate of evolution at each site is calculated using the Bayesian method [155] and the LG evolutionary model. For mutational analysis, hinge residues that displayed the highest ConSurf conservation score (8 or 9 on a 1-9 scale) are chosen.

3.1.6 Assessing Mutation Outcomes *In Silico* by Perturbation Response Analysis

The mutation of a residue which contributes to an allosteric network with role in functional mechanism is expected to change the dynamics and functionality of the protein [156]. Thus, experimental methods such as site-directed mutagenesis and functional assays are useful in assessment of the residues' role in function via exploring effects of its mutation on the protein. However, as protein size increases the number of target residues for mutation is likely to increase, not to mention the other nineteen amino acids the target residue can be converted. This causes high costs in material and time. Therefore, preliminary evaluation of mutation outcomes by computational methods is valuable in reducing these costs by reducing the targeted residues. In this project a perturbation response analysis based on ANM-LD simulation method is proposed, whose details are described below.

In the proposed approach, *in silico* mutations are generated with PyMOL 1.3 [157] in both initial and target conformations. The mutant structures are then energy minimized using the Amber 11-Sander biomolecular simulation package [149], and the resulting initial and final mutant conformations were used to perform ANM-LD simulations.

The resulting mutant ANM-LD transition pathways are then assessed with respect to the sampled bond/backbone rotation angles (ϕ) of hinge residues (and their contact residues) that enable the movement performed by the mostly used ANM mode. To quantify a mutant's deviation (from WT), a cumulative angular deviation

score (AD) is calculated (3.12) as the sum of absolute differences between the mutant and mean WT (μWT) angles over each frame and each hinge of the relevant ANM mode (and their immediate neighbors in either sequence or 3D space). Large AD scores indicate that the mutation is likely to cause a relatively larger perturbation, hence a good candidate for experimental evaluations.

$$AD(mutant) = \Sigma_{\substack{\text{ANM1} \\ \text{hinges}}} \Sigma_{\text{frame}} |(\phi_{\text{mutant}}(resno, frame) - \mu WT(resno, frame))| \quad (3.12)$$

3.1.7 Advances in ANM-LD Algorithm

As part of this thesis, various advancements are proposed on ANM-LD algorithm in order to improve its sampling capacity. These changes are focused on the mode selection format, representation of the studied system and combination with experimental restraints. On these aspects the performed amendments on ANM-LD algorithm are summarized below.

3.1.7.1 Restricted ANM-LD Simulations. ANM-LD simulations are able to identify intrinsic modes of motion that are key for a given conformational transition. To evaluate the contribution of a specific mode of motion to the conformational transition, a restricted ANM-LD simulation protocol is proposed. In this restricted protocol, a specific ANM mode (or set of modes) is disallowed if the overlap (dot product) between this mode of motion and any calculated mode at each cycle is above a certain threshold t_{rest}. In this study t_{rest} is used as 0.3.

In summary, in a restricted ANM-LD simulation, an ANM mode of interest is blocked, and the effects of this blocking on the dynamics of the transition are evaluated. Such restrictions may lead to alternative pathways of conformational transitions, and in the case of an irreplaceable mode of motion, a particular restriction will impede the

conformational transition.

3.1.7.2 Monte Carlo-Based Mode Selection. Using solely best overlapping mode with the difference vector might be a factor for poor sampling of the energy landscape, since it might oblige a shortcut toward the target conformation. Therefore, a Monte Carlo based mode selection algorithm was added to ANM-LD simulations in order to introduce stochasticity into the perturbation direction of the system. In this algorithm, the ANM mode to be used for perturbation of the structure among all available modes is selected based on a probability distribution proportioned by its overlap with the difference vector. So, if a random variable r_{MC} corresponds to the range of probability value for an ANM mode, this mode will be selected to perturb the structure. An additive approach to this is to combine Monte Carlo based sampling with the best overlap algorithm in consecutive or intermittent ANM-LD cycles.

3.1.7.3 All-Atom ANM-LD Simulations. Regular ANM-LD algorithm uses C_α atoms as nodes of the elastic network to define protein conformation. A more detailed depiction of this network is possible, in which other backbone and side chain heavy atoms might be treated as nodes. This is especially useful if the resulting trajectories will be assessed based on an all-atom parameter such as presence/absence of salt bridges and H bonds, since all-atom approach might be more suitable to provide side chain movements and rotations enabling the relevant dynamic behavior. Thus, each heavy atom from both backbones and side chains of the protein residues is taken as a node of the conformational network for ANM calculations in the present work to refine the transition trajectory. The nodes within $R_c = 5$ Å are thought to be interacting with each other. The rest of the calculations are similar with regular ANM calculations and following energy minimization and LD steps [62].

3.1.7.4 Ligand-Bound ANM-LD Simulations. The initial and target conformations used in regular ANM-LD simulations are directly extracted from ligand-bound states (if necessary) without the presence of ligand molecules and any change on the structure except energy minimization. Although the functionally important conformational ensemble of the protein is already assumed to pre-exist in the ligand-free protein, the presence of ligands might stabilize certain conformations by increasing their probability in the conformational ensemble [2, 27, 158] obtained by a very short LD in each ANM-LD cycle, accumulating the effects towards the end of the simulation. Besides, during the minimization and LD steps, the energetics of the sampled intermediate conformations might be improved with the explicit presence of the ligand and its interactions with relevant amino acid side chains. Additionally, the ligand effect in the obtained conformation after undergoing an LD part might cause the selection of different ANM mode in the following ANM-LD cycle, hence affecting resulting conformational trajectory. For non-targeted ANM-LD simulations in which only some collective variable describing the movement of ligand is used, the effect of explicit ligand presence might be even more important for narrowing the sampled path and hence easing the functional conformation change while requiring a smaller number of parallel simulations.

Based on its configuration, ligand molecule will be represented as interacting nodes in ANM calculation in addition to protein nodes and full-atom representation in minimization and LD steps.

3.1.8 Visualization

All PDB-based protein figures were created using PyMOL 1.3 [157].

3.2 Experimental Methods

In this thesis, observables from two experimental methods used in studying protein structure and dynamics are exploited to assist the ANM-LD simulations. These are two different techniques with distinct working mechanisms: atomic force microscopy (AFM) and small-angle X-ray scattering (SAXS). Both methods are summarized below.

3.2.1 Atomic Force Microscopy (AFM)

Unlike other conventional microscopy techniques AFM does not use supplements such as light or electrons to be reflected by the surface in order to operate and obtain picture but uses a probe to sample and physically sense the surface in three dimensions. The properties of the scanning probe (tip) determine the resolution of the image. The cantilever, an apparatus which the tip is mounted to, reflects the beam obtained from laser to a photo diode to be detected. A piezo scanner, which underlies the sample, is used for the positioning of the sample and the tip. In this way the tip can create a 3D map and finally an image of the system being examined [159]. In addition to imaging of biomolecules, the mechanistic features or interactions of these systems with other molecules can be examined in experimental setups such as pulling experiments. In these experiments the biomolecules are attached to tip and/or surface and as the tip is elevated, the resulting force is measured in the events of protein unfolding or breaking of a protein-protein interaction [159]. The most advantageous features of AFM for studying biomolecules are its availability to be performed under physiological conditions for the system to be studied and its applicability at single molecule level [160] while other methods only provide ensemble averaged results. This is also the reason of the choice of AFM for studies on non-globular proteins as demonstrated by the studies on intrinsically disordered α-synuclein [161], aggregation behavior of amyloid beta (Aβ) proteins [162] and membrane proteins [160]. On the other hand, the system of interest should be obtained in purified state to be scanned by AFM which might stand as a complication depending on the system of interest.

In general, AFM experiments support protein dynamics studies by providing 3D images of stable or transition structures of target proteins. In case of manipulation of the system such as induced mutation or temperature change it is also possible to observe the effect of perturbation on binding characteristics and/or flexibility of the target system by pulling experiments in addition to capturing structural differences.

3.2.2 Small Angle X-ray Scattering (SAXS)

Small angle X-ray scattering (SAXS) is a biophysical characterization method used to study the general shape and structures of biological macromolecules in solution. SAXS provides low-resolution information about the shape, structure and complex state of proteins, nucleic acids, and various macromolecular complexes [163]. Experiments can be carried out in solution without the need for crystallization, which allows for experiments in different pH environments and concentrations of molecules to provide useful information about the structure-function relationship [164].

The setup of an SAXS experiment is conceptually simple: a particle solution, usually placed in a quartz capillary, is illuminated by an aggregated monochromatic X-ray beam, with the intensity of the scattered X-rays recorded by an X-ray detector [165]. The scatter pattern of the pure solvent is also collected and subtracted from the sample (blank) solution scatter, leaving only the signal from the relevant particles [166]. The scattering pattern obtained is related to the general shape and size of the particles studied. Many characteristic parameters such as molecular weight, particle volume, maximum size ($D_{\max}$) and gyration radius (R_g) of the sample under investigation can be obtained from the experimental scattering model [167].

The CRYSOL program can predict SAXS solution scatter profiles from known atomic structures of macromolecules [168]. As an input to the program, a PDB file with an X-ray or NMR structure of a biomolecule can be used. This is useful when an experimental SAXS profile is available and need to be evaluated. The program uses multipole expansion of scattering amplitude and considers the hydration shell to

calculate the spherical mean scattering pattern. CRYSOL can fit the theoretical scattering curve by evaluating the SAXS experimental data and minimizing the discrepancy (χ^2 value, chi-square). This fitting is accomplished by changing three parameters: (1) the average displaced solvent volume per atom group, (2) the contrast of the hydration shell, (3) the relative background [168]. Within the scope of this project, not the experimental data but CRYSOL-predicted SAXS profiles are used to lead the ANM-LD simulations, as a preliminary study.

3.3 Integration of ANM-LD Simulations with Experimental Restraints

As described earlier, both AFM and SAXS observables were used as experimental restraints to guide and improve sampling of ANM-LD simulations. Both integration approaches are discussed below.

3.3.1 Integration with SAXS Data

Both volume and SAXS scatter profiles can be used as collective variables extracted from SAXS data to guide the ANM-LD simulations. CRYSOL [168] algorithm was used to predict SAXS scatter profiles and volume data out of X-ray crystal structures in this study. For volume to be applied as a leading collective variable, the ANM modes which increase or decrease the volume according to the volume change between current and target conformations are chosen. For the application of SAXS scatter profiles, the ANM modes which decrease the χ^2 between the resulting perturbed conformation and the target conformations are selected during the simulation. Both variables can be used in combination with the best overlap decision criterion on mode selection.

3.3.2 Integration with AFM Restraints

Despite the variability of observables from AFM experiments, such as interaction and unfolding forces [169], protein flexibility [170], the topographic images are focus of this study as the source of the obtained restraints. In the present work, the topological AFM images of BtuCD and c-Src kinase proteins were determined by our group (in collaboration with Hamdi Torun, Ph.D.), using AFM contact mode of Dimension Edge, Bruker Nano (Santa Barbara, CA, USA) with Bruker's Sharp Microlever (MSNL-10) probes, later to be applied in developing hybrid ANM-LD simulations combined with experimental restraints. Both physical adsorption and 2.5% glutaraldehyde application were performed for the attachment of proteins to the mica surface. The PID parameters were optimized before each scan by controlling and minimizing the oscillation of the system. The images were processed in Gwyddion [171] software.

In order to interpret the prospective AFM images, potential positionings of BtuCD-F protein complex (PDB code: 2QI9 [122]) and c-Src kinase (PDB codes 2SRC [107] and 1Y57 [108] on mica surface were evaluated based on the charge of surface residues. The residue charges were determined by calculation of electrostatic potentials in kcal/(mol.e) using PBEQ Solver [172].

Depending on the protein of study, different solvents were used, and distinct preliminary preparations were performed. For a complete tetramer BtuCD, BtuC protein His-tagged on its N-terminal and BtuD protein are mixed to form a complex structure in FOS-12 detergent. The resulting mixture in concentrations of 0.25-1mg/ml range was incubated on the freshly cleaved mica for 30 minutes, in separate experiments. Height profiles of the protein are obtained with and without its capping protein BtuF in 0.25 mg/ml concentration.

For c-Src kinase, 0.25-1 mg/ml protein solutions were prepared in 50 mM Tris buffer including 100mM NaCl (pH 8.0). The solution was incubated on the freshly cleaved mica for 30 minutes. Height profiles of the protein are obtained with and without 1 mM ATP.

After the topographic images are obtained, they were used to extract relevant collective variables for guiding the ANM-LD algorithm. Since the x and y axes of the images are relatively low quality due to tip broadening, principal axes of the obtained images were defined as the collective variables to maximize the information obtained from these images. For this purpose, principal component analyses (PCA) of the obtained images of BtuF-bound BtuCD and ATP-bound c-Src kinase were performed and first three principal axes were determined. These axes were then used to guide the simulations from apo conformations for both BtuCD and c-Src kinase by choosing the most overlapping ANM modes to these principal axes.

4. EVALUATION OF POTENTIAL ANM-LD APPLICATIONS ON MEMBRANE PROTEIN DYNAMICS

Membrane proteins encompass 15 to 39 % of human proteome according to the predictions based on the unique physicochemical characteristics of their residue composition, such as hydrophobicity, charge and secondary structural propensity toward a-helix [173]. Among those, only around 15% have their structure determined by experiments [174–176]. Nevertheless, membrane proteins perform vital cellular functions such as molecular transfer and are particularly important to sense the environmental cues and production of the relevant response. Thus, understanding their dynamic properties serve to different purposes, such as exploration and manipulation of cellular signaling, drug development and more. Nonglobular topology and high hydrophobicity characteristics of the membrane proteins along with their lipid-based native environment, make them a tough target for protein dynamics studies. Most experimental techniques such as NMR and X-ray crystallography exploit the solubility of examined biomolecules in water, which is difficult for membrane proteins. On the other hand, computational sampling methods have difficulty mimicking the effects of surrounding lipid environment on protein dynamics and long-time scaled activation/inactivation dynamics of membrane proteins, due to limitations on the available computational power. These factors demand enhanced sampling methods to target membrane protein dynamics. In this chapter, ANM-LD was evaluated as a potential enhanced sampling method to study transition dynamics between known energetically stable conformations of membrane proteins. Having applied to various membrane systems (e.g. ClyA, FimD) with large conformational changes up to 30 Å between their stable states and resulted a conformational trajectory reaching the target conformation using only internal dynamics of the system as guidance, ANM-LD is a promising tool. In this work, ANM-LD results were particularly evaluated on ATP-binding cassette transporters MalFGK_2, PglK and BtuCD, with a more elaborate focus on the latter.

4.1 ANM-LD Renders Biologically and Physically Meaningful Dynamic Information

BtuCD is a membrane-embedded transporter, responsible for ATP-dependent intake of vitamin B12 in *E. coli* cells. Its functional form is a tetramer consisting of 1152 amino acid residues with two homodimers corresponding to transmembrane domain (TMD) and nucleotide binding domain (NBD) whose constituent monomers are made of 326 and 249 residues, respectively. TMD takes the vitamin from a substrate binding protein (SBP) named BtuF and transfers it to the cytoplasm. This transfer is ATP-dependent and requires a coupling between TMD and NBD dynamics [177]. These dynamics alter conformations between outward (periplasm) and inward (cytoplasm) facing states of the protein, while being subjected to binding and release of SBP and nucleotides and vitamin B12 transfer into the cell [178]. Thus, the transition dynamics between different states are of importance to understand B12 transport mechanism.

Studying the transition dynamics of BtuCD between its stable functional states is difficult by conventional Molecular Dynamics (MD) which can reach 200-μs duration in maximum in contrast to hundreds of milliseconds required for the real-life conformational changes of BtuCD [179]. ANM-LD serves as a potential tool to target this problem, since enhanced sampling of the conformational space is achieved through perturbations in the ANM mode directions related with fluctuations leading to the target conformation, despite the explicit time definition is lacked. Additionally, the damping effect of lipid environment on the protein dynamics is considered by using a small collision frequency value ($\gamma_{LD} \leq 5\ ps^{-1}$ while collisions of water molecules are represented by $\gamma_{LD} \approx 50 ps^{-1}$) for the implicit solvent-based energetic calculations in LD part of the ANM-LD algorithm. So, although transition rates cannot be determined, a trajectory of conformations between distinct functional and/or energetically stable states can be obtained by which the order of events that underlie the transition pathway (s) could be followed. Using the intermediate —relatively short lived and less stable— conformations, many dynamic properties can be extracted such as the progress of root mean square deviation (RMSD) from the

initial to the target conformation, the residue mobilities and their dynamic coupling, represented by mean square fluctuations (MSF) and dynamic cross-correlation maps (DCCM) respectively.

In the present work, the dynamics of BtuCD were simulated by ANM-LD in four different transitions: 1. Apo (PDB code: 1L7V [119]) to ATP-bound (PDB code: 4R9U, [120]), 2. Apo to BtuF-bound (PDB code: 2QI9 [122]), 3. Apo to ATP BtuF-bound (PDB code: 4FI3 [121]) 4. ATP-bound to ATP BtuF-bound conformations. Starting from RMSDs of up to 3.7 Å between the initial and target conformations, the RMSD between the reached final conformation (of the simulation) and the target conformation were in the range of 1.0-1.5 Å in these simulations (Figure 4.1), indicating to a sufficient approach to the target.

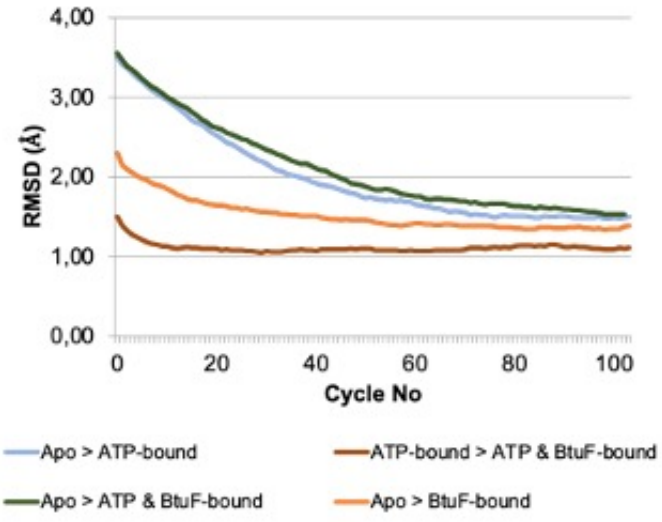

Figure 4.1 RMSDs from the target of four different conformational transitions in BtuCD: apo to ATP-bound, apo to ATP & BtuF-bound, apo to BtuF-bound and ATP-bound to ATP & BtuF-bound conformations.

The resulting conformational trajectories were firstly used to calculate MSFs for each residue of BtuCD, in all simulated conformational transitions. MSF shows how much a residue fluctuates during a certain time window of the simulation resulting a non-equilibrium dynamic parameter. While high-fluctuating residues are corresponding to major dynamic contributors to the relevant transition, low-fluctuating residues indicate to the region with underlying mechanistic roles in the transition. The equation of MSFs is given in Material and Methods section (Eq.

3.9).

MSF values disclose that the docking of BtuF to apo BtuCD has almost no effect on the mobility of residues on NBDs and mainly mobilizes the periplasmic gate residues and surrounding BtuF-binding sites (Figure 4.2). These findings are supported by the experimental data demonstrating that the docking of BtuF induces conformational changes mostly at the periplasmic side of the TMD [180, 181], and has a small effect on ATP hydrolysis rates [179, 182]. On the other hand, ATP binding motions increase the mobility of a wide range of residues located both on NBDs (including ATP sites) and TMDs (particularly cytoplasmic parts of helices 4, 5, 7, 8, and 10 yet also BtuF sites). This observation is consistent with the EPR/DEER and single-molecule spectroscopic measurements underscoring the global effect of ATP binding [84, 180, 181]. On the other hand, when BtuF docks to ATP-bound BtuCD, it further advances the mobility in BtuF binding and the periplasmic gate residues. When it binds to apo conformation along with ATP, a similar augmenting effect is revealed by the extended and increased mobilities of TMD residues induced by ATP binding. The consistency between experimentally determined dynamic features of BtuCD and MSF results from ANM-LD simulations support the applicability of ANM-LD to extract physically and biologically meaningful dynamic information.

Another dynamic feature to be obtained from ANM-LD trajectories is dynamic coupling between various residue fluctuations, as defined by DCCM of the conformational trajectory corresponding to a virtual time window (i.e. number of ANM-LD cycles). For equilibrium dynamics, the DCCM analysis is a widely used method, however just like MSFs extracted from ANM-LD trajectories, DCCMs calculated here correspond to non-equilibrium dynamic behavior of the protein for the relevant conformational transition. The extracted information from non-equilibrium dynamics is useful particularly for studying anisotropic aspects of protein dynamics such as energy transduction [80].

The four simulated transitions give opportunity to study the required communication between different regions of BtuCD to perform mechanistic steps

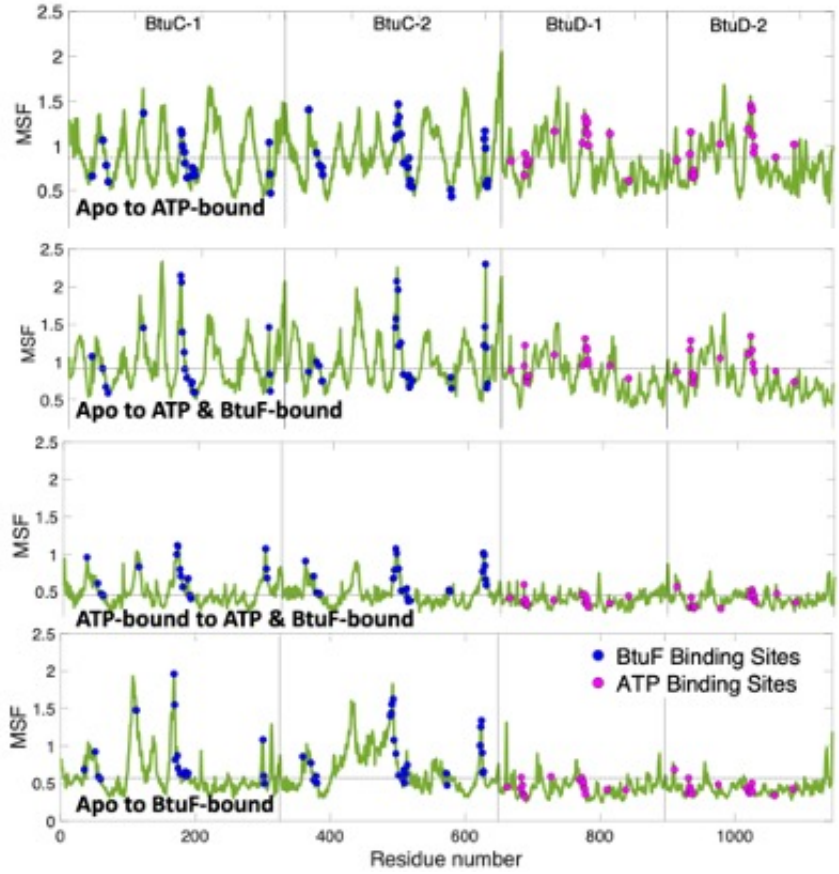

Figure 4.2 MSFs of BtuCD residues during four different conformational transitions: apo to ATP-bound, apo to ATP BtuF-bound, apo to BtuF-bound and ATP-bound to ATP BtuF-bound conformations.

following BtuF and ATP binding events by DCCM analysis. As an ABC transporter, allosteric communication and information transfer between distant BtuCD domains are crucial, regarding the nucleotide state of the NBD -i.e. bound or unbound, and BtuF state of the TMD - i.e. docked or undocked [183–185]. This allosteric communication is essential for the co-orchestration of the B12 transport by local and global conformational changes related with BtuF docking, ATP hydrolysis, opening and closing of gates and substrate release.

As depicted on Eq. 3.10, the positional deviation vectors from mean coordinates of each residue (as obtained from ANM-LD simulation trajectory for a virtual time window) were used to calculate DCCMs of residual fluctuations. The resulting cross-correlations range between -1 and 1 after normalization with multiplication of standard deviations in residue fluctuations, indicating to maxima of

negative and positive correlations, respectively. Positively correlated sites indicate to parallel and synchronous movements of two residues, while negatively correlated sites indicate to concurrent antiparallel movement.

DCCM of apo to BtuF-bound transition (full trajectory) discloses weak correlations among the protein upon BtuF binding Figure 4.3. This suggests that BtuF docking does not trigger long-range allosteric interactions in the transporter, inducing only a local coupling among BtuF sites. On the contrary, binding of ATP sets off strong global coupling (both positive and negative correlation of movements) in both the TMD and NBD. These correlations correspond to both intradomain and interdomain coupling in BtuCD (i.e., between BtuC-1 and BtuD-1 and between BtuC-2 and BtuD-2). Besides these global effects, binding of ATP also causes the strongest dynamic coupling between specific distant functional regions (i.e. high correlations between ATP binding residues and the BtuF-docking site, and between the ATP binding sites and the cytoplasmic gates and coupling helices). According to the presently accepted working mechanisms of BtuCD (ter Beek et al., 2014; Locher, 2016; Lewinson and Livnat-Levanon, 2017; Yang et al., 2018) these allosteric communication between the functional sites is expected. The disclosure of these prominent allosteric interactions between experimentally identified functional sites by ANM-LD without any previous knowledge other than crystal structures attest to the applicability of the ANM-LD approach on membrane proteins.

While BtuF alone has a local effect on protein dynamics, it works as a tuner on allosteric coupling accompanying ATP. This is observed in apo to ATP/BtuF bound simulations by the increased intradomain coupling mainly in TMDs and slightly in NBDs and also by high interdomain cooperativity between TMD and NBDs, extending to the periplasmic BtuC sites in comparison with the correlations observed in apo to ATP-bound simulations (Figure 4.3, black rectangles). The presence of BtuF together with ATP has also a prominent effect on the translocation pore dynamics. While ATP alone is bound to BtuCD, the movement of TMD helices 5a and 5b is positively correlated in space and time (i.e. at the same ANM-LD cycles), indicating to a synchronous parallel movement (Figure 4.3, black ellipses).

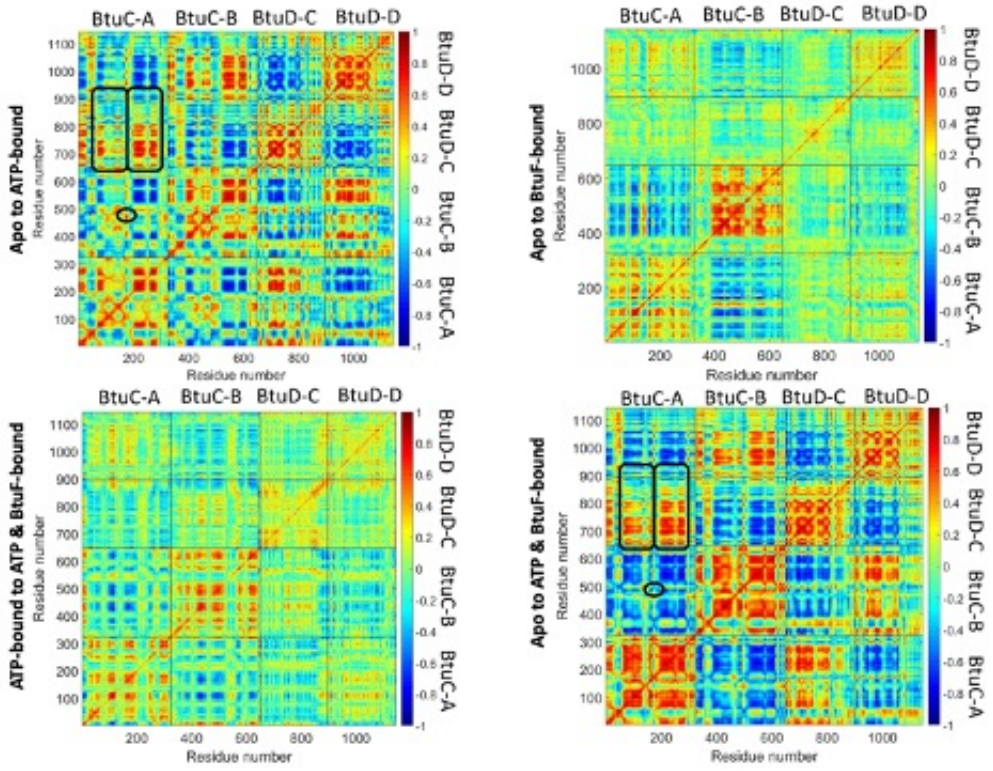

Figure 4.3 DCCM of BtuCD residues from four different conformational transitions: apo to ATP-bound, apo to ATP & BtuF-bound, apo to BtuF-bound and ATP-bound to ATP & BtuF-bound conformations.

This co-movement does not allow the opening of translocation cavity. On the other hand, the presence of BtuF contribute to the appearance of negative correlations between TMD helices 5a and 5b (Figure 4.3). This antiparallel (yet synchronized) correlation of movements of helices 5a and 5b enables opening of the translocation pathway to allow the passage of vitamin B12.

Consequently, both MSF and dynamic cross correlation results show that ATP binding is the major player of protein dynamics in BtuCD, while BtuF acts as a "fine-tuner", augmenting the effects of ATP, and has specific local effects that are essential to the transport cycle and substrate extraction. All these findings are in agreement with the experimental knowledge on BtuCD's working mechanism and support the potential of ANM-LD as a useful enhanced sampling method for membrane embedded systems.

4.2 ANM-LD Allows Deciphering Order of Events via Observed Transient Conformations

The uptake of vitamin B12 into the cell by BtuCD necessitates an initial step; the opening of the periplasmic gate in order to fetch the vitamin from docked BtuF into the translocation pore. The periplasmic gate consists of L172 residue pairs from both BtuC monomers and blocks the entrance due to the steric clash with the vitamin. All available crystal structures of BtuCD –for apo (OF), ATP-bound (OF), BtuF-bound (IF) and ATP & BtuF-bound (IF) states– have closed periplasmic gate. This observation raises the question that if all these energetically stable conformations are closed to B12 entrance, then by which mechanism the vitamin B12 is delivered to the translocation cavity from BtuF. A potential answer would be a transient opening of the L172 gate and a following re-closure after B12 is taken inside. Therefore, the deduction of transient conformations between these stable structures is important. Intermediate conformations obtained from ANM-LD simulations of above mentioned four transitions were explored in this aspect. From the ANM-LD simulations, the distance between the L172 gate residues was extracted (Figure 4.4). As these distances indicated, BtuF binding leads to the further closure of L172 gate, whereas ATP binding causes it to open. However, when both bind simultaneously as in 3rd transition, the two residues move apart after ∼20th cycle and remain open for ∼30 cycles before rapidly closing back to their initial positions. These observed distances suggest that during the transition from the apo to the ATP BtuF-bound state, ATP binding drives the transient opening of the periplasmic gate (to allow delivery of vitamin B12) and BtuF drives its re-closure (to prevent escape of vitamin B12) [84]. Consequently, it was shown that the transitional conformations with open periplasmic gate can be deduced using ANM-LD method.

After taking vitamin B12 into the translocation cavity, the vitamin is pushed further inwards by a squeezing mechanism. In order to delve to this mechanism and disclose the contributing dynamic steps, the sequence of events that occur during the transition from the apo state to the BtuF/ATP-bound state was studied. Four mirror

residue pairs were chosen as representatives of upper TMDs, bottom TMDs, NBDs and cytoplasmic gates, Q109, R138, S67 and S143 residue pairs respectively. The C_α-C_α distances of these residue pairs were monitored throughout the transition (Figure 4.4). Among these distances, Q109 pair initiates the conformational changes with a steep decrease corresponding to the closing of upper TMD possibly due to BtuF docking. This is followed by S67 pair drawing together after 10 cycles as an indication of NBDs closing due to ATP binding. At 20th cycle, while these movement continue, the L172 residues started to move apart, allowing access of vitamin B12. After 20 cycles, the bottom TMD represented by R138 starts widening accompanying with the cytoplasmic gate opening. As demonstrated by the behavior of TM 5 helices, by cycle 35 the translocation cavity was wide enough to accommodate vitamin B12. This expansion happens along with the closing of upper TMD (Q109 residues had already closed by $\sim$8 Å), possibly locking B12 in the protein, since bottom TMD residue R138 and S143 moved very little. The sequential opening and then closing of TM 5 helices surrounding the translocation cavity from upper TMD to bottom TMD, generates a squeezing movement pushing B12 to the cytoplasm. It is important to note that these dynamics are observed only when both ATP and BtuF are present, as shown by the distances in other three transitions (Figure 4.4). The observed dynamics are supported by the observed FRET (Förster resonance energy transfer) intensities of the same residue pairs upon ATP and BtuF binding (by Lewinson group) which correspond to their remoteness from each other [84]. In the end the resulting order of events demonstrates the mechanistic steps required for ATP-orchestrated B12 transport from highly affinitive BtuF, through a translocation cavity to the cytoplasm.

4.3 ANM-LD Enables Dissection of Complex Transition Dynamics into Elemental Normal Modes

During an ANM-LD simulation, certain modes from an available pool of slow ANM modes is utilized to guide the conformation transition from an initial conformation toward a target conformation. The repetitively selected modes

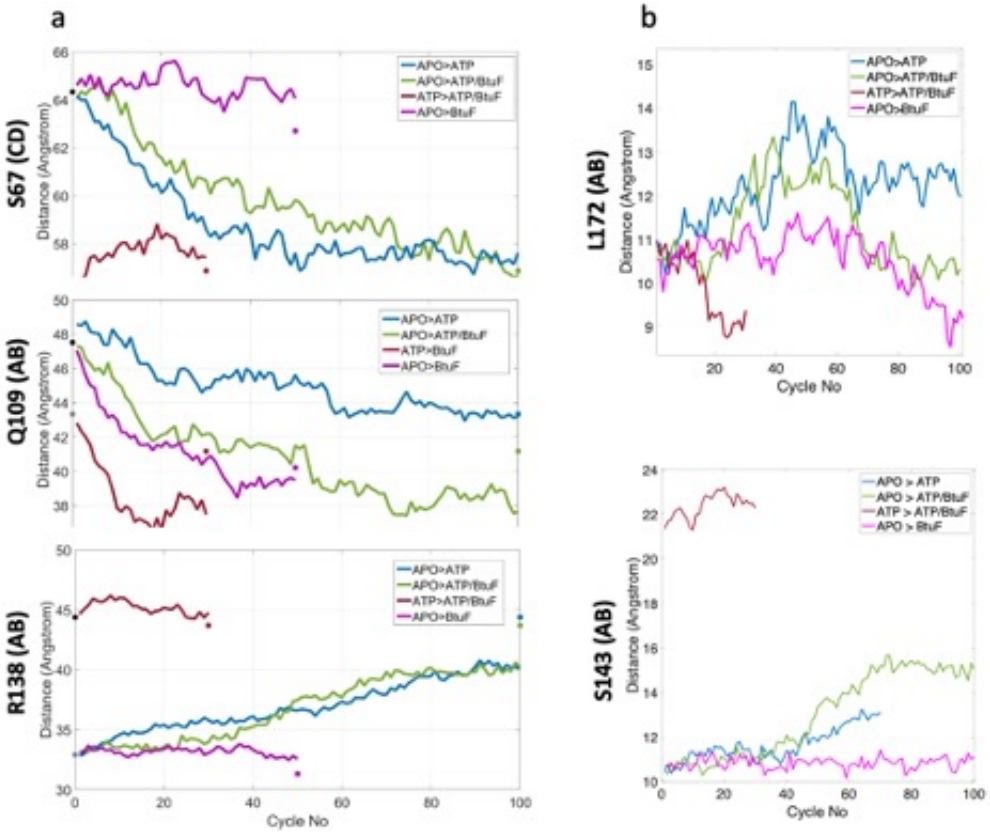

Figure 4.4 The distances between the pairs of (a) S67 (CD), Q109 (AB) and R138 (AB) (b) L172 (AB) and S143 (AB) for all four transitions.

correspond to the specific global movements in BtuCD that enable the conformational transitions hence substrate delivery in real life. In parallel simulations of apo to the ATP BtuF-bound transition, four slow ANM modes (ANM modes 1, 3, 6 and 10) were repeatedly observed. ANM mode 1 corresponds to the twisting movement of the protein with anti-parallel rotation of the NBD and the TMD dimers. ANM mode 3 is responsible of a motion of the periplasmic region that leads to opening of the translocation channel from periplasm. ANM mode 6 corresponds to an expansion and contraction movement of BtuCD along the vertical axis, comprising the closing and opening of the NBDs, opening/closing of cytoplasmic gates. In this mode of movement, the L172 pair moves up in parallel, increasing the steric clash with BtuF-bound vitamin B12 and abolishing its binding site thus enabling the dismantling of B12 from BtuF. Finally, ANM mode 10 couples the closure of the NBD with the opening of the periplasmic gate and closure of the cytoplasmic gate. These ANM modes make above-discussed order of events related with the functional

mechanism possible.

In apo to ATP-bound transition, these four modes are also observed yet in lower selection frequencies. Additionally, another mode - ANM mode 7 - appears in this transition to be used frequently. In this mode of movement, the L172 pair moves in an antiparallel manner, leading to the opening of the gate accompanying with the closing of NBDs indicating to the preparation of periplasmic site for B12 extraction in response to the ATP binding (in contrast with ANM mode 6 observed in apo to ATP & BtuF-bound transition). Thus, this mode of motion couples ATP binding with periplasmic movement. Another important aspect of this movement is the downward and outward tilt of the periplasmic loop that connects TM helices 2 and 3 which prepares the docking site for BtuF's helices 3 and 8 [122]. Combined, all these modes of movements act as a claw crane: first, the claws open, and then swoop upward to clutch B12 from BtuF and close. Later the B12 is squeezed through a translocation cavity toward the cytoplasmic opening.

Deciphering these ANM modes responsible in the transition, one can extract the underlying dynamic components as means to control and manipulate the transition, affecting dynamic properties and order of events. This becomes a useful approach for understanding and allaying protein malfunction due to mutations and molecular modifications. Finally, the ANM modes information can be used for advancements in ANM-LD algorithm to obtain alternative pathways (see Mode Restricted ANM-LD Simulations in Chapter 5).

4.4 ANM-LD Provides a Computational Platform for Construction of a Dynamic Allosteric Network

A mode of motion to appear, an underlying infrastructure made of residues functioning as hinge sites are of significance. These hinge residues are mechanistically key sites and generally conserved through evolution while making an allosterically and

dynamically coupled network of residues [186].

In order to determine the hinge sites of the prominent ANM modes deduced from ANM-LD simulations, these modes were mapped onto a 1D dynamic space defined by GNM [61,131]; based on their similarity to GNM mode shapes. GNM is more robust in the prediction of MSFs and thus their minima points as hinge sites, due to its simplistic and orientation dependent energy function compared to ANM [124]. Combining with evolutionary information, by extracting the residues with the conservation scores larger than 6 on ConSurf [153] resulted network of residues with potential mechanistic role in the B12 uptake, some overlapping with functional sites (i.e. cytoplasmic gates, coupling helix, BtuF and ATP sites). The extracted network of residues, combine mechanistically key sites with ligand binding sites and cytoplasmic gates, indicating to a dynamically orchestrated mechanism of action. Using this approach, 15 amino acids (i.e., TMD: Gly67, Leu90, Ala121, Gly124, Leu172, Met211, Arg232, Gly244, Cys279 and Asp291; NBD: Glu27, Gly48, Gly54, Ile186 and Ala200; Figure 4.5) other than known functional sites (i.e. cytoplasmic gates, coupling helix, BtuF and ATP sites) are determined that serve as key hinges for the dynamic communication network of BtuCD. With the hinge sites overlapping these functional sites, the determined sites orchestrate the mechanism of B12 uptake by the allosteric communication they perform throughout the protein.

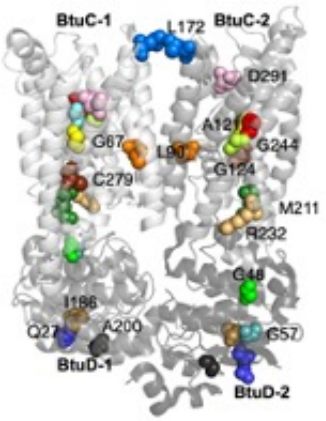

Figure 4.5 Positions of key dynamic sites on BtuCD, to be targeted by single-point mutation study.

4.5 ANM-LD Predicts Potential Perturbation Response of Mutations

Mutation studies are useful for deciphering the sites with functional and allosteric significance. Instead of random mutations, in the present work the residues belong to a dynamic allosteric network underlying a certain mode of motion deduced by ANM-LD were used as mutation target sites. By changing the amino acid at the hinge position, its local characteristics and interactions are altered, in the end affecting the hinge function and hence global dynamics. To maximize the mutation effect, radical amino acid changes were proposed (e.g. change of side chain charge and size).

In the present work, *in silico* single-point mutations were inserted at the deduced positions of both twin monomers. Then, the conformation transition between resulting mutant conformations of both apo and ATP BtuF-bound structure was simulated by ANM-LD. The resulting conformational transition pathways are compared with WT in terms of virtual rotation angles for a preselected set of residues. For this, the rotation angles between the plane formed by three consecutive C_α carbons ($C_{\alpha i}$, $C_{\alpha i+1}$, and $C_{\alpha i+2}$) and the plane generated by the next three C_α carbons ($C_{\alpha i+1}, C_{\alpha i+2}$, and $C_{\alpha i+3}$) were calculated. Then the behavior of these angles during the transitions was evaluated. The mutants with the largest deviation from WT are expected to cause the maximum functional perturbation to BtuCD.

To economize the calculations for comparing the bond angles of WT and mutant BtuCDs, the bond angles formed by hinge residues of the most frequently selected mode (i.e. ANM mode 1) and their nearest neighbors were focused instead of the whole protein. ANM mode 1 is essential for the simulated conformational change, describing the twisting movement of the protein with antiparallel rotation of the NBD and the TMD dimers, and any mutant BtuCD unable to perform this mode of motion might loses its activity. Therefore, for each mutant a mutation outcome score was defined as the sum of angular deviations from WT. To this aim, the deviations of the rotational

angles of the mutant from the means of the rotational angles of WT BtuCD (mWT) were summed over all the ANM-LD steps and for all the 57 rotational bond angles (Eq. 3.12).

High angular deviation scores would indicate a deformation effect on structural dynamics and thus function, while the lower angular deviation scores would indicate an acceptable plausible fluctuation around the WT functional activity. The resulting order of the mutant scores (Figure 4.6) was evaluated experimentally by site-directed mutagenesis and functional growth assay of these mutants [83]. Mutants with relatively low angular deviations (AD) values (e.g., A121H) are predicted to display WT-like behavior, while those with relatively high AD values (e.g., G244R) are expected to deviate from WT activity. Among 15 mutants, L90R, A121W, G244R and L172G show functional loss in these experiments. These four mutants have also the highest angular deviation scores. If other amino acids were proposed in these positions such as L90G, A121C/G/I/L/R/T/V, L172A, G244D, and G244A mutations, the resulting angular deviation scores are relatively low. This shows that the proposed method for evaluating mutation outcome is amino-acid specific. When these mutants were tested experimentally, the function of BtuCD was shown not to be affected as expected for all these mutants except G244D. The discrepancy between computational and experimental evaluation of G244D may arise from the nature of the mutational effect that might be differently affecting the overall conformational transition pathway from the one predicted with angular deviation score. The mutation might cause local perturbations that may interfere with gating, substrate interactions, interactions with the BtuF, or local helix packing; with the latter offering a plausible explanation in the case of the G244D mutation. Alternatively, another mode of motion other than ANM mode 1 might be related with the observed mutation effect. Therefore, for future studies, the sets of hinge residues for deducing rotational bond angles might be expanded.

In conclusion, 25 (out of 26) point mutations show a very good correlation between the ANM-LD predictions (angular deviation scores) and the *in vivo* vitamin B12 uptake assay. To compare the predictions based on the angular deviation

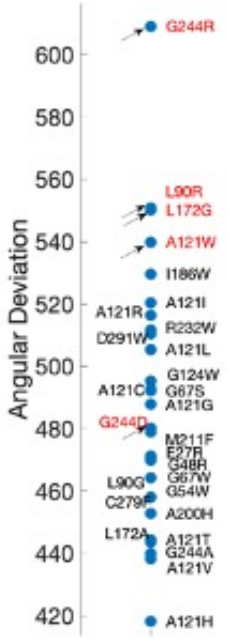

Figure 4.6 The cumulative angular deviation values calculated for the *in silico* mutants. The mutants whose activity was shown to be diminished by experiments are given in red.

calculations with the predictions that are based on thermodynamic stability calculations (generally used for mutation outcome prediction [187–189]), we used FoldX [190] and calculated the $\Delta\Delta$G of the mutations for both the initial (apo) and target (ATP/BtuF-bound) conformations. G67W, G54W and A200H were predicted to be the most un-stabilizing mutations with potential detoriation on the function. However, in functional assays these mutants show similar B12 transport with WT. These results indicate that ANM-LD is a better predictor of the activity than $\Delta\Delta$G scores for both conformations, possibly due to the kinetic information it extracts from the relevant conformational transition.

4.6 Inferring Directionality of Allosteric Signaling in Conformational Transitions

The allosteric signaling between various sites of a protein is a key aspect of protein function. Thus, studying the nature of this communication is crucial to a complete understanding of functional mechanism. The direction of the allosteric

signal is an important component, determining the source of the transferred information, intermediate communication sites/pathways and the final affected site. By directional nature of this signaling, perturbations at key sites such as ligand binding and posttranslational modification sites, are transferred to the distant functional regions including the active site. Deciphering these sites stand as a useful approach to understand and modulate protein function.

ANM-LD simulations offer an opportunity to examine the directional features of allosteric communication taking place in a specific conformational transition. If residues A and B are allosterically connected during the transition, does the conformational changes of A or B initiate the communication? In other words, which residue initiates the movement, and which is driven? To investigate these allosteric relations in BtuCD, we re-analyzed the ANM-LD trajectories by calculation of time-delayed cross-correlations. For this, a time delay τ, expressed as in cycles of the simulation, was introduced to the correlations between residue fluctuations.

Given a time delay, if the correlation C ($A_{t+\tau}$, B_t) is larger than C ($B_{t+\tau}$, A_t), the change in A will drive that of B, at the same time meaning A's movement must precede B's. For BtuCD, the main initiator of allosteric communication is of question: is it ATP binding or BtuF binding residues? Therefore time-delayed correlations are used to determine if the conformational changes that occur in the ATP-binding residues precede (and thus possible trigger) those that occur in the residues that are involved in BtuF binding and other regions, or vice versa. The time-delayed cross-correlation maps demonstrate that the correlations between ATP and BtuF sites peaks if ATP-driven conformational changes occur previous to the fluctuations of BtuF sites, not vice versa (Figure 4.7a, black ellipses). Here, ATP-binding sites drive the movement of BtuF-docking sites (black ovals) and are driven by the movement of MalE-docking sites (black ovals. In $PglK_2$, ATP-binding sites drive the movement of external helices (black ovals), yet the fluctuations of ATP binding sites are also preceded by coupling helix and neighboring helix, indicating to a cyclic allosteric signaling instead of unidirectionality observed in BtuCD and MalFGK_2.For example, the correlation between the ATP-binding residue S163 (located in the D loop - NBD) and the periplasmic gate residues

V170, R173, Q174, and Y177 is only present if the conformational change in S163 precedes that of these residues. Similarly, the BtuF-binding residues (residues L249, L298, and A300-L303) show high correlation with the ATP-binding sites (NBD), but only if their movement is preceded by that of the NBD-located residues. This means that the ATP induced conformational changes occur before the BtuF-driven ones, and that binding of ATP initiates allosteric connectivity in BtuCD. Here, time delayed DCCMs indicate to an order and direction in the allosteric signaling which might imply a causality imposed by ATP sites on protein dynamics. In order to elaborate on the causal nature of allosteric signaling and to understand its physical basis, the GNM transfer entropy calculations were performed as described in next section.

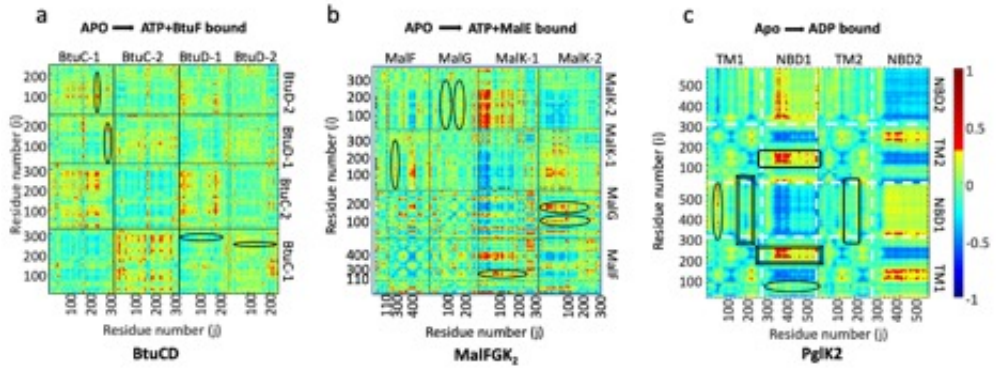

Figure 4.7 DCCMs with an imposed time-delay (τ) between residues i and j, showing the correlation of residue i at time t with residue j at time t+ τ; for BtuCD (a), MalFGK_2 (b) and PglK_2 (c).

For evaluating the relationship between direction of allosteric signal deduced by ANM-LD and working mechanisms of ABC transporters, the results obtained for BtuCD were compared with two other transporters MalFGK_2 and in PglK. While BtuCD is a type II sub-class [122] ABC importer, the maltose transporter MalFGK_2 belongs to type I sub-class ABC importer [191] and PglK, a lipid-linked oligosaccharide flippase is an ABC exporter [127]. To compare the allosteric network of these transporters with distinct topologies and functional mechanisms, ANM-LD simulations were run for each system and DCCM analyses (both time-independent and time-delayed) were performed. For MalFGK_2, the ANM-LD trajectories and resulting DCCMs were kindly provided by Burcu Aykaç Fas.

In general, the cross-correlation maps are similar between MalFGK_2 and BtuCD upon transition from the apo to the ATP and SBP-bound state (MalE for MalFGK_2) in which the NBD moves as a single dynamic unit. As in BtuCD, the TMDs are fragmented to smaller dynamic units, indicating that there are more hinges governing the dynamics in the TMDs than in the NBDs and the ATP-binding sites and the SBP-docking sites are highly correlated. Despite this similarity, important mechanistic differences are observed between BtuCD and MalFGK_2. Overall, the TMDs of MalFGK_2 are more mobile than those of BtuCD, especially apparent in the SBP-docking sites, where the MalE-docking sites are much more mobile than their BtuF counterparts. These differences indicate that docking of BtuF requires less rearrangement than is required for MalE, and suggest that, BtuF docks via a lock-and-key mechanism (geometrical complementarity) while MalE docks via an induced fit mechanism. This is supported by the observation of unchanged conformation of BtuF upon B12 binding [192]. On the contrary, MalE undergoes significant conformational changes upon binding of maltose [193, 194].

When directionality of allosteric signaling was compared between these two importers, two opposite behavior was observed. In contrast to BtuCD, SBP is the initiator for the conformational changes in MalFGK_2, including ATP binding sites. These differences between MalFGK_2 and BtuCD are well correlated with the different mechanism used by each system: in MalFGK_2, the ATP-binding sites are not formed unless maltose-loaded MalE docks [195, 196] and the basal ATP hydrolysis rate is very low and is stimulated by docking of maltose-bound MalE [197]. In contrast, in BtuCD, binding of ATP alone is sufficient to induce the formation of the ATP-binding sites, and this event is BtuF and B12 independent [84], while the basal ATP hydrolysis rate is high and is only mildly stimulated by BtuF [179, 182]. In conclusion the ANM-LD simulations were able to capture the mechanistic difference between these two importers belonging to the different types.

Having distinguished two types of importers in terms of dynamic features, another challenge for ANM-LD simulations is to target an exporter and compare with the studied importers. Thus, another ABC transporter, PglK, a lipid-linked

oligosaccharide flippase which operates by an atypical export mechanism where the substrate never fully enters the transmembrane translocation cavity, and the transporter remains in the OF conformation during the transport cycle [127], was studied with the same approach. To investigate the allosteric connectivity in PglK, ANM-LD simulations of the conformational transition between its apo conformation (IF) and its ADP-bound conformation (OF) was performed [127].

The time independent cross-correlations of PglK deviate from the two importers. First of all, its NBDs move as a whole dynamic unit and TMDs form larger and more continuous dynamic units than the importers. These observations indicate that, in the exporter, there are fewer hinges forming more global and rigid motions. In addition, a long-distance, strong allosteric connection between the substrate-binding "external helix" and the ATP-binding sites were observed (black ellipses) disclosing a communication on substrate occupancy to the ATP sites. Using time-delayed correlations the intricate nature and directionality of PglK appears to be different from the two importers. While a unidirectional signaling mechanism appears in both importers, in PglK, the allosteric network is best described by two nested allosteric cycles, with the allosteric signal propagating between the NBDs and external helices in a cyclic manner. In the "big" signaling cycle, the external helix of chain A drives the movement of TM helices 2-3 of opposite chain and TM helices 4-5 of same chain. Importantly, these regions harbor the positively charged residues R86 and R260, which have been proposed to be in direct interaction with the substrate [127]. From these regions, the signal is transmitted to the coupling helix, and then to the NBDs of opposite chain (black rectangles), continuing to the external helix. From the external helix of chain B, the signal continues as described above for external helix A, via TM2-3 A/TM4-5 B, to coupling helix A, to NBD A, and finally back to external helix A, concluding the external helix-NBD communication of the big allosteric cycle. In the "small" allosteric cycle, the structural rearrangements of the external helix (chain A) are transmitted to TM6 of opposite chain. TM6 harbors the positively charged residues R302 and R309, which together with R86 and R260 of TM2-3/TM4-5 participate in substrate recognition. From TM6, the signal is transmitted to the NBD (of chain A) and then back to the external helix of the same

chain, concluding the external helix-NBD signaling of the small allosteric cycle. This external helix-NBD crosstalk that is present in both allosteric cycles creates an inter-dependence between ATP-related and substrate binding/release.

The distinct dynamic behavior and allosteric mechanisms of different transporters could be extracted by using analyses based on only ANM-LD trajectories [83]. These findings indicate the potential of ANM-LD as a tool to disclose functional mechanisms and allosteric interactions of proteins.

4.7 GNM Transfer Entropy as a Measure to Decipher Internally Embedded Directionality of Allosteric Signal

The directionality of allosteric signal observed for the simulated conformation transitions might arise from the used ANM modes of motion leading the transition pathways, hence embedded in the structure. Therefore, in order to investigate the physical source of the causality relations of observed opposite allosteric communications in BtuCD and MalFGK_2, the GNM correspondents of frequently selected ANM modes were used to calculate transfer entropy of each system [139]. If the transfer entropy from residue A to residue B leads to a decrease in the degree of uncertainty for residue B, i.e., the dynamics of A affects the dynamics of B. Using only the most fundamental modes of motion of BtuCD and MalFGK_2, the transfer entropy that underlie these cooperative motions were computed. The resulting maps display the net transfer entropy from residue i to j Figure 4.8. Transfer entropies were calculated by considering only the GNM modes that correspond to the most-frequently used ANM modes identified in the ANM-LD simulations, in order to decipher the causal effect of intrinsic dynamics on the observed directionality of the allosteric signal. On Figure 4.8 red colors indicate positive transfer entropy (sending information), and blue colors indicate negative transfer entropy (receiving information). Magenta spheres and black triangles correspond to SBP-binding and ATP-binding residues in BtuCD (4.8a) and MalFGK_2 (4.8b). In BtuCD, the net

transfer entropies between the ATP binding residues and BtuF-docking residues are positive when the information flow is from the former to the latter (upper-left quadrants) and negative (lower-right quadrants) for the opposite direction. Three such prominent examples are highlighted by black ovals. This means that in BtuCD the direction of causality is from the ATP binding residues to the SBP-docking residues. MalFGK manifests the opposite behavior; positive transfer entropies occur in the SBP-to-ATP binding direction (lower-right quadrants) and negative in the ATP-to-SBP binding direction (upper-left quadrants). This means that in MalFGK_2, the direction of causality is from the SBP-docking residues to the ATP binding residues.The calculations show that the ATP-binding sites in BtuCD decrease the uncertainty of motion of the BtuF binding residues (Figure 4.8a). This observation supports the notion that in BtuCD the dynamics of the ATP sites precedes, and that of the BtuF sites follow. More importantly, these results show that the ATP sites behave as transfer entropy sources and causally affect the dynamics of BtuF binding residues. This indicates to the role of ATP as an initiator of allosteric signaling hence functional mechanism.

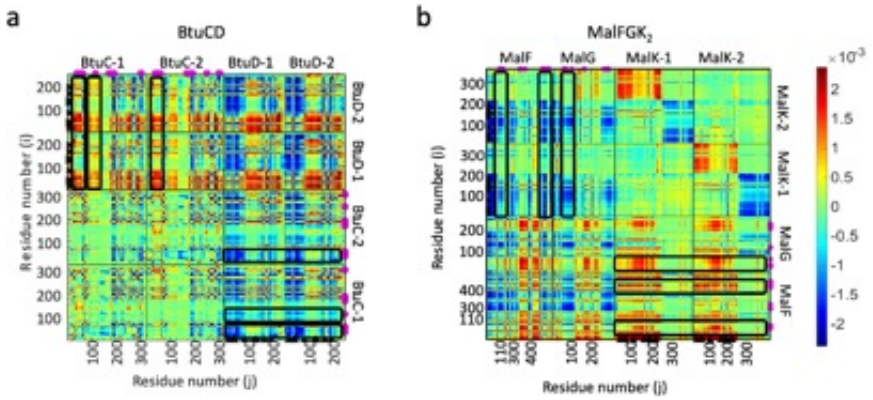

Figure 4.8 The net transfer entropy (TE(i,j)-TE(j,i)) maps calculated for the outward-facing conformations of BtuCD (a) and MalFGK_2 (b) are shown here.

In contrast with BtuCD, the MalE-binding sites in MalFGK_2 decrease the uncertainty of motion of the ATP-binding residues (Figure 4.8b), suggesting that in MalFGK_2 the dynamics of the MalE-docking sites precedes, and that of the ATP sites follow. Here, SBP sites are the initiator of the allosteric signaling toward ATP

sites as opposed to BtuCD, which indicate to different working mechanisms of two importers, BtuCD with constant and MalFGK_2 with conditional ATP activity, as previously discussed.

In conclusion, the directionality inferred from the time-delayed correlations arise from the contact topology of the initial structures and this embedded causality can be predicted using GNM Transfer Entropy.

4.8 Conclusion

Using ABC transporters as case systems, the efficiency of ANM-LD method on deducing dynamic information related with functional mechanisms of membrane proteins was evaluated in the present study, along with the biological and physical relevance of the extracted results in comparison with FRET and mutations experiments.

From the obtained conformational trajectories of BtuCD from ANM-LD simulations, the computed fluctuations of the individual residues and the coupling between them implied that ATP-binding affects the protein dynamics globally while BtuF-docking mobilizes only periplasmic region. This observation is in agreement with the previous experimental knowledge on ATP being the main player of the BtuCD's functional mechanism and BtuF as a fine-tuner [84, 180, 181]. Additionally, the observed dynamical coupling among functional sites (i.e. cytoplasmic gates, coupling helix, periplasmic gate, BtuF and ATP-binding sites) manifested from ANM-LD trajectories is supporting the presently accepted working mechanism of BtuCD requiring this communication [183–185].

The ANM-LD simulations are able to disclose mechanistic steps of functional conformation changes, by disclosing transient conformations and the order of events due to the dynamic capacity of used slow modes to lead the protein toward the large conformational changes with minimum energy [60]. These modes are evolutionarily

exploited to impose certain order of dynamic events whose composition manifests as the working mechanism of the protein. In BtuCD, the used slow modes to forward the conformational changes toward the target conformation resulted sequential dynamic events which lags the opening of cytoplasmic gates to control the vitamin release and corresponds to a squeezing mechanism, in agreement with the FRET intensities [84].

As slow modes are responsible for the cooperative movement of the protein and the needed long-range allosteric communication, dissecting these modes provide an important advantage of ANM-LD simulations, treating complex transition dynamics as a superimposition of simpler modes of motions. Determining these modes and their underlying hinge network, the key dynamical sites of allosteric interaction can be disclosed, forming a potential allosteric communication network. Once these positions are revealed, then the effect of their potential manipulation such as mutations can be predicted successfully by ANM-LD as supported by single-point mutation experiments and functional assays of the mutants [83].

Having provided a tool to predict dynamic network of residues, the directionality and causality features of the signaling among this network can be disclosed using ANM-LD and GNM transfer entropy calculations together [83]. Determining the source of the allosteric signal, one can target the protein to modulate and control the function which is useful in drug development and targeting. Finally, the mechanistic differences between various systems can be uncovered by ANM-LD simulations as shown for the three different transporters and distinguishing between different systems support the studies of side effect free (caused by targeting various proteins instead of one) drug development.

In conclusion, ANM-LD stands as an efficient computational tool, particularly to complement experimental techniques to study protein dynamics toward the aim of understanding and controlling the functional mechanisms of proteins.

5. SAMPLING CONFORMATIONAL TRANSITION DYNAMICS OF PROTEINS BY ADVANCED ANM-LD SIMULATIONS

The modular nature of ANM-LD algorithm enables further exploitation and advancement of the method to attain an enriched tool for studying protein dynamics. First of all, ANM mode selection, a significant part of the algorithm shaping the transition pathway, is suitable for tinkering and diversifying the set of ANM modes to guide the simulations [83]. This specialty broadens the sampled conformational space and enables the detection of rare events [131]. Consequently, instead of choosing the easiest path between two conformations, bypaths can be generated resulting a more complete sampling of the energy landscape [83].

Two approaches were proposed for varying the set of ANM modes to be selected. In the first method, called mode restriction, popularly selected ANM modes were eliminated from the mode pool so that the otherwise underrepresented ANM modes can be selected for guiding the transition. In the second approach, the mode pool was kept complete, yet a Monte Carlo (MC) algorithm was used to decide on the selection of the modes to guide the conformational transition instead of the most overlapping modes with the difference vector. These two extensions create alternative transition pathways on the conformational landscape, as shown below for c-Src kinase and BtuCD proteins.

Another modification can be done on the node resolution of ANM networks representing protein structures, on which nodes are defined as C_α atoms in regular ANM-LD [62, 83, 84]. Particularly, if an all-atom variable is known and will be used to assess the biological and physical relevance of the simulated trajectory, an all-atom approach might be more suitable to provide side chain movements and rotations for a high-resolution representation of the conformational change. Thus, in the present work, each heavy atom of the protein was taken as a node of the representing network

for ANM calculations to refine the transition pathway of c-Src kinase.

The fourth advancement was to include ligand molecules in the ANM-LD simulations. The initial and target conformations used in regular ANM-LD simulations are directly extracted from ligand-bound states (if necessary) without any change on the structure except energy minimization. Yet during the minimization and LD steps, the energetics of the sampled intermediate conformations might be improved with the explicit presence of the ligand and its interactions with relevant amino acid side chains. Although the functionally important conformational ensemble of the protein is already assumed to pre-exist in the ligand-free protein, the presence of ligands might stabilize certain conformations by increasing their probability in the conformational ensemble [2, 27, 158] obtained by a very short LD in each ANM-LD cycle, yet accumulating the effects towards the end of the simulation. In the end, the ligand effect in the obtained conformation after undergoing an LD part might cause the selection of different ANM mode in the following ANM-LD cycle, hence affecting resulting conformational trajectory. For untargeted ANM-LD simulations in which only some collective variable describing the movement of ligand is used, the effect of explicit ligand presence might be particularly important for narrowing the sampled path and hence easing the functional conformation change while requiring a smaller number of parallel simulations. While untargeted simulations were performed on BtuCD alone, targeted ANM-LD simulations were run for both c-Src kinase and BtuCD.

Finally, in order to ensure a biological and physical relevance of the transition pathway sampled by ANM-LD, observed collective variables obtained from atomic force microscopy (AFM) and small-angle X-ray scattering (SAXS) experimental methods were used to guide the ANM-LD simulations. All these modifications were employed individually in ANM-LD algorithm on globular c-Src kinase and membrane embedded BtuCD, for serving to coverage of different aspects of protein dynamics.

5.1 Mode Restriction in ANM-LD Simulations

In regular ANM-LD algorithm, number of slow ANM modes (modemax) is predefined among which the mode to perturb the conformation is selected for proceeding the simulation. The selection criterium is to maximize the overlap (dot product) of the mode vector with the difference vector between the immediate and target conformation. So the ANM mode with highest overlap is used to lead the simulation in the relevant ANM-LD cycle. Therefore, the set of used ANM modes is quasi-determined and vary only due to the randomness in LD part of the method from the initial cycle onward [83, 84]. In order to surface underrepresented ANM modes for sampling rare transition events and conformations, the dominantly selected ANM modes are eliminated so that the resulting mode pool is restricted [83]. Toward this aim, mode restriction in ANM-LD was applied to two protein systems, globular and non-globular; namely c-Src kinase and BtuCD.

5.1.1 Mode Restricted ANM-LD Simulations of c-Src Kinase Dynamics

c-Src kinase is a globular protein functioning in the phosphorylation of tyrosine residues on its downstream protein targets in cellular signaling cascades of cell proliferation and differentiation [198]. It is a widely studied system due to its potential role in cancer [106, 115, 199]. The c-Src kinase activation requires a global conformational change with 23 Å RMSD from the apo conformation as depicted on Figure 2.1 [107, 108]. This large RMSD difference between initial and target conformations and the availability of previous knowledge on structure and function of c-Src kinase make it a good case system for evaluating the results of implementations in ANM-LD algorithm.

Firstly, the activation dynamics of c-Src kinase (inactive to active, with 2SRC to 1Y57 PDB codes respectively) were sampled by regular ANM-LD simulations. The final conformation approached within 3.2 Å of the target conformation (Figure 5.1). This successful transition into the active conformation was achieved by using the appropriate

composition of ANM modes used for guiding the simulations. The set of used ANM modes were dominated by repetitive use of slow ANM modes 1 and 6 as highlighted with green rectangles on Figure 5.1, particularly in the first half of the transition during which the most global changes occur. Yet supported by other slow modes (i.e. slow ANM modes 2-4, 7 and 9) these modes were thought to dominate the major collective changes of the protein and blocked separately in following ANM-LD simulations.

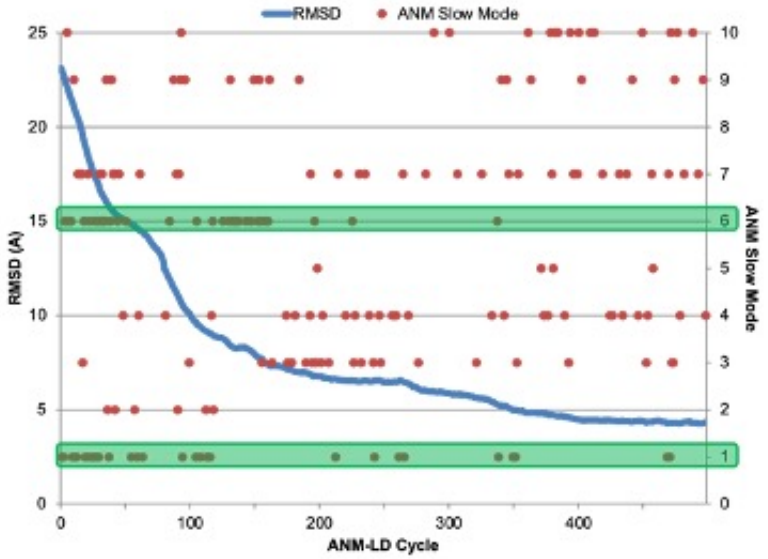

Figure 5.1 RMSD from the target (active-1Y57) obtained from the ANM-LD trajectory of inactive to active conformational transition. Selected modes in the first ten ANM modes are given as red spheres.

The restriction of ANM mode 1 results in a slower RMSD decay toward the target conformation indicating to a possible alternative transition pathway (Figure 5.2). On the other hand, the blockage of ANM mode 6 hinders the transition resulting a trajectory away from the target conformation (Figure 5.2). This shows that ANM mode 6 is a vital mode for activation, while ANM mode 1 has a supporting role. Therefore, the environmental changes due to phenomena such as mutations, around the hinges underlying ANM 6 might be riskier for functional maintenance yet the hinges of ANM 1 might also contribute to the functional mechanism.

The above-mentioned observation is supported by the presence of oncogenic mutation sites around and at these hinge residues for both modes (Figure 5.3). Among 127 oncogenic mutation sites, 51% reside within 7 Å of C_α atoms of these hinge sites.

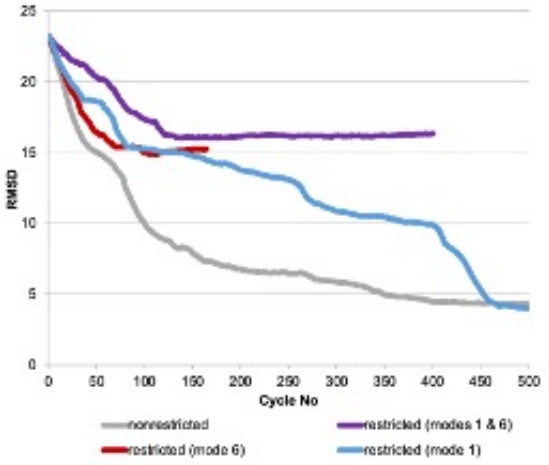

Figure 5.2 RMSD from the target (active-1Y57) of inactive to active conformational transition with nonrestricted (gray) and restricted on ANM modes 1 (blue), mode 6 (red) and both (purple).

When the opposite conformational transition and the modes of motion and underlying hinge residues enabling this structural change (active to inactive) are considered, the percentage increases to 70 %. This means that 70 % of known oncogenic mutation sites are likely to affect global dynamics of the system as described by the given modes of motion, if a mutation disrupting its hinge function occurs. This indicates the potential role of global dynamics in oncogenesis due to mutations with missense translation on proteins and applicability of ENM-based tools to target the formation mechanism of these disastrous effects.

Following the promising results on c-Src kinase, mode restriction was evaluated also on a membrane protein. The membrane embedded BtuCD was selected as a case system due to the accumulated knowledge on its functional dynamics in our group and in the available literature.

5.1.2 Mode Restricted ANM-LD Simulations of BtuCD Dynamics

On BtuCD, the selected ANM modes in apo to ATP-bound transition were derived from four parallel ANM-LD simulations. Among first ten slow modes, ANM mode 7 was repeatedly selected and followed by ANM mode 6 in selection frequency

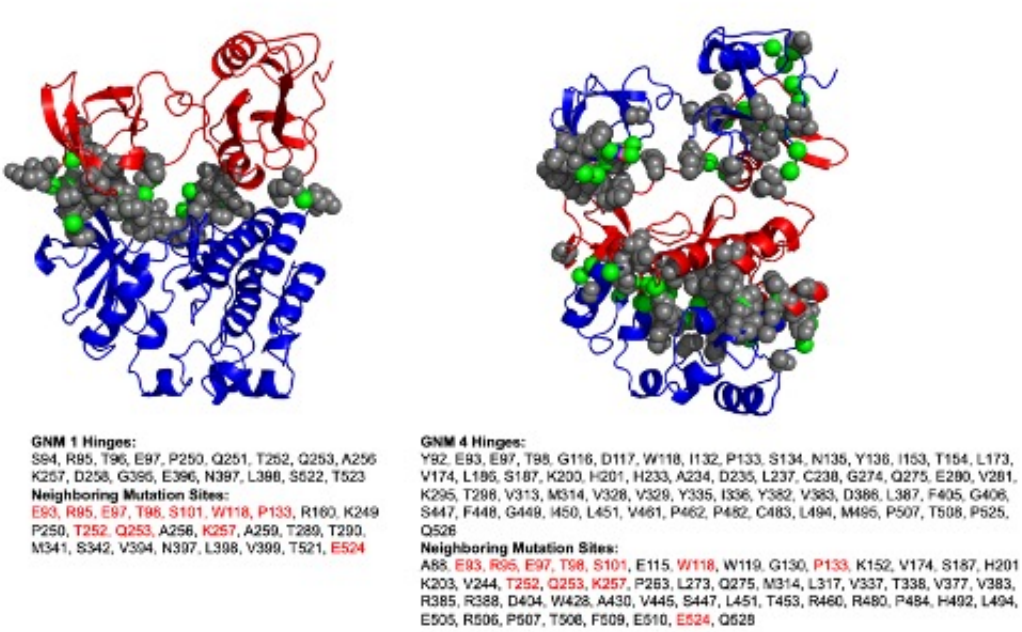

Figure 5.3 Hinge residues of slow GNM modes 1 and 4 in green and neighboring mutation sites in gray. Overlapping neighboring mutation sites to both GNM mode hinges are given in red.

in this transition. ANM mode 6 corresponds to a squeeze and release motion of the protein along the vertical axis, which closes NBDs while opening cytoplasmic gates and stretching periplasmic gate toward the periplasm and vice versa. On the other hand, ANM mode 7 couples the closure of the NBD with the opening of the periplasmic gate and closure of the cytoplasmic gate, while also ensuring the coupling of opposite behavior. In order to test the indispensability of these two modes, three ANM-LD simulations with mode-restriction were performed. Firstly, ANM mode 6 was blocked separately, then ANM Mode 7 was restricted alone and finally two modes were eliminated together. Resulting RMSDs from the target conformation display the changes in the conformational transition trajectories of restricted runs in comparison to those of regular ANM-LD simulations. Restricting ANM mode 6 alone hinders the approach to the target conformation, while ANM mode 7 restriction only delays the transition reaching a similar final RMSD to the target conformation with the regular ANM-LD trajectory. On the other hand, if these modes are blocked together, the transition trajectory approach toward the target conformation by no means (Figure 5.4).

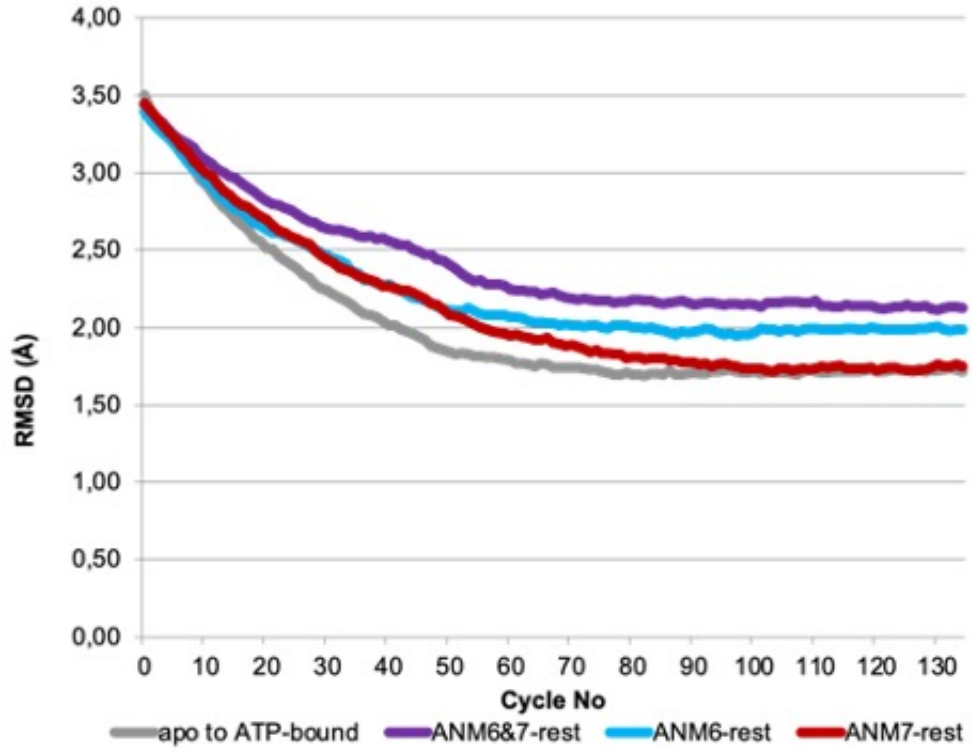

Figure 5.4 RMSD from the target (holo-4R9U) of apo to ATP-bound conformational transition with nonrestricted (gray) and restricted on ANM modes 6 (blue), mode 7 (red) and both (purple).

In order to understand, how mode restriction affects the dynamic coupling observed in the simulation trajectories, the cross correlations were calculated for each restriction run from the initial one-third of the conformational transition trajectories (for extracting the most collective changes that appear mostly in the initial part of the simulation) and compared with the results of regular ANM-LD simulations. Two correlated pairs were found to vary between apo>ATP and apo>ATP/BtuF-bound transitions (5.5): 1) The correlation of TM3-4 helices (92-139AB) region with neighboring ATPase domain in apo>ATP/BtuF transition and 2) the correlation of upper region of TM5 helix (155-166AB) with neighboring TMDâs same region in apo>ATP-bound transition. As shown of Figure 5.5, the absences of ANM mode 6 and 7 both separately and together during the transition appear to break the coupling between TM 5 helices of opposite BtuCs, affecting the opening behavior of periplasmic gate located on the upper part of this helix, required for taking B12 inside [83]. Yet the restriction of ANM mode 6 did not change apo>ATP transition-like behavior much. On the other hand, individual restriction of ANM

mode 7 and restriction of both ANM 6 and 7 resulted a similar correlation pattern with apo>ATP/BtuF-bound transition.Besides, the restriction of ANM mode 7 alone or together with ANM mode 6 also causes a coupling between TM helices 3 and 4 and neighboring NBD to appear. ANM mode 7 couples NBD closing with periplasmic TMD opening, which enables the preparation of BtuF docking site in response to the ATP binding. Surprisingly, when ANM mode 7 is blocked, the apparent cross-correlation resembles the correlations observed in apo to ATP/BtuF-bound transition. This displays that the presence of BtuF suppresses the ANM mode 7 and this mode functions as a dynamic switch to govern conformational transitions required for B12 transport. Overall, these findings indicate to the role of ANM modes 6 and 7 in the pattern of dynamic coupling in apo to ATP-bound transition.

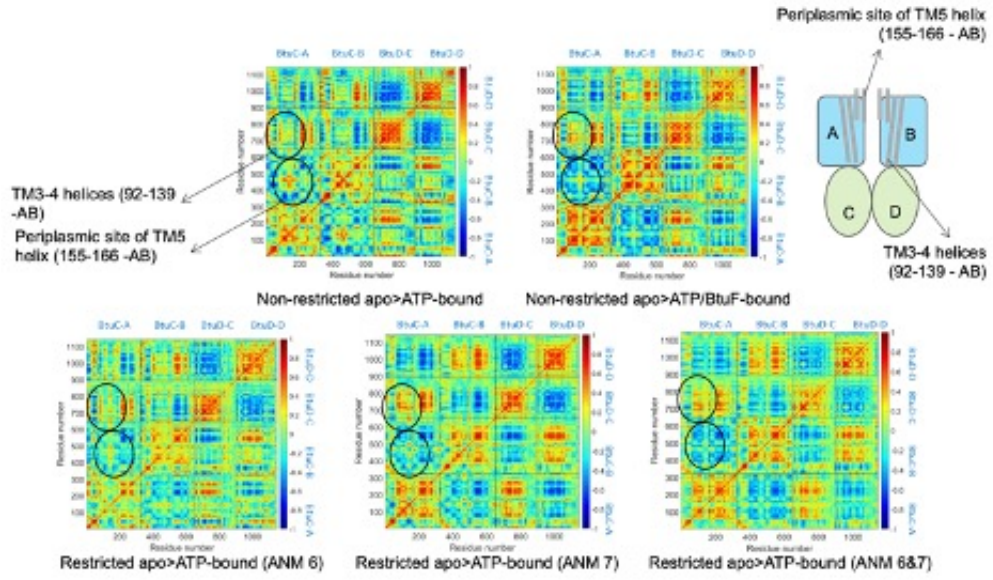

Figure 5.5 DCCMs of normal and restricted simulations of apo>ATP-bound transition and normal apo>ATP/BtuF-bound transition.

Another way to evaluate the effect of mode restriction on the obtained ANM-LD trajectories, is to calculate and compare pair-wise C_α distances between twin residues of S67 (NBD), Q109 (TMD) and R138 (TMD) since the movements of these residues summarize the general movement of the protein in NBDs, periplasmic part and cytoplasmic TMD region respectively [84]. For S67CD and Q109AB pair distances, restriction of ANM mode 7 is responsible to retard NBD closing and to facilitate closing of TMDs, resembling to the distances appear in apo to

ATP/BtuF-bound transition (Figure 5.6). This again shows the effect of ANM mode 7 on the coupling between TMD and NBDs. On the other hand, the restriction of ANM mode 6 delays cytoplasmic TMD region opening since R138AB pair distances increase slower than regular runs. Again, the observed R138AB pair distances are more similar to the distances in regular apo to ATP/BtuF-bound transition. Herewith, ANM modes 6 and 7 appear as dynamic switches which can be modulated in response to absence and presence of BtuF. All these observations indicate to the intriguing potential of mode restriction in ANM-LD for protein dynamics study.

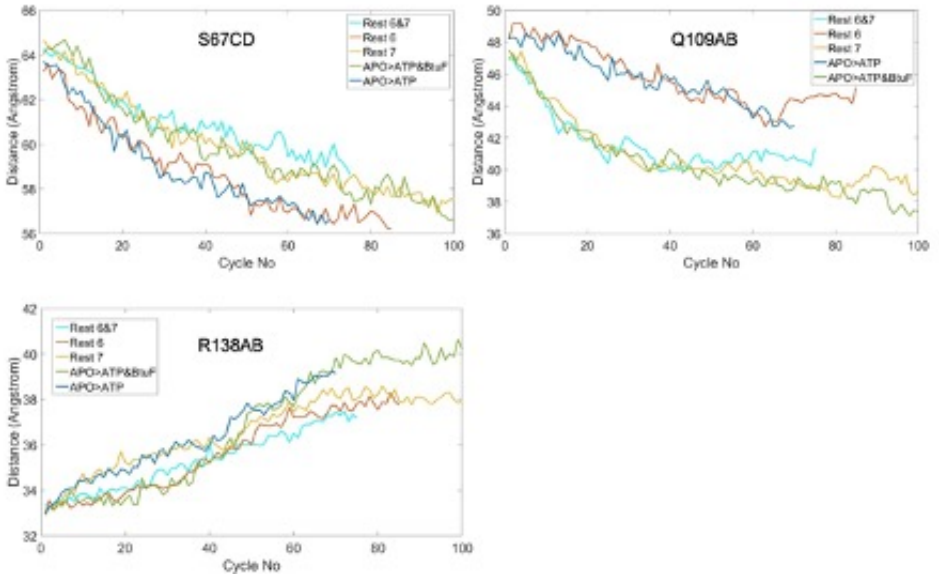

Figure 5.6 The pair C_α distances for S67CD, Q109AB and R138AB residues in regular and restricted apo to ATP-bound and regular apo to ATP/BtuF-bound transitions.

5.2 Monte-Carlo Based Mode Selection in ANM-LD

Although mode restriction is useful for obtaining alternative transition pathways and capturing rare events, it cannot be applied to the ANM modes whose restriction is detrimental for the transition as in the case of c-Src kinase activation transition (inactive-to-active) and apo to ATP-bound BtuCD both in which the blockage of ANM mode 6 hinders the transition. Thus, another approach to change the pool of selected ANM modes is to change the selection criterium. Instead of choosing the ANM mode with highest overlap with the difference vector, a Monte

Carlo (MC) based approach is used. Firstly, the modemax number of available ANM modes is distributed between 0 and 1, each having a tendency to be selected in proportion with the overlap between the relevant mode vector and the difference vector. After this, a pseudo-random generator of MATLAB (rand function) is used to supply a number between 0 and 1. The ANM mode whose range comprises the resulting number is selected to perturb the conformation. While highly overlapped modes have a high probability of being selected, there is still stochasticity due to the nature of MC approach. This randomness provides the spontaneity leading to alternative transition pathways whose cumulative effects might also contribute to the prospective changes in the pathways.

In order to diversify selected ANM modes in c-Src kinase inactivation simulations (active to inactive), MC was applied to the mode selection part in two different ways. Among all modes (modemax= 3N-6), in the first approach MC was used in each cycle of the simulation. In the second approach MC was applied intermittently having in one cycle mode selection based on the best overlap criterion while in the following cycle having MC administration. In order to assess the obtained trajectories, the RMSDs from the target conformation are compared with the regular ANM-LD trajectory in which only the best overlap criterion is used. In addition, two known collective variables, which are widely used in the literature to define conformational landscape of c-Src kinase: 1- df: the difference in the atomic distances between E310(CD)-R409(CZ) and E310(CD)-K295(NZ) and 2- RMSD of A-loop, were calculated to see how MC affects the conformational sampling.

When RMSDs to the target conformation are considered, ANM-LD with consistent MC sampling lagged behind and converged around 8 Å at 900th cycle in contrast with the regular ANM-LD which converged around 3 Å at 200th cycle. When only first 400 cycles of the trajectories of both simulations are considered, the folding of A-loop was shown to dominate the initial movements in simulations with consistent MC sampling while in regular run the distance difference between E310(CD)-R409(CZ) and E310(CD)-K295(NZ) atoms due to H bond shift is observed to be affected in the first place (Figure 5.7). The conformations from first 400 cycles

of ANM-LD trajectory (highlighted in green) were used to calculate collective variables df and RMSD of A-loop to define the transition pathway in distinct simulations.

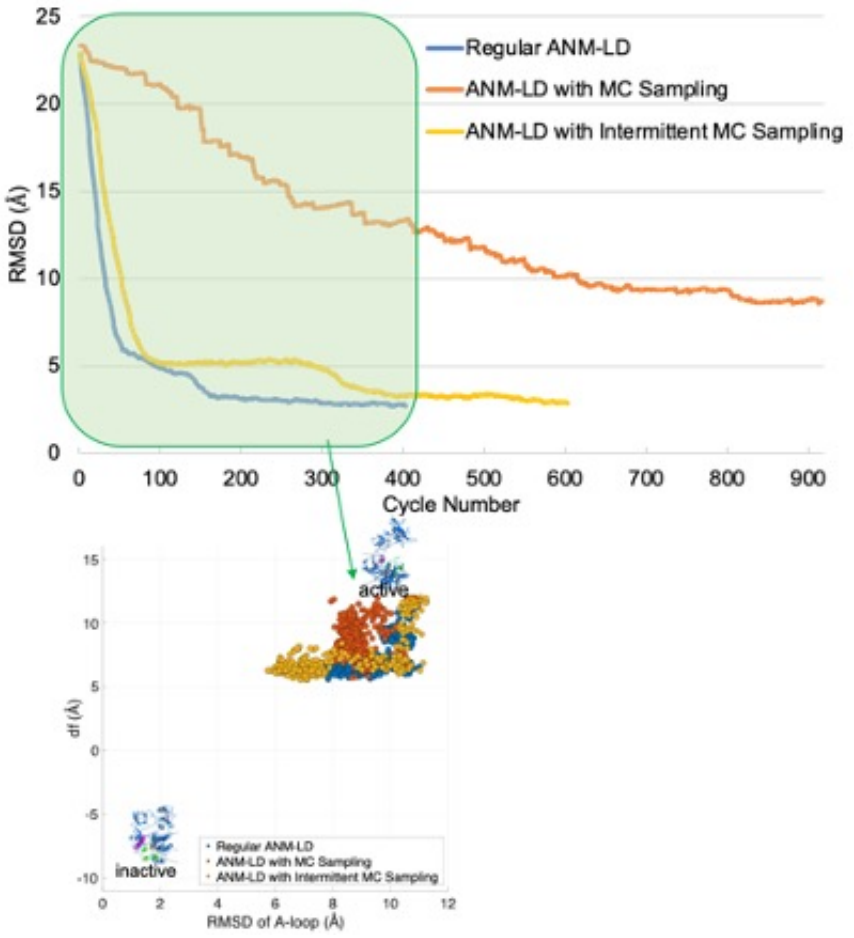

Figure 5.7 RMSD from the target (inactive) and energy landscapes for regular ANM-LD simulations and ANM-LD simulations with consistent and intermittent MC sampling approaches.

In order to reduce the stochasticity in performed perturbations imposed by the selection of faster modes by MC algorithm, the selection criterium was used as combination of MC and best overlap with difference vector in another ANM-LD simulation. This combination provides an optimum solution between sampled space and final RMSD, since the RMSD to the target conformation reaches the similar final values with the regular run despite of 200 cycles delay yet with more structured A-loop as proposed in the literature [200].

Although combining mode selection criteria improved the sampled conformational space, the expected ⅃-shaped pathway from active toward the inactive conformation could not be achieved by only introducing MC sampling in mode selection. Yet the high-resolution atomistic characteristic of the studied collective variables might be difficult to be captured by a hybrid approach such as ANM-LD. Thus, combination of different advancement approaches might be efficient in targeting c-Src kinase transition dynamics by ANM-LD simulations.

5.3 All-Atom ANM-LD Simulations

Regular ANM-LD algorithm uses C_α atoms as nodes of network to define a protein conformation for the ANM calculations. A more detailed depiction of this network is possible, in which other backbone and side chain atoms might also be treated as nodes. This is especially useful if the resulting trajectories will be assessed by an all-atom parameter such as presence/absence of salt bridges and H bonds as in c-Src kinase collective variables described above, since all-atom approach might be more suitable to provide side chain movements and rotations enabling the observed behavior. Toward the aim of refinement in the ANM-LD approach and resulting trajectories, each heavy atom from both backbones and sidechains of the protein residues was taken as a node of the conformational network for ANM calculations in the present work. The resulting algorithm is called all-atom ANM-LD simulations. Since the included heavy atoms are in closer proximity with each other than neighboring C_α atoms, the atoms within $R_c = 5$ Å are thought to be interacting and treated as connected by harmonic springs instead of 13 Å (as it was in regular ANM-LD simulations).

As described in the previous section, the conformational landscape of c-Src kinase had been defined and summarized by two atomistic variables df and RMSD of A-loop in the literature [113, 114, 200], both related with H bond shifts. These variables require high-resolution sampling of the conformational transitions hence all-atom ANM-LD simulations of c-src kinase were performed between inactive and active conformations in both directions. As shown on Figure 5.8, resulting

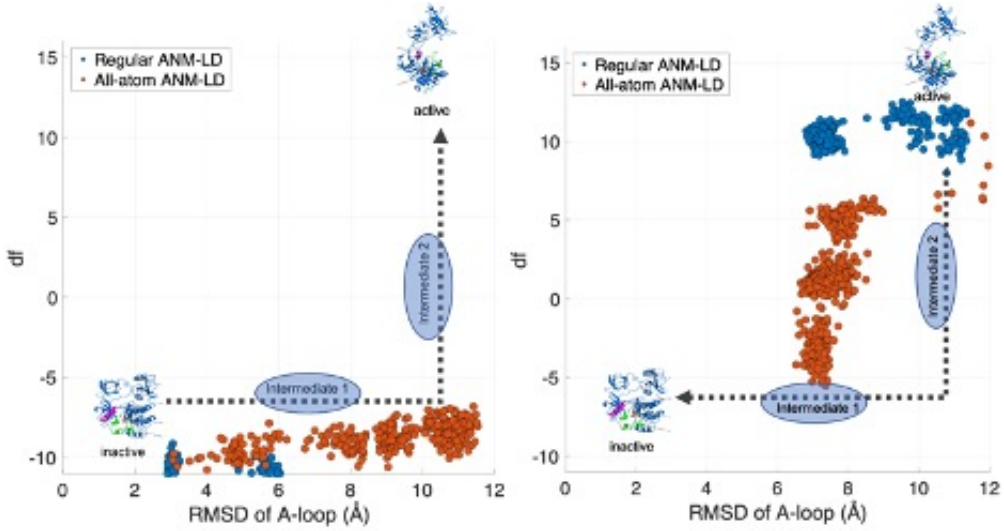

Figure 5.8 The conformations from ANM-LD trajectories of regular and all-atom ANM-LD simulations were used to calculate df and RMSD of A-loop to define the transition pathway in both directions.

trajectories of all-atom simulations indicated to a better sampling on the energy landscape defined by these variables in both directions. Yet in terms of these variables, the final conformations of the simulations could not reach the target conformation, although they succeeded an approach within 1.5 Å RMSD from the target. This might be due to the absence of ATP and Mg atoms in the simulations, since the H bonds related with the defined collective variables are in close proximity and interaction with these atoms. In order to evaluate this hypothesis, ANM-LD simulations were performed between explicitly ligand-bound conformations on both c-Src kinase and BtuCD.

5.4 ANM-LD Between Ligand-Bound Conformations

Binding ligand molecules such as nucleotides, metal ions, glucose and more is thought to change the distribution of states in the conformational ensemble of a protein [2]. Thus, the presence of ligands might also affect the sampling of conformational landscape particularly by untargeted simulations. In case of regular ANM-LD, the presence of target coordinates for leading the mode selection imposes the generation of conformational trajectory relevant for the functional transition of

interest. Additionally, although the protein is stripped from its bound ligands in regular ANM-LD, the exact contact topology of protein is kept, and this allows the relevant internal dynamic modes to be extracted and to lead the simulation. However, for cases in which the ligand effect due to its interactions with relevant amino acid side chains is very significant or unknown, explicit presence of the ligand in minimization and LD and/or its depiction as a node in ANM calculations might improve the sampling and optimize the sampled trajectory. Although the guiding of the conformational trajectory toward the target conformation is embedded in the ANM-LD algorithm, the presence of explicit ligand might ease the relevant transition by affecting the energetic calculations in minimization and LD steps (and increasing the probability of certain conformations) thus influencing the ANM mode to be selected in the following cycle. Additionally, the minor dynamic effects imposed by explicit ligand presence might accumulate toward the end of the simulation. In order to assess the effect of explicit ligand in sampled trajectories, c-Src kinase and BtuCD were selected as case systems, one globular and one membrane system, respectively.

5.4.1 ANM-LD Simulations on ATP-Bound c-Src Kinase

On c-Src kinase, the limited sampling of the conformational landscape by regular ANM-LD arises the question whether the explicit presence of ATP molecule and two accompanying Mg^{++} ions improve the sampling on the above discussed energy landscape defined by df and RMSD of A-loop. Toward answering this problem, ATP was represented as three nodes and each Mg^{++} ion was accepted as one node for ANM calculations during the ANM-LD simulations of inactive to active and active to inactive transitions.

In both directions, including ATP molecule explicitly in the simulations greatly expanded the sampled conformational area on the given landscape (Figure 5.9). Furthermore, in inactive to active simulations, incorporating explicit ATP atoms enabled to sample two known intermediate states (intermediates 1 and 2 on Figure 5.9) from the literature [200], although it could not reach the final

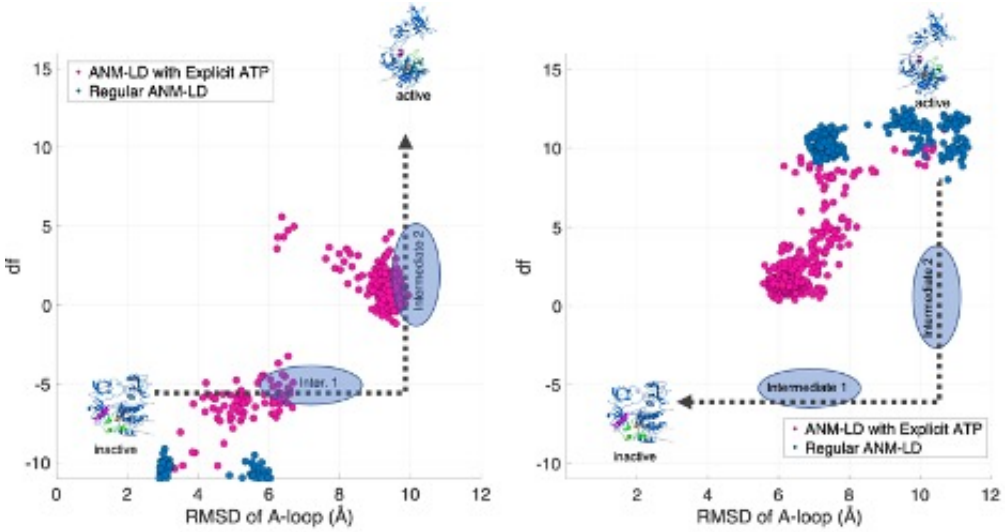

Figure 5.9 The conformations from ANM-LD trajectories of regular and explicit ATP including ANM-LD simulations were used to calculate df and RMSD of A-loop to define the transition pathway in both directions.

conformation completely. This might be due to the fact that at this point ATP and Mg^{++} might be already used and removed from the protein. This might also explain why in active to inactive transition, known intermediate states 1 and 2 could not be captured. Thus, enabling the chemical changes such as from ATP to ADP, or physical changes such as removal of a partner molecule during the ANM-LD simulations might be a good target for future studies. On the other hand, the stabilization of active state is thought to require Y416 residue on A-loop to be phosphorylated [113] and without it, the active state is visited transiently. This transience might require to increase the number of parallel simulations to observe the active state without the relevant phosphorylation.

Consequently, among other advancements, including ligand had the major effect on the quality of obtained trajectory in terms of biophysical relevance and expanded sampling. In the case of c-Src kinase, this improvement yielded a straightforward method to sample conformational transitions between the inactive and active states with validation of biological and physical relevance of the trajectory. In the literature, the complete transition of full-length kinase was sampled by targeted MD simulations [201] yet the resulting trajectories were not assessed in terms of biological and physical relevance either basing on the widely used H-bond related collective variables (i.e.

df and RMSD of A-loop) or other means. Targeted MD simulations might generate unrealistic and high-energy conformations [202] thus require the energetic assessment of the resulting trajectory. On the other hand, the studies presenting the energetics of the conformational transition based on these variables had to combine set of enhanced sampling methods (i.e. pathway methods, umbrella sampling calculations and Markov state models) yet didn't provide the the complete transition pathway between fully-open and fully-closed c-Src kinase [114], corresponding to 2SRC and 1Y57 crystal structures respectively. Thereby, the ligand incorporating ANM-LD simulations stand as a promising tool to study transition dynamics of c-Src kinase systems in addition to other systems with large conformational changes among functional states.

5.4.2 ANM-LD Simulations on Vitamin B12 Docked BtuCD

Vitamin B12 is the substrate of BtuCD transporter, which is imported to the cytoplasm through a translocation cavity located at the interface of BtuC dimer. It is thought to be transferred inside the cell via a peristaltic movement of the protein following the vitamin's release from BtuF [84,121]. The compactness of the cavity [203] might require tight interactions of the vitamin with the surrounding amino acids in the cavity. Due to these interactions and large size of the vitamin B12 (up to 15 Å), the explicit depiction of the vitamin in ANM-LD simulations might affect the sampling and be beneficial for disclosing how the vitamin leaves the protein.

The effect of explicit substrate presence on BtuCD's conformational fluctuations was initially studied by two types of ANM-LD simulations. First type of simulations were run between vitamin B12-docked BtuCD conformation as initial (corresponding to ATP BtuF-bound structure-4FI3, [121]) and a substrate-free modeled IF conformation (based on Hi1470/71 of *Haemophilus influenzae* [204]) as target states, the latter representing the transient structure after the substrate release. The starting conformation of the simulation is ATP BtuF-bound structure because it is the state in which substrate binding protein BtuF is bound to BtuCD and B12 potentially resides in the translocation cavity. The vitamin B12 was docked

to 4FI3 structure but excluded in IF-model conformation since it was expected to be released at this conformation. Second version of simulations were run untargeted from two different initial conformations, apo [119] and ATP BtuF-bound [121] conformations separately in order to eliminate the artifacts which might be imposed by using a model target conformation. The simulations starting from apo conformation were performed in order to disclose preexisting modes of motion in the free state of the protein which enable the mechanistic motions of B12 transfer.

The substrate is incorporated into the ANM-LD algorithm at various sections. Firstly, the vitamin B12 molecule was represented by 8 nodes corresponding to atoms Co, C3, C8, C13, C17 from corrin ring and atoms P from phosphate, C3R from ribose and C5B from dimethyl benzimidazole parts of the molecule in the ANM calculation step. Followingly, in minimization and LD steps, B12 molecule and its interactions were considered in full atoms style. For simplicity and computational efficacy, BtuF was not included in the simulations.

5.4.2.1 Targeted ANM-LD Simulations with B12 Molecule. For targeted simulations, obtained trajectories and observables were compared with results from ANM-LD simulations of same transition (ATP/BtuF-bound to IF) without the vitamin B12. In these simulations, at the mode determination step to be used for the perturbation of the conformation, the overlap between difference vector and ANM modes was calculated by considering only C_α atoms of the protein, skipping the vitamin B12 since it is missing in the target conformation.

ATP BtuF-bound > IF-model transition is expected to occur via multiple changes on ATP BtuF-bound conformation when initial and target conformations are compared. Firstly, cytoplasmic gate II (residues 83-85) should open for enabling the B12 transition through the translocation cavity of BtuCD into cytoplasm while ATP sites move away from each other due to ATP hydrolysis and ADP release. In addition, the periplasmic part of BtuCD is further closed by reinforcing inward-facing behavior. In order to evaluate the resulting ANM-LD trajectories in terms of these dynamic

behavior, the pair C_α distances of cytoplasmic gate II residues P84, NBD residue S67 and C_α distance between G38 and G129 ATP sites, pair periplasmic residues of V170 and L172 were calculated both B12 including and excluding simulations and compared correspondingly in order to see the effect of vitamin B12's presence on the functional conformational transition.

In terms of the RMSD from the target conformations, the simulations with vitamin B12 provided slightly better (0.2 Å smaller) final RMSDs in comparison with regular ANM-LD simulations, in all five parallel runs of both. This observation indicates that the presence of vitamin B12 in the simulations limitedly facilitate the transition to reach its final state. The exception is the simulation performed with 0.8 Å deformation factor (DF) in B12-including simulations. In this simulation, the transition was achieved by large hence coarse perturbations, which might hinder the local changes to be sampled efficiently. Therefore, DF values smaller than 0.8 Å should be chosen for future ANM-LD simulations of BtuCD system.

In addition to final RMSDs, pair C_α distance calculations for cytoplasmic gate II residue P84 also indicated to an improvement upon B12 involvement in the simulation in terms of reaching target distance in IF-model (Figure 5.10). The best matching distance value with target conformation was observed when B12 is included, all modes were considered for perturbation and DFs between 0.1-0.5 Å is applied (DFs descending as going from slow to fast ANM modes). These observations show that B12 presence might facilitate cytoplasmic gate opening.

The pair C_α distance of L172 observed in both B12 including and excluding simulations could not reach final distance corresponding to a more closed periplasmic state. The reason might be the ATP and/or BtuF-dependence of periplasmic gate closing. Thus, in future studies, ATP and BtuF might be included in the ATP BtuF-bound conformation. On the other hand, V170 pair distance decreased more than its expected distance especially in case of B12 absence. The presence of B12 resulted a more similar distance with target distance by keeping this region (lower part of periplasmic region) relatively more open. These results might also be

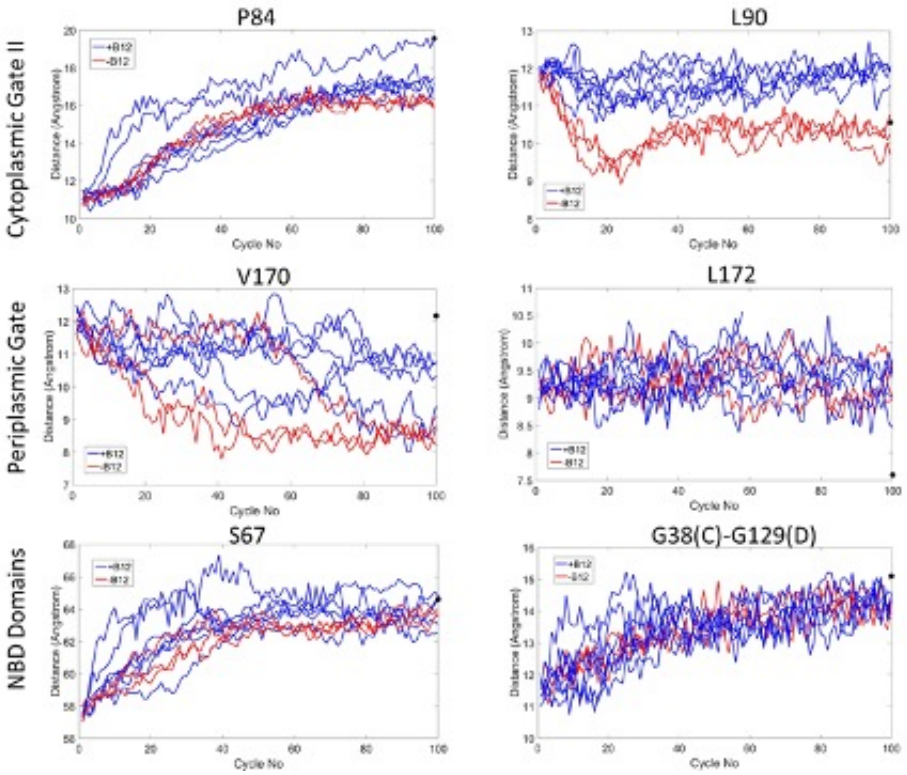

Figure 5.10 C_α distances from ATP BtuF-bound>IF-model transition are given for twin residues P84, L90, V170, L172, S67 and residues, G38-G129). Black spheres represent target distances.

potentially improved by consideration of ATP and BtuF in the simulations.

The opening of opposite nucleotide binding domains (NBDs), as observed in the pair C_α distance of S67, could be achieved in both B12 including and excluding simulations (Figure 5.10) with slightly better results when B12 is used. On the other hand, the opening of ATP sites which can be depicted by the C_α distance between ATP sites residues G38 and G129 of opposite domains could not be fully obtained in both B12-included and B12-excluded runs except for one B12 included simulation. Here again, the presence of B12 provide a slight improvement, but it might be further improved by consideration of ATP in ANM-LD simulations, since its binding is the main driver of this movement [83]. Here the BtuF is not expected to cause a large impact as observed from previous results that BtuF has only local effect on the protein [83].

In addition to showing a mediocre improvement in BtuCD dynamics captured by ANM-LD simulations, B12-including simulations also disclosed a TMD residue,

namely L90, which as a twin pair is located near the cytoplasmic gate II and block the cytoplasmic exit of translocation cavity, locking the vitamin inside (Figure 5.11). This pair has to open sideways for B12 to move toward the cytoplasm. The pairwise C_α distance between L90 residues through ANM-LD simulation of ATP BtuF-bound conformation to IF-model might provide useful information on mechanism of B12 release since L90 gate is not open in neither the initial nor the target conformations of BtuCD. Since IF is an *in silico* model, the target L90 pair distance must be considered with caution. But assuming it to be the correct model, then opening of L90 should be observed at intermediate conformations of the simulation. Although the target distance value is approximated better in B12-excluding simulations, the persistence of L90 pair openness is observed only when B12 is considered by hindering its closing (Figure 5.10, top right panel) in the simulations. Yet even the largest distance (12.5 Å) is still smaller than 15 Å which is approximately the size of vitamin B12. This result is consistent with the observation obtained from all B12-including simulations that B12 release to the cytoplasm was not achieved yet. This might also indicate to the potential flaws of the IF-model.

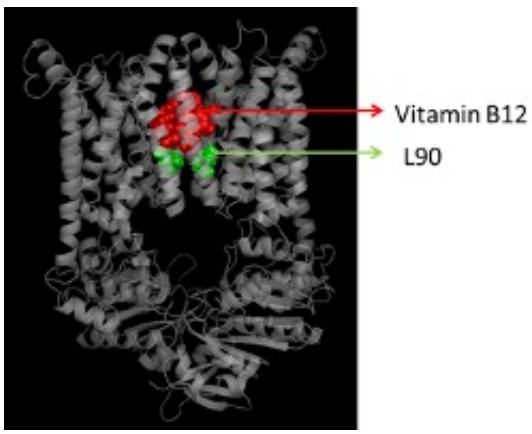

Figure 5.11 The last conformation obtained from ANM-LD trajectories of ATP BtuF-bound to IF-model transition. Vitamin B12 in red is occluded by L90 pair residues in green.

5.4.2.2 Untargeted ANM-LD Simulations with B12 Molecule. When the IF-model conformation was used as target in ANM-LD simulations, the movements of B12 molecule were restricted to the upper region of cytoplasmic gate II, due to the blockage by L90 residue pairs. Thus, the downward B12 transfer was failed to be

sampled with the consideration of difference vector between initial and target conformations only. However, the structure corresponding to B12-released BtuCD conformation in its early steps is not known and thus cannot replace IF-model target, yet the modularity of ANM-LD can at least be used to decipher potential ANM modes, which pushes B12 molecule toward the cytoplasm and provides L90 opening, for characterizing the B12 release dynamics. Therefore, only the ANM modes moving vitamin B12 downward were selected as modes to perturb the structure, based on their overlap with the B12's difference vector between periplasmic (initial) and cytoplasmic (target) positions obtained from B12 docked structures. No other target structural information was used to guide the simulations. Although the generated modes include 3N+2 (3N-6+8, N: residue number) elements, only 8 elements of the mode vector corresponding to B12 molecule was used for overlap calculation with the difference vector (between current initial and target structures) to decide best overlapping mode to perturb the structure.

In the vitamin B12 movement-directed ANM-LD simulations, B12 reaches its expected position at the cytoplasmic end of the BtuC, however resulting BtuCD conformations are distorted. Therefore, the simulations with smaller DF values and faster modes in mode pool for perturbations, might be better to minimize these distortions. Yet the deformation factors (DFs) smaller than 0.5 were not sufficient to generate enough structure perturbation to allow B12 transfer to be sampled. Among simulations with DF$\geq$0.5 Å, if the number of accessible modes (modemax) is equal to or larger than 30, the BtuCD protein acquires a compact shape. If protein was allowed to move only in first 10 or 20 modes, then the structure was extended at the NBD part. Final conformations of two runs with modemax=10, 30 and DF=1, 0.5 respectively are given on Figure 5.12, indicating to the observed distortions on the protein. The simulations were stopped once the vitamin B12 reaches its target position. Using set of ANM modes from a pool with less than 30 slow modes result in the opening of NBDs indicating to a rather disruptive departing of two BtuDs on ATP sites. On the other hand, using more than 30 slow modes results in a more compact and less distorted conformation which indicates to fine tuning roles of faster modes in BtuCD dynamics.

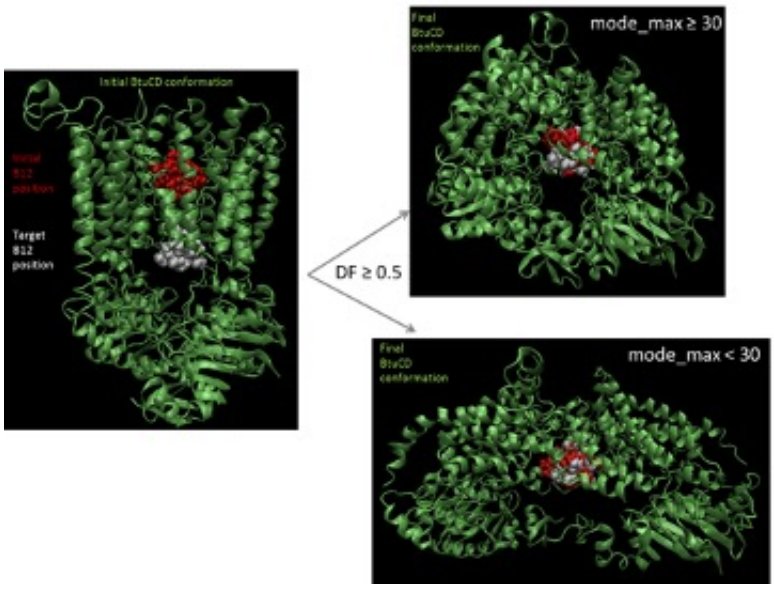

Figure 5.12 Initial and final conformations of two untargeted simulations with modemax=10, 30 and DF=1, 0.5 Å respectively are shown.

For the purpose of avoiding large distortions in BtuCD structures during ANM-LD simulations, mode pool for perturbation consisting of all available ANM modes was also tried. Indeed, using faster ANM modes intermittently during ANM-LD simulations improved the obtained conformations in terms of distortions by restricting the distortions to secondary structure of few sites. This might indicate to the need of local adjustments accompanying the global changes in order to concert B12 release to the cytoplasm.

In order to extract pre-existing modes of motion in the apo state of the protein which enables the B12 transfer, ANM-LD simulations were also performed on apo (1L7V) conformation with docked vitamin B12. As observed in ANM-LD simulations initiated from ATP/BtuF-bound conformation, large DF and small mode number available for mode selection cause NBD opening in extended form, while using relatively faster modes improves the conformations in the generated trajectory. This observation displays that the extreme opening of NBDs is not due to the ligand condition of the initial conformation.

Although the simulations on BtuF/ATP bound BtuCD and apo conformations provide similar results, how these results are achieved dynamically differ based on the utilized ANM modes. On Figures 5.13 and 5.14, the used ANM modes in respective B12 simulations are given for BtuF/ATP and apo conformations respectively. ANM modes 4 and 7 of B12 docked 4FI3 conformation are frequently selected in all parallel runs while restricting the mode pool to first 10 modes increases the use of mode 7 in ATP/BtuF-bound conformation (4FI3) to IF-model run. A relatively faster mode ANM 25 is also popularly selected which might have effect on the observed differences between the runs with different observations in compactness and extended NBD opening. Here it is important to note that these are the modes of B12 docked 4FI3 conformation. ANM modes 4 and 7 of this conformation correspond to 3 and 6 modes of apo (B12less) conformation. For apo (1L7V) to IF-model run, it is important to note that these are the modes of B12 docked apo conformation. ANM modes 6 and 8 of B12 bound conformation correspond to 6 and 7 modes of B12less apo conformation. Consequently, alow ANM mode 6 (as mapped to B12less apo structure) was common in both simulations. However, ANM mode 7 (according to B12less apo structure) was specific to the simulations on apo conformation. This mode is also intriguing since it is not reachable if BtuF is bound to the conformation as discussed in previous sections.

These findings indicate that three slow modes of motion have role in the downward movement of B12, ANM modes 3, 6 and 7 of apo conformation. However, for a proper release of B12, the counteracting faster modes of motion are also required to keep the protein intact and functional by local adaptations of the conformation. Further experimental outcomes on missing intermediate conformation corresponding to modeled IF BtuCD conformation, which just-released the vitamin B12, might be beneficial for further understanding of the substrate release by BtuCD protein.

In conclusion, the presence of B12 slightly improved the sampling of ATP/BtuF bound $>$ IF-model transition, indicating to the potential importance of incorporating substrate and ligands to ANM-LD simulations. In addition, as suggested above, dissecting modes of motion required for B12 push of BtuCD is an

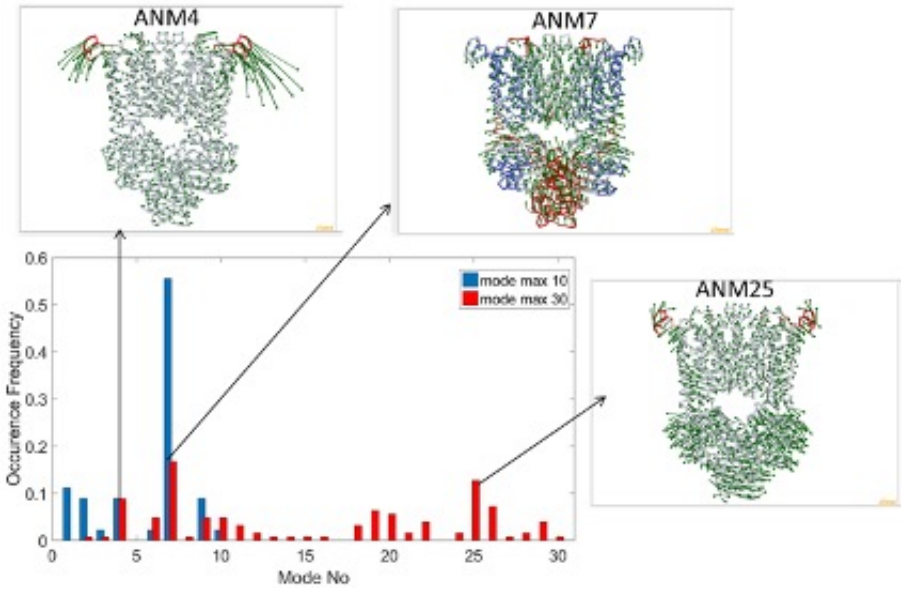

Figure 5.13 Selected ANM modes in ATP/BtuF-bound conformation (4FI3) to IF-model run. ANM modes 4 and 7 of the B12 docked conformation are frequently selected.

indication of another powerful aspect for ANM-LD simulations with ligands. For future studies, this might provide a way, by comparing ligand-free and ligand-including simulations, for deciphering functional modes of motion with potential roles other than substrate push. This might enable disclosing functional sites which also moves during these ligand movement or hinge sites making this movement possible.

5.5 Integration of Experimental Restraints into the ANM-LD Simulations

As observed from ANM-LD simulations of B12 docked BtuCD, the presence of correct experimental structures in the target as well as in the initial is beneficial for ensuring a biologically and physically relevant trajectory. However, the number of proteins in nature far exceeds the number of available structures [205]. Therefore, all available experimental information obtained from various methods are valuable for complementing computational tools to sample protein dynamics efficiently. Yet, resolution of the experimental data is also important for correct leading of molecular

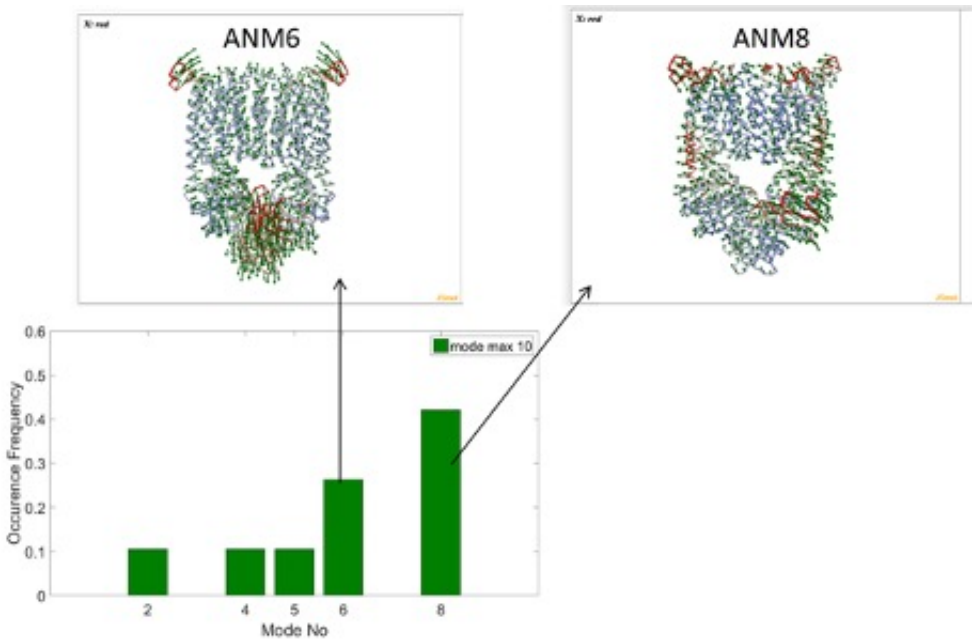

Figure 5.14 Selected ANM modes in apo (1L7V) to IF-model run (mode max 10). ANM modes 6 and 8 of B12 docked apo conformation are frequently selected.

simulations. In case of low resolution, collective variables representing the structural data can be used instead of the direct image or full atomistic structure of target conformation, to guide the simulations. In the present work, such collective variables obtained from atomic force microscopy (AFM) and small-angle X-ray scattering (SAXS) were implemented to the ANM-LD algorithm due to their working conditions similar to proteins' native environment, increasing popularity in recent years on protein studies hence resulting availability of the experimental variables [42, 159, 166].

5.5.1 Integration with AFM Restraints

In the present work, the AFM images of c-Src kinase and BtuCD proteins, as two case systems of this study, one as a globular and the other as a membrane system, were determined by our group (in collaboration with Hamdi Torun, Ph.D.) to be used in developing hybrid ANM-LD simulations combined with experimental restraints.

5.5.1.1 Obtained AFM Restraints for c-Src Kinase. Before performing the AFM experiments, the possible positionings of c-Src kinase on negatively charged mica surface were predicted for c-Src kinase using the surface charge of the protein as predicted by PBEQ Solver [172]. Using this algorithm, both active and inactive forms of the protein were used to calculate the electrostatic potentials, since both structures can be present in the sample solution. In the active form (PDB:1Y57), SH2 and kinase domains (KD) are possibly attached to the mica surface due to their relatively positive charge on their surface (Figure 5.15ab). In the inactive form (PDB: 2SRC) KD has more positively charged residues on the surface than other regions, thus KD might be responsible for mica attachment (Figure 5.15c). For these predictions, it is important to note that these calculations only consider protein-mica electrostatic interactions, however other environmental parameters such as salt concentration and temperature might have a role in determining the attached region to the mica surface.

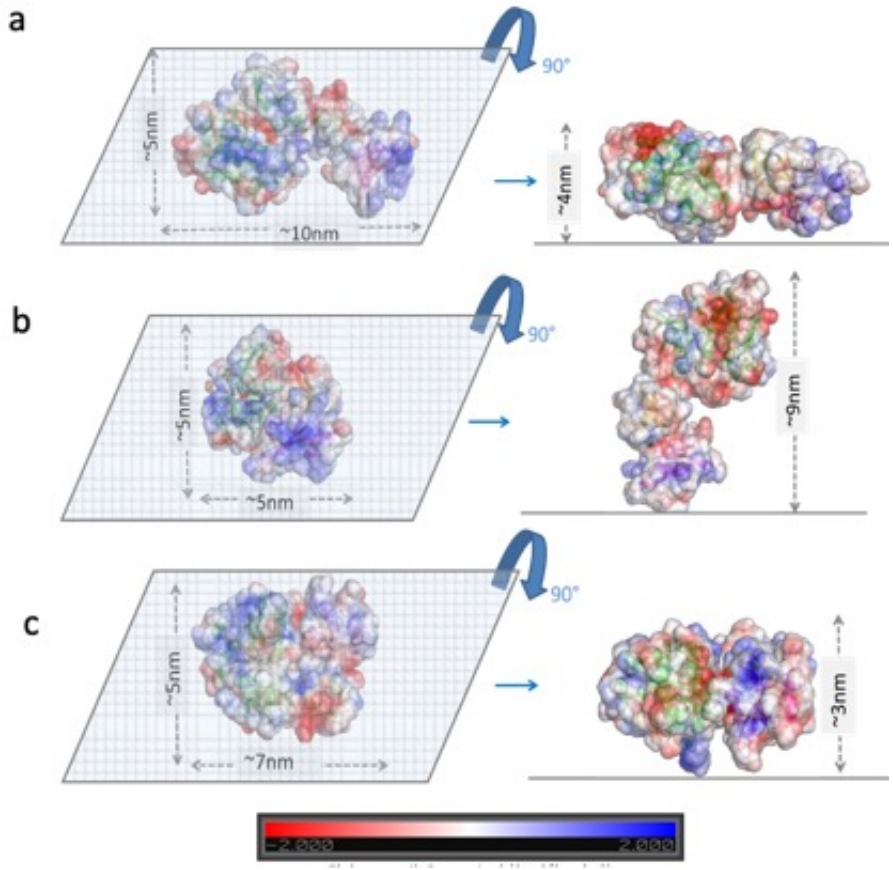

Figure 5.15 Surface representation of c-Src kinase, colored by electrostatic potentials (kcal/(mol.e)). a-b) For active conformation and c) inactive conformation the expected heights are given beside.

Having predicted heights in 3, 4 and 9 nm for different positionings of c-Src kinase (Figure 5.15), AFM imaging experiments were performed for 0.25 mg/mL untagged c-Src kinase which located on the mica surface either with using 2.5% glutaraldehyde fixation or physical adsorbance. 125 μM ATP was used to activate protein. The scanning of the surface was carried out by AFM contact mode of Dimension Edge, Bruker Nano (Santa Barbara, CA, USA). The resulting images are shown on Figure 5.16.

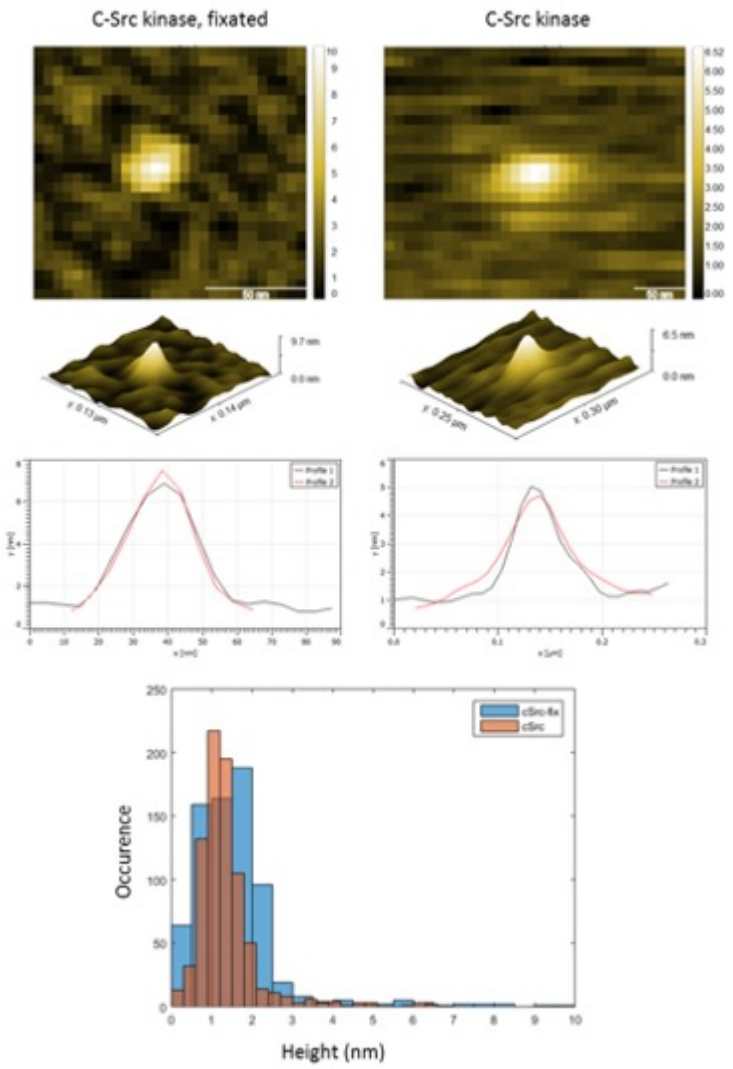

Figure 5.16 AFM images and height profiles obtained from c-Src kinase protein solution on mica surface.

The obtained AFM images showed 4nm height for c-src kinase which was one of the expected values in the z axis. At this axis, the AFM provides data in atomistic detail, so it can be concluded that the scanned protein is in the active conformation

and contacted the surface in both KD and SH2 domain. On the other hand, the x and y axes are challenging due to tip broadening. Indeed, observed x and y values around 30 nm are much larger than the expected width (5-7 nm). The application of %2.5 glutaraldehyde fixation improved the image, while providing x and y values around 10 nm, closer to the real values but still larger than expected due to tip broadening.

Since the resolution of the image is not high due to tip broadening, the observables which requires atomistic detail such as distances between residue C_α atoms cannot be extracted as collective variables. Thus, principal axes of c-Src kinase images were determined. The ANM modes which overlap most with the first three PC axes were selected in the simulations to guide the transition. Using first principal axis of c-Src kinase image (which corresponds to the active form) in ANM-LD simulations starting from inactive conformation provided 14 Å RMSD closeness to the active crystal conformation with PDB code 1Y57 (starting from 23 Å). When second and third principal axes are made use of, both second alone and second, third together, RMSD from the target conformation was decreased 5 Å (Figure 5.17). This indicates that the second principal component axis contributes to the observed improvement, disclosing first two principal components beneficial to guide the ANM-LD simulations. However, since the full approach to the target conformation could not be achieved, principal components of obtained AFM images might better be complementary to other observable variables than be used alone. Still, the improvement of the image might enhance the results. In conclusion, principal components can be useful for detecting collective variables from experimental images of proteins (from AFM and other techniques such as electron microscopy) to direct the simulations particularly for large conformational changes.

5.5.1.2 Obtained AFM Restraints for BtuCD. As done for c-Src kinase, in order to interpret the prospective AFM images, potential positioning of BtuCD-F protein complex (PDB code: 2QI9) on mica surface was evaluated based on the charge of surface residues. The residue charges were determined by calculation of electrostatic potentials (kcal/mol.e) using PBEQ Solver [172]. The resulting charges

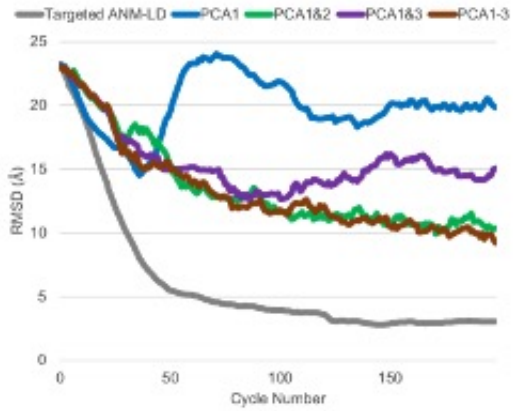

Figure 5.17 RMSD values of integrated ANM-LD simulations based on AFM variables for c-Src kinase.

indicate to an asymmetric positively charged area on BtuC possibly due to BtuF-binding. Combining with His-tags on the N-termini of BtuCs, the protein complex is likely to attach the mica surface through this area. Then the potential heights would be around 8 nm and if BtuCD binds to BtuF, then the heights are around 10 nm.

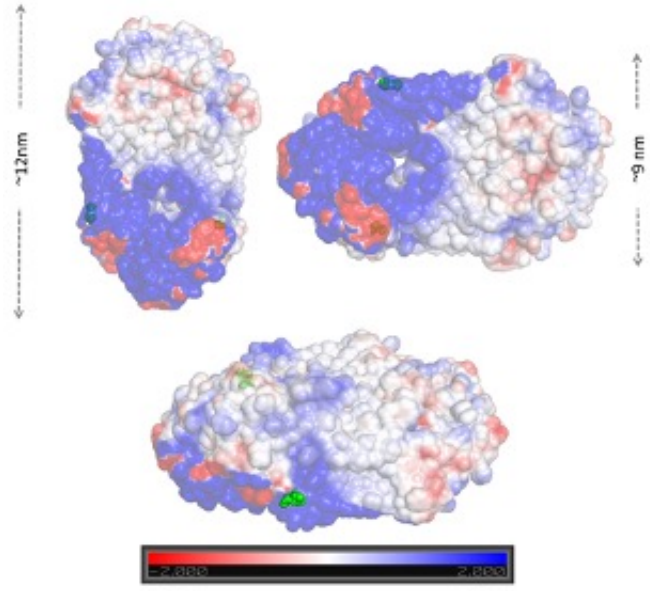

Figure 5.18 Surface representation of BtuCD-F complex colored by electrostatic potentials (green sphere: His-tag).

For AFM imaging, 1 mg/ml BtuC protein His-tagged on its N-terminal and 1 mg/ml untagged BtuD protein were mixed to form a complex structure for BtuCD in

FOS-12 detergent. The resulting mixture is incubated on the freshly cleaved mica for 30 minutes. Height profiles of the protein were obtained with and without its capping protein BtuF in 0.25 mg/ml concentration (Figure 5.19) using AFM contact mode of Dimension Edge, Bruker Nano (Santa Barbara, CA, USA). The resulting images are given on Figure 5.19).

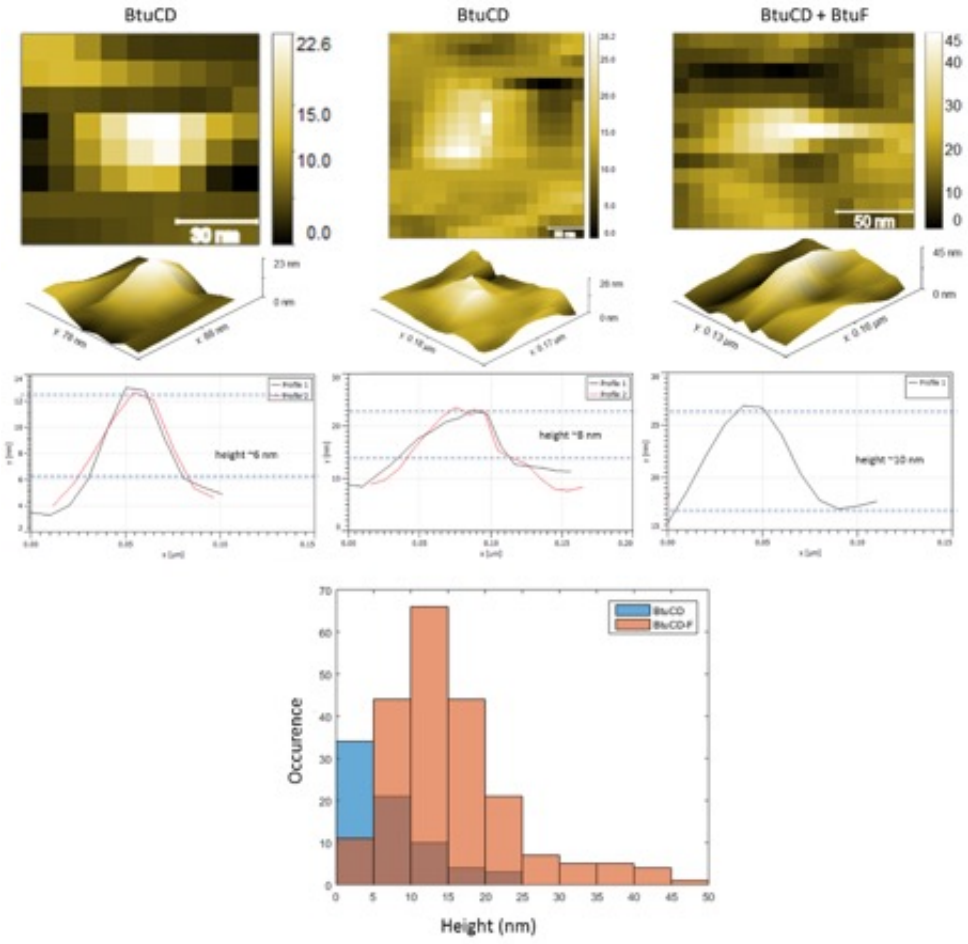

Figure 5.19 AFM images and height profiles of BtuCD and BtuCD-F complexes.

In the present work, obtained AFM images of both BtuCD and BtuCD-F display heights in accordance with the expected values. The heights consistent with the available data are indication of correct imaging. Additionally, from the obtained images the binding of BtuF could be captured from the observed heights of the complex as an increase from 8nm to 10 nm upon the addition of BtuF to the surface. After evaluation of the heights, the resulting images were used to extract PCA-based collective variables for the ANM-LD algorithm.

Principal component analyses (PCA) of the obtained images of BtuF bound BtuCD were performed and first three principal axes were determined. These axes were used to guide the simulations in the most overlapping ANM modes from apo to BtuF-bound conformation. However, no decrease was observed in RMSDs from the target conformation. This might be due to the already low RMSD (3.5 Å) between the apo BtuCD and BtuCD-F conformations which might give similar principal axes for first three components. Thus, higher principal axes might be beneficial in this regard to reflect the differences between initial and target conformations. Obtaining higher resolution images might also improve the effect of the experimental restraints in future studies. The presence of detergent instead of lipid environment might be another reason for low-resolution images and liposome embedded BtuCD-F might be an alternative for AFM imaging in future studies.

5.5.2 Integration with SAXS Data

From SAXS experiments of a protein, both its volume information and SAXS intensity profiles of the scattered rays from its sample can be extracted and used as collective variables for leading ANM-LD simulations. Although SAXS equipment is not available at our laboratory, Crysol [168] algorithm, which predicts SAXS intensity profiles and volume data out of available X-ray crystal structures, can be useful in providing ideal observables for developing hybrid simulations with ANM-LD method. Toward this aim, the active state of c-Src kinase was simulated by ANM-LD simulations guided by the volume and SAXS profile changes between active and inactive states separately, instead of the conventionally used difference vector. The ANM modes providing the proposed change in these two parameters are selected for perturbation of the protein and minimization and LD parts follow this step as usual.

Volume data for guiding ANM-LD simulation was shown not to be efficient, since no decrease in reached RMSDs from the target conformation was reached. This parameter may be insufficient to summarize complex conformational changes occur during the functional transition. For example, the volume data might not change

much in the initial cycles yet increase/decrease toward the target conformation. Having imposed the volume change in the beginning of the simulations might cause the inefficacy of this parameter in leading the simulations.

Another parameter to be used in SAXS-hybrid ANM-LD simulations is the change in SAXS profiles. The resemblance of a conformation to a SAXS profile is defined by the χ^2 calculations. Low χ^2 values correspond to high resemblance. Thus, in this study ANM modes which lead to the target SAXS profiles by decreasing χ^2 values of sampled intermediate conformations were used to guide the simulations. Only 3 Å drop is observed from the initial RMSD. These findings show that both volume and SAXS profile data are not sufficient to guide ANM-LD simulations toward a target conformation. However, they can be used to supplement regular ANM-LD simulations to diversify the sampled trajectory on the energy landscape. With this in mind, change in SAXS intensity profiles was used in intermittent cycles in combination with difference vector. Although it has no effect on performance in terms of reached RMSD to the target conformation, yet using this parameter changes the transition dynamics, by delaying the transition as shown on Figure 5.20. This might provide sampling of alternative pathways, which cannot be detected using target structure alone. Other than RMSD, χ^2 can also be used to assess the trajectory obtained from combined ANM-LD simulations with both SAXS profile and difference vector. The general profiles of both RMSD and χ^2 are similar, and they converge at the similar cycles (180 cycles). Since the final χ^2 is low (near zero), the trajectory is approved by another term other than RMSD to reach the final conformation. However, χ^2 profile differs from RMSD plot at 20-40 and 80-90 cycles with high χ^2 values in contrast with decreasing RMSD. The conformations at these χ^2 peaks as transient deviations from target conformation might correspond to two potential intermediate states mentioned in the literature [200].

5.6 Conclusion

The integration approaches discussed above, set foundation for further manipulations of the algorithm to use ever increasing experimental data for guiding

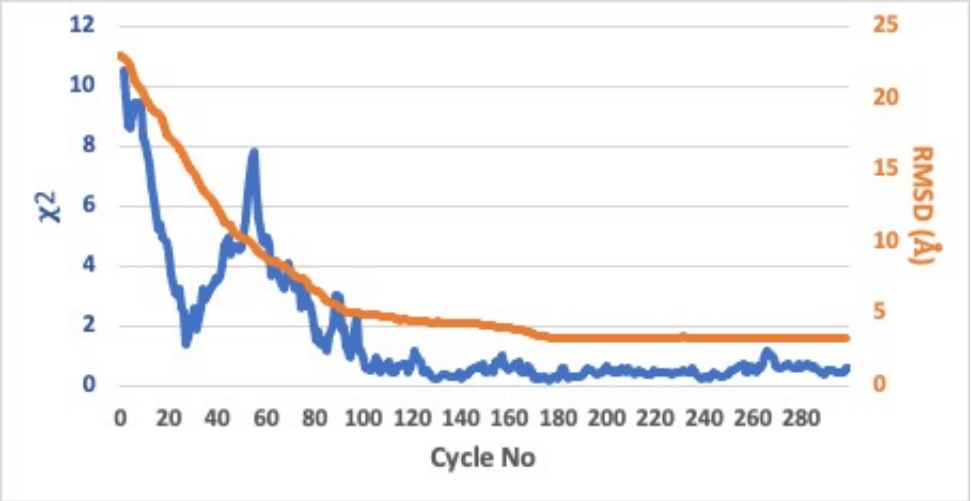

Figure 5.20 RMSD and χ^2 plot of combined ANM-LD simulations with SAXS profiles and target conformations.

the ANM-LD simulations. In future works, maximum entropy principle which aims the selection of the simulation model that contains the least amount of information might be applied for the ANM-LD and experimental data integration.

6. CONCLUSION

ANM-LD is an in-house developed enhanced sampling method to study conformational changes of proteins up to large-scaled and long-time motions. Driven only by intrinsic modes of motion, it enables to study functional conformational transition pathways between two known structures and to disclose inherent working mechanism of proteins. In first part of this study, the applicability of ANM-LD on various protein systems, including the ones deviating from globular features, were evaluated and then assessed in the example of membrane-embedded vitamin B12 importer BtuCD. Possible offerings of the ANM-LD results on the functional mechanism of proteins were estimated in this case system. Having validated ANM-LD as a methodological platform to infer knowledge on protein dynamics, the second part of the thesis focused on the advancement of ANM-LD algorithm to meet the requirements for covering multifarious aspects of protein dynamic studies such as broadening the sampled free energy landscape, detection of rare conformational states, capturing alternative transition pathways, increasing the resolution of conformational representations, incorporating ligand molecules and ensuring a biological and physical relevance of the transition pathway.

ANM-LD simulations of BtuCD importer between its different functional states showed that ANM-LD was capable of extracting biologically and physically meaningful dynamic information for even non-globular systems, as the obtained dynamical features agreed well with the experimental findings. As shown by the previous experiments, the residual fluctuations along the sampled trajectory and their coupling among each other indicated that ATP has a global effect on BtuCD dynamics, whereas substrate binding protein BtuF has local impact upon binding. Once the quality of the obtained ANM-LD trajectory is assured, the resulting set of transient conformations can be used to infer functional dynamic features of the protein such as order of events and mechanism of action.

Apart from yielding a conformational trajectory, due to its working setup, an ANM-LD simulation also discloses a set of ANM modes corresponding to complex transition dynamics of a protein, since the simulation is led by ANM modes. This feature enables ANM-LD to dissect complex transition dynamics into elemental normal modes of motion. The underlying dynamic infrastructure of these modes, which is formed by the relevant hinge residues, serves as a dynamic allosteric network of residues that have role in the conformational transition hence functional mechanism. Additionally, a perturbation response score (i.e., cumulative angular deviation score) was developed to assess the dynamic deviations from WT protein caused by *in silico* mutations on these hinge sites. Conformational trajectories of mutant and WT protein are needed to predict and rank the functional effect of mutations. The most deviant mutants are expected to be the most functionally affected. Herein, the trajectories were provided by ANM-LD simulations of mutants and WT proteins, yet they could also be provided from other techniques such as MD. This approach provides computational means to guess the dynamical and functional outcomes of single point mutations. Both extracted network of hinge residues and estimated outcomes of their possible mutations are useful for saving resources and time as providers of preliminary data instead of random point mutations to detect functional residues.

Another dynamical feature to be deduced from the obtained conformational trajectory is the directionality in allosteric signaling extracted by assuming a time delay between correlated residue fluctuations. The sources of allosteric signal can be good targets for functional control and manipulation of the protein. The appearing directionality is a result of the causal interactions between residues as imposed by the physicodynamic characteristics of used normal modes during the simulation. This causality can be calculated by GNM transfer entropy method which made available on an online server (http://safir.prc.boun.edu.tr/gnmtransentropy/home.html) as a contribution to scientific community.

Besides the versatility of its results that allow many dynamic properties of proteins to be studied, the modularity of ANM-LD algorithm also makes it available

to be advanced in various aspects. Firstly, knowing the normal modes guiding the relevant transition provide means to manipulate the transition dynamics by blocking certain modes of motion. This might also help to decipher how crucial a mode is to a conformational transition. Because the absence of some modes hinders the transition, while of some only changes its characteristics. Another way of enhancing the sampling of transition dynamics is by changing the mode selection algorithm in ANM-LD. Instead of choosing best overlap with the difference vector between initial and target conformations, a Monte Carlo based selection algorithm can be used to introduce stochasticity in the sampling algorithm. Using different sets of modes for guiding the simulations, parallel transition pathways can be acquired.

Another advancement on ANM-LD algorithm was to increase resolution of the structure representations used for ANM calculation step. By using all heavy atoms in ANM calculations, the resulting modes make local adjustments possible while performing a collective conformational change during the simulations. In case of c-Src kinase, using all-atom representation improved the observed H-bond shifts in favor of the respective patterns in the literature. Additionally, in order to enable direct comparison between ANM and GNM mode shapes, all-atom GNM algorithm was also developed and made available online (http://safir.prc.boun.edu.tr/allatomgnm/) for future applications.

An additional upgrade on the algorithm was to incorporate ligand molecules in ANM-LD simulations. Although, proteins are potentially capable of expressing all fluctuations in their apo form, certain conformational transitions might become more favorable in the presence of a ligand. Therefore, the ligand incorporation might facilitate certain dynamic events to be captured during the simulation. Herein adding ATP molecules in c-Src kinase simulations greatly expanded the sampled free energy landscape defined by H-bonding patterns and A-loop flexibility. Besides, both of two intermediate conformations between inactive and active conformations defined in the literature could be captured by consideration of ATP and Mg^{++}. In another system BtuCD, presence of the substrate vitamin B12, improved the observed translocation cavity dynamics (opening and closing behavior to let the vitamin pass through) in

terms of sizes of periplasmic and cytoplasmic openings. These results indicated that adding ligands enhanced sampling capacity of ANM-LD simulations. Consequently, instead of high number of parallel simulations to observe a dynamic event, a ligand introduction might facilitate it due to the accumulation of energetic changes during LD parts of each cycle.

A major improvement in ANM-LD algorithm was to make it compatible for working with experimental restraints. In the absence of a high-resolution target conformation, these restraints can be of good use to guide the simulations in a biologically meaningful transition pathway. Even if the target conformation is present, these restraints can still be valuable to achieve alternative transition pathways to the ones observed in regular ANM-LD trajectories. Herein both AFM and SAXS data were incorporated into ANM-LD algorithm, yet other experimental methods could also be utilized.

In addition to development of an efficient sampling algorithm, this thesis also provided more information on functional dynamics of two studied case systems c-Src kinase and BtuCD. In c-Src kinase, the normal modes of motion shown to be relevant for functional conformational transitions appeared to be rendered by hinge residues overlapping with cancer associated mutation sites. The hinge sites were also shown to be functional in BtuCD, as their mutations resulted in impairment of function. This supports the role of dynamics and the relevant infrastructure enabling these fluctuations in function and hence present them as potential targets to control function and repair it when necessary. Besides, protein dynamics are responsive, from a pool of potential modes of motion, they undergo the ones related to its current environment (i.e., ligand presence, partner interaction, pH or temperature). This adaptability is encoded in the structure. In BtuCD, changing mode pool to guide ANM-LD simulations by blocking a functional mode, without changing the target conformation (ATP-bound), resulted an alternative transition pathway and dynamic behavior leading to another conformation (ATP & BtuF-bound) as if it was used as target. This shows that the relevant mode to lead to this target conformation is already present in apo conformation yet becomes prominent in response to the environment by recruiting different modes of motion. This

orchestration of modes, while some emphasized and some underrepresented, changing roles occasionally, enable proteins to perform mechanistic steps required for function.

In conclusion, ANM-LD is a prosperous enhanced sampling algorithm providing information on various aspects of protein dynamics and functional mechanism. Its modularity enables the use of different approaches for obtaining alternative pathways of transition and improving sampling. Thus for future studies, it will be a good base as an enhanced sampling method for further implementation such as addition of lipids as a membrane environment and use of other experimental data. Finally, ANM-LD can also be useful for studying other biological systems than proteins, including RNA, lipids and carbohydrates.

7. FUTURE WORK

This thesis mainly focused on to show and expand the applicability of the enhanced sampling method ANM-LD, particularly on proteins deviating from globular proteins. Having shown its potential use, in future studies, the number and variety of nonglobular proteins can be increased along with other biological macromolecules such as nucleotides to be tested by ANM-LD simulations. This aim will be easier as ANM-LD algorithm will be available as an online server, open to the use of other scientists in addition to another server for GNM transfer entropy calculations.

The algorithm itself can be improved by incorporating energetic criteria into the mode selection and structure perturbation parts of the method and by definition of explicit lipid molecules as a surrounding membrane for related systems. For hybrid ANM-LD simulations, experimental restraints obtained from methods other than AFM and SAXS can be used. Using successful restraints might also put the necessity of a target conformations away for performing ANM-LD simulations.

8. LIST OF PUBLICATIONS PRODUCED FROM THE THESIS

1. Structural insight into host plasma membrane association and assembly of HIV-1 Matrix protein, H. Ciftci, H. Tateishi, K. Koiwai, R. Koga, K. Anraku, K. Monde,C. Dag, E. Destan, B. Yuksel, E. Ayan, G. Yildirim, M. Yigin,F. B. Ertem, A. Shafiei, O. Guven, S. O. Besler, R. G. Sierra, C. H. Yoon, Z. Su, M. Liang, B. Acar, T. Haliloglu, M. Otsuka, F. Yumoto, M. Fujita, T. Senda, H. DeMirci, *to be published*, 2021.

2. Distinct Allosteric Networks Underlie Mechanistic Speciation of ABC Transporters, B. Acar, J. Rose, B. A. Fas, N. Ben-Tal, O. Lewinson, T. Haliloglu, *Structure*, Volume 28, Issue 6, pp. 651-663.e5, 2 June 2020.

3. Single-molecule probing of the conformational homogeneity of the ABC transporter BtuCD, M. Yang, N. L. Levanon, B. Acar, B. A. Fas, G. Masrati, J. Rose, N. Ben-Tal, T. Haliloglu, Y. Zhao, O. Lewinson, *Nature Chemical Biology*, Vol. 14, pp. 715â722, 2018.

4. Collective Dynamics Underlying Gating Mechanism of Repeat Protein FimD, B. Acar, B. A. Fas, T. Haliloglu, *2nd NGP-NET Symposium*, pp. 72, 15-17 September 2016.

5. Allostery through Collective Dynamics in PapC Usher, B. Acar, B. A. Fas, T. Haliloglu, *1st NGP-NET Symposium*, pp. 55, 6-9 October 2015.

REFERENCES

1. Vendruscolo, M., "Determination of conformationally heterogeneous states of proteins," *Current Opinion in Structural Biology*, Vol. 17, pp. 15–20, Feb 2007.

2. Boehr, D. D., R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," *Nature Chemical Biology*, Vol. 5, no. 11, pp. 789–796, 2009.

3. Zhuravlev, P. I., and G. A. Papoian, "Protein functional landscapes, dynamics, allostery: A tortuous path towards a universal theoretical framework," *Quarterly Reviews of Biophysics*, Vol. 43, pp. 295–332, Aug 2010.

4. Osawa, M., K. Takeuchi, T. Ueda, N. Nishida, and I. Shimada, "Functional dynamics of proteins revealed by solution nmr," *Current Opinion in Structural Biology*, Vol. 22, pp. 660–669, Oct 2012.

5. Campitelli, P., T. Modi, S. Kumar, and S. B. Ozkan, "The role of conformational dynamics and allostery in modulating protein evolution," *Annual Review of Biophysics*, Vol. 49, pp. annurev–biophys–052118–115517, May 2020.

6. Henzler-Wildman, K., and D. Kern, "Dynamic personalities of proteins," *Nature*, Vol. 450, pp. 964–972, Dec 2007.

7. Klepeis, J. L., K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," *Current Opinion in Structural Biology*, Vol. 19, pp. 120–127, Apr 2009.

8. Villali, J., and D. Kern, "Choreographing an enzyme's dance," *Current Opinion in Chemical Biology*, Vol. 14, pp. 636–643, Oct 2010.

9. Wand, A. J., "The dark energy of proteins comes to light: Conformational entropy and its role in protein function revealed by nmr relaxation," *Current Opinion in Structural Biology*, Vol. 23, pp. 75–81, Feb 2013.

10. Chen, Z., and A. P. Zeng, "Protein engineering approaches to chemical biotechnology," *Current Opinion in Biotechnology*, Vol. 42, pp. 198–205, Dec 2016.

11. Campbell, E., M. Kaltenbach, G. J. Correy, P. D. Carr, B. T. Porebski, E. K. Livingstone, L. Afriat-Jurnou, A. M. Buckle, M. Weik, F. Hollfelder, and et al., "The role of protein dynamics in the evolution of new enzyme function," *Nature Chemical Biology*, Vol. 12, pp. 944–950, Nov 2016.

12. Lee, Y., R. Lazim, S. J. Y. Macalino, and S. Choi, "Importance of protein dynamics in the structure-based drug discovery of class a g protein-coupled receptors (gpcrs)," *Current Opinion in Structural Biology*, Vol. 55, pp. 147–153, Apr 2019.

13. Perperopoulou, F., F. Pouliou, and N. E. Labrou, "Recent advances in protein engineering and biotechnological applications of glutathione transferases," *Critical Reviews in Biotechnology*, Vol. 38, pp. 511–528, May 2018.

14. Sledz, P., and A. Caflisch, "Protein structure-based drug design: from docking to molecular dynamics," *Current Opinion in Structural Biology*, Vol. 48, pp. 93–102, Feb 2018.

15. Mannige, R. V., "Dynamic new world: Refining our view of protein structure, function and evolution," *Proteomes*, Vol. 2, pp. 128–153, Mar 2014.

16. Ng, P. C., and S. Henikoff, "Predicting the effects of amino acid substitutions on protein function.," *Annual review of genomics and human genetics*, Vol. 7, pp. 61–80, Jan 2006.

17. Kollmar, M., "Fine-tuning motile cilia and flagella: Evolution of the dynein motor proteins from plants to humans at high resolution," *Molecular Biology and Evolution*, Vol. 33, pp. 3249–3267, Dec 2016.

18. Thorvaldsen, S., and O. Hössjer, "Using statistical methods to model the fine-tuning of molecular machines and systems," *Journal of Theoretical Biology*, Vol. 501, p. 110352, Sep 2020.

19. Daniels, K. G., Y. Suo, and T. G. Oas, "Conformational kinetics reveals affinities of protein conformational states," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 112, pp. 9352–9357, Jul 2015.

20. Ha, J. H., and S. N. Loh, "Protein conformational switches: From nature to design," *Chemistry - A European Journal*, Vol. 18, pp. 7984–7999, Jun 2012.

21. Plattner, N., and F. Noé, "Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and markov models," *Nature Communications*, Vol. 6, pp. 1–10, Jul 2015.

22. Das, N., and P. Sen, "Structural, functional, and dynamical responses of a protein in a restricted environment imposed by macromolecular crowding," *Biochemistry*, Vol. 57, pp. 6078–6089, Oct 2018.

23. Steinberg, B., and M. Ostermeier, "Environmental changes bridge evolutionary valleys," *Science Advances*, Vol. 2, p. e1500921, Jan 2016.

24. Guarnera, E., and I. N. Berezovsky, "Allosteric sites: Remote control in regulation of protein activity," *Current Opinion in Structural Biology*, Vol. 37, pp. 1–8, Apr 2016.

25. Lee, M. J., and M. B. Yaffe, "Protein regulation in signal transduction," *Cold Spring Harbor Perspectives in Biology*, Vol. 8, p. a005918, Jun 2016.

26. Li, G., D. Magana, and R. B. Dyer, "Anisotropic energy flow and allosteric ligand binding in albumin," *Nature communications*, Vol. 5, p. 3100, Jan 2014.

27. Tsai, C.-J., S. Kumar, B. Ma, and R. Nussinov, "Folding funnels, binding funnels, and protein function," *Protein Science*, Vol. 8, pp. 1181–1190, Jan 1999.

28. Damry, A. M., M. M. Mayer, A. Broom, N. K. Goto, and R. A. Chica, "Origin of conformational dynamics in a globular protein," *Communications Biology*, Vol. 2, pp. 1–10, Dec 2019.

29. Maisuradze, G. G., A. Liwo, and H. A. Scheraga, "Relation between free energy landscapes of proteins and dynamics," *Journal of Chemical Theory and Computation*, Vol. 6, pp. 583–595, Feb 2010.

30. Kabir, K. L., N. Akhter, and A. Shehu, "From molecular energy landscapes to equilibrium dynamics via landscape analysis and markov state models," in *Journal of Bioinformatics and Computational Biology*, Vol. 17, World Scientific Publishing Co. Pte Ltd, Dec 2019.

31. Sfriso, P., A. Emperador, L. Orellana, A. Hospital, J. L. GelpÃ, and M. Orozco, "Finding conformational transition pathways from discrete molecular dynamics simulations," *Journal of Chemical Theory and Computation*, Vol. 8, pp. 4707–4718, Nov 2012.

32. Gershenson, A., L. M. Gierasch, A. Pastore, and S. E. Radford, "Energy landscapes of functional proteins are inherently risky," *Nature Chemical Biology*, Vol. 10, pp. 884–891, Nov 2014.

33. Guo, Y., M. Duan, and M. Yang, "The observation of ligand-binding-relevant open states of fatty acid binding protein by molecular dynamics simulations and a markov state model," *International Journal of Molecular Sciences*, Vol. 20, Jul 2019.

34. Chong, S. H., and S. Ham, "Distinct role of hydration ater in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis," *Accounts of Chemical Research*, Vol. 48, pp. 956–965, Apr 2015.

35. Cooper, D. R., P. J. Porebski, M. Chruszcz, and W. Minor, "X-ray crystallography: Assessment and validation of proteinsmall molecule complexes for drug discovery," *Expert Opinion on Drug Discovery*, Vol. 6, pp. 771–782, Aug 2011.

36. Smyth, M. S., and J. H. Martin, "x ray crystallography," *Journal of Clinical Pathology - Molecular Pathology*, Vol. 53, no. 1, pp. 8–14, 2000.

37. Rambo, R. P., and J. A. Tainer, "Accurate assessment of mass, models and resolution by small-angle scattering," *Nature*, Vol. 496, pp. 477–481, Apr 2013.

38. Tikole, S., V. Jaravine, V. Y. Orekhov, and P. Güntert, "Effects of nmr spectral resolution on protein structure calculation," *PLoS ONE*, Vol. 8, p. e68567, Jul 2013.

39. Callaway, E., "Revolutionary cryo-em is taking over structural biology," *Nature*, Vol. 578, p. 201, Feb 2020.

40. Casanal, A., S. Shakeel, and L. A. Passmore, "Interpretation of medium resolution cryoem maps of multi-protein complexes," *Current Opinion in Structural Biology*, Vol. 58, pp. 166–174, Oct 2019.

41. Sarkar, R., "Native flexibility of structurally homologous proteins: Insights from anisotropic network model," *BMC Biophysics*, Vol. 10, Jan 2017.

42. De Pablo, P. J., "Introduction to atomic force microscopy," *Methods in Molecular Biology*, Vol. 783, pp. 197–212, 2011.

43. de Pablo, P. J., *The application of atomic force microscopy for viruses and protein shells: Imaging and spectroscopy*, Vol. 105, pp. 161–187. Academic Press Inc., Jan 2019.

44. Acbas, G., K. A. Niessen, E. H. Snell, and A. G. Markelz, "Optical measurements of long-range protein vibrations," *Nature Communications*, Vol. 5, pp. 1–7, Jan 2014.

45. Wei, L., L. Yu, H. Jiaoqi, H. Guorong, Z. Yang, and F. Weiling, "Application of terahertz spectroscopy in biomolecule detection," *Frontiers in Laboratory Medicine*, Vol. 2, pp. 127–133, Jun 2019.

46. Opella, S. J., and F. M. Marassi, "Applications of nmr to membrane proteins," *Archives of Biochemistry and Biophysics*, Vol. 628, pp. 92–101, Aug 2017.

47. Coskuner, O., and V. N. Uversky, *Intrinsically disordered proteins in various hypotheses on the pathogenesis of Alzheimer's and Parkinson's diseases*, Vol. 166, pp. 145–223. Elsevier B.V., Jan 2019.

48. Godwin, R., W. Gmeiner, and F. R. Salsbury, "Importance of long-time simulations for rare event sampling in zinc finger proteins," *Journal of Biomolecular Structure and Dynamics*, Vol. 34, pp. 125–134, Jan 2016.

49. Zhou, H., and P. Tao, "Dynamics sampling in transition pathway space," *Journal of Chemical Theory and Computation*, Vol. 14, pp. 14–29, Jan 2018.

50. Adcock, S. A., and J. A. McCammon, "Molecular dynamics: Survey of methods for simulating the activity of proteins," *Chemical Reviews*, Vol. 106, pp. 1589–1615, May 2006.

51. Fersht, A. R., "Profile of martin karplus, michael levitt, and arieh warshel, 2013 nobel laureates in chemistry.," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 110, no. 49, pp. 19656–19657, 2013.

52. Baweja, L., and J. Roche, "Pushing the limits of structure-based models: Prediction of nonglobular protein folding and fibrils formation with go-model simulations," *Journal of Physical Chemistry B*, Vol. 122, pp. 2525–2535, Mar 2018.

53. Robustelli, P., S. Piana, and D. E. Shaw, "Developing a molecular dynamics force field for both folded and disordered protein states," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 115, pp. E4758–E4766, May 2018.

54. Best, R. B., W. Zheng, and J. Mittal, "Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association," *Journal of Chemical Theory and Computation*, Vol. 10, pp. 5113–5124, Nov 2014.

55. Nerenberg, P. S., B. Jo, C. So, A. Tripathy, and T. Head-Gordon, "Optimizing solute-water van der waals interactions to reproduce solvation free energies," *Journal of Physical Chemistry B*, Vol. 116, pp. 4524–4534, Apr 2012.

56. Alves Da Silva, R., L. Degrève, and A. Caliri, "Lmprot: An efficient algorithm for monte carlo sampling of protein conformational space," *Biophysical Journal*, Vol. 87, pp. 1567–1577, Sep 2004.

57. Kiderat, A., *Enhanced conformational sampling in Monte Carlo simulations of proteins: Application to a constrained peptide (entropy-sampling Monte Carlo/scaled-collective-variable Monte Carlo/cell-adhesive Arg-Gly-Asp sequence)*, Vol. 92, 1995.

58. Aiello, O. E., and M. A. Da Silva, "New approach to dynamical monte carlo methods: Application to an epidemic model," *Physica A: Statistical Mechanics and its Applications*, Vol. 327, pp. 525–534, Sep 2003.

59. Tian, P., K. Lindorff-Larsen, W. Boomsma, M. H. Jensen, and D. E. Otzen, "A monte carlo study of the early steps of functional amyloid formation," *PLOS ONE*, Vol. 11, p. e0146096, Jan 2016.

60. Haliloglu, T., and I. Bahar, "Adaptability of protein structures to enable functional interactions and evolutionary implications," *Current Opinion in Structural Biology*, Vol. 35, pp. 17–23, Dec 2015.

61. Haliloglu, T., I. Bahar, and B. Erman, "Gaussian dynamics of folded proteins," *Physical Review Letters*, Vol. 79, no. 16, pp. 3090–3093, 1997.

62. Atilgan, A., S. Durell, R. Jernigan, M. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," *Biophysical Journal*, Vol. 80, pp. 505–515, Jan 2001.

63. Kim, M. K., R. L. Jernigan, and G. S. Chirikjian, "Efficient generation of feasible pathways for protein conformational transitions," *Biophysical Journal*, Vol. 83, no. 3, pp. 1620–1630, 2002.

64. Lee, B. H., S. Seo, M. H. Kim, Y. Kim, S. Jo, M.-k. Choi, H. Lee, J. B. Choi, and M. K. Kim, "Normal mode-guided transition pathway generation in proteins," *PLOS ONE*, Vol. 12, p. e0185658, Oct 2017.

65. Mitsutake, A., Y. Mori, and Y. Okamoto, "Enhanced sampling algorithms," *Methods in Molecular Biology*, Vol. 924, pp. 153–195, 2013.

66. Papaleo, E., "Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: Strength in unity," *Frontiers in Molecular Biosciences*, Vol. 2, May 2015.

67. Pietrucci, F., "Strategies for the exploration of free energy landscapes: Unity in diversity and challenges ahead," *Reviews in Physics*, Vol. 2, pp. 32–45, Nov 2017.

68. Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, Vol. 220, pp. 671–680, May 1983.

69. Berhanu, W., P. Jiang, and U. H. E. Hansmann, *Enhanced Sampling for Biomolecular Simulations*, pp. 249–267. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

70. Ekimoto, T., and M. Ikeguchi, "Multiscale molecular dynamics simulations of rotary motor proteins," *Biophysical Reviews*, Vol. 10, pp. 605–615, Apr 2018.

71. Fyta, M., and M. Fyta, *Multiscale, hybrid and coarse-grained methods*. IOP Publishing, 2016.

72. Wu, K., S. Xu, B. Wan, P. Xiu, and X. Zhou, "A novel multiscale scheme to accelerate atomistic simulations of bio-macromolecules by adaptively driving coarse-grained coordinates," *Journal of Chemical Physics*, Vol. 152, p. 114115, Mar 2020.

73. Krieger, J. M., P. Doruker, A. L. Scott, D. Perahia, and I. Bahar, "Towards gaining sight of multiscale events: utilizing network models and normal modes in hybrid methods," *Current Opinion in Structural Biology*, Vol. 64, pp. 34–41, Oct 2020.

74. Schlick, T., "Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules," *F1000 Biology Reports*, Vol. 1, p. 51, Jul 2009.

75. Gur, M., J. D. Madura, and I. Bahar, "Global transitions of proteins explored by a multiscale hybrid methodology: Application to adenylate kinase," *Biophysical Journal*, Vol. 105, pp. 1643–1652, Oct 2013.

76. Uyar, A., N. Kantarci-Carsibasi, T. Haliloglu, and P. Doruker, "Features of large hinge-bending conformational transitions. prediction of closed structure from open state," *Biophysical Journal*, Vol. 106, pp. 2656–2666, Jun 2014.

77. Costa, M. G. S., P. R. Batista, P. M. Bisch, and D. Perahia, "Exploring free energy landscapes of large conformational changes: Molecular dynamics with excited normal modes," *Journal of Chemical Theory and Computation*, Vol. 11, no. 6, pp. 2755–2767, 2015. PMID: 26575568.

78. Lima, A. N., R. J. de Oliveira, A. S. K. Braz, M. G. de Souza Costa, D. Perahia, and L. P. B. Scott, "Effects of ph and aggregation in the human prion conversion into scrapie form: a study using molecular dynamics with excited normal modes," *European biophysics journal : EBJ*, Vol. 47, p. 583â590, July 2018.

79. Peng, X., Y. Zhang, Y. Li, Q. Liu, H. Chu, D. Zhang, and G. Li, "Integrating multiple accelerated molecular dynamics to improve accuracy of free energy calculations," *Journal of Chemical Theory and Computation*, Vol. 14, no. 3, pp. 1216–1227, 2018. PMID: 29394067.

80. Wang, W. B., Y. Liang, J. Zhang, Y. D. Wu, J. J. Du, Q. M. Li, J. Z. Zhu, and J. G. Su, "Energy transport pathway in proteins: Insights from non-equilibrium molecular dynamics with elastic network model," *Scientific Reports*, Vol. 8, pp. 1–13, Dec 2018.

81. Jalalypour, F., O. Sensoy, and C. Atilgan, "Perturb−scan−pull: A novel method facilitating conformational transitions in proteins," *Journal of Chemical Theory and Computation*, Vol. 16, no. 6, pp. 3825–3841, 2020. PMID: 32324386.

82. Fas, A., and Haliloglu, "To be published," 2021.

83. Acar, B., J. Rose, B. Aykac Fas, N. Ben-Tal, O. Lewinson, and T. Haliloglu, "Distinct allosteric networks underlie mechanistic speciation of abc transporters," *Structure*, Apr 2020.

84. Yang, M., N. Livnat Levanon, B. Acar, B. Aykac Fas, G. Masrati, J. Rose, N. Ben-Tal, T. Haliloglu, Y. Zhao, and O. Lewinson, "Single-molecule probing of the conformational homogeneity of the abc transporter btucd," *Nature Chemical Biology*, Vol. 14, pp. 715–722, Jul 2018.

85. Kessel, A., and N. Ben-Tal, *Introduction to Proteins*, Chapman and Hall/CRC, Mar 2018.

86. Levitt, M., and C. Chothia, "Structural patterns in globular proteins," *Nature*, Vol. 261, pp. 552–558, Jun 1976.

87. Cilia, E., R. Pancsa, P. Tompa, T. Lenaerts, and W. F. Vranken, "From protein sequence to dynamics and disorder with dynamine," *Nature Communications*, Vol. 4, pp. 1–10, Nov 2013.

88. Devine, P. W., H. C. Fisher, A. N. Calabrese, F. Whelan, D. R. Higazi, J. R. Potts, D. C. Lowe, S. E. Radford, and A. E. Ashcroft, "Investigating the structural compaction of biomolecules upon transition to the gas-phase using esi-twims-ms," *Journal of the American Society for Mass Spectrometry*, Vol. 28, pp. 1855–1862, Sep 2017.

89. Do Viet, P., D. B. Roche, and A. V. Kajava, "Tapo: A combined method for the identification of tandem repeats in protein structures," *FEBS Letters*, Vol. 589, pp. 2611–2619, Sep 2015.

90. Li, J., A. Mahajan, and M. D. Tsai, "Ankyrin repeat: A unique motif mediating protein-protein interactions," *Biochemistry*, Vol. 45, pp. 15168–15178, Dec 2006.

91. Ramirez-Alvarado, M., J. W. Kelly, and C. M. Dobson, *Protein Misfolding Diseases: Current and Emerging Principles and Therapies*, John Wiley and Sons, Jul 2010.

92. Dunker, A. K., J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, and et al., "Intrinsically disordered protein," *Journal of Molecular Graphics and Modelling*, Vol. 19, no. 1, pp. 26–59, 2001.

93. Uversky, V. N., "Intrinsically disordered proteins and their "mysterious" (meta)physics," *Frontiers in Physics*, Vol. 7, p. 10, Feb 2019.

94. Wright, P. E., and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nature Reviews Molecular Cell Biology*, Vol. 16, pp. 18–29, Dec 2015.

95. Hansen, J. C., X. Lu, E. D. Ross, and R. W. Woody, "Intrinsic protein disorder, amino acid composition, and histone terminal domains," *Journal of Biological Chemistry*, Vol. 281, pp. 1853–1856, Jan 2006.

96. Jensen, M. R., M. Zweckstetter, J. R. Huang, and M. Blackledge, "Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using nmr spectroscopy," *Chemical Reviews*, Vol. 114, pp. 6632–6660, Jul 2014.

97. Ferreiro, D. U., and E. A. Komives, "The plastic landscape of repeat proteins," *Proceedings of the National Academy of Sciences*, Vol. 104, pp. 7735–7736, May 2007.

98. Farabella, I., T. Pham, N. S. Henderson, S. Geibel, G. Phan, D. G. Thanassi, A. H. Delcour, G. Waksman, and M. Topf, "Allosteric signalling in the outer membrane translocation domain of papc usher," *eLife*, Vol. 3, pp. 1–19, Oct 2014.

99. Morishita, H., and T. Yagi, "Protocadherin family: diversity, structure, and function,"

100. Benke, S., D. Roderer, B. Wunderlich, D. Nettels, R. Glockshuber, and B. Schuler, "The assembly dynamics of the cytolytic pore toxin clya," *Nature Communications*, Vol. 6, pp. 1–15, Feb 2015.

101. Lodish, H., A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, "Collagen: The fibrous proteins of the matrix," 2000.

102. Monsellier, E., M. Ramazzotti, N. Taddei, and F. Chiti, "Aggregation propensity of the human proteome," *PLoS Computational Biology*, Vol. 4, p. e1000199, Oct 2008.

103. Ardito, F., M. Giuliani, D. Perrone, G. Troiano, and L. L. Muzio, "The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (review)," *International Journal of Molecular Medicine*, Vol. 40, pp. 271–280, Aug 2017.

104. Von Rau, F., A. De Ruiter, and T. A. Leonard, "A switch in nucleotide affinity governs activation of the src and tec family kinases," *Scientific Reports*, Vol. 7, pp. 1–14, Dec 2017.

105. Burnham, M. R., P. J. Bruce-Staskal, M. T. Harte, C. L. Weidow, A. Ma, S. A. Weed, and A. H. Bouton, "Regulation of c-src activity and function by the adapter protein cas," *Molecular and Cellular Biology*, Vol. 20, pp. 5865–5878, Aug 2000.

106. Sen, B., and F. M. Johnson, "Regulation of src family kinases in human cancers," *Journal of Signal Transduction*, Vol. 2011, pp. 1–14, 2011.

107. Xu, W., A. Doshi, M. Lei, M. J. Eck, and S. C. Harrison, "Crystal structures of c-src reveal features of its autoinhibitory mechanism," *Molecular Cell*, Vol. 3, no. 5, pp. 629–638, 1999.

108. Cowan-Jacob, S. W., G. Fendrich, P. W. Manley, W. Jahnke, D. Fabbro, J. Liebetanz, and T. Meyer, "The crystal structure of a c-src complex in an active conformation suggests possible steps in c-src activation," *Structure*, Vol. 13, pp. 861–871, Jun 2005.

109. Cantley, L. C., K. R. Auger, C. Carpenter, B. Duckworth, A. Graziani, R. Kapeller, and S. Soltoff, "Oncogenes and signal transduction," *Cell*, Vol. 64, pp. 281–302, Jan 1991.

110. MacAuley, A., and J. A. Cooper, "Structural differences between repressed and derepressed forms of p60c-src.," *Molecular and Cellular Biology*, Vol. 9, pp. 2648–2656, Jun 1989.

111. Alexandropoulos, K., and D. Baltimore, "Coordinate activation of c-src by sh3-and sh2-binding sites on a novel, p130(cas)-related protein, sin," *Genes and Development*, Vol. 10, no. 11, pp. 1341–1355, 1996.

112. Moarefi, I., M. LaFevre-Bernt, F. Sicheri, M. Huse, C. H. Lee, J. Kuriyan, and W. T. Miller, "Activation of the src-family tyrosine kinase hck by sh3 domain displacement," *Nature*, Vol. 385, pp. 650–653, Feb 1997.

113. Meng, Y., and B. Roux, "Locking the active conformation of c-src kinase through the phosphorylation of the activation loop," *Journal of Molecular Biology*, Vol. 426, pp. 423–435, Jan 2014.

114. Fajer, M., Y. Meng, and B. Roux, "The activation of c-src tyrosine kinase: Conformational transition pathway and free energy landscape," *Journal of Physical Chemistry B*, Vol. 121, pp. 3352–3363, Apr 2017.

115. Griffiths, G. J., M. Y. Koh, V. G. Brunton, C. Cawthorne, N. A. Reeves, M. Greaves, M. J. Tilby, D. G. Pearson, C. J. Ottley, P. Workman, and et al., "Expression of kinase-defective mutants of c-src in human metastatic colon cancer cells decreases bcl-xl and increases oxaliplatin- and fas-induced apoptosis," *Journal of Biological Chemistry*, Vol. 279, pp. 46113–46121, Oct 2004.

116. Miyazaki, K., T. Senga, S. Matsuda, M. Tanaka, K. MacHida, Y. Takenouchi, Y. Nimura, and M. Hamaguchi, "Critical amino acid substitutions in the src sh3 domain that convert c-src to be oncogenic," *Biochemical and Biophysical Research Communications*, Vol. 263, pp. 759–764, Oct 1999.

117. Snyder, ', M. A., J. Michael Bishop, J. P. Mcgrath, A. D. Levinson2, and G. W. Hooper, *A Mutation at the ATP-Binding Site of pp6Ov-src Abolishes Kinase Activity, Transformation, and Tumorigenicity Downloaded from*, Vol. 5, 1985.

118. Tate, J. G., S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, and S. A. Forbes, "COSMIC: the Catalogue Of Somatic Mutations In Cancer," *Nucleic Acids Research*, Vol. 47, pp. D941–D947, 10 2018.

119. Locher, K. P., A. T. Lee, and D. C. Rees, "The e. coli btucd structure: A framework for abc transporter architecture and mechanism," *Science*, Vol. 296, pp. 1091–1098, May 2002.

120. Korkhov, V. M., S. A. Mireku, D. B. Veprintsev, and K. P. Locher, "Structure of amp-pnp-bound btucd and mechanism of atp-powered vitamin b12 transport by btucd-f," *Nature Structural and Molecular Biology*, Vol. 21, pp. 1097–1099, Dec 2014.

121. Korkhov, V. M., S. A. Mireku, and K. P. Locher, "Structure of amp-pnp-bound vitamin b 12 transporter btucd-f," *Nature*, Vol. 490, pp. 367–372, Oct 2012.

122. Hvorup, R. N., B. A. Goetz, M. Niederer, K. Hollenstein, E. Perozo, and K. P. Locher, "Asymmetry in the structure of the abc transporter - binding protein complex btucd-btuf," *Science*, Vol. 317, pp. 1387–1390, Sep 2007.

123. Tirion, M. M., "Large amplitude elastic motions in proteins from a single-parameter, atomic analysis," *Phys. Rev. Lett.*, Vol. 77, pp. 1905–1908, Aug 1996.

124. Chapman, and H. . Crc, *Normal Mode Analysis*, Chapman and Hall/CRC, Dec 2005.

125. Bastolla, U., "Computing protein dynamics from protein structure with elastic network models," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Vol. 4, pp. 488–503, Sep 2014.

126. Bussi, G., and M. Parrinello, "Accurate sampling using langevin dynamics," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 75, p. 056707, May 2007.

127. Perez, C., S. Gerber, J. Boilevin, M. Bucher, T. Darbre, M. Aebi, J.-L. Reymond, and K. P. Locher, "Structure and mechanism of an active lipid-linked oligosaccharide flippase," *Nature*, Vol. 524, pp. 433–438, Aug 2015.

128. Lezon, T. R., I. H. Shrivastava, Z. Yang, and I. Bahar, *Elastic Network Models For Biomolecular Dynamics: Theory and Application to Membrane Proteins and Viruses*, pp. 129–158. Dec 2009.

129. Mishra, S. K., and R. L. Jernigan, "Protein dynamic communities from elastic network models align closely to the communities defined by molecular dynamics," *PLOS ONE*, Vol. 13, p. e0199225, Jun 2018.

130. Meireles, L., M. Gur, A. Bakan, and I. Bahar, "Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins," *Protein Science*, Vol. 20, pp. 1645–1658, Oct 2011.

131. Bahar, I., A. R. Atilgan, and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential," *Folding and Design*, Vol. 2, pp. 173–181, Jun 1997.

132. Eyal, E., L. W. Yang, and I. Bahar, "Anisotropic network model: Systematic evaluation and a new web interface," *Bioinformatics*, Vol. 22, pp. 2619–2627, Nov 2006.

133. Petrone, P., and V. S. Pande, "Can conformational change be described by only a few normal modes?," *Biophysical Journal*, Vol. 90, pp. 1583–1593, Mar 2006.

134. Doruker, P., A. R. Atilgan, and I. Bahar, "Dynamics of proteins predicted by molecular simulations and analytical approaches: Application to α-amylase inhibitor," *Proteins: Structure, Function and Genetics*, Vol. 40, pp. 512–524, Aug 2000.

135. Mahajan, S., and Y. H. Sanejouand, "On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins," *Archives of Biochemistry and Biophysics*, Vol. 567, pp. 59–65, Feb 2015.

136. Echave, J., "Why are the low-energy protein normal modes evolutionarily conserved?," *Pure and Applied Chemistry*, Vol. 84, no. 9, pp. 1931–1937, 2012.

137. Fuglebakk, E., S. P. Tiwari, and N. Reuter, "Comparing the intrinsic dynamics of multiple protein structures using elastic network models," *Biochimica et Biophysica Acta - General Subjects*, Vol. 1850, pp. 911–922, May 2015.

138. Emekli, U., D. Schneidman-Duhovny, H. J. Wolfson, R. Nussinov, and T. Haliloglu, "Hingeprot: Automated prediction of hinges in protein structures," *Proteins: Structure, Function and Genetics*, Vol. 70, pp. 1219–1227, Mar 2008.

139. Hacisuleyman, A., and B. Erman, "Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions.," *Proteins*, Vol. 85, no. 6, pp. 1056–1064, 2017.

140. Schreiber, T., "Measuring information transfer," *Physical Review Letters*, Vol. 85, pp. 461–464, Jul 2000.

141. Karplus, M., and J. Kuriyan, "Molecular dynamics and protein function.," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102, pp. 6679–85, May 2005.

142. Karplus, M., and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature Structural Biology*, Vol. 9, no. 9, pp. 646–652, 2002.

143. Karplus, M., and G. A. Petsko, "Molecular dynamics simulations in biology," *Nature*, Vol. 347, no. 6294, pp. 631–639, 1990.

144. Hollingsworth, S. A., and R. O. Dror, "Molecular dynamics simulation for all," *Neuron*, Vol. 99, pp. 1129–1143, Sep 2018.

145. Shaw, D. E., K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, and et al., "Millisecond-scale molecular dynamics simulations on anton," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09*, p. 1, Association for Computing Machinery (ACM), 2009.

146. Wu, X., and B. R. Brooks, "Self-guided langevin dynamics simulation method," *Chemical Physics Letters*, Vol. 381, pp. 512–518, Nov 2003.

147. Izvekov, S., and G. A. Voth, "Modeling real dynamics in the coarse-grained representation of condensed phase systems," *Journal of Chemical Physics*, Vol. 125, p. 151101, Oct 2006.

148. Paquet, E., and H. L. Viktor, "Molecular dynamics, monte carlo simulations, and langevin dynamics: A computational review," *BioMed Research International*, Vol. 2015, 2015.

149. Case, D.A. , T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, P.A. Kollman, "Amber 11," *University of California*, 2010.

150. Fiser, A., and A. Sali, "Modloop: automated modeling of loops in protein structures.," *Bioinformatics (Oxford, England)*, Vol. 19, pp. 2500–1, Dec 2003.

151. Loncharich, R. J., B. R. Brooks, and R. W. Pastor, "Langevin dynamics of peptides: The frictional dependence of isomerization rates of n-acetylalanyl-n'-methylamide," *Biopolymers*, Vol. 32, no. 5, pp. 523–535, 1992.

152. Izaguirre, J. A., D. P. Catarello, J. M. Wozniak, and R. D. Skeel, "Langevin stabilization of molecular dynamics," *The Journal of Chemical Physics*, Vol. 114, no. 5, pp. 2090–2098, 2001.

153. Glaser, F., T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, and N. Ben-Tal, "Consurf: identification of functional regions in proteins by surface-mapping of phylogenetic information.," *Bioinformatics (Oxford, England)*, Vol. 19, pp. 163–4, Jan 2003.

154. Eddy, S. R., "Profile hidden markov models.," *Bioinformatics (Oxford, England)*, Vol. 14, no. 9, pp. 755–63, 1998.

155. Mayrose, I., D. Graur, N. Ben-Tal, and T. Pupko, "Comparison of site-specific rate-inference methods for protein sequences: empirical bayesian methods are superior.," *Molecular biology and evolution*, Vol. 21, pp. 1781–91, Sep 2004.

156. Carlson, G. M., and A. W. Fenton, "What mutagenesis can and cannot reveal about allostery," *Biophysical Journal*, Vol. 110, pp. 1912–1923, May 2016.

157. Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.8." November 2015.

158. Ma, B., S. Kumar, C. J. Tsai, and R. Nussinov, "Folding funnels and binding mechanisms," *Protein Engineering*, Vol. 12, pp. 713–720, Sep 1999.

159. Eaton, P., and P. West, *Atomic Force Microscopy*, Vol. 9780199570454, Oxford University Press, May 2010.

160. Müller, D. J., and A. Engel, "Atomic force microscopy and spectroscopy of native membrane proteins," *Nature Protocols*, Vol. 2, pp. 2191–2197, Sep 2007.

161. Corvaglia, S., B. Sanavio, R. P. Hong Enriquez, B. Sorce, A. Bosco, D. Scaini, S. Sabella, P. P. Pompa, G. Scoles, and L. Casalis, "Atomic force microscopy based nanoassay: A new method to study α-synuclein-dopamine bioaffinity interactions," *Scientific Reports*, Vol. 4, pp. 1–9, Jun 2014.

162. Lovas, S., Y. Zhang, J. Yu, and Y. L. Lyubchenko, "Molecular mechanism of misfolding and aggregation of aβ (13-23)," *Journal of Physical Chemistry B*, Vol. 117, pp. 6175–6186, May 2013.

163. Kikhney, A. G., and D. I. Svergun, "A practical guide to small angle x-ray scattering (saxs) of flexible and intrinsically disordered proteins," *FEBS Letters*, Vol. 589, pp. 2570–2577, Sep 2015.

164. Brosey, C. A., and J. A. Tainer, "Evolving saxs versatility: solution x-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology," *Current Opinion in Structural Biology*, Vol. 58, pp. 197–213, Oct 2019.

165. Grishaev, A., "Sample preparation, data collection, and preliminary data analysis in biomolecular solution x-ray scattering," *Current Protocols in Protein Science*, Vol. 1, p. Unit17.14, Nov 2012.

166. Jacques, D. A., and J. Trewhella, "Small-angle scattering for structural biology - expanding the frontier while avoiding the pitfalls," *Protein Science*, Vol. 19, pp. 642–657, Apr 2010.

167. Mrozowich, T., S. McLennan, M. Overduin, and T. R. Patel, "Structural studies of macromolecules in solution using small angle x-ray scattering," *Journal of Visualized Experiments*, Vol. 2018, p. 58538, Nov 2018.

168. Svergun, D., C. Barberato, and M. H. Koch, "Crysol - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates," *Journal of Applied Crystallography*, Vol. 28, pp. 768–773, Dec 1995.

169. Garcia, R., and R. Proksch, "Nanomechanical mapping of soft matter by bimodal force microscopy," 2013.

170. Martinez-Martin, D., E. T. Herruzo, C. Dietz, J. Gomez-Herrero, and R. Garcia, "Noninvasive protein structural flexibility mapping by bimodal dynamic force microscopy," *Physical Review Letters*, Vol. 106, p. 198101, May 2011.

171. Nečas, D., and P. Klapetek, "Gwyddion: an open-source software for spm data analysis," *Open Physics*, Vol. 10, pp. 181–188, Jan 2012.

172. Jo, S., M. Vargyas, J. Vasko-Szedlar, B. Roux, and W. Im, "Pbeq-solver for online visualization of electrostatic potential of biomolecules.," *Nucleic acids research*, Vol. 36, no. Web Server issue, 2008.

173. Almen, M. S., K. J. Nordström, R. Fredriksson, and H. B. Schiöth, "Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin," *BMC Biology*, Vol. 7, p. 50, Aug 2009.

174. Kozma, D., I. Simon, and G. E. Tusnády, "Pdbtm: Protein data bank of transmembrane proteins after 8 years," *Nucleic Acids Research*, Vol. 41, Jan 2013.

175. Tusnady, G. E., Z. Dosztanyi, and I. Simon, "Pdb_tm: Selection and membrane localization of transmembrane proteins in the protein data bank," *Nucleic Acids Research*, Vol. 33, Jan 2005.

176. Tusnady, G. E., Z. Dosztányi, and I. Simon, "Transmembrane proteins in the protein data bank: identification and classification," *BIOINFORMATICS*, Vol. 20, no. 17, pp. 2964–2972, 2004.

177. Pan, C., J. Weng, and W. Wang, "Atp hydrolysis induced conformational changes in the vitamin b12 transporter btucd revealed by md simulations," *PLoS ONE*, Vol. 11, pp. e0166980–e0166980, Nov 2016.

178. Lewinson, O., A. T. Lee, K. P. Locher, and D. C. Rees, "A distinct mechanism for the abc transporter btucd-f revealed by the dynamics of complex formation," *Nature Structure Molecular Biology*, Vol. 17, pp. 332–338, Mar 2010.

179. Borths, E. L., B. Poolman, R. N. Hvorup, K. P. Locher, and D. C. Rees, "In vitro functional characterization of btucd-f, the escherichia coli abc transporter for vitamin b12 uptake.," *Biochemistry*, Vol. 44, pp. 16301–9, Dec 2005.

180. Joseph, B., G. Jeschke, B. A. Goetz, K. P. Locher, and E. Bordignon, "Transmembrane gate movements in the type ii atp-binding cassette (abc) importer btucd-f during nucleotide cycle.," *The Journal of biological chemistry*, Vol. 286, pp. 41008–17, Nov 2011.

181. Joseph, B., V. M. Korkhov, M. Yulikov, G. Jeschke, and E. Bordignon, "Conformational cycle of the vitamin b12 abc importer in liposomes detected by double electron-electron resonance (deer).," *The Journal of biological chemistry*, Vol. 289, pp. 3176–85, Feb 2014.

182. Tal, N., E. Ovcharenko, and O. Lewinson, "A single intact atpase site of the abc transporter btucd drives 5% transport activity yet supports full in vivo vitamin b12 utilization.," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 110, pp. 5434–9, Apr 2013.

183. ter Beek, J., A. Guskov, and D. J. Slotboom, "Structural diversity of abc transporters.," *The Journal of general physiology*, Vol. 143, pp. 419–35, Apr 2014.

184. Locher, K. P., "Mechanistic diversity in atp-binding cassette (abc) transporters.," *Nature structural & molecular biology*, Vol. 23, pp. 487–93, Jun 2016.

185. Lewinson, O., and N. Livnat-Levanon, "Mechanism of action of abc importers: Conservation, divergence, and physiological adaptations," *Journal of Molecular Biology*, Jan 2017.

186. Zheng, W., and M. Tekpinar, "Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model," *BMC Structural Biology*, Vol. 9, pp. 1–17, Jul 2009.

187. Yin, S., F. Ding, and N. V. Dokholyan, "Modeling backbone flexibility improves protein stability estimation," *Structure*, Vol. 15, pp. 1567–1576, Dec 2007.

188. Worth, C. L., R. Preissner, and T. L. Blundell, "Sdm - a server for predicting effects of mutations on protein stability and malfunction," *Nucleic Acids Research*, Vol. 39, Jul 2011.

189. Montanucci, L., E. Capriotti, Y. Frank, N. Ben-Tal, and P. Fariselli, "Ddgun: An untrained method for the prediction of protein stability changes upon single and multiple point variations," *BMC Bioinformatics*, Vol. 20, p. 335, Jul 2019.

190. Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, "The foldx web server: An online force field," *Nucleic Acids Research*, Vol. 33, pp. W382–W388, Jul 2005.

191. Oldham, M. L., and J. Chen, "Crystal structure of the maltose transporter in a pretranslocation intermediate state," *Science*, Vol. 332, pp. 1202–1205, Jun 2011.

192. Liu, M., T. G. Sun, J. P. Hu, W. Z. Chen, and C. X. Wang, "Study on the mechanism of the btuf periplasmic-binding protein for vitamin b12," *Biophysical Chemistry*, Vol. 135, pp. 19–24, Jun 2008.

193. Sharff, A. J., L. E. Rodseth, J. C. Spurlino, and F. A. Quiocho, "Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis.," *Biochemistry*, Vol. 31, pp. 10657–63, Nov 1992.

194. Quiocho, F. A., J. C. Spurlino, and L. E. Rodseth, "Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor.," *Structure (London, England)*, Vol. 5, pp. 997–1015, Aug 1997.

195. Oldham, M. L., D. Khare, F. A. Quiocho, A. L. Davidson, and J. Chen, "Crystal structure of a catalytic intermediate of the maltose transporter.," *Nature*, Vol. 450, pp. 515–21, Nov 2007.

196. Orelle, C., T. Ayvaz, R. M. Everly, C. S. Klug, and A. L. Davidson, "Both maltose-binding protein and atp are required for nucleotide-binding domain closure in the intact maltose abc transporter," *Proc. Natl. Acad. Sci. U.S.A.*, Vol. 105, pp. 12837–12842, Sep 2008.

197. Davidson, A. L., H. A. Shuman, and H. Nikaido, "Mechanism of maltose transport in escherichia coli: transmembrane signaling by periplasmic binding proteins.," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 89, pp. 2360–4, Mar 1992.

198. Shah, N. H., J. F. Amacher, L. M. Nocka, and J. Kuriyan, "The src module: an ancient scaffold in the evolution of cytoplasmic tyrosine kinases," *Critical Reviews in Biochemistry and Molecular Biology*, Vol. 53, pp. 535–563, Sep 2018.

199. Wheeler, D. L., M. Iida, and E. F. Dunn, "The role of src in solid tumors," *The Oncologist*, Vol. 14, pp. 667–678, Jul 2009.

200. Shukla, D., Y. Meng, B. Roux, and V. S. Pande, "Activation pathway of src kinase reveals intermediate states as targets for drug design," *Nature Communications*, Vol. 5, Mar 2014.

201. Yoon, H. J., S. Lee, S. J. Park, and S. Wu, "Network approach of the conformational change of c-Src, a tyrosine kinase, by molecular dynamics simulation," *Scientific Reports*, Vol. 8, no. 1, p. 5673, 2018.

202. Ma, J., and M. Karplus, "Molecular switch in signal transduction: Reaction paths of the conformational changes in ras p21," *Proceedings of the National Academy of Sciences*, Vol. 94, no. 22, pp. 11905–11910, 1997.

203. Locher, K. P., and E. Borths, "Abc transporter architecture and mechanism: Implications from the crystal structures of btucd and btuf," in *FEBS Letters*, Vol. 564, pp. 264–268, No longer published by Elsevier, Apr 2004.

204. Pinkett, H. W., A. T. Lee, P. Lum, K. P. Locher, and D. C. Rees, "An inward-facing conformation of a putative metal-chelate-type abc transporter," *Science*, Vol. 315, pp. 373–377, Jan 2007.

205. Liu, S., X. Xiang, X. Gao, and H. Liu, "Neighborhood preference of amino acids in protein structures and its applications in protein structure assessment," *Scientific Reports*, Vol. 10, Dec 2020.