論文 / 著書情報
Article / Book Information

題目(和文)	PaCS-MD と MSM による p53-DBD/DNA 複合体の 解離過程と結合自由エネルギーの解析
Title(English)	Investigating dissociation process and binding free energy of p53-DBD/DNA complex by PaCS-MD and MSM
著者(和文)	SOBEHMohamed Marzouk
Author(English)	Mohamed Marzouk Sobeh
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11722号, 授与年月日:2022年3月26日, 学位の種別:課程博士, 審査員:北尾 彰朗,伊藤 武彦,田口 英樹,村上 聡,山田 拓司
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11722号, Conferred date:2022/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Doctoral Dissertation

博士論文

Investigating dissociation process and binding free energy of p53-DBD/DNA complex by PaCS-MD and MSM

(PaCS-MD と MSM による p53-DBD/DNA 複合体の解離過程と結合自由エネルギーの解析)

A Dissertation Submitted for the Degree of Doctor of Philosophy

February 2022

School of Life Science and Technology,
Tokyo Institute of Technology

Mohamed Marzouk Sobeh
モハメド　マルゾーク　ソベ

Doctoral Dissertation

東京工業大学
Tokyo Institute of Technology

Investigating dissociation process and binding free energy of p53-DBD/DNA complex by PaCS-MD and MSM

A Dissertation Submitted for the degree of Doctor of Philosophy

February 2022

School of Life Science and Technology,
Tokyo Institute of Technology

Mohamed Marzouk Sobeh

Acknowledgements

First and foremost, I would like to praise Allah the Almighty, the Most Gracious, and the Most Merciful for His blessing given to me during my study and in completing this thesis. I am asking from Allah continuous support for further success in my future scientific work.

I would like to express my greatest appreciation to my supervisor, **Prof. Akio Kitao** for his supervision, support, valuable comments, suggestions, discussions and kind help throughout the research steps for preparing and completing this dissertation with good quality in the final form. In addition, I would also thank him for his immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life.

I would like to extend my deeply grateful to **Dr. Takemura Kazuhiro** for the thoughtful discussions, advice throughout my research, and answering all my "Why" questions regarding science and Japanese live style. In addition, my sincere thanks to **Dr. Hiroaki Hata**, and **Dr. Kenichiro Takaba** the alumni of Kitao lab for their support in the first period of my Ph.D.

I would like to express my sincere gratitude for my colleagues, **Tegar-san** and **Darin-san** for their help, comments during my research work and writing my dissertation. Also, my special thanks to all members of Kiato Lab; Dr. Duy Phuoc Tran, Taira-san, Ogawa-san, Misu-san, Yoshioka-san, Osawa-san, Nakaya-san, Ikizawa-san, and all other memebrs for their comments and valuable research suggestions in our group seminars. In addition, I would like to pay my regards to **Shibayama Hidemi-san** for the help and support of the document works, allowing me to concentrate on my research.

I cannot dream about this dissertation without the tremendous understanding, encouragement, and strong moral support of my lovely wife **Asmaa**, my sweet daughter **Karma**, and new hero **Yusuf**.

Finally, I would like to express my gratitude to my parents, my sisters and my brother, they are every thing in my live.

MOHAMED MARZOUK SOBEH

Abstract

The binding of biological molecules is critical for understanding their functions, disease mechanisms, and the development of new drugs. The binding of a transcription factor p53 protein to DNA with its DNA binding domain (p53-DBD) is required for its defensive function against cancer as a "guardian of the genome." The p53-DBD binds to the minor and major grooves of DNA in a sequence-specific way in response to diverse stresses including DNA damage, and is important to initiate transcription of genes involved in cell cycle arrest, apoptosis, and DNA repair. Accumulation of mutations in p53-DBD has a potential to destabilize and unfold the structure, limiting capacity of p53 to identify and bind to its target sequence. Investigating how essential residues of the binding surface stabilize the p53-DBD/DNA complex structure and they contribute to energy of binding is crucial for understanding p53-DBD recognition of certain DNA sequences. These mechanisms are currently unclear, and more study is necessary to identify the essential residues that enable sequence-specific p53-DBD DNA interactions.

Molecular dynamics (MD) simulation is an advantageous technique for studying biomolecular systems because it can generate trajectories of molecular behavior along time scale at atomic resolution. Despite recent advances in computing speed and sophistication, determining the phase space of a biomolecule remains a difficult challenge. This "sampling challenge" is connected to the large temporal gap between MD and biophysical processes. Our lab has developed a technique called parallel cascade selection molecular dynamics (PaCS-MD) that may be utilized to accelerate the detection of these biologically slow processes and overcome this sampling difficulty. PaCS-MD was recently utilized to build dissociation pathways for a variety of protein-ligand complexes and then used to derive binding free energy and kinetic rates, showing remarkable agreement with experimental results. However, applying this strategy to larger systems and analyzing the dissociation pathways remains difficult.

Specifically, in this thesis, I investigate the dissociation process of the p53-DBD from the DNA duplex that contains a consensus sequence, which is the particular target of the p53-DBD, by combining the dissociation PaCS-MD (dPaCS-MD) with the Markov state model (MSM). Based on all-atom model including explicit solvent, I first simulated the p53-DBD dissociation processes by 75 trials of dPaCS-MD, which required average simulation time of 11.2 ± 2.2 ns per trial. By setting the axis of the DNA duplex as the Z-axis and the binding side of p53-DBD on DNA as the +side of the X-axis, dissociations took place along the +X and -Y directions (namely, -Y directions) in 93 % of the cases, while 7 % moved along +X and +Y directions (+Y directions). Toward the -Y directions, p53-DBD dissociates first from the major groove and then detached from the minor groove, while unbinding from the minor groove occurred first in dissociations along the +Y directions. Analysis of the free energy landscape by MSM showed that the loss of the minor groove interaction with p53-DBD toward the +Y directions is relatively high (1.1 kcal/mol), whereas the major groove detachment more frequently occurred with lower free energy costs. The standard

binding free energy calculated from the free energy landscape was -10.9 ± 0.4 kcal/mol, which agrees with an experimental value of –11.1 kcal/mol. These results indicate that this combination, dPaCS-MD/MSM, can be a powerful tool to investigate dissociation mechanisms of two large molecules. The minor groove binding is mainly stabilized by R248, identified as the most important residue that tightly binds deep inside the minor groove. Analysis of the p53 key residues for the DNA binding indicates high correlations with the cancer-related mutations, which confirms that impairments of interactions between p53-DBD and DNA can be frequently related to cancer.

dPaCS-MD/MSM is a promising combination that can be used to investigate multiple pathways during the dissociation of two large molecules, as well as to identify critical residues for major dissociation pathways and to quantitatively calculate the complex's binding free energy, which can aid future studies in explaining the effects of mutations on binding free energy and binding process.

Contents

List of Figures

List of Tables

List of Publication

- Mohamed Marzouk Sobeh, and Akio Kitao:, **Paper Title**, **"Dissociation pathways of p53 DNA binding domain from DNA and critical roles of key residues elucidated by dPaCS-MD/MSM."**, Accepted at Journal of Chemical Information and Modeling (2022) https://doi.org/10.26434/chemrxiv-2021-z08zx

Abbreviations

LAH	**L**ist **A**bbreviations **H**ere
3D-MSM	**T**hree-dimensional **M**arkov **S**tate **M**odel
ACE	**Ace**tyl
AFM	**A**tomic **F**orce **M**icroscopy
AMBER	**A**ssisted **M**odel **B**uilding and **E**nergy **R**efinement
CHARMM	**C**hemistry at **Har**vard **M**acromolecular **M**echanics
COM	**C**enter **o**f **M**ass
CVs	**C**ollective **V**ariables
dPaCS-MD	**D**issociation **P**arallel **C**ascade **S**election **M**olecular **D**ynamics
FF	**F**orce **F**ield
FEL	**F**ree **E**nergy **L**andscapes
GPUs	**G**raphical **P**rocessing **Unit**s
GROMOS	**GRO**ningen **Mo**lecular **S**imulation
H-REMD	**H**amiltonian **R**eplica **E**xchange **M**olecular **D**ynamics
ITC	**I**sothermal **T**itration **C**alorimetry
ITS	**I**mplied **T**ime **S**cale
KCl	Potassium Chloride
LSH	**L**oop-**S**heet-**H**elix
MD	**M**olecular **D**ynamics
MDM2	**M**urine **D**ouble-**M**inute **C**lone **2** Protein
MM-PBSA	**M**olecular **M**echanics **P**oisson-**B**oltzmann **S**urface **A**rea
MSM	**M**arkov **S**tate **M**odel
MTTK	Martyna-Tuckerman-Tobias-Klein
NME	**N**-**me**thlyamide
OPLS	**O**ptimized **P**otentials for **L**iquid **S**imulations
p53-DBD	**p**53's **D**NA **B**inding **D**omain
PaCS-MD	**P**arallel **C**ascade **S**election **M**olecular **D**ynamics
PBC	**P**eriodic **B**oundary **C**onditions
PCA	**P**rinciple **C**omponent **A**nalysis
PDB	**Pr**otein **D**ata **B**ank
PLIP	**P**rotein **L**igand **I**nteraction **P**rofiler

PME	**P**article-**M**esh **E**wald
PMEMD	**P**article-**M**esh **E**wald **M**olecular **D**yanmics
PMF	**P**otential **M**ean **F**orce
RMSD	**R**oot-**M**ean-**S**quare **D**eviation
RMSF	**R**oot-**M**ean-**S**quare **F**luctuation
SMD	**S**teered **M**olecular **D**ynamics
T-REMD	**T**emperature **R**eplica **E**xchange **M**olecular **D**ynamics
VMD	**V**isual **M**olecular **D**ynamics
WHAM	**W**eighted-**H**istogram **A**nalysis
ZAFF	**Z**inc **A**MBER **F**orce **F**ield

Chapter 1

Introduction and Literature review

1.1 General Introduction

All living organisms comprise various biological macromolecules such as proteins, lipids, carbohydrates, and nucleic acids, which are involved in many biological functions required to maintain life. These macromolecules are engaged in a variety of biological processes necessary for survival. Proteins are biopolymers composed of amino acids linked together into long chains, similar to beads hung on a string. There are twenty distinct amino acids found naturally in proteins. Proteins are playing a critical role to regulate the cellular environments such as **enzymatic catalysis** of chemical conversions in and around the cell, **regulatory proteins** controlling gene expression, and **receptor proteins** (which sit in the lipid membrane) accepting inter-cellular signals that are often transmitted by hormones, which are proteins as well[1].

One of the most important activities of the proteins is to regulate the progression of the cell cycle. The cell cycle acts as a "highway" to replace the worn-out cells with new cells and to maintain our bodies running basically continually and in good health. The cell cycle is pre-programmed to control the timing of cell division. This kind of regulatory system is referred to as the homeostatic mechanism, which may be considered as the "traffic controllers" of the cell since it is mostly dependent on signals to determine whether to stay in or exit the cell cycle.

The proto-oncogenes are responsible for coding for the "go" signals that command the cell to remain in the cell cycle and to continue to divide further. Oncogenic tumor suppressor genes generate proteins (such as the tumor suppressor protein p53), which operate as "stop" signals that

command the cell to leave the cell cycle and cease proliferating. If these "stop" or "go" signals are not received by the cells, the cells lose their ability to maintain homeostasis. As a consequence of accumulating DNA damage (also known as mutations), cells may lose their capabilities to respond to or produce "stop" signals, which may result in the development of cancer.

The generation of cancer is a complicated process that takes a long time to complete because it necessitates the accumulation of damage in the cell growth-controlling genes, which may include damage to proto-oncogenes and tumor suppressor genes. The tumor suppressor protein "p53" plays important roles in maintaining the integrity of our genome. Thus, it is the primary focus of my research. In the next part, I will go over the p53 structure and function in more depth.

1.2 p53 structure, functions, and mutation

p53 was first discovered and considered as an oncogene in 1979, but its main role was identified in the 1990s to be as a tumor suppressor[2, 3]. p53 is a tetrameric multidomain transcriptional factor that regulates cell response to a variety of stresses, including DNA damage, hypoxia, UV irradiation, oncogene activation, and other stress signals[4–6]. When acting as a transcription factor, p53 binds to DNA promoters in a sequence-specific way, resulting in the transcriptional activation of a variety of genes that encode proteins involved in cell cycle arrest, cell apoptosis, and DNA repair.[7, 8]. Binding of p53 to specific DNA sequences is necessary to activate gene expression[6, 9], indicating the importance of the p53 DNA binding domain (p53-DBD) (residues 94–292) to its function compared to the other functional domains of p53: the transactivation (1–44), tetramerization (325–356), and the C-terminal (357–393) domains[10]. p53-DBD is composed of a large immunoglobulin-like sandwich that forms a compact barrel-like structure. This serves as the basic framework for the DNA-binding surface. This binding surface, rich in basic amino acids, consists of a loop-sheet-helix motif (LSH) containing Loop 1 (L1) and two large loops (L2 and L3). A zinc ion stabilizes these two loops, which is coordinated by one His and three Cys residues. The LSH motif and these loops on the p53 DNA-binding surface establish contact with the minor and major grooves of DNA, as well as the intermediate areas between them. However, the accumulation of mutations in p53-DBD may result in the instability and unfolding of the structure, resulting in loss of p53 function as a “guardian of the genome” by inhibiting the ability of p53 to recognize and bind to its target, specific sequence, which has the general form RRRCWWGYYY (R =

A/G, W=A/T, Y=C/T)[10] and is conserved in all organisms (consensus sequence). Due to the fact that p53-DBD identifies the consensus sequence uniquely, mutations in p53-DBD might result in oncogenicity[11, 12].

About a half of all human tumors are thought to contain mutations in the p53 gene. According to the IARC TP53 Database https://p53.iarc.fr[13], p53 has many mutations (approximately 30,000), of which $\sim$ 75% are single missense mutations. In addition, $\sim$95% of the p53 mutations occur in p53-DBD[13–15], indicating the potential importance of understanding the roles of p53-DBD key residues in DNA binding. Based on the structure of p53-DBD, the p53-DBD mutations may be divided into two groups. One group of mutations has an effect on the 3D-structure of the p53 native conformation, which is critical for the protein's protective function; these alterations are referred to as "structural mutations". The other group of mutations impacts the direct interaction of p53 with DNA, and these are referred to as "contact mutations". Reportedly, 30% of all p53-DBD mutations occur in six well-known "hotspot" mutation sites as shown in Fig. 1.1 and 1.2. Of these, R248 (L3) and R273 (LSH) are contact mutations and R175 (L2), G245 (L3), R249 (L3), and R282 (LSH) are structural mutations[10, 16]. The two contact mutations have the highest mutation rates and are associated with more aggressive malignancies, which indicates the vital role of retaining the binding ability of p53-DBD to DNA to prevent oncogenesis[9, 17]. Consequently, the functions played by these residues in maintaining the p53-DBD/DNA complex structure and the energy required to bind the complex are thus critical for understanding how p53-DBD recognizes the particular DNA consensus sequence. These processes are still a mystery. Therefore, more thorough information is necessary on the critical residues that allow p53-DBD to retain its sequence-specific binding to DNA throughout time. Furthermore, knowing the association/dissociation processes of p53-DBD with DNA will give valuable information on the mechanisms of binding as well as the critical residues that regulate binding to the DNA molecule.

Therefore, the structural and dynamical properties of p53 are not only important for understanding the fundamental mechanism of its functions, but they are also relevant for industrial applications, such as in the field of therapeutic medicine, where p53 has been implicated in a variety of cancer types. A large number of 3D structures of p53 with atomic resolution have been solved recently utilizing a variety of experimental approaches, such as X-ray crystallography and nuclear magnetic resonance etc., providing new insights into their biological functions. Concerning protein dynamics, the p53 protein is an innately dynamic molecule that undergoes conformational changes

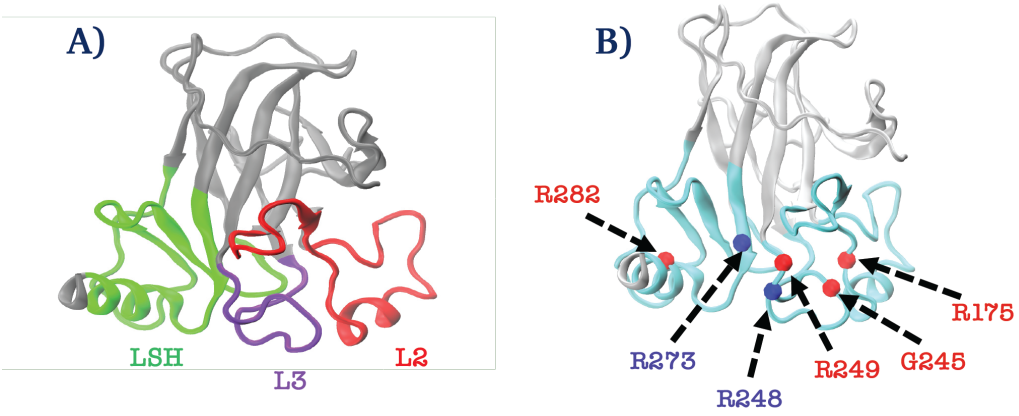

FIGURE 1.1: A) The p53-DBD with the DNA binding surface's main constituents; L2, L3, and LSH. B) The positions of the six hotspot mutations residues on the binding surface are shown by blue contact mutations and red structural mutations.

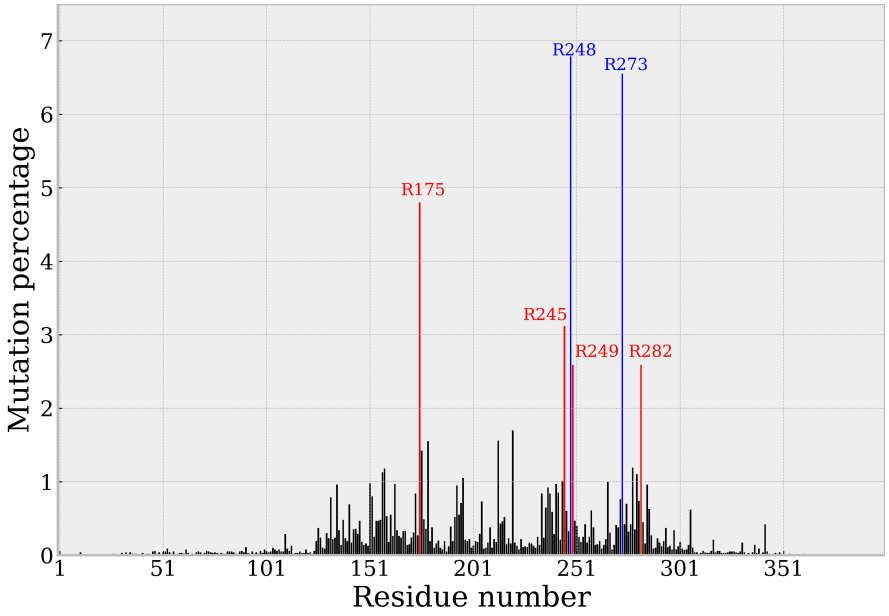

FIGURE 1.2: The percentage of mutation for each p53-residues with the six hotspot mutations residues highlighted; contact mutations in blue and structural mutations in red

that occur over a broad variety of timescales as we can see the typical timescale of proteins in Fig.1.3[18]. These dynamical behaviors must be thoroughly investigated. It may be difficult to obtain the appropriate detailed information on the underlying conformational ensembles via experimental approaches, but they may give insights into their dynamical features. To better understand the conformational transitions of proteins, molecular dynamics (MD) has been extensively used, giving time-dependent information on protein fluctuation at atomic resolution.

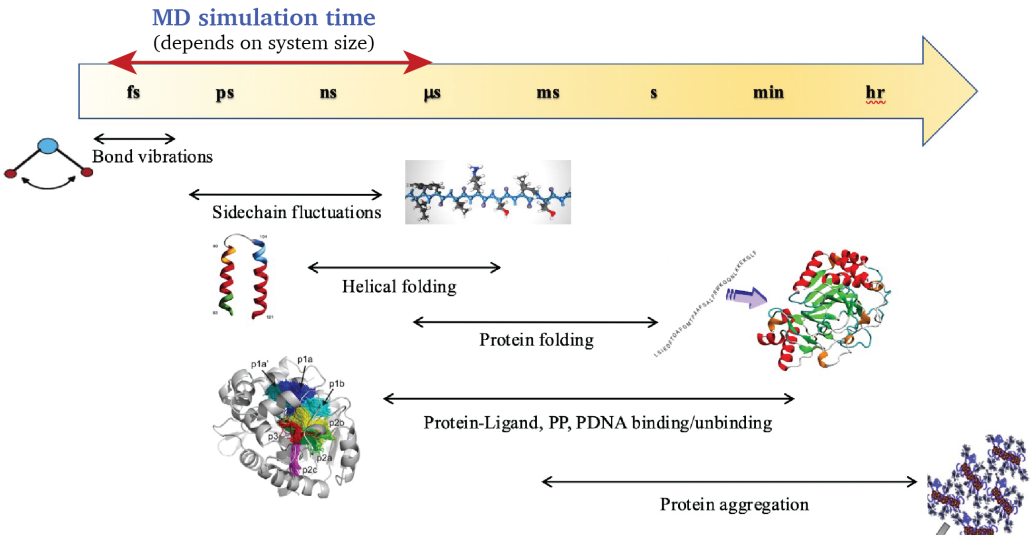

FIGURE 1.3: The typical timescale of protein dynamics.

1.3 MD Simulation and Enhanced Sampling Techniques

MD simulation is a standard technique to study the structure and function of biological macromolecules, to comprehend the mechanisms underlying complex processes, to analyze experimental findings, and to make predictions[19–21]. To investigate protein-protein, protein-DNA, and protein-ligand interactions, MD simulation can provide a comprehensive and quantitative explanation. The computation of free energy landscapes (FEL) is particularly useful in MD simulation because it enables researchers to quantify interactions between biological molecules.[22]. A typical MD simulation with atomistic empirical force fields with explicit solvent molecules (all-atom MD) can investigate a variety of biological processes that occur within a relatively short timescale, including conformational change[23–25], ligand binding[26], and fast folding events of proteins[27, 28]. In contrast, other critical biological events, such as protein-ligand, protein-protein, and protein-DNA binding and unbinding, the majority of protein folding events, and protein aggregation, frequently occur over much longer timescales and are therefore difficult to observe using all-atom MD simulations. This is because MD simulation is fundamentally confined to conformational sampling around certain local energy minima, which is incapable of overcoming large energy barriers to nearby configurational states in a computationally feasible amount of time.[29, 30]. Consequently, biological processes with longer timescales are now beyond the capabilities of conventional all-atom MD simulations.

With specialized state-of-the-art hardware or exascale cloud computing infrastructure, millisecond timescale motion in biomolecular systems of enormous size may be seen. Regrettably, public access to such supercomputers is restricted. A another strategy to compel high-performance computing is to expedite the observation of these slow processes and infrequent phenomena via the use of a vast range of novel "enhanced sampling methods"[19, 29–32]. Some of these methods apply artificial bias to enhance sampling, and include free-energy perturbation[33], umbrella sampling[34, 35], replica exchange umbrella sampling[36], metadynamics[37, 38], steered MD[39], accelerated MD[40, 41], and adaptive biasing force[42]. Other methods can enhance sampling to explore rare events during simulation without applying any biasing force, such as parallel cascade selection molecular dynamics (PaCS-MD)[43, 44], milestoning[45, 46], weighted ensemble[47], and forward flux sampling[48]. All of these strategies, with or without the addition of an artificial bias, enhance conformational transitions between the system's two well-defined molecular states A and B throughout the majority of biological events, making the rate of transition between them more predictable. In general, these states may be two chemical species, distinct forms of crystal structures, folded and unfolded protein conformations, and others, spanning a broad variety of spatial and temporal scales[49].

In particular, PaCS-MD is a powerful technique for efficiently estimating the standard binding free energy difference (ΔG^o) of various biological complexes using the Markov state model (MSM)[44, 50–52]as a tool for analyzing the PaCS-MD-generated trajectories. PaCS-MD encompasses cycles of multiple independent parallel short all-atom MD simulations accompanied with selection of initial structures for the next cycle based on a certain quantity. Here, I used the inter-center of mass distance between p53-DBD and DNA (Inter-COM distance, d) as a quantity for ranking the snapshots generated in each cycle. By repeating a series of cycles for the selected top ranked snapshots, dissociation PaCS-MD (dPaCS-MD) generates structures with larger Inter-COM distances than those found in the previous cycle, which significantly enhances the probability of transitions from the bound to unbound states[43, 53, 54]. Short MD trajectories, a series of molecular configurations, connect the initial bound and final unbound states along the dissociation pathways, and the MD trajectories from many trials of dPaCS-MD can be combined to generate different dissociation pathways, which mutually overlap in conformational space. Using these trajectories, I can construct an MSM which describes the dynamics of a biochemical process as a sequence of transitions between metastable conformational states (microstates)[55–57]. Our

group recently established the dPaCS-MD/MSM combination to speed the detection of protein-ligand[44] and protein-protein[50, 51] dissociation processes (slow processes), and to accurately calculate ΔG^o in good agreement with experimental values. As a result, I aimed to expand the use of the dPaCS-MD/MSM combination to larger biomolecular complexes, such as protein/DNA systems. In particular, I focus on the p53-DBD/DNA complex dissociation process (slow process) in my study.

1.4 Literature Review

Previously, critical residues and regions of the p53-DBD/DNA complex[58, 59] conformational change in p53-DBD[59, 60], and relative binding energy of the complex[58] were investigated.

Barakat et al[58], conducted atomistic MD simulations on p53-DBD/DNA to show a few crucial residues, specifically R248, S241, and N239, contribute to binding energy. Using the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) process, the binding energy of mutant p53-DBD at different temperatures was calculated relative to the wild-type binding energy at 300 K[58].

Pradhan et al[60], conducted extensive MD simulations of the wild-type and destabilizing mutants of the DBD (E258V, R110L, R175H and R248Q) and showed that different p53-DBD conformational activities are induced in a turn region (S6/S7). They found that the wild-type p53-DBD is largely characterized by a 'closed' conformation of the S6/S7 turn, whereas the p53-DBD mutants preferentially adopt a 'open' conformation and possess a novel druggable pocket that could be stabilized by some tested ligands. Upon probing the structural mechanisms underlying these conformation differences, it was found that there were distinct differences between the dynamics of specific amino acids in the wild-type and mutant p53-DBD; these residues were part of a network of interactions and govern the conformational state of the S6–S7 turn. These residues belong to the loop L2, helix H1 and **beta**-strands S6 and S7. This causes the loss of binding function of the p53 in the destabilizing mutants of the p53-DBD. They used a proper distance as the reaction coordinate to improve sampling between the closed and open states of the pocket using the umbrella sampling technique. Moreover, virtual screening of the FDA approved drugs for this pocket identified compounds that stabilize the p53-DBD close to its functional wild-type conformation[60].

Barros et al[59], investigated the effects of the Y220C mutation (the most frequent p53 cancer mutation observed outside the DNA-binding interface) using MD simulation, MSM, and NMR relaxation studies and showed the formation of a deep hydrophobic pocket within L6 (221-230, also termed S7/S8 loop). Their model indicated that the mutation not only affects the conformational landscape of L6 but also modulates that of the L1 loop, which is essential for DNA interactions. This result indicates the existence of allosteric communication between the two loops. They also identified a novel cryptic pocket nestled in the extended conformation of L6 that could be exploited for transpacific drug design effort[59].

Therefore, there is a demand for more information and details on the essential residues that maintain tight binding of p53-DBD to DNA. In addition, non-native contacts formed in intermediate structures during the dissociation of the p53-DBD/DNA complex also give important insights about dissociation and association processes. Furthermore, calculating the free energy profile and landscape along the pathway of the dissociation still computationally challenging, since accurate (absolute) free energy calculation requires sufficient sampling of configurational space along this reaction coordinate. The difficulty of obtaining adequate sampling and convergence of the measurement prevented simulation dissociation process of large complexes like the p53-DBD/DNA complex, however, PaCS-MD/MSM overcome these challenges and enables to calculate binding free energy of the p53-DBD to DNA which was the main target of this dissertation.

1.5 Objectives of The Proposed Research

I can summarize my objectives of the proposed research work into three main points:

1. The use of dPaCS-MD to simulate the dissociation process of the p53-DBD/DNA complex in order to investigate this important biological process. To the best of my knowledge, this is the first time to be achieved, at least using the enhanced sampling techniques with all-atom models with explicit solvent

2. Identification of the key residues of the p53-DBD/DNA binding interface. These residues play major roles in dissociation process, having an impact on the binding free energy of the complex.

3. Generation of preferred dissociation pathways and calculation of binding free energy, which are still computationally challenging.

1.6 Thesis Organization

The thesis consists of four chapters and is organized as follows. It starts from Chapter 1 that gives an introduction and literature survey to the research work for p53-DBD/DNA complex. This includes the role of p53 as a transcription factor, the ranking of the top missense mutations affect p53-DBD, the limitation in conventional MD simulation to simulate the biologically rare events, survey of the recent studies in order to understand the problem statement, and presenting the objectives of the research.

- **Chapter** 2: This chapter comprises all of the specifics on the computational approaches that were employed in this doctoral dissertation. On the topic of MD simulations, I will cover the basics of the discipline, as well as theoretical aspects and algorithms such as the concept of force fields and numerical integration, as well as thermostats and barostats. Apart from that, I go into further depth on the enhanced sampling techniques, which include both biased and unbiased methods. Finally, generic analytical approaches for MD simulations will be discussed.

- **Chapter** 3: This is the thesis's primary chapter, in which I employed PaCS-MD and MSM to investigate the dissociation pathways of the p53-DBD/DNA complex. PaCS-MD was used to build 75 dissociation pathways, and then several MSM models were developed to determine the free energy landscape. Then, the critical residues and main dissociation pathways were determined.

- **Chapter** 4: This chapter provides a summary of the thesis, as well as the findings that I have reached as a result of my study. It is my goal to demonstrate the scientific contributions that this research activity has made. Finally, I propose many potential study directions for consideration in future work.

Chapter 2

Overview of Simulation and Analysis Methods

2.1 Introduction

This chapter outlines the computational methods and algorithms used for the research in this thesis. The concept of Molecular Dynamics (MD) was introduced by Alder and Wainwright in late 50's to find out the interactions of hard spheres by using MD Simulation[61]. After around 20 years, the first simulation for the protein was performed by McCammon et al[62]. They succeeded in simulating all atom MD of Bovine Pancreatic Trypsin Inhibitor which consist of 58 residues for less than 10 ps. Over the past several decades, improvements in algorithms, software, and computer hardware have allowed simulations of microsecond to millisecond timescales for systems with tens of thousands of atoms in solvated conditions. Therefore, MD simulations has become increasingly important for chemists, physicists, bio-scientists, and engineers. Nowadays, all-atom MD simulations are considered as a powerful tool for studying the behavior of biological macromolecules (e.g., proteins, RNA, and DNA). Simulating the three-dimensional motion of these biomolecules could be used to elucidate the conformational transitions of biomolecular systems, and can facilitate drug design by identifying and quantifying how small-molecule compounds interact with potential drug targets[19, 20, 63].

The central idea behind MD simulations to study the time-dependent behavior of biological macromolecules is solving the second-order differential equations represented by Newton's second law in equation 2.2 [19]:

$$\mathbf{F}_i(t) = -\frac{\partial U(\mathbf{x}(t)}{\partial \mathbf{x}_i(t)} \tag{2.1}$$

$$\mathbf{F}_i(t) = m_i \mathbf{a}_i(t) \quad = m_i \frac{\partial \mathbf{v}_i(t)}{\partial t} \quad = m_i \frac{\partial^2 \mathbf{x}_i(t)}{\partial t^2} \tag{2.2}$$

where $\mathbf{F}_i(t)$ is the net force acting on the i^{th} atom of the system at a given point in time t. $\mathbf{a}_i(t)$ is the corresponding acceleration, and m_i is the mass. The acceleration as in equation 2.2 is the first derivative of velocity $\mathbf{v}_i(t)$ with respect to time. The velocity is defined as the derivative of position vector $\mathbf{x}_i(t)$ of i^th atom of the system with respect to time. Therefore, the instantaneous configuration of the system, represented by the vector $\mathbf{x}_i(t)$, describes the position of the N interacting atoms in the Cartesian space ($\mathbf{x} = \{x_1, y_1, z_1, \quad x_2, y_2, z_2, \quad ..., \quad x_N, y_N, z_N\}$). As a result, in order to determine the configuration of the system, we must first determine the force acting on it, as specified in equation2.2. Typically, in MD simulations, we use a classical mechanics description of the forces, which limits us to the motion of the nuclei. The net force $\mathbf{F}_i(t)$ is given by the negative gradient of the potential-energy function with respect to the position of atom *i* as shown in equation 2.1. To do this an empirical potential energy function is introduced and the simplified representation resulting from this is known to as the force field (FF), which I will discuss briefly in the next sections. Prior to that, I would want to briefly describe the major steps of the MD simulation before digging into them in greater detail.

Main steps of a MD simulation

The 3D structure of the biomolecular system required for all atoms MD simulation is often prepared from x-ray crystallography, nuclear magnetic resonance (NMR), cryo-electron microscopy, or homology modeling. Experimentally solved protein structures are collected and distributed by the Worldwide Protein Data Bank (wwPDB), which stores more than 184,700 structures to date and $\approx 10,000$ new structures are deposited annually[64].

Section 2.2 is a general introduction to the basics of conventional MD simulations. While each system analyzed will provide its own set of obstacles and concerns, the basic procedure for performing

a MD simulation is as follows:

- **System preparation**:

  System preparation is concerned with construction of the initial state of the simulation system as input for a suitable simulation package, which may include generating a starting structure, solvation into water (if required), and preparing a FF. Section 2.2.1 will discuss FF and the simulation packages in depth. This is the most essential stage of the simulation and varies significantly depending on the system composition and the available knowledge about the initial structure. The worst-case scenario is that the prepared system is not what you intended (e.g., it contains incorrect molecules or protonation states) but chemically valid and well described force field allows you to proceed without error through the remaining steps — and indeed, this is a frequent outcome of system preparation problems.

- **Minimization/Relaxation** :

  The goal of minimization, or relaxation, is to move the initial structure to a local energy minimum in order to prevent the molecular dynamics simulation from instantly "exploding" (i.e., the forces on certain atoms are too large so that the atoms move an unreasonable distance in a single time step). This is achieved by standard minimization algorithms such as steepest descent.

- **Assignment of velocities** :

  For MD simulation, we require not just atomic positions, but also velocities. Nevertheless minimization only gives a final set of positions. As a result, random initial velocities must be given to atoms in order to achieve the appropriate Maxwell-Boltzmann distribution at the specified temperature. Then, using one of the appropriate techniques that I shall discuss in Section 2.2.2, we may begin numerical integration of the equations of motion.

- **Equilibration** :

  Typically, we are interested in selecting the most probable configurations in a certain thermodynamic ensemble (e.g., the NVE or NVT ensemble) at a given state point (e.g. temperature, and pressure). However, if we begin in a less stable configuration, a significant portion of our equilibration time may be spent relaxing to reach the more relevant configuration space. In sections 2.2.3 and 2.2.4, I will discuss the various statistical ensembles and the algorithms that control temperature and pressure during MD simulations.

- **Production** :

  After completing equilibration, data collection for analysis may begin. Typically, this stage is referred to as "production." The primary distinction between equilibration and production simulations is that we want to keep and evaluate the data obtained during the production simulation.

The analysis of production entails calculating expectation values that may provide insights about the atomistic mechanisms underlying significant biological phenomena, such as conformational change[23, 25], ligand binding[26], and protein folding[27, 28]. However, some other biological processes often take place in much longer timescales and cannot generally be observed by all-atom MD simulations such as protein-ligand, protein-protein and protein-DNA binding/unbinding, and protein aggregation. In section 2.3 the details of the different techniques used to enhance the sampling during the simulation to facilitate investigating such rare events will be introduced. Then in section 2.4 I will describe some of the computational techniques to analyze trajectory data and finally the important tool in computational biology will be introduced in section 2.5.

2.2 Conventional Molecular Dynamics Simulations

2.2.1 Force Fields

The set of potential energy functions from which atomic forces are derived is called force fields. Typically, these energy functions are composed of a large number of parametrized components based on experimental and/or quantum mechanical studies of tiny molecules or fragments. Such parameters are thought to be transferable to the larger biomolecule of interest[29]. The precise decomposition of the FF terms is defined as the term of the energy of bonded and non-bonded term as in equation 2.3:

$$U_{total} = U_{bonded} + U_{non\_bonded} \tag{2.3}$$

The component of bonded and non-bonded depends on the covalent and non-covalent bond interactions expressed by the following equations:

$$U_{bonded} = U_{bonds} + U_{angles} + U_{dihedrals} \tag{2.4}$$

$$U_{non\_bonded} = U_{electrostatic} + U_{vanderwalls} \tag{2.5}$$

The commonly used empirical force fields for biomolecular systems simulation which use the similar functional form of bonded and non-bonded terms are AMBER (Assisted Model Building and Energy Refinement)[65], CHARMM (Chemistry at Harvard Macromolecular Mechanics)[66], OPLS (Optimized Potentials for Liquid Simulations)[67], and GROMOS (GROningen MOlecular Simulation)[68]. A general functional form of AMBER force field is shown in the following equation [69]:

$$\begin{aligned} U_{total} = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ + \sum_{i<j}[\frac{A_{ij}}{{R_{ij}}^{12}} - \frac{B_{ij}}{{R_{ij}}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}] \end{aligned} \tag{2.6}$$

The first three terms in equation 2.6 represent intramolecular interactions of the atoms or bonded term of the FF. They describe variations in potential energy as a function of bond stretching (U_{bonds}), angle-bending (U_{angles}), and dihedral angles ($U_{dihedrals}$) between atoms directly involved in bonding relationships as shown in Fig.2.1[19, 70]. The bond-stretching and angle-bending terms, first and second term in equation 2.6, are expressed as a harmonic potentials with bond lengths (r), angles (θ), force constants k_b and k_θ, and reference values r_0 and θ_0 respectively. The dihedral (torsional) angle is comprised of four consecutively connected atoms. Therefore, Fourier type expansion as expressed in the third term of equation 2.6 is used to characterize the dihedral potential. Here V_n is the barrier to free rotation for the "natural" bond, n is the periodicity of the rotation (number of cycles in 360°), ϕ is the dihedral angle and γ is the angle where the potential passes through its minimum value.

The non-bonded interactions in FF are calculated pairwise between two atoms, denoted i and j, and commonly include van der Waals ($U_{vanderwaals}$) and electrostatic ($U_{electrostatic}$) interactions represented by fourth and fifth terms in equation 2.6, respectively and shown in Fig.2.1. The van der

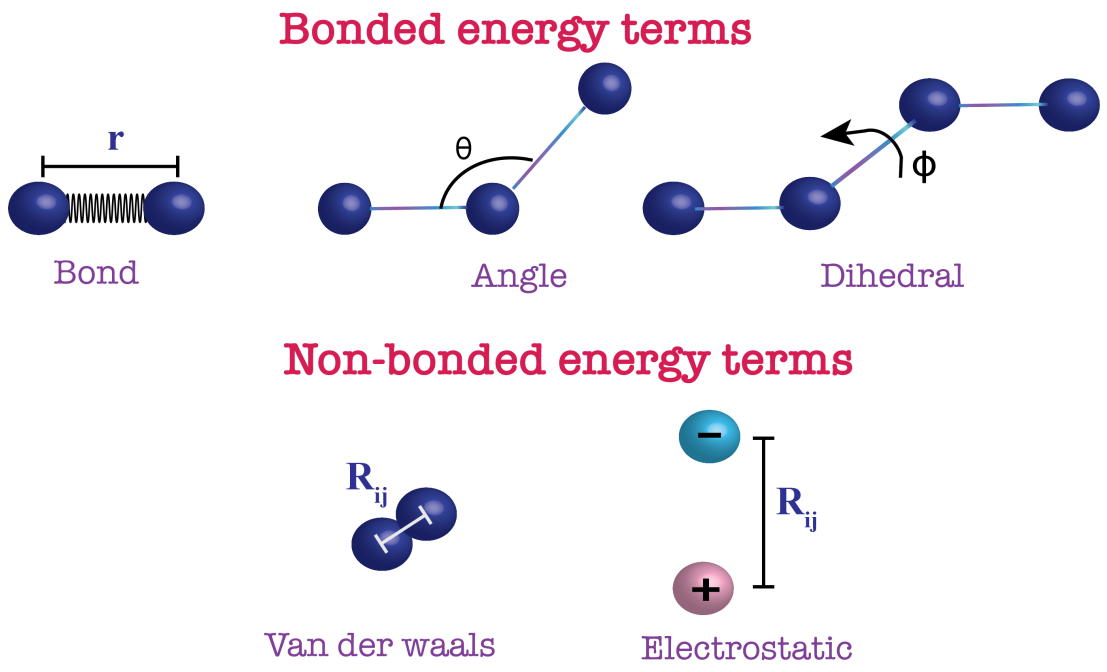

FIGURE 2.1: Schematic illustration of bonded and non-bonded energy terms of the empirical force field.

Waals interactions are generally treated with a 12-6 Lennard-Jones potential where the A_{ij} and B_{ij} parameters control the depth and position (interatomic distance) of the potential energy well for a given pair of non-bonded interacting atoms and R_{ij} is the interatomic distance. For the electrostatic interactions which modeled by a Coulombic interaction of atom-centered point charges, q_i and q_j are the point charges on atoms and ε is the dielectric constant.

The calculations of non-bonded pairwise interactions are the most computationally expensive portion of a MD simulation. As the number of atoms (N) in a system increases, the number of van der Waals and electrostatic interactions will grow as the square of that number (N^2), potentially resulting in a prohibitively large number of interactions to evaluate. For that reason in earlier MD simulations A *spherical truncation* scheme applied to cutoff these non-bonded interactions at certain distance, for instance 8 Å. This means the interactions beyond that cutoff distance were ignored to limit the maximum number of interactions to reduce the computational cost. While this can be accepted for van der Waals interactions because it is quickly decay with respect to distance R_{ij}^{-6}, in electrostatic interactions this is not accepted. Electrostatic interactions are fundamentally long-range acting forces and play a dominant role in protein structural stability and are also crucial determinants in the initial encounter of many association processes[29]. Hence severe errors can arise from neglecting electrostatic interactions beyond some cutoff distance. Therefore, in many

types of simulations particle-mesh Ewald (PME) scheme can be used to include all long-range electrostatics with a comparable simulation cost. PME is a method to efficiently calculate the infinite range Coulomb interaction under periodic boundary conditions (PBC), repeating copies of the system. Because the Coulomb interaction has infinite range, under PBC particle ***i*** within the unit cell interacts electrostatically with all other particles ***j*** within the cell, as well as with all the periodic images of ***j***. It also interacts with all of its own periodic images[71].

2.2.2 Integration Algorithms

Because molecular systems often include a large number of particles, all atom positions (3N) in a biomolecular system are a function of sophisticated in nature potential energy. These equations have no analytical solution and can only be solved numerically. Many numerical algorithms have been developed for integrating the Newton's equations of motion. The velocity Verlet algorithm[72], which is a variant of the original Verlet method[73], is one of the most extensively used numerical integrators. In addition, another algorithm called leap-frog[74, 75] can also used to numerically integrate the equations of motion.

All algorithms assume that the positions and dynamic properties (velocities, accelerations, etc.) can be approximated as Taylor series expansions ,as in equation 2.7, after dividing the integration step into small steps, each separated by a fixed time step Δt

$$\begin{aligned}
\mathbf{r}(t+\Delta t) &= \mathbf{r}(t) + \Delta t\, \mathbf{v}(t) + 1/2\ \Delta t^2\ \mathbf{a}(t) + 1/6\ \Delta t^3\ b(t) + 1/24\ \Delta t^4\ c(t) + \ldots \\
\mathbf{v}(t+\Delta t) &= \mathbf{v}(t) + \Delta t\, \mathbf{a}(t) + \frac{1}{2}\Delta t^2\, b(t) + \frac{1}{6}\Delta t^3\, c(t) + \ldots \\
\mathbf{a}(t+\Delta t) &= \mathbf{a}(t) + \Delta t\, b(t) + \frac{1}{2}\Delta t^2\, c(t) + \ldots \\
\mathbf{b}(t+\Delta t) &= b(t) + \Delta t\, c(t) + \ldots
\end{aligned} \tag{2.7}$$

where **v** is the velocity (the first derivative of the position with respect to time), **a** is the acceleration (the second derivative), **b** is the third derivative and so on.

2.2.2.1 The Verlet algorithm

The most commonly used time integration algorithm in MD simulation and computer graphics is Verlet algorithm[73]. The main idea in this algorithm is to write two third-order Taylor expansions for the positions one forward $\mathbf{r}(t + \Delta t)$ and one backward $\mathbf{r}(t - \Delta t)$ in time as in equations 2.8 and 2.9

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t\, \mathbf{v}(t) + 1/2\ \Delta t^2\ \mathbf{a}(t) + 1/6\ \Delta t^3\ b(t) + 1/24\ \Delta t^4\ c(t) \quad (2.8)$$

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \Delta t\, \mathbf{v}(t) + 1/2\ \Delta t^2\ \mathbf{a}(t) - 1/6\ \Delta t^3\ b(t) + 1/24\ \Delta t^4\ c(t) \quad (2.9)$$

Adding these two equations 2.8 and 2.9 and rearrangement gives us:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \Delta t^2\ \mathbf{a}(t) \quad (2.10)$$

Take in consideration, higher-order terms in the Taylor expansion are ignored. One of the drawbacks of Verlet algorithm is that the velocities do not appear as an explicit term but we can calculate velocities in variety of ways using equation 2.11 :

$$\mathbf{v}(t) = [\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)]\ /\ 2\Delta t \quad (2.11)$$

Also, the velocity can be estimated at the half step as follows:

$$\mathbf{v}(t + \Delta t) = [\mathbf{r}(t + \Delta t) - \mathbf{r}(t)]\ /\ \Delta t \quad (2.12)$$

However, the velocities are not available until the positions have been computed at the next step. Another disadvantage, Verlet algorithm is not a self-starting algorithm; the new positions are obtained from the current positions $\mathbf{r}\,(t)$ and the positions of the previous time step $\mathbf{r}(t - \Delta t)$ which we do not have at t = 0. Because of all of these limitations, several variants of the Verlet algorithm have been developed to address some of the original Verlet algorithm's shortcomings.

2.2.2.2 The leap-frog algorithm

To overcome the lack of explicit velocity, leap-frog algorithm use an approximation for derivative, one should consider the velocity v_i at the midpoint between times (t) and $(t + \Delta t)$

$$\mathbf{v}\left(t + \frac{1}{2}\,\Delta t\right) = \frac{\mathbf{r}\,(t + \Delta t) - \mathbf{r}\,(t)}{\Delta t} \tag{2.13}$$

The equation 2.13 can be solved in term of $r\,(t + \Delta t)$ and give us :

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t\ \ \mathbf{v}\left(t + \frac{1}{2}\,\Delta t\right) \tag{2.14}$$

$$\mathbf{v}\left(t + \frac{1}{2}\,\Delta t\right) = \mathbf{v}\left(t - \frac{1}{2}\,\Delta t\right) + \ \ \Delta t\ \ \mathbf{a}(t) \tag{2.15}$$

To apply leap-frog algorithm, firstly we calculate velocity at time $\left(t + \frac{1}{2}\Delta t\right)$ from accelerations at time (t) and velocities at time $\left(t - \frac{1}{2}\Delta t\right)$. Then the positions at time $(t + \Delta t)$ can be inferred from the velocities just calculated at time $\left(t + \frac{1}{2}\Delta t\right)$ and the positions at time (t) using equation 2.15. Then, the velocities at time (t) can be calculated from

$$\mathbf{v}(t) = \frac{1}{2}\ \ [\mathbf{v}\left(t + \frac{1}{2}\,\Delta t\right) + \mathbf{v}\left(t - \frac{1}{2}\,\Delta t\right)] \tag{2.16}$$

This means that the velocities **leap frog** over the positions to give their values at time $\left(t + \frac{1}{2}\Delta t\right)$ then the positions **leap frog** over the velocities to give their new values at time $(t + \Delta t)$, which will be used for the next velocities at time $\left(t + \frac{3}{2}\Delta t\right)$, and so on. As a result, the leap-frog algorithm addresses the Verlet algorithm's explicit velocities disadvantages. However, it still has issues such as synchronized positions and velocities, as well as the inability to self-start.

2.2.2.3 The Velocity Verlet algorithm

The Velocity Verlet was developed in order to solve the associated problem with the Verlet algorithm. The velocity and position are determined at the same time period in this algorithm, which is similar to the leapfrog approach. The velocity Verlet algorithm gives the positions and

dynamic properties (velocities, accelerations) at the same time as in equations 2.17 and 2.18.

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{v}(t) + \frac{1}{2}\Delta t^2 \mathbf{a}(t) \tag{2.17}$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{1}{2}\Delta t[\mathbf{a}(t) + \mathbf{a}(t + \Delta t)] \tag{2.18}$$

As shown in equation 2.18, the velocity Verlet algorithm is applied as three steps.

- **First :** The positions at time $(t + \Delta t)$ are calculated by equation 2.17 using velocities and acceleration at time (t).
- **Second :** The velocities at time $\left(t + \frac{1}{2}\Delta t\right)$ determined using :

$$\mathbf{v}\left(t + \frac{1}{2}\,\Delta t\right) = \mathbf{v}(t) + \frac{1}{2}\Delta t\;\; \mathbf{a}(t) \tag{2.19}$$

  Then net forces are computed from the current positions to give acceleration at time $(t + \Delta t)$
- **Third :** The velocities at time $(t + \Delta t)$ are determined using:

$$\mathbf{v}(t + \;\Delta t) = \mathbf{v}\left(t + \frac{1}{2}\,\Delta t\right) + \frac{1}{2}\Delta t\;\; \mathbf{a}(t + \;\Delta t) \tag{2.20}$$

It's worth noting that the Velocity Verlet and Leapfrog methods outperform the semi-implicit Euler technique in the long run. Up to a half-velocity time step, these algorithms are nearly equivalent. In the semi-implicit Euler technique, the Velocity Verlet differs solely when the midway velocity is used as the final velocity. The method has a second-order error, similar to the midway method.

2.2.3 Statistical Ensembles

To correctly imitate the experimental settings, numerous physical parameters, such as pressure and temperature, may be easily considered in the simulations. Dynamic properties of macromolecules are measured as an ensemble of billions of molecules called systems, rather than as direct observations. A **microstate** is a specific microscopic configuration of a system, while an **ensemble** is a collection of all potential systems with different microscopic states but all belonging to the same macroscopic or thermodynamic state. To put it another way, ensemble refers to a large group of

microscopically described states (microstates) of a system with certain set of constant macroscopic properties (such as temperature, pressure, and volume) that will govern the complete simulation of the system[25, 29, 76].

The microstate can be characterized by positions **q** and momenta **p** variables of the system. The Hamiltonian, H, which describes the total energy of a microstate is given by:

$$H(\boldsymbol{q},\boldsymbol{p}) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + U(\boldsymbol{q}) \tag{2.21}$$

where the first term is the the kinetic energy and the second term is potential energy of the system. There are different ensembles are used in MD simulation as following:

- **Microcanonical ensemble (NVE)** A system (solid, liquid, or gas) that has a constant number of particles (N) and energy and is totally isolated from volume (V) variations. The interchange of kinetic and potential energy with the total energy constant is seen in this ensemble. All microstates in this ensemble have the same probability, which is given by:

$$P\left(\boldsymbol{q},\boldsymbol{p}\right) \propto \delta\left[H\left(\boldsymbol{q},\boldsymbol{p}\right) - E\right] \tag{2.22}$$

- **Canonical ensemble (NVT)** The canonical ensemble is the collection of all systems whose thermodynamic state is characterized by a fixed number of particles (N), at fixed volume (V), with a contact of heat bath with constant temperature (T). Hence, MD simulation in NVT ensemble is also referred as Constant Temperature Molecular Dynamics. As the system is coupled to an heat reservoir, the microstates of the systems can have different energies with probability of finding a microstate *i* with energy E_i is given by :

$$P\left(\boldsymbol{q},\boldsymbol{p}\right) \propto \exp\left(-\beta E_i\right) \tag{2.23}$$

  where $\beta = \frac{1}{k_B T}$ is the inversed temperature and k_B is the Boltzmann constant.

- **Isothermal-isobaric ensemble (NPT)** Isothermal-isobaric ensemble is a statistical mechanical ensemble with a constant number of particles (N), temperature (T), and pressure (P). The isothermal-isobaric ensemble, commonly known as the NPT ensemble, is one of the most extensively used ensembles since the majority of practical experiments are conducted under

tightly controlled temperature and pressure settings. The probability of finding a microstate *i* with energy E_i is given by :

$$P\left(\boldsymbol{q},\boldsymbol{p}\right)\left\{-\beta\left(E_i+PV\right)\right\} \tag{2.24}$$

- **Grand canonical ensemble (μVT)** Grand canonical ensemble is the statistical ensemble where the volume (V) , temperature (T), and chemical potential (μ) are fixed, but the number of particles and energy can exchange with the surrounding bath. This ensemble is particularly applicable to systems such as chemical reactions where the number of particles varies. The probability of finding a microstate *i* with energy E_i is given by :

$$P\left(\boldsymbol{q},\boldsymbol{p}\right)\left(-\beta E_i+\mu\beta N_i\right) \tag{2.25}$$

To model biomolecular systems, I must generally imitate experimental parameters such as constant temperature and pressure. In the next section, I will give more details about controlling such conditions

2.2.4 Controlling Temperature and Pressure

Typically, thermodynamic characteristics of interest are examined in a laboratory under open-air circumstances, which means they are examined at practically constant temperature and pressure (on a short time scale). If I want to control temperature and pressure, I must use a thermostat and a barostat algorithm.

2.2.4.1 Thermostats

A thermostat is a Newtonian MD system modification that generates a statistical ensemble at a constant temperature. It can replicate experimental environments, modify temperature in algorithms, and prevent energy drifts due to numerical errors[25, 76].

The temperature during the MD simulation can be related to the kinetic energies by the equipartition theorem which is given by:

$$\left\langle \sum_{i=1}^{N} \frac{1}{2} m_i {v_i}^2 \right\rangle = \frac{3}{2} N k_B T \tag{2.26}$$

where the angle brackets indicate the time-averaged quantity. The temperature estimated from a single snapshot is referred to as the instantaneous temperature, which is not necessarily equal to the desired temperature but fluctuates around it.

Several popular and historic thermostats which are used in MD will be briefly explored:

- **Velocity rescaling thermostat[77]** Velocity rescaling thermostat is one of the easiest thermostats to implement. In this thermostat, rescaling the velocity at every time steps achieves by fixing kinetic energy (isokinetic) to match the temperature of MD. Therefore, the temperature reaches the desired temperature without any fluctuations and the generated ensemble is isokinetic ensemble rather than the canonical ensemble.

- **Berendsen thermostat[78]** The Berendsen thermostat which also known as the weak coupling thermostat is similar to the simple velocity rescaling thermostat but instead of rescaling velocities completely and abruptly to the target kinetic energy, it includes a relaxation term to allow the system to more slowly approach the target. The idea of Berendsen thermostat, when a system at a specific average temperature T comes in contact with a heat bath at a different temperature T_0 (the target temperature), the rate of temperature change is given by :

  $$\frac{dT}{dt} = \frac{1}{\tau} \left(T_0 - T \right) \tag{2.27}$$

  where τ should be the relaxation time which determines how tightly the temperature bath and the system are coupled together. The scaling factor is given by:

  $$\lambda = \left\{ \frac{\left(3N - 1 \right) k_B T_0}{\sum_i m_i {v_i}^2} \right\}^{\frac{1}{2}} \tag{2.28}$$

  where N is the total number of particles, and (3N-1) is the number of degrees of freedom of the system. Although the Berendsen thermostat allows temperature fluctuations through τ, the system will never reach the appropriate value for a canonical ensemble.

- **Langevin thermostat[79]** The Langevin thermostat controls the temperature to a fixed reference temperature by inserting friction and stochastic terms in the equation of motion. The

modified equations of motion in Langevin dynamics is given by:

$$\mathbf{F}_i = m_i \frac{d^2\mathbf{r}_i}{dt^2} + m_i\gamma_i \frac{d\mathbf{r}_i}{dt} - R_i\left(t\right) \tag{2.29}$$

the first term is the standard interactions calculated during the simulation. The second term is the damping used to tune the friction of the implicit bath where γ_i is the collision frequency which determines the strength of the coupling to the heat bath. The third term is the stochastic force $R_i\left(t\right)$, effectively gives random collisions with solvent molecules and assumed to be uncorrelated with the positions and velocities of the particles.

Although the Langevin equation is known to converge to the canonical ensemble, the collision frequency must be carefully examined in order to maintain the temperature without significantly perturbing the system's dynamics, which will become microcanonical when γ_i = 0.

- **Nosé-Hoover thermostat[77]** The Nosé-Hoover thermostat extracts away the thermal bath from the Langevin thermostats and condenses it into a single extra degree of freedom. This fictitious degree of freedom has a "mass" that may be modified to interact with the particles in the system in a predictable and reproducible manner while preserving the canonical ensemble. The "mass" of the fictitious particle is significant since it affects the fluctuations that will be noticed. Although the Nosé-Hoover thermostat is one of the most extensively developed and utilized thermostats, it should be noted that ergodicity can be a concern in small systems.

2.2.4.2 Barostats

Similar to temperature, constant pressure can be controlled at constant value by a barostat that modifies the volume of the simulated system.

The pressure during the MD simulation can be measured using the virial theorem:

$$\mathbf{P} = \frac{2}{3\mathbf{V}}\left(\langle\mathbf{E}_k\rangle - \langle\frac{1}{2}\sum_{i<j}\mathbf{r}_{ij} \bullet \mathbf{F}_{ij}\rangle\right) \tag{2.30}$$

where $\mathbf{V}$ is the volume of the system, E_k is the kinetic energy, $\mathbf{F}_{ij}$ is the force on particle *i* due to particle *j*, and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. Again as in thermostat, form this formula in equation 2.30 a

time-averaged quantity of the pressure can be obtained and the instantaneous pressure is calculated using a single snapshot. The pressure will not always be equal to the target pressure but undergoes fluctuations around the target pressure. There are several methods used for pressure control which are analogous to those used for temperature control.
I will briefly explore popular and historic barostats used in MD :

- **Volume rescaling barostat[25, 76]** The easiest way is to rescale the volume of the system every time this barostat is executed where the volume of the system is modified such that the instantaneous pressure is exactly equal to the target pressure. This approach does not properly generate the isothermal-isobaric ensemble. Also, it does not smoothly approach the target pressure either, which might cause very unphysical issues with the system during integration.

- **Berendesen barostat[78]** Berendesen barostat is analogous to the Berendsen thermostat where the pressure is weakly coupled to a "pressure bath" which slowly approaches to the target pressure. To maintain the system to a desired pressure P_0, the atomic coordinates and volume are scaled periodically by a scaling factor, such that the rate of change of pressure is given by :

$$\frac{dP}{dt} = \frac{1}{\tau}(P_0 - P) \tag{2.31}$$

where τ should be the relaxation time which determines how tightly the pressure bath and the system are coupled together. The scaling factor is given by:

$$\mu = 1 - \frac{\beta \Delta t}{3\tau}(P_0 - P) \tag{2.32}$$

where β is the compressibility which may not be accurately known. The compressibility does not have to be precisely known as the compressibility and the relaxation time appear only as a ratio in the dynamics. In practice, compressibility of liquid water is often used. While the Berendsen barostat approaches the target pressure realistically, the ensemble from which it samples is not clearly defined and so cannot be assured to be NPT ensemble. Berendsen may be advantageous during the first phases of equilibration, but it should not be utilized for production sampling.

- **Andersen barostat[80]** The Andersen barostat is an extended system algorithm similar to the Nosé-Hoover thermostat. By introducing an additional degree of freedom to the equations of motion, the system is coupled to a fictitious pressure bath. This behaves as though an isotropic piston is acting on the system. The right ensemble is sampled by this barostat. It is, however, isotropic in nature, therefore applying anisotropic pressures to different sections of the system is not conceivable.

- **Parrinello-Rahman[81]** This barostat has roughly the same features as the Andersen barostat, but also supports anisotropic scaling of the simulation box's size and shape, which makes it particularly helpful for solid simulations.

- **Martyna-Tuckerman-Tobias-Klein (MTTK)[82]** The MTTK barostat has a high degree of resemblance to the Parrinello-Rahman and Andersen barostats. When it was determined that Parrinello-Rahman's equations of motion held true only for big systems, the MTTK barostat included additional equations of motion to sample the ensemble appropriately for smaller systems as well. Thus, MTTK is often regarded as a superior algorithm than Parrinello-Rahman for such systems.

2.3 Enhanced Sampling Techniques

The dynamics of biomolecules is directly linked to their functions, with major conformational changes typically taking place on timescales ranging from microseconds to seconds[83]. The main obstacle to use all-atom MD simulation to observe such rare biological events is that the achievable time scale is too short (usually less than a microsecond), even if significant computational resources are used. Recently, a new method introduced to improve the efficiency of MD simulations which referred to as enhanced sampling techniques. A wide variety of enhanced sampling techniques have been proposed to capture the long-timescale biological events such as protein-ligand, protein-protein, and protein-DNA binding/unbinding, protein aggregation, domain motion and protein folding. Some of these methods apply artificial bias to enhance sampling, while others can enhance the sampling without applying any biasing force[19, 29, 30]. Here I will introduce details of the most widely used methods in both categories.

2.3.1 Biased enhanced sampling techniques

2.3.1.1 Metadynamics

Metadynamics [37, 38] is a powerful technique that has been successfully developed by Parrinello and coworkers to enhance the sampling of rare events. Metadynamics enhances the rare events occurrence by discouraging the system from revisiting the same phase space by introducing a repulsive bias potential to the original potential. In Metadynamics simulation, an external history-dependent bias potential is imposed on the system during the simulation. The bias potential which is a function of the collective variables (CVs) can be written as a sum of Gaussians added along the CVs space to encourage the system to visit configurations which have not already been sampled as shown in figure (2.2).

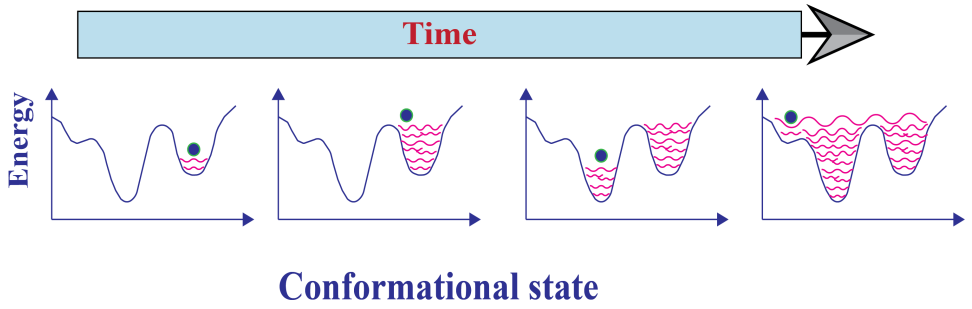

FIGURE 2.2: Schematic representation of the progressive filling of the potential by means of the Gaussians deposited along the trajectory.

If S considered as a set of d functions of the microscopic coordinates R of the system. At time t, the Gaussian-type bias potential $V_G(S,t)$ can be expressed as:

$$V_G(S,t) = \int_0^t dt' \omega \quad exp\left(-\sum_{i=1}^{d} \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2} \right) \tag{2.33}$$

where ω is an energy rate and σ_i is the width of the Gaussian for the *i*th CV.

As shown in figure (2.2), the idea in Metadynamics to escape local minima is adding a small Gaussian hill to the potential energy of the current region of state space periodically. The local minimum is gradually filled in by the addition of additional Gaussian hills, forcing the system to explore a variety configurations. After a sufficiently enough simulation, the total potential energy

will become flat as all the minima are filled by accumulated Gaussian hills. As a result, this approach is dependent on certain characteristics such as hill height, width, and frequency of addition, which might be adjusted carefully to make sure that they do not affect the statistics. While this strategy is successful for exploring a small number of CVs, its performance degrades fast as the number of CVs increases, as the computational effort needed to discourage the visiting phase space visits increases.

Metadynamics not only has found widespread applications in the field of biomolecular dynamics simulations[84–86] but it also successfully used in Material science[87] and chemical reactions[88].

2.3.1.2 Steered MD

Steered MD (SMD) simulations are basically an enhanced sampling approach that incorporates time-dependent external forces for the purpose of investigating the free energy profiles in a chosen direction as shown in Fig 2.3[89, 90].

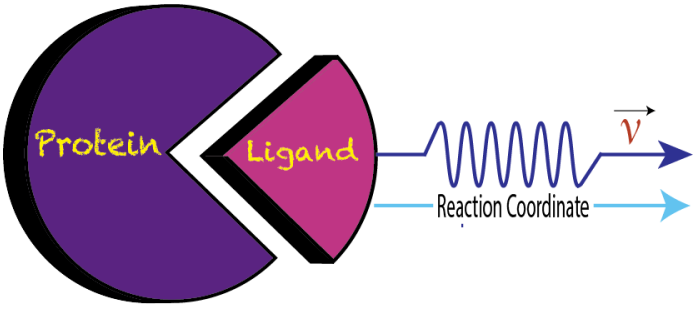

FIGURE 2.3: Constant velocity SMD simulations for pulling a ligand out of its complex

SMD is inspired by experimental techniques that use mechanical forces to accelerate transitions between different energy minima and facilitate the study of biomolecules' binding properties, such as single-molecule pulling experiments[91], atomic force microscopy (AFM)[92], optical tweezers[93], and biomembrane force probe[94]. External forces may be applied to a protein-ligand complex by constraining the ligand at a specific place in space (constraint point) with an external harmonic potential. Then, by repositioning the constraint point in a desired direction, the ligand is forced to migrate away from its original location in the protein, enabling it to explore other contacts along its unbinding route. Assuming that we have a single reaction coordinate in the x direction and an external potential equal to $U = K\left(x - x_0\right)^2/2$, then the external force acting on the system

can be given by

$$\mathbf{F} = K\left(x_0 + \mathbf{V}t - x\right) \tag{2.34}$$

where *K* is the stiffness of the restraint, and x_0 is the initial position of the restraint point moving with a constant velocity **v**. In addition, Jarzynski's equality[95] may be used to determine the potential of mean force along the conformational change route in SMD dissociation simulations, where the transition is defined by the amount of work necessary to cause it.

$$\Delta\mathbf{F} = -\frac{1}{\beta}\ln\left\langle e^{-\beta W}\right\rangle \tag{2.35}$$

where W is the work performed on the system by external force, and $\beta = 1/k_B T$ is the inverse temperature and k_B is the Boltzmann constant. the average bracket in equation 2.35 implies that we can directly calculate the equilibrium information (binding free energy ΔG) from the ensemble of non-equilibrium quantity (work acts on the system) that can be calculated from integration of the force applied on the system in equation 2.34.

Thus, recording applied forces and ligand location over time offers structural information regarding the ligand-receptor complex's structure-function correlations, binding routes, and processes governing enzyme selectivity. SMD simulations have gained prominence in recent years and have already generated considerable qualitative insights into a range of biologically relevant topics, including protein folding/unfolding[39, 96], binding/unbinding[97], drug transport across membrane channels[98], and other biochemical processes[99].

2.3.1.3 Temperature replica exchange MD

Elevating the temperature of a simulated system, i.e., introducing a kinetic bias, is a straightforward technique to enhance the velocity of movements. Because the transition rates between local minima rise at increasing temperatures, a broader phase space may be examined in less computing time. Therefore, such techniques (that change the temperature) avoid the trapping of biomolecules in local energy-minimum configurations as simulated by typical MD simulations (single low temperature) and facilitate barrier crossing events in the biomolecules[32, 100].

As illustrated in Figure 2.4[100], temperature replica exchange MD (T-REMD) which also referred to as parallel tempering are based on running multiple copies (replicas) of parallel MD

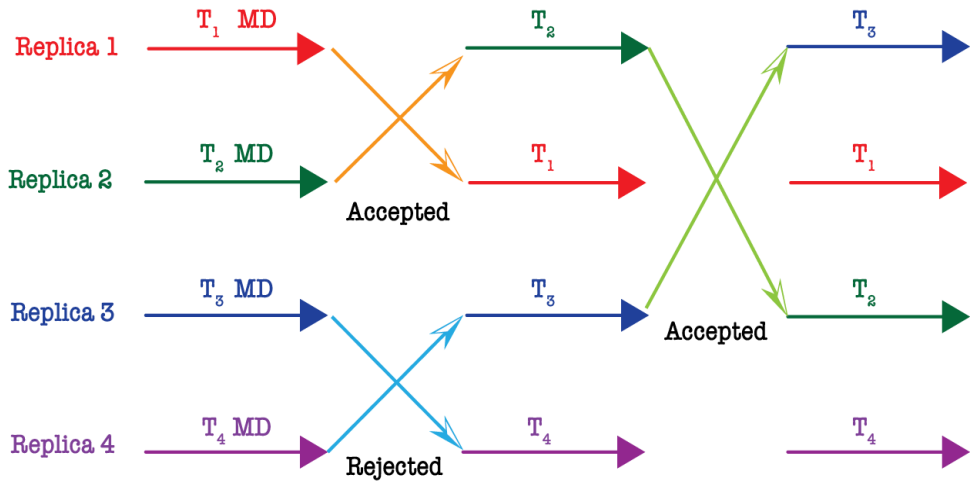

FIGURE 2.4: Schematic illustration of T-REMD with temperature exchange between 4 non-interacting replicas of MD runs. At a certain MD simulation time interval(represented by colored arrows), exchange between each pair of replicas with neighboring temperatures at a probability that meets the Metropolis criterion can occur.

simulations at different temperatures. Exchanges of system configurations (temperature swapping) between neighboring replicas are attempted at defined time intervals to achieve enhanced sampling. The replica exchanges can be accepted or rejected based on an energetic criterion, which ensures that the generalized ensemble of the system satisfies a detailed balance and converges towards the canonical distribution. Typically, this routine is based on a Metropolis criterion, where the exchange probability $P\left(T_k \rightarrow T_l\right)$ depends on the reference temperature of two replicas $\left(T_k, T_l\right)$ and their potential energies $\left(V_k, V_l\right)$:

$$P\left(T_k \rightarrow T_l\right) = \begin{cases} 1, & \Delta \leq 0 \\ \exp(-\Delta), & \Delta > 0 \end{cases} \tag{2.36}$$
$$\Delta = \left(\beta_k - \beta_l\right)\left(V_k - V_l\right),$$

with $\beta_k = 1/k_B T_k$, $\beta_l = 1/k_B T_l$ being the inverse of the temperature T_k and T_l multiplied by the Boltzmann constant k_B. As a result of the nature of this technique, a computational setup is required in which numerous simulations may be run concurrently with sufficiently quick and frequent communication between the computing nodes. Although the use of specialized MD engines on graphics processing units (GPUs) significantly reduces the computational cost[101], the challenge becomes more difficult for large biomolecular systems because the number of required replicas is estimated to increase with $N^{1/2}$ for a system with N degrees of freedom[102]. Nevertheless, given the availability of the required high-performance computing environment, the key

problem is how to choose the number and interval of replicas, as well as the whole temperature range. Consider that the highest temperature should be sufficiently high to prevent replicas from being trapped in local energy minima, and that each replica should spend an equal amount of time at each temperature, implying that temperatures should be set in such a way that neighboring replicas have a comparable acceptance ratio. It was demonstrated that exchange acceptance probability of about 20% yields optimal performance[103]. Estimating the ideal temperature distribution, the number and spacing between T-REMD are not straightforward, and numerous ideas for adjusting these parameters have been made[104, 105]. However, these parameters should be re-evaluated for user-specific MD engine and force field combinations.

T-REMD is more prevalent in practice because to its simplicity of implementation and the absence of weighting elements to tune. Additionally, T-REMD simulations result in substantial temperature sampling, which may give useful information in addition to increased conformational sampling. Moreover, by doing a weighted-histogram analysis (WHAM), it is feasible to aggregate data from numerous replicates and determine the anticipated value of any physical quantity at any temperature[106]. Since the introduction of T-REMD in the remarkable research of penta-peptide Met-enkephalin folding[106], several studies have been conducted to better explain the protein folding issue, determine the binding affinities of cyclic peptides, and many other biologically rare events. It should also be noted that Replica exchange, like T-REMD, may also be conducted using order parameters other than temperatures, such as the use of various Hamiltonians (H-REMD)[102].

2.3.2 Unbiased enhanced sampling techniques

2.3.2.1 PaCS-MD

Parallel cascade selection molecular dynamics (PaCS-MD), is one of the enhance sampling methods which enhances the conformational transitions to increase the possibility of biologically rare events without external perturbations using cycles of multiple independent MD simulations conducted in parallel. The flow chart of PaCS-MD algorithm is shown in Fig. 2.5, while its procedure of the structural sampling can be briefly described as follows:

1. A short MD simulations for the selected starting conformation (typically 0.1 ns) with canonical ensemble (NVT or NPT), this refered as preliminary cycle (cyc0).

2. Ranking the structure of the output MD trajectory of cyc0 based on selection criteria such as Inter-COM distance(d) between the two bio molecules.

3. Selecting the initial structures (seeds or replicas) of the first (next) cycle ,cyc1, (n_{rep} typically ranges from 10 to 100) so that snapshots of higher rank (higher d) are selected.

4. Regenerating the initial velocities of the selected replicas using Maxwell-Boltzmann distribution, and run parallel short MD simulations for them.

5. Ranking the output MD trajectory's snapshots of this cycle based on the selection criteria.

6. Repeating steps 3 - 5 until the value of the chosen criteria reach a determined threshold to .

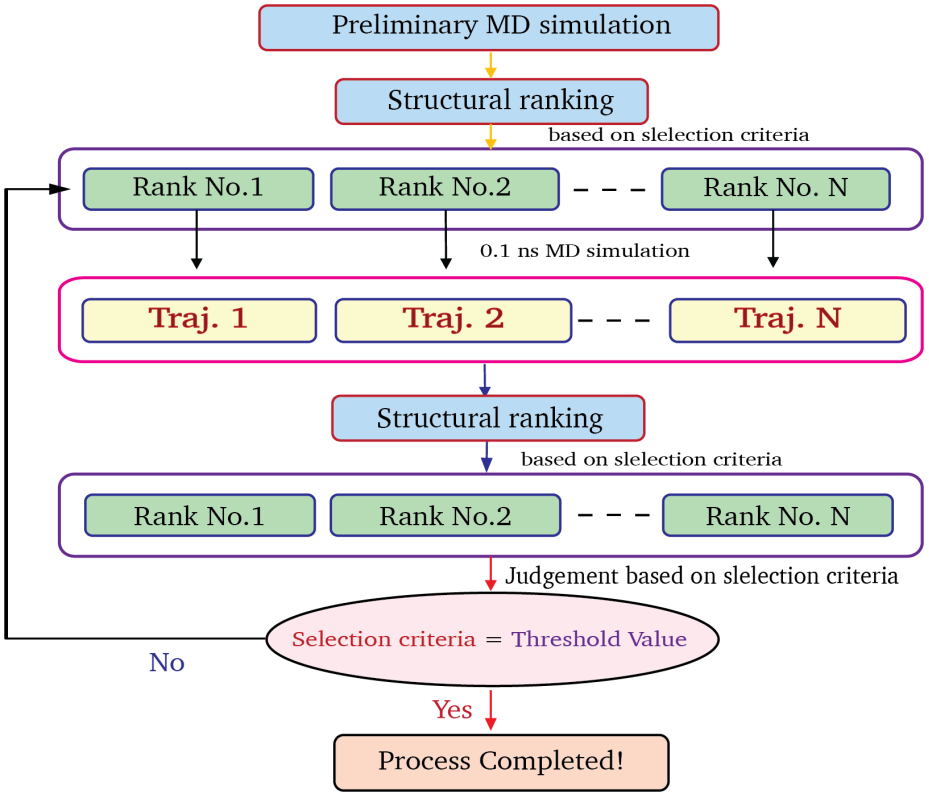

FIGURE 2.5: The flowchart of PaCS-MD algorithm

Take in consideration that, one round of independent short MD simulations followed by structure selection is regarded as one cycle of PaCS-MD while the whole process from initial cycle to the final cycle in which the threshold has been reached is regarded as one trial of PaCS-MD.

2.4 General Analysis Methods For MD Simulations

2.4.1 Root-mean-square deviation and root-mean-square fluctuation

Calculating the root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) is a typical way to estimate the equilibrium of a biomolecules simulation [107, 108]. The RMSD of the backbone atoms compared to the starting conformation can be calculated using the following equation

$$RMSD = \sqrt{\frac{1}{N}\sum_{N}(r_i(t_0) - r_i(t))^2} \tag{2.37}$$

where N is the number of atoms, $r_i(t)$ the position coordinate of atom i, at time t, and $r_i(t_0)$ the coordinate of the reference structure. RMSD gives an indication of local equilibrium and useful to monitor the structural change during the simulation. However, before RMSD analysis the simulated system need to be aligned onto the starting conformation to remove the overall translation and rotation. While RMSD provides an overall estimate for the entire system to assess the degree of movement of individual residues is better to compute the RMSF as in equation 2.38. The root mean squared distance between an atom's average position in a given set of structures might be defined as RMSF.

$$RMSF = \sqrt{\frac{1}{T}\sum_{T}(r(t) - \bar{r})^2} \tag{2.38}$$

where T is the total simulation time (or number of snapshots) and $\bar{r}$ is the average position. In addition, RMSF can be related to the B-factor used in crystallography by multiplying by $\frac{8}{3}\pi$.

2.4.2 Clustering

Typically hundreds of MD trajectories and millions of conformations can be generated. In order to understand biological processes, one of the biggest issues is how to quickly and reliably extract meaningful information from these vast output data. The use of conformational clustering is a useful strategy for dealing with this problem. Similar MD conformations are clustered together based on a geometric criterion, such as the RMSD between them. Following clustering,

which reduces the millions of MD conformations to few dozen, numerous evaluations of the examined biological macromolecules thermodynamic and kinetic characteristics may be carried out with ease. Numerous methods have been used to cluster molecular dynamics trajectories into two main categories: hierarchical and partitional clustering algorithms.[109, 110].

Hierarchical clustering is a technique for creating a hierarchy of clusters based on a certain linking criteria (similarities or dissimilarities between different clusters). For instance, bottom-up agglomerative hierarchical clustering algorithms begin with each MD conformation in its own cluster "one element-one cluster" and iteratively combines two nearest clusters using predefined linkage criteria until specific requirements are fulfilled to end this process (e.g. a desired number of clusters are reached). Agglomerative hierarchical clustering algorithms are classified into numerous subtypes depending on the linking criteria used, including the single-linkage and average-linkage methods. Single-linkage examines the shortest distance between two cluster members, while average-linkage analyzes the average of all distances between cluster members. Average-linking is predicted to be one of the most helpful linkage method available.

When you use partitional clustering algorithms, all of the MD conformations are simply split into non-overlapping clusters. These algorithms don't assume that all of the conformations and clusters are in a hierarchy. Partitional clustering methods may be further classified into several subgroups based on their algorithm architecture. The most prominent kind of clustering algorithm is center-based clustering, which includes the K-Means method[111]. K-Means clustering is a widely used clustering technique that aims to minimize the objective function specified by:

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \tag{2.39}$$

where *i* is the cluster index, k is the defined number of clusters, x_j is the data point belonging to cluster C_i, and μ_i is the centroid of clusterC_i. $\|x_j - \mu_i\|^2$ is a chosen distance measure between point x_j and μ_i. The The K-means clustering algorithm begins by designating k random points as cluster centroids, and then assigns all MD conformations to the cluster centroid that is closest to them. The centroid of each cluster is then recalculated, and data points are reallocated to the newly defined cluster centroid in order to minimize the objective function. This technique is continued until a stopping requirement is satisfied, such as when the centroids of freshly generated clusters remain constant or the maximum number of iterations is achieved. The K-means technique does not

always discover the ideal solution, but rather approximates the objective function's local minima. K-Means clustering has been frequently used to cluster MD datasets in order to generate states for the construction of MSM. Additionally, the technique is sensitive to the initial cluster centers, which are randomly chosen, and it is advisable to execute the procedure numerous times with various random seeds to lessen initial reliance. K-means++[112], a variant of the regular K-means algorithm, may be used to improve the first estimate of cluster center locations.

2.5 Markov state models

The Markov state model (MSM) is widely used in computational biology as a technique to predict stationary and kinetic information of biomolecular mechanisms from MD simulation data[55]. This can be achieved by coarse-graining the high-dimensional configuration space from MD trajectories into *n* discrete microstates $S_{i=1,2\cdots n}$ by clustering, principle component analysis (PCA), and time-lagged independent component analysis(TICA). Then A conditional transition probability matrix, termed the transition matrix$\mathbf{T} \equiv \{T_{ij}\}$, is estimated using the maximum likelihood estimation from the simulation trajectories **x**.

$$T_{ij}(\tau) = \text{Prob}\left(\mathbf{x}_{t+\tau} \in S_j \mid \mathbf{x}_t \in S_i\right) \tag{2.40}$$

where $\mathbf{x}_t$ and $\mathbf{x}_{t+\tau}$ represent the coordinates at time t and $t + \tau$. The transition matrix describes the chance of jumping from one state to another in some time interval τ which is referred to as the lag time. The MSM assumes that the transition between states is memoryless, which means the transition to the next state depend only on the current state and does not depend on where the system was in the previous states. The eigen-decomposition of transition matrix $\mathbf{T}$ yields a set of eigenvectors and corresponding eigenvalues. The eigenvalues are related to the relaxation timescales of kinetic processes, and eigenvectors indicate the associated structural change occurring at these timescales. The relaxation timescale, also known as the implied timescale, is given by:

$$t_i = -\frac{\tau}{\ln \lambda_i} \tag{2.41}$$

where t_i is the relaxation timescale corresponding to the i^{th}ith eigenvalue λ_i. The largest eigenvalue, λ_1 is always 1 for a model where networks are connected and satisfies the detailed balance

condition. Thus, the corresponding eigenvector represents the equilibrium distribution, i.e. stationary distribution. The stationary distribution of π can be easily obtained by solving the equation

$$\pi = \pi \mathbf{T} \tag{2.42}$$

The potential of mean force (PMF) of microstate *i* can be obtained as a natural logarithm of the stationary distribution as:

$$\Delta \mathrm{G}_{PMF(i)} = -k_B T \ln \pi_{(i)} \tag{2.43}$$

However, to relate the output binding energy to the experimental values, I have to calculate the standard binding free energy difference, $\Delta \mathrm{G}^o$, which I can obtain from the Potential of Mean Force following the methodology of reference [113]. The $\Delta \mathrm{G}_{PMF}$ should be corrected by adding the volume correction ,$\Delta \mathrm{G}_V$, which corresponds to the free energy of taking the system form the standard-state volume $V^o = 1661\,\text{Å}^3$ (1M concentration) to the sampled unbound volume V_u.

$$\Delta \mathrm{G}^\circ = \Delta \mathrm{G}_{\mathrm{PMF}} + \Delta \mathrm{G}_{\mathrm{V}} \tag{2.44}$$

The $\Delta \mathrm{G}_V$ of the free energy difference could be calculated by

$$\Delta G_V = -k_B T \ln \frac{V_u}{V^0} \tag{2.45}$$

where k_B is the Boltzmann constant, and T is the temperature. While the $\Delta \mathrm{G}_{PMF}$in general case between the bound and unbound states is given by

$$\Delta \mathrm{G}_{PMF} = -k_B T \ln \frac{Q_b}{Q_u} \tag{2.46}$$

Where, Q_b and Q_u are the partition functions for the bound and unbound regions, respectively. The ratio of the two patriation functions given by

$$\frac{Q_b}{Q_u} = \frac{\int_b e^{-\beta W(r)} dr}{\int_u e^{-\beta W(r)} dr} \tag{2.47}$$

Where $\beta = \frac{1}{k_B T}$, and $W_{(r)}$ is the three-dimensional PMF. In this method they considered the lowest value of W(r) is set to be zero for simplicity, so that the integration $\int_b e^{-\beta W(r)} dr = V_b$, the

bound volume. Also the difference between the lowest point of the PMF, equal zero, minus the exponential average of the PMF over the entire unbound region defined as PMF depth, ΔW. As a result $\int_u e^{-\beta W(r)} dr = V_u e^{-\beta \Delta \mathrm{W}}$. Therefore, the ratio of the partition function could be expressed as:

$$\frac{Q_b}{Q_u} = \frac{V_b}{V_u e^{-\beta \Delta W}} \tag{2.48}$$

And the $\Delta \mathrm{G}_{PMF}$ can be expressed as

$$\Delta \mathrm{G}_{PMF} = -k_B T \ln \frac{V_b}{V_u \, e^{-\beta \Delta W}} \tag{2.49}$$

Substituting the expressions in equation 2.44, where V_u cancels out, I can get this expression which match the equation 15 in reference [113] as I did not apply any restraint in the binding complex.

$$\Delta G^o = -\Delta W - k_B T \ln \frac{V_b}{V^o} \tag{2.50}$$

Finally, the $\Delta \mathrm{G}^o$, is estimated based on the equilibrium distribution by the following equation[114]:

$$\Delta G^o = -k_B T \ln \frac{P_b}{P_u} - k_B T \ln \frac{V_u}{V^o} \tag{2.51}$$

where P_b, P_u are the probabilities of the bound and unbound states respectively and the second term is the free energy of the volume correction.

Chapter 3

Dissociation pathways of p53-DBD from DNA and critical roles of key residues elucidated by dPaCS-MD/MSM

3.1 Introduction

The tumor suppressor p53 reacts to DNA damage by activating regulatory genes that inhibit cell development until the damage is repaired, as well as by activating other factors that trigger apoptosis. When p53 is mutated, cell proliferation may go unchecked, resulting in tumor formation. As a consequence, p53 is the most extensively researched human gene ever discovered. However, several unanswered concerns remain about the control of p53 activity in response to cellular stresses. The purpose of this chapter is to shed light on one of the critical mechanisms that govern p53 activity, such as the dissociation process. Using dPaCS-MD, I investigated the dissociation process of the p53-DBD/DNA complex, and native and non-native contacts released/formed during the dissociation. In section 3.2 I will show how the combination of dPaCS-MD/MSM overcomes challenges in simulating dissociation process of large complexes and enables to calculate binding free energy of the p53-DBD to DNA. Then, in section 3.3.4, I will demonstrate how critical residues such as R248 and R280, determined by their contact probability during the dissociation process, contribute to the tight binding of p53-DBD to DNA. In the next section 3.3.5, I will provide the data acquired from FEL generated by MSM that reveal the processes of the two main dissociation

pathways of the p53-DBD/DNA complex. Afterwards, the predicted value of ΔG°, which was derived using free energy analysis and was shown to be in acceptable agreement with the experimental value, will be provided. The ability of dPaCS-MD/MSM to investigate the directions of dissociation and reproduce the experimentally measured binding free energy can open the door for future study, relating effects of mutations to the dissociation pathways and the binding free energy value[115].

3.2 Materials and Methods

3.2.1 Interactions between p53-DBD and DNA

As the core domain of p53 is the most highly conserved across species and is responsible for the protein's DNA-binding specificity, it is the most important domain to study. In this work, I investigated the fundamental mechanisms of DNA consensus sequence recognition by p53-DBD and thus focused on the interactions of the p53-DBD monomer with DNA. An earlier study revealed that p53-DBD (residues 80–290) binds to the consensus sequence only as four monomers[116]. I searched the p53-DBD/DNA complex structures deposited in the Protein Data Bank (PDB) and selected one (PDB ID: 1TSR[10]) as a suitable structure for our purpose. 1TSR contains a p53-DBD trimer. One monomer is bound extensively with the consensus sequence (Chain B), a second monomer binds to a non-consensus site on the DNA, and the third does not bind to DNA but makes protein-protein contacts, stabilizing the crystal packing. This complex is thus called a monomeric p53-DNA complex. Of the other complex structures in the database (PDB IDs: 2GEQ[117], 2ADY[118], 3EXL[119], and 3KMD[120]), p53-DBD binds with DNA as a dimer in 2GEQ and a dimer of dimers in the other complexes. I chose Chain B of 1TSR, p53-DBD residues 94–289, as it forms the most native-like monomer interactions with both the major and minor grooves of the DNA[10]. This p53-DBD monomer binds to a DNA duplex containing the consensus sequence and the complementary strand. The sequences are

$$5' - TTTCCT\mathbf{AGACTTGCCC}AATTA - 3'$$

$$3' - AAGGATCTGAACGGGTTAATA - 5'$$

where bold characters indicate the decamer consensus sequence. It should be noted that the notation of the DNA residue numbers in this study starts from the 5′-(T1) to 3′-end (A21) of the upper strand and continues to the 5′-end of another strand (A22) and finally reaches to the 3′-end (A42). The aforementioned 6 hotspot mutations (Contact mutations: R248 and R273, structural mutations: R175, G245, R249, and R282) are indicated in Fig. 3.1A. The interactions on the binding interface between p53-DBD and DNA comprises three types: (1) major groove contacts with LSH; (2) minor groove contacts with L3; and (3) phosphate contacts with L3 and LSH (Fig. 3.1B). L2 does not make significant interactions with the DNA in the crystal structure, but it makes some stable interactions after equilibrium. With the DNA major groove, R280 and R283 of LSH form hydrogen bonds while K120 of LSH forms both hydrogen bonds and -cation interactions. R248 of L3 forms hydrogen bonds with the minor groove as well as a salt bridge with a phosphate group, which is consistent with that R248 is the most frequently mutated p53 residue in human cancer, and is widely assumed to be involved in DNA binding[10, 120]. The phosphate groups of the DNA backbone interact with p53-DBD residues such as K120, R273, and R283 (LSH) and S241 (L3) by forming salt bridges as shown in Fig. 3.1B

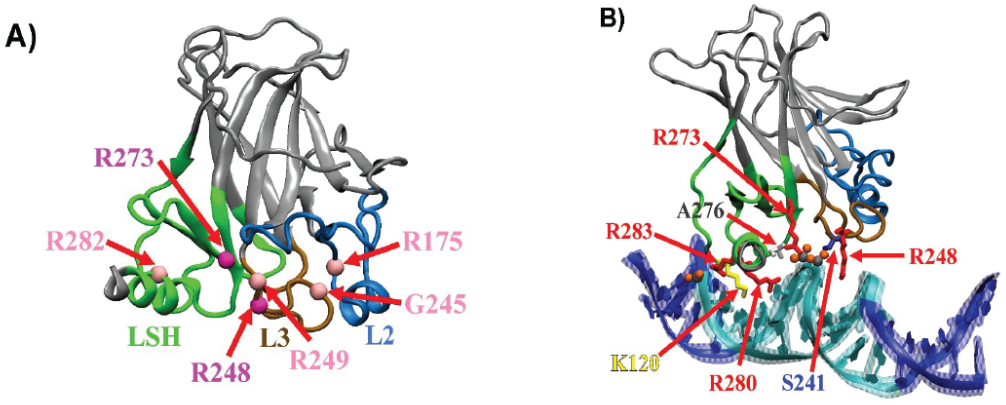

FIGURE 3.1: (A) Crystal structure of p53-DBD (Chain B of PDB ID: 1TSR[10]) and six hotspot mutation sites indicated by spheres: two contact (magenta) and four structural mutation sites (pink). The regions important for DNA binding (LSH (green) L2 (brown), L3 (light blue)) are shown. B) Chain B of p53-DBD in complex with the DNA duplex in the crystal structure (PDB ID: 1TSR). The region of the DNA duplex containing the consensus sequence and the complementary strand, and other regions, are shown in cyan and blue, respectively. The orientation of p53-DBD is different from that in A, so as to best visualize the interactions between p53-DBD and DNA. Key residues are shown in red (R residues), yellow (K), blue (S), and white (A). Grey and orange spheres represent phosphorus and oxygen atoms of the phosphate groups interacting with these amino acid residues, respectively. VMD 1.9.3 was used to create all the structural images shown in this work[121]

3.2.2 Conventional MD simulations

To generate the initial model of p53-DBD, the N- and C- termini were capped with acetyl (ACE) and N-methlyamide (NME) groups using CHARMM-GUI[122]. Next, I solvated the complex in a 124 × 109 × 94 Å^3 box filled with 4343 TIP3P water molecules[123]. The total charge of the system was then neutralized with 150 mM potassium chloride (KCl) ions to mimic physiological conditions. The axis of the DNA duplex was aligned to the Z-axis so that the $5'$-end of the DNA strand with the consensus sequence was oriented toward the -Z direction and p53-DBD was bound to the +X direction of the DNA. As the accuracy of the force field model is fundamental to successful application of computational methods, the AMBER ff14SB[65] and OL15[124] force fields were used for the protein and DNA respectively. Due to the presence of a zinc ion in this complex, I used the recently developed Zinc AMBER Force Field (ZAFF)[125] to simulate the zinc, which is covalently linked with its coordinate residues. A previous study showed that DNA-binding activity and accurate folding of p53-DBD rely on zinc coordination[126]. These preparation steps for the system were performed using the LEaP program of the Amber16 package[71].

To start the simulation, the system was equilibrated in three successive steps. (1) Energy minimization: the energy of the prepared system was minimized by the steepest descent method followed by the conjugate gradient method. (2) NVT equilibration: the temperature of the system was equilibrated at 300 K for 10 ns using a Langevin thermostat[79]. This step was conducted under isothermal and isochoric conditions (NVT ensemble). (3) NPT equilibration: the pressure was adjusted to 1 bar using the iso-thermal-isobaric (NPT) ensemble for 10 ns with a Berendsen barostat[78]. I applied positional restraints (force constant: 10 $kcal/mol\text{Å}^2$) on the heavy atoms of p53-DBD and DNA during the NVT and NPT equilibration steps. After equilibration, the box dimensions of the system were 118 × 104 × 89 Å^3. During the MD simulation, the SHAKE algorithm[127] was used to constrain covalent bonds involving hydrogen atoms, and the SETTLE algorithm[128] was employed to keep the water molecules rigid. The electrostatic interactions were calculated with a real-space cutoff of 10 Å using the particle mesh Ewald (PME) method[129].

After equilibration, NPT MD simulation of the complex was further conducted for 1 μs without any positional restraints to sufficiently sample the conformational space. The MD trajectory was used to inspect the stability of the complex and the distance between p53-DBD and the DNA during the simulation time. Next, clustering was performed using the hierarchical algorithm to

choose the most populated cluster as the starting structures for PaCS-MD. GPU implementation of the PMEMD module of the Amber16 package[71] was used to perform all the MD simulations, with the integration of equation of motion every 2 fs. The analyses were performed using the cpptraj module of the Amber Tools 16 package[71].

To observe sufficient dissociation, the simulation box of the most populated cluster of the 1 μs conventional MD simulation was extended to a box of 176 × 136 × 136 $\mathrm{\AA}^3$ and the gap was filled with ∼95,000 TIP3P[123] water molecules and 150 mM KCl. The total number of atoms in the system was ∼290,000 atoms, including water molecules and ions. Then, short energy minimization was performed, followed by equilibration at 300 K and 1 bar for 10 ns. Afterwards, the size of the simulation box was 171 × 132 × 132 $\mathrm{\AA}^3$.

3.2.3 Dissociation simulation by dPaCS-MD

In contrast to conventional conformational sampling techniques, PaCS-MD [43] is an efficient conformational sampling approach that may create conformational transition pathways utilizing cycles of parallel short MD simulations without the use of external forces. Each cycle begins by the selection of starting conformations, each of which is used for a distinct replica. For each replica, short-time MD simulation is started with regeneration of random initial velocities with Maxwell-Boltzmann distribution. These cycles are repeated until complete dissociation is achieved. Our group recently succeeded in generating the dissociation pathways of protein/ligand complexes such as tri-N-acetyl-d-glucosamine from hen egg white lysozyme[44], the transactivation domain of the p53 protein from murine double-minute clone 2 protein (MDM2)[50], and a fragment of the flagellar motor protein FliM from the chemotaxis signaling protein CheY[51]. In the current work, I extended this method to a larger biomolecular complex, p53-DBD/DNA.

The number of parallel MD simulations replicas (n_{rep}) typically ranges from 10 to 100. I selected n_{rep} = 10, as this was expected to give sufficient data to build the MSM if trajectories from multiple PaCS-MD trials were merged. The snapshots of the trajectories of the current cycle were ranked based on the Inter-COM distance between interface residues of p53-DBD and those of DNA, d, and the top 10 snapshots were selected for the next cycle. To decide the optimal choice of defining the interface residues, I conducted trials with different heavy atom-heavy atom distances and found that 5 $\mathrm{\AA}$ efficiently allowed dissociation. When the Inter-COM distance was defined as

the distance between whole molecules, dissociation was also observed but required longer cycles because this distance includes the effect of fluctuations in parts of the protein and DNA other than the interface. In this case, I also observed sliding of p53-DBD along the DNA, but this is beyond the scope of this work and I intend to investigate this observation further in the future.

As the initial structures for PaCS-MD, five snapshots were selected from the last 5 ns trajectory of the equilibration MD in the large box with 1 ns gap. In other words, I conducted the dissociation simulation from five different initial conformations to obtain more statistics to build the MSM. During dissociation by PaCS-MD, I applied positional restraints (force constant: 1 $kcal/mol\text{Å}^2$) for the three base pairs at each end of the DNA strands. This treatment mimics p53-DBD binding to a longer DNA duplex, where fluctuations of the end of the simulated DNA are expected to be less pronounced than in shorter DNA sequences. For each of the top 10 snapshots, a 0.1 ns MD simulation was performed, and the trajectory was recorded every 200 fs, and thus each trajectory contained 500 frames. Finally, I repeated the cycles until d = 70 Å to complete the dissociation as we can see in the PaCS-MD flowchart in Fig. 3.3 . For each snapshot, I conducted 15 trials of PaCS-MD dissociation. Therefore, the total number of trials collected was 75, which is equal to 5 (restart files) × 15 (the number of PaCS-MD trials for each restart file).

3.2.4 Analysis of p53-DBD/DNA interactions

Since the inter-molecular interactions at the binding interface of p53-DBD and DNA are essential for stabilizing the complex, I compared the interactions of all five different conformations in solution with those in the crystal structure. I also investigated the interactions for the intermediate conformations during the dissociation process, which might give insights into key residues that bind longer with the DNA than other residues. I used Protein Ligand Interaction Profiler (PLIP) to define these interactions and easily identify nonbonded interactions between p53-DBD and DNA[130]. Also, I checked the top 50 missense mutations, as ranked by their frequencies in diverse human cancers derived from human cancers in the IARC TP53 database R18[13, 131] to check whether losing any of these interactions might cause cancer.

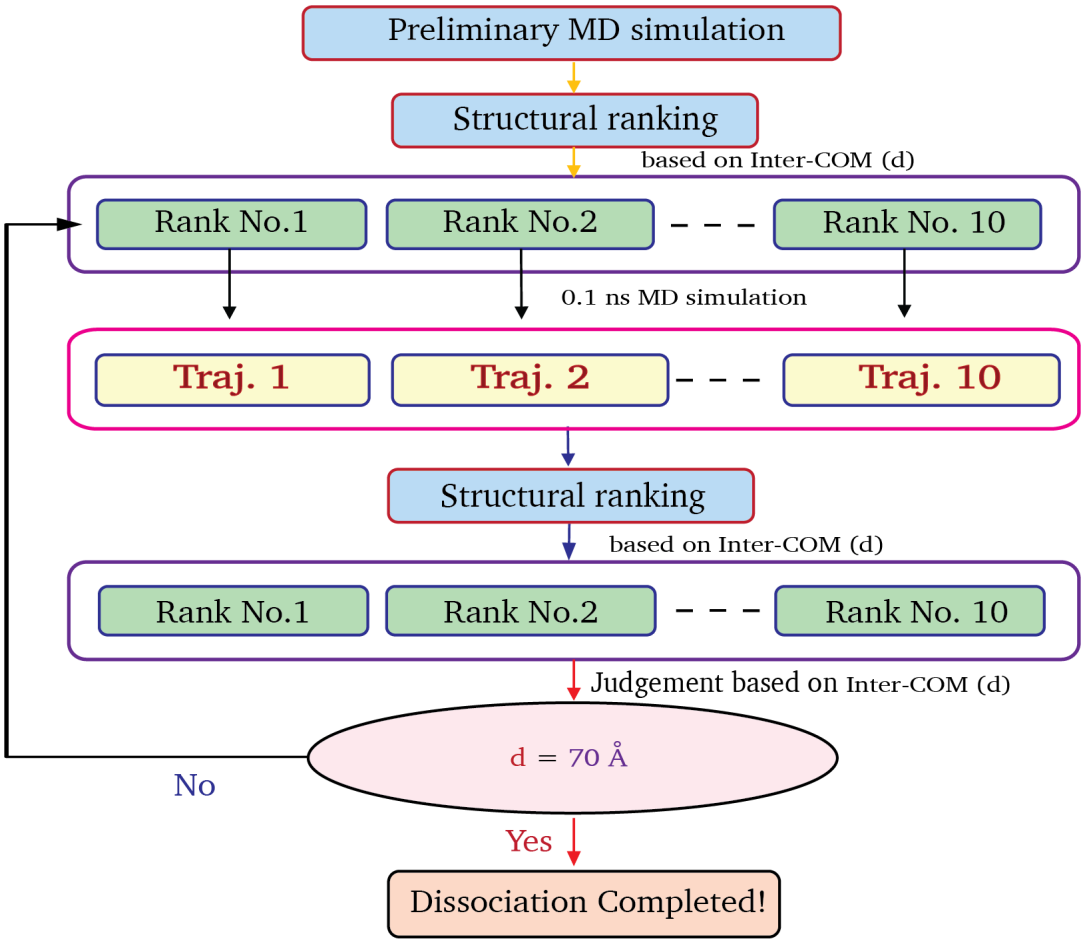

FIGURE 3.2: The flowchart of PaCS-MD based on the selected criteria (d), the number of replica (10), and the threshold value of the selection criteria to assure complete dissociation of p53-DBD/DNA complex

3.2.5 Free energy analysis by MSM

In computational biology, MSM is often used because it may reveal insights on the mechanisms of biological processes by emphasizing transitions between microstates, which can help to understand how they work.[55, 132]. In my case, by constructing MSMs using MD simulation data, I properly describe the dissociation process and gain insight into possible dissociation pathways. The dissociation pathways of p53-DBD from the DNA generated by PaCS-MD were analyzed by MSM, as shown previously by our research group[50]. Since MD simulations of a new cycle originate from snapshots of the previous cycle, the unbiased trajectories from one trial of PaCS-MD overlap in conformational space along the dissociation pathway. By generating a large number of PaCS-MD pathways, trajectories from different trials also closely overlap, enabling construction of an MSM that covers larger conformational space. Here, for the merged trajectories of 75 trials of

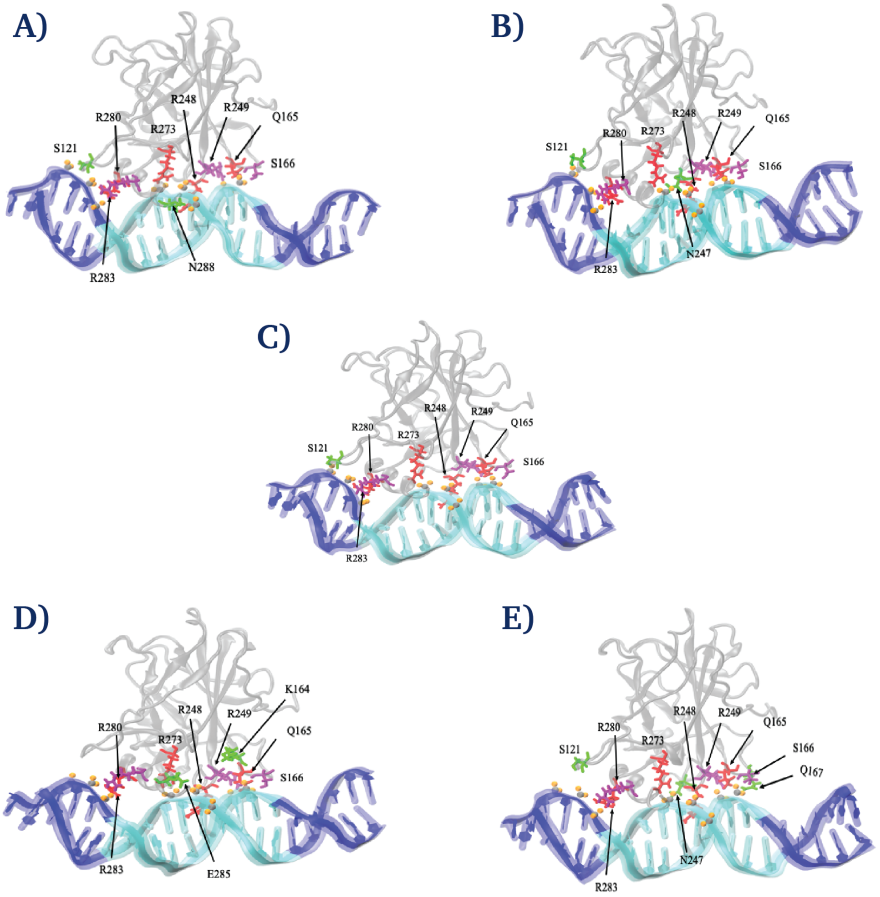

FIGURE 3.3: The five starting conformations with p53-DBD (Transparent gray) is depicted as a New Cartoon, whereas DNA (cyan indicates consensus residues, blue indicates leftover residues) is represented as a New Ribbons. The binding residues are labeled red, purple, and green. The VDW representation used for the p53-DBD binding DNA nucleotides' phosphate atoms (gray spheres) and oxygen (orange).

PaCS-MD, I built an MSM model based on three-dimensional (3D) Inter-COM vector coordinates between p53-DBD and DNA interface residues, hereafter referred to as 3D-MSM. In other words, the COM position of the DNA interface residues was employed as the origin of the 3D coordinates, and the relative COM position of the p53-DBD interface residues was used as the coordinates for the MSM. The 3D-MSM protocol is established, and experimental values of ΔG^o and kinetic rates are well reproduced[50]. I examined the root-mean-square deviation (RMSD) of p53-DBD and DNA during the dissociation simulations. The RMSD values were mostly in the range of 1–2 Å, indicating that the conformational changes of both p53-DBD and DNA were relatively small. This result also supports our assumption that the free energy change of dissociation can be mainly characterized by the position of p53-DBD relative to the DNA.

After choosing 3D COM as the metric, the next step for building the MSM as I explained in

more details in section 2.5 is clustering the MD snapshots into microstates using the appropriate clustering algorithm. Construction of the 3D-MSM used the snapshots with Inter-COM distances d $\leq$ 65 Å. Here, I employed k-means[111] with an initial guess of the cluster center position using k-means++[112]. A reasonable result was attained after testing with various numbers of cluster centers until I reached an optimal number of 800 cluster centers. The optimal lag time (or observation interval) for the MSM model was determined by building MSMs with several lag times and examining the relationship between the implied time scales (ITS) and the lag time. ITS of the slowest 100 processes at different lag times is shown in Fig.3.4. According to the results of this investigation, a lag time of 50 ps was determined to be the optimal amount for achieving Markovian behavior. Next, the transition probability matrix was estimated, followed by calculation of the stationary probabilities of the microstates, as described in section2.5. This 3D-MDM provides the stationary probability of the microstate i in 3D space and is converted to the FEL as in equation 2.42 and 2.43. For the construction of 3D-MSM, I utilized the PyEMMA package version 2.5.7[133].

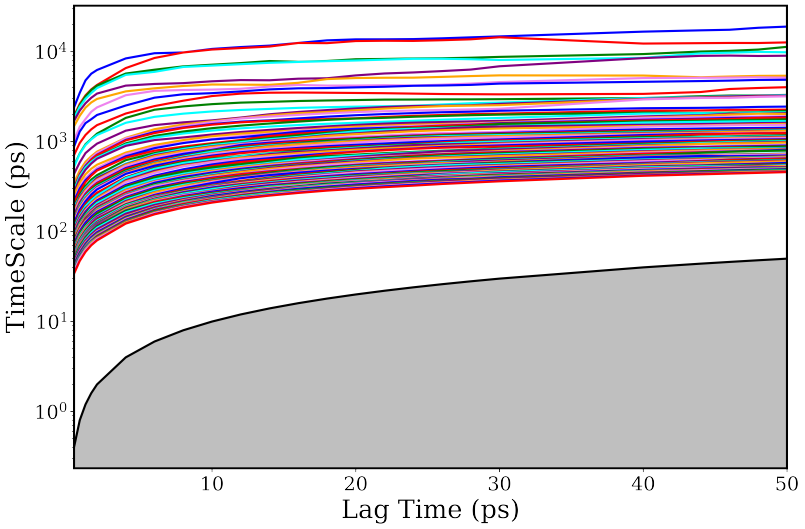

FIGURE 3.4: Implied time scales of the 50 slowest processes for different Markov state models at different lag times. The black line separates the area where the dynamics of the processes is resolvable (white) from the non-resolvable area (grey).

Free energy difference between two states is solely dependent on the initial state and the final state. As a result, binding free energy is defined as the difference in free energy between the bound and unbound states between two states. Although the bound state is a well-defined complex state, the unbound state can be any state in which interactions between p53-DBD and DNA are negligible.

Even in the cases where p53-DBD dissociated to very different positions, all dissociated states are equivalent in free energy and all can be considered as unbound states as long as interactions between the protein and DNA can be ignored. The calculation methods for standard binding free energy based on the 1D free energy profile are well established[114, 134]. The bound and unbound states are defined from the FEL. Using the probabilities of the bound (Pb) and unbound (Pu) states, the binding free energy difference from the potential of mean force ΔG_{PMF} was calculated by

$$\Delta G_{PMF} = -k_B T \ln \frac{P_b}{P_u} \tag{3.1}$$

To calculate the standard free energy difference of binding, ΔG^o, ΔG_{PMF} is corrected by adding the volume correction[114], which corresponds to the free energy of taking p53-DBD from the standard-state volume V^o = 1661 $\mathring{A}^3$ (1 M solute concentration) to the sampled unbound volume V_u. The value of ΔG^o is estimated from equation 2.51. In this equation, the second term is the free energy of the volume correction. The volume of the convex hull defined by the 3D COM coordinates of p53-DBD in the unbound state relative to the DNA was used as the value of V_u using Qhull[135]. The obtained values of V_u, P_u and P_u were used for the ΔG^o calculation.

3.3 Results and Discussion

3.3.1 Structure of the p53-DBD monomer/DNA complex in equilibrium

I first examined the stability of the p53-DBD monomer/DNA complex during a 1 μs MD. The RMSD of the heavy atoms of p53-DBD and DNA from the starting conformation during the 1 μs MD was calculated, as shown in Fig.3.5A. The RMSD of p53-DBD slightly fluctuated around 2–3 $\mathring{A}$, indicating the stability of p53-DBD during the simulation, whereas the RMSD of DNA was larger (3–5 $\mathring{A}$). This is consistent with the B-factor values of DNA being roughly 3-fold that of p53-DBD in 1TSR, indicating larger fluctuations of DNA by 1.7-fold (square root of 3) in amplitude. Larger fluctuation of DNA compared to that of p53-DBD were also reported in the previous works[58, 136]. The larger conformational change of DNA may be related to the removal of two monomers in the crystal form of 1TSR, which contains trimers packed together around the DNA, as well as to the solution environment in the simulation compared to that in the crystal. A 20^o bend in DNA is induced by binding of the p53-DBD dimer[117]. Also, partial disordering of the DNA

ends was reported during a structure refinement calculation upon solvation of the crystal structure of a p53 core domain tetramer assembled on full consensus sites without any constraints on the DNA[120]. It should be noted that a significant change in the DNA RMSD was observed around 240–310 ns and is correlated with a sudden change of the Inter-COM distance between the interface residues of p53-DBD and DNA, d, from $\sim$ 13 Å to a stable value of around 8 Å during the remainder of the simulation (Fig. 3.5B). This indicates tighter binding of p53-DBD with the DNA after the transition. Interestingly, the RMSD of p53-DBD did not particularly change in this time range, showing that p53-DBD binds with the DNA deeper after the transition, changing the DNA structure and interactions between p53-DBD and DNA without significant change in the p53-DBD conformation. The corresponding changes in intermolecular interactions are reported in section 3.3.2.

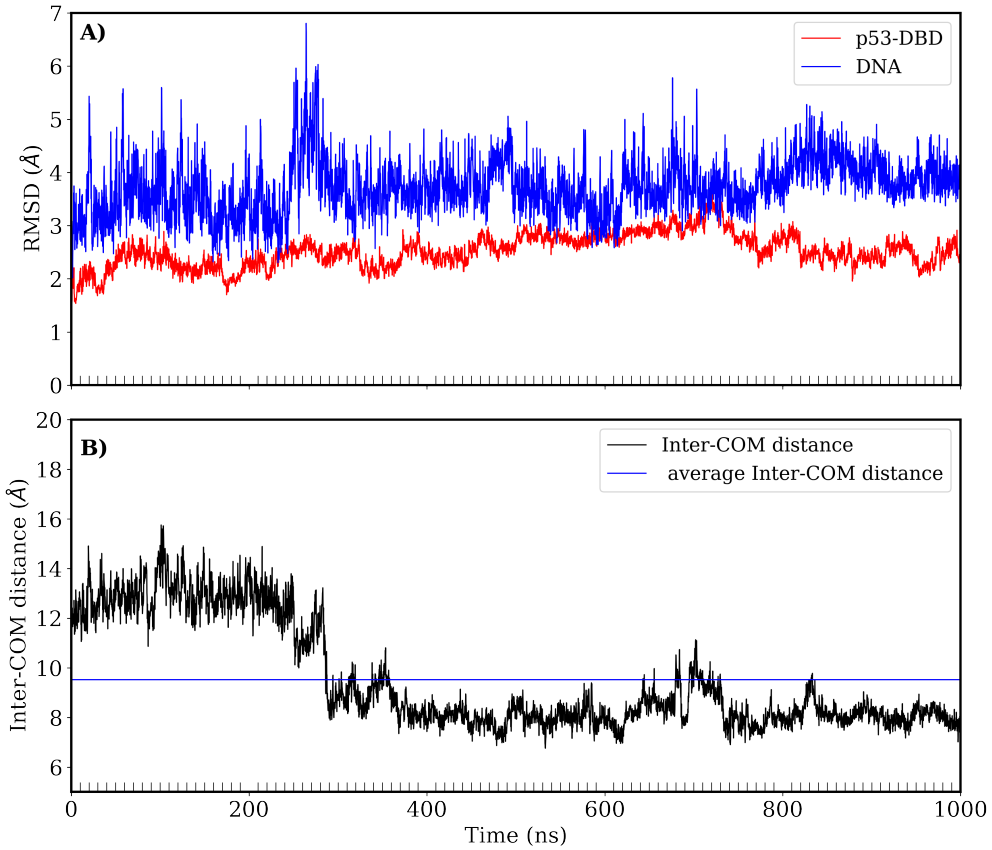

FIGURE 3.5: A) Heavy atom RMSD of p53-DBD (red) and DNA (blue) during the 1 μs MD. B) Inter-COM distance between the interface residues of p53-DBD and DNA during the 1 μs MD. The straight line (blue) in B indicates the average value of 9.5 Å.

3.3.2 p53-DBD/DNA interactions before dissociation

After the 1 μs conventional MD simulation, the structure of the most populated cluster was equilibrated in an extended box with sufficient space to accommodate the dissociation of p53-DBD from the DNA, as described in section 3.2.2. Five different conformations around the most stable structure during the MD simulation were then chosen as starting conformations for PaCS-MD, as mentioned in section 3.2.3.

Since multiple starting conformations should be used for efficient sampling[25], I chose different conformations for the initial structure of dPaCS-MD rather than relying only on different initial velocities for one conformation to sample a wider range of p53-DBD dissociation pathways from DNA. Before starting the dissociation process, I identified the key residues in the binding interface of p53-DBD/DNA by PLIP[130], which revealed differences from the crystal structure.

As shown above 3.3.1, the structure of the p53-DBD/DNA complex was slightly changed compared to the crystal structure. I investigated the interactions between p53-DBD and DNA for the five starting conformations of PaCS-MD by PLIP and compared them to those in the crystal structure, as shown in Table 3.1. These changes mainly occurred as the result of the aforementioned tighter binding of p53-DBD with the DNA at around 240–310 ns. R248, R273, R280, and R283 bind to the DNA during MD as well as in the crystal, suggesting their importance for binding. At the major groove, K120 was altered by its neighbor S121 in the MD simulation. This result is consistent with other p53-DBD structures, in which K120 and other residues of the L1 loop in LSH do not make a significant contribution to DNA binding[117, 120]. At the minor groove, in addition to R248, R249 interacts with the DNA, and this is expected to play an important role in binding because the R249S mutation (PDB ID: 2BIO) changes the structure of the L3 loop changed, and DNA binding affinity is drastically reduced[137, 138]. Q165 and S166 make contacts with the phosphates but no mutation related to these residues is listed among the 50 most common missense mutations[131]. Below, I discuss the roles of these two residues.

3.3.3 p53-DBD dissociation pathways from DNA

I successfully generated 75 different dissociation pathways of p53-DBD from DNA by conducting 15 trials of dPaCS-MDs from each of five starting conformations. Figure 3.6 shows d as a

TABLE 3.1: Comparison of p53-DBD residues that make contact with DNA in at least four out of five starting conformations obtained by MD simulation and in the crystal structure, 1TSR.

	Type of contact		
	Major groove	**Minor groove**	**Phosphate**
Both	R280, R283	R248	R248, R273, R283
Only in MD	S121	R249	Q165, S166
Only in 1TSR	K120	-	K120, S241

function of the number of PaCS-MD cycles. To assure complete dissociation, each trial was continued until d reached 70 Å, which required an average of 112 ± 22 cycles (the value after '±' shows the standard deviation) and 11.2 ± 2.2 ns of dPaCS-MD time. Since each cycle contains 10 parallel MD simulations for 0.1 ns, the accumulated computational cost is 8.4 μs (0.1 ns × 10 × 112 × 75). The variation in the number of cycles is relatively large due to the use of 10 replicas, as our group previously showed[44]. The dissociation of the p53-DBD/DNA complex can be divided into three stages: bound, partially bound, and unbound states. In the bound state (d $\leq$15 Å; below the dotted line in Fig.3.6), major key interactions between p53-DBD and DNA were maintained, whereas a part of p53-DBD dissociated from the DNA in the partially bound state (15 $\geq$ d > 35 Å, between the dotted and dashed lines in Fig.3.6). In the unbound state (d > 35 Å; above the dashed line), p53-DBD completely dissociated, and d increased linearly until it reached the threshold value.

To visualize the 75 dissociation pathways of p53-DBD from the DNA, I used the COM positions of the p53-DBD interface residues relative to those of DNA, as shown in Fig.3.7. Inspection of the dissociation pathways indicates that the sampled space formed a cone-like shape around the DNA. Around 93% of the sampling pathways (70 pathways) dissociated along the +X and -Y directions (namely, the -Y directions), while the other pathways moved along +X and +Y directions, which only occupy 7% (5 pathways), hereafter called the +Y directions. Therefore, the X-axis is considered to correspond to the main reaction coordinate for dissociation, while the other two coordinates (corresponding to the Y- and Z-axes) are considered as secondary coordinates. I did not observe clear correlation between dissociation and rotation of p53-DBD but rotation tends to be associated with dissociation processes. For example, the upper part of one of the Movies (during visual inspection) shows that p53-DBD rotated clockwise around the X-axis upon dissociation along the -Y directions while the lower part of another movie indicates a clockwise rotation around the Y-axis upon dissociation along the +Y directions. To search for a specific DNA sequence, a linear diffusion mechanism defined by either translation along DNA coupled with

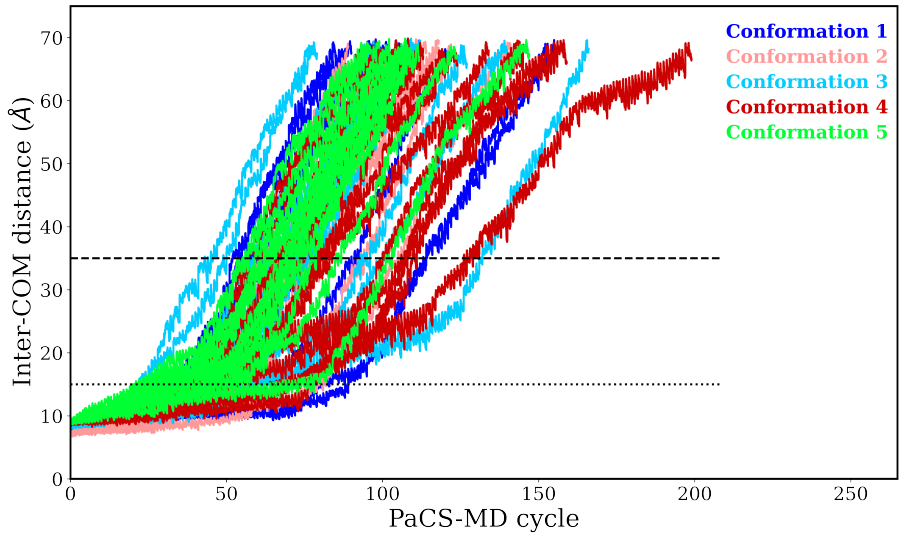

FIGURE 3.6: Inter-COM distance between p53-DBD and DNA, d, as a function of the number of PaCS-MD cycles for 75 trials. The values of d are plotted only for the replica per cycle whose change in d is the largest among the 10 replicas. Each of the five starting conformations is colored differently, with these colors matching the colors used for the dissociation pathways shown in Fig. 3.7. The dotted and dashed lines indicate the borders between the bound, partially bound, and unbound states.

rotation (rotation-coupled sliding) and sliding not tightly coupled to rotation (rotation-uncoupled sliding, sometimes referred to as jumping or hopping) were proposed based on coarse-grained MD simulations[139, 140]. Both rotation-coupled and -uncoupled sliding processes were recently observed using fluorescence microscopy[141].

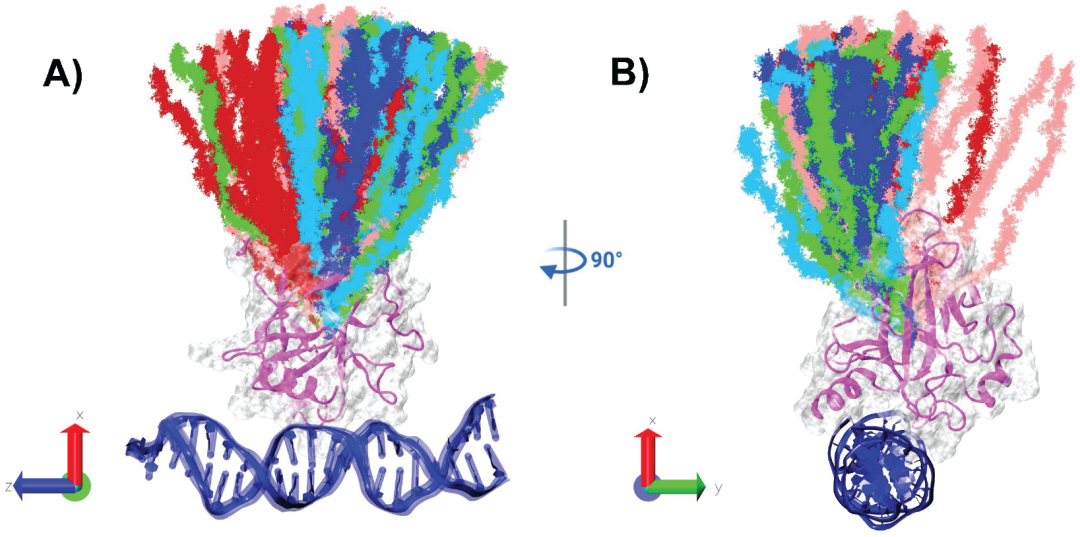

FIGURE 3.7: Dissociation pathways of 75 PaCS-MD trials of p53-DBD (pink cartoon model) from DNA (blue) represented by the COM positions of p53-DBD relative to the DNA in the trajectories. The coloring of the pathways is identical to that in Fig. 3.6. A) Front view and B) view rotated by 90^o around the X-axis (in the YZ-plane).

To investigate the main differences between the -Y and +Y directions, I inspected the trajectories of both directions in detail. The first trajectory represents a typical dissociation pathway along the -Y directions, while the second trajectory represents that along the +Y directions. I noticed that toward the -Y directions, p53-DBD dissociated from the DNA major groove and subsequently dissociated from the minor groove. In contrast, along the +Y directions, p53-DBD first dissociated from the minor groove and later dissociated from the major groove.

3.3.4 Key interactions during the dissociation process and their relation to cancer mutations

To obtain insights into the essential interactions between p53-DBD and DNA during dissociation, I calculated the contact probabilities for 41 contact pairs of p53-DBD and DNA residues as a function of the Inter-COM distance (d). These contact pairs were selected from the pairs of p53-DBD and DNA residues whose inter-atom distances were within 3.5 Å for at least one pair of atoms in at least one snapshot among the equilibrated trajectories just before the 1^{st} cycle of the PaCS-MD trials. To calculate the contact probabilities, I generated a reactive trajectory for each trial of PaCS-MD, comprising a series of molecular configurations connecting the initial bound and final unbound states along the dissociation pathway, and calculated the probabilities from all the 75 trials with a bin size of 1 Å as a function of d. The result for all 41 pairs is shown in Fig.3.9 for the range of bound and partially bound states (d $\leq$ 35 Å). Of these pairs, I selected 19 native contact pairs that include the pairs that started with high probabilities (> 80%) and which occupy 46% of all the pairs (Fig.3.8A) and two transient contact pairs whose probabilities were low in the bound state $(< 20\%)$ but increased to around $30 \sim 40\%$ before and after the border of the bound and unbound states, as shown in Fig.3.8B. I noticed that 63% of the native contact pairs (12 pairs) contain an Arginine residue of p53-DBD. Arginine is an amino acid residue essential for binding that has a positively charged guanidinium group at the end of a polar region, making it suitable for binding a phosphate anion, a main constituent of DNA strands. This also explains why five of the top six hotspot mutation sites are Arginine residues, all of which are situated in the p53-DBD/DNA interface bound to the DNA consensus sequence (Fig.3.1A).

R248 and R273 are the two most highly mutated sites in the p53 protein (6.79 and 6.55 % respectively)[13] and both are considered as contact residues, as shown in Fig.3.1. I found that R248

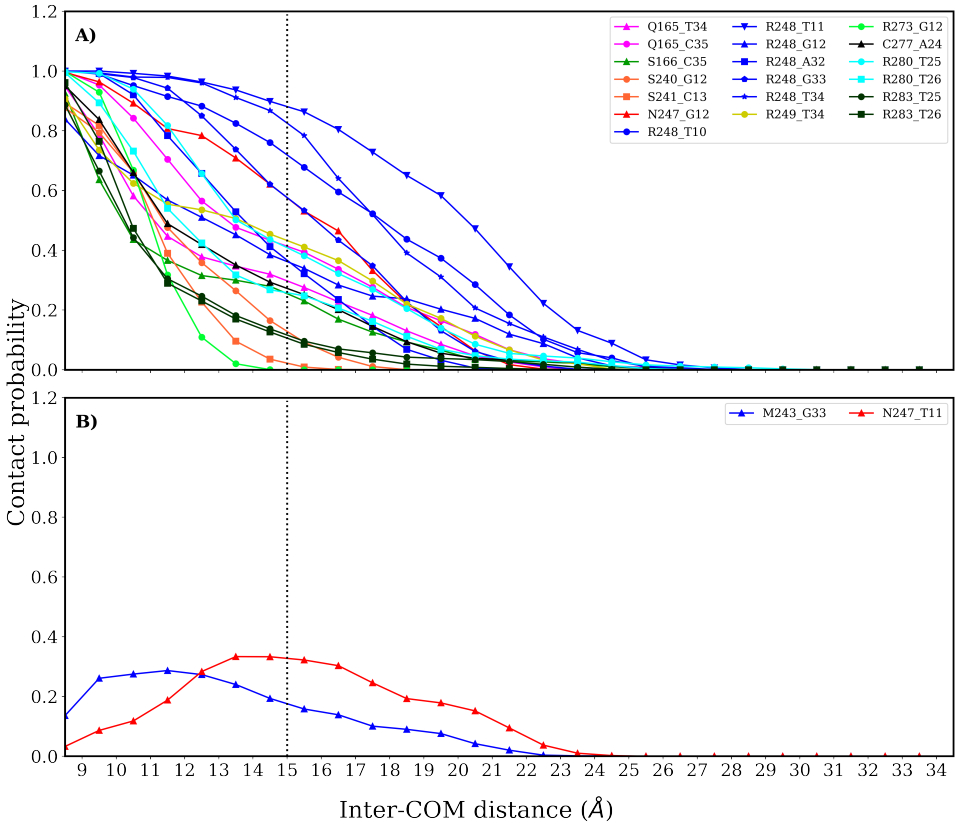

FIGURE 3.8: Contact probabilities as a function of the Inter-COM distance between p53-DBD and DNA, d. A) 19 native contact pairs whose initial probabilities were greater than 80%. B) Two transient pairs whose initial probabilities were smaller than 20% but showed a significant increase during the dissociation process.

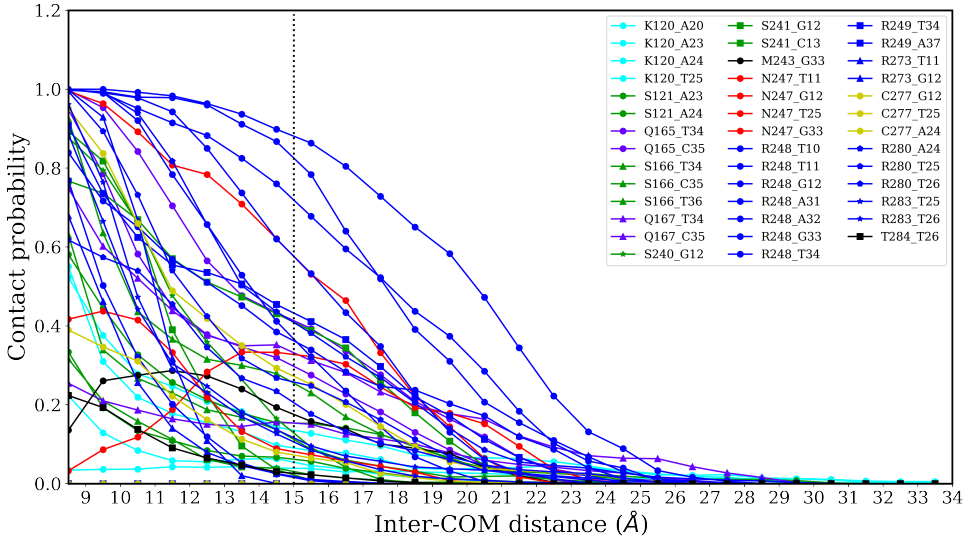

FIGURE 3.9: Contact probabilities of all 41 pairs of native contacts as a function of Inter-COM distance between p53-DBD and DNA with a bin size of 1 Å.

maintained binding with the DNA base pairs for a much longer time before complete dissociation. R248 was identified as the last residue to dissociate from DNA in 75% of the trials (56 cases out of

75) as shown in Table3.2. On the other hand, R273 completely dissociated at d = 14 Å within the bound state. This indicates the major role that R248 plays to maintain the tight binding of p53-DBD to the DNA until the final step of the dissociation process. As mentioned in section 3.2.1, R248 is situated in L3 and is bound to a deep part of the DNA minor groove, and is stabilized by a variety of non-bonded interactions with the DNA. This residue is situated in this position, contributing to the stabilization of L3[142]. A closer look revealed that R248 binds to the DNA minor groove by seven different contacts with three nucleotide residues (T10, T11, and G12) from one strand and with four (A31, A32, G33, and T34) from the other strand (Fig. 3.9). Six of these contacts are considered as the native contacts. It should be noted that T10–G12 are in the middle of the consensus sequence (A7–C16), indicating the central role of R248 in sequence recognition. Three of the contacts (R248–T10, R248–G12, and R248–T34) are maintained with probabilities around 30% until d = 20 Å, while the other three interactions (R248–T11, R248–A32, and R248–G33 tend to be lost at shorter d. These findings could explain why the R248Q mutation, which is both a contact and a structural mutation[143], causes a decrease in binding affinity to DNA. The result of MD simulation suggested that the p53 binding affinity to DNA is reduced in the R248Q mutant but that the complex is not dramatically destabilized[58, 144]. In contrast, R273 binds with the DNA major groove[120, 145]. R273 only has two nonbonded interactions with DNA (R273-T11 and R273-G12) and only one of them showed a probability greater than 80% and quickly dissociated. This could explain why mutations of R273 (R273H and R273C) do not contribute as much to binding affinity as the mutation of R248 (R248W), but these two mutations together are directly associated with impaired DNA binding, probably due to loss of the arginine guanidinium group[146].

Although the binding of p53-DBD to the minor groove mainly relies on R248, the neighboring R249 was also identified to bind with DNA in all the five starting conformations. In the 1TSR crystal structure, R249, which is situated in L3, interacts with other parts of the protein, such as L2, L3, and strands of the β sandwich, playing a role in stabilizing the p53-DBD structure[10]. In my simulations, due to its neighboring position to R248 in the minor grove, R249 also interacts with T34 of DNA, maintaining $\sim 40\%$ of the binding probability at the beginning of the partially bound state and also playing roles in maintaining binding to the DNA. R249S mutations are extremely common in liver cancer in some developing countries[17]. A change in the size of the side chain and absence of the positively charged guanidinium group in R249S were reported to be the main

reasons for the disruption of the R249 hydrogen bond network, ultimately leading to the loss of DNA binding[147].

R280 and R283 contribute to the binding of p53-DBD with the DNA major groove. R280 forms two hydrogen bonds with the bases of T25 and T26, maintaining contact probabilities around $20 \sim 40\%$ at d = 17 Å. This residue was identified as the last residue to dissociate from DNA in 20% of trials (15 cases out of 75) as shown in Table 3.2. This percentage shows that this is the second most frequent residue to be the last residue to dissociate from DNA (the most frequent residue is R248). R283 bound with T25 and T26 dissociated faster (probabilities of < 10% at the same distance). In some crystal structures, R280 also binds to bases of the major groove and the backbone of DNA and makes invariant contacts with the conserved guanine base[120, 145]. As a result, R280 is also classified as a contact mutation site, the same as mutations of R248 and R273. Although R280 is absent from the listed of the hotspot mutations, prior findings showed that R280 is important in direct DNA recognition and that mutation of this residue (R280K) impairs DNA transcription, resulting in various forms of cancer[148, 149]. I examined the list of the top 50 missense p53 mutations ranked by their frequencies in diverse human cancers[13, 131] and found that two R280 mutations are below the top 25 and the total low percentage is only around 0.8%. In contrast, R248Q and R248W are ranked in the top five, with a frequency of around 7.9 %.

TABLE 3.2: The order of dissociation for the last four dissociated residues during the dissociation process. The important binding residues that make consistence with these residues takes blue color, while the common binding residues of the 5 starting conformations includes the residues colored blue and violet

Dissociation Order	R248	R249	N247	Q165	C277	R280	R283	R273	S240	S241
Number of appearance in last 4	**75**	43	54	40	22	**38**	13	1	6	8
As last dissociated residue	**56(75%)**	0	0	0	1	**15(20%)**	1	0	0	2
As 2nd last dissociated residue	**10(13%)**	10	**20(27%)**	**12(16%)**	9	**9(12%)**	0	1	0	2
As 3rd last dissociated residue	4	15	15	14	9	8	7	0	3	2
As 4th last dissociated residue	5	18	19	14	3	6	5	0	3	2

The native contacts were gradually lost during the dissociation process, but two transient interactions were formed, as shown in Fig.3.8B. The contact probabilities of M243 and N247 were maximum at d = 11.5 Å and 13.5 Å, respectively. N247 maintained binding with G12 of the DNA for a relatively long period of time until it completely lost the contact at d = 25 Å in the middle of the partially bound state. Although the N247 interaction was only detected in one of the five starting conformations (conformation 5), it can form a transient interaction with G33 at around d = 13–16 Å with a probability > 30%. Transient interactions may suppress an abrupt energy change

during the dissociation process by forming interactions to support binding. Another transient interaction between M243 and G33 was detected with a probability < 20% before the end of the bound state. When p53-DBD tetramers are bound to DNA as dimers of dimer (PDB IDs: 2ADY[118] and 3KMD[120]), M243 and several surface residues are situated in the L3 hairpin, forming a shell of non polar interactions and stabilizing the p53 dimer interface called “core dimer"[118]. Therefore, the transient M243 interaction may be only observed in the monomer dissociation and might not happen if the dimer dissociates together.

C277 maintained binding with a 20% probability until d = 17 Å. This residue is a prime binding target for some anti-cancer compounds that attempt to reactivate mutant p53[150]. In the 3KMD crystal structure, C277 contacts with bases of the major groove and the backbone of DNA by different non-bonded interactions[120]. The contacts of C277 to the major groove vary depending on the DNA sequence[145].

3.3.5 Free energy landscape (FEL) of dissociation and two dissociation directions

In this section, I aimed to convert VMD-visualized pathways (see Fig.3.8) from simple conformations to microstates capable of connecting the complex’s bound and unbound states. Then, using statistics, to quantify the probability of transitioning from one state to another, providing more insight into the dissociation process. To satisfy all of these demands, I developed the MSM model, which is capable of integrating multiple short MD simulations trajectories of PaCS-MD into a single model of the protein conformational ensemble that incorporates critical thermodynamic properties while also preserving the system’s atomic level details. Particular attention is paid to the most favorable dissociation pathway connecting the bound state to the totally unbound state, as well as the expected value of the complex’s binding free energy.

I examined the stability of p53-DBD and DNA during the dissociation simulations by calculating the RMSD as shown in Fig.3.10 and 3.11, respectively. The RMSD values of p53-DBD were mostly in the range of 1–2 Å, while those for DNA were somewhat more than 2, demonstrating that both p53-DBD and DNA structures are stable under conditions of relatively minor conformational changes.

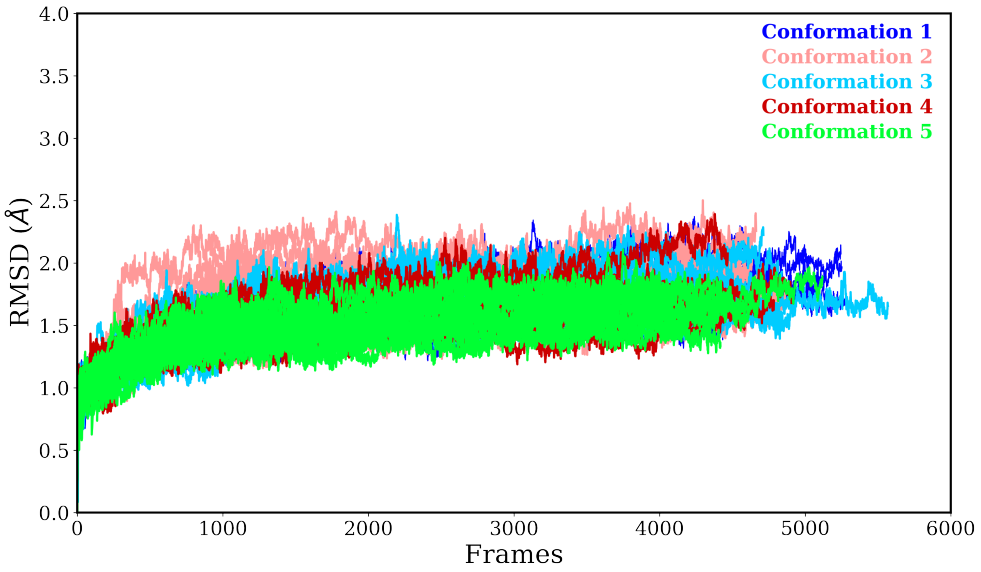

FIGURE 3.10: All atoms RMSD of p53-DBD for all the 75 dPaCS-MD trials during the dissociation process

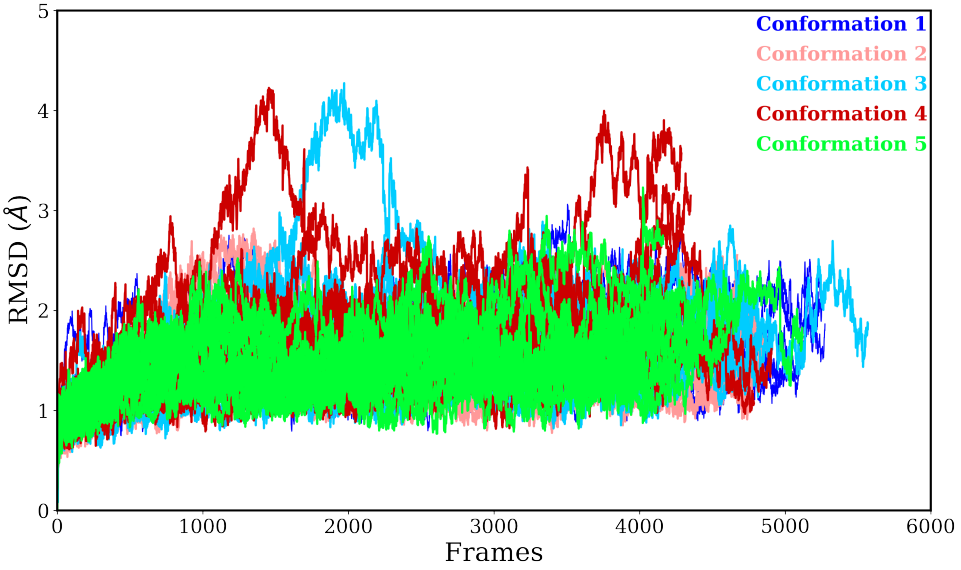

FIGURE 3.11: All atoms RMSD of DNA for all the 75 dPaCS-MD trials during the dissociation process

I obtained the FEL of p53-DBD dissociation from DNA, as described in section 3.2.5, and showed the projections onto the XY-, XZ-, and YZ-planes, as shown in Figs.3.12A, B, and C, respectively. The global minimum of the FEL agrees with the bound state. The bound state is a very deep free energy minimum (red color in FEL) and no other clear minimum is found along the dissociation pathways. As shown in Fig.3.7 and I visually inspected movies, dissociation mainly (93%) occurred along the -Y directions on the XY-plane and other dissociations (7%) occurred along the +Y directions. Therefore, I chiefly focused on the FEL of the XY-plane, shown as Fig.3.12A. Although the -Y directions are the major dissociation pathways, free energy changes along these directions are steeper compared to those along the + Y directions in Fig.3.12A. This tendency is also seen as a left-right asymmetry in the FEL of the YZ-plane (Fig.3.12C). Compared to Fig.3.12A, Fig.3.12B shows better vertical symmetry around Z = ~ 0 in the FEL of the XZ-plane, indicating equivalence of dissociations along a variety of Z directions.

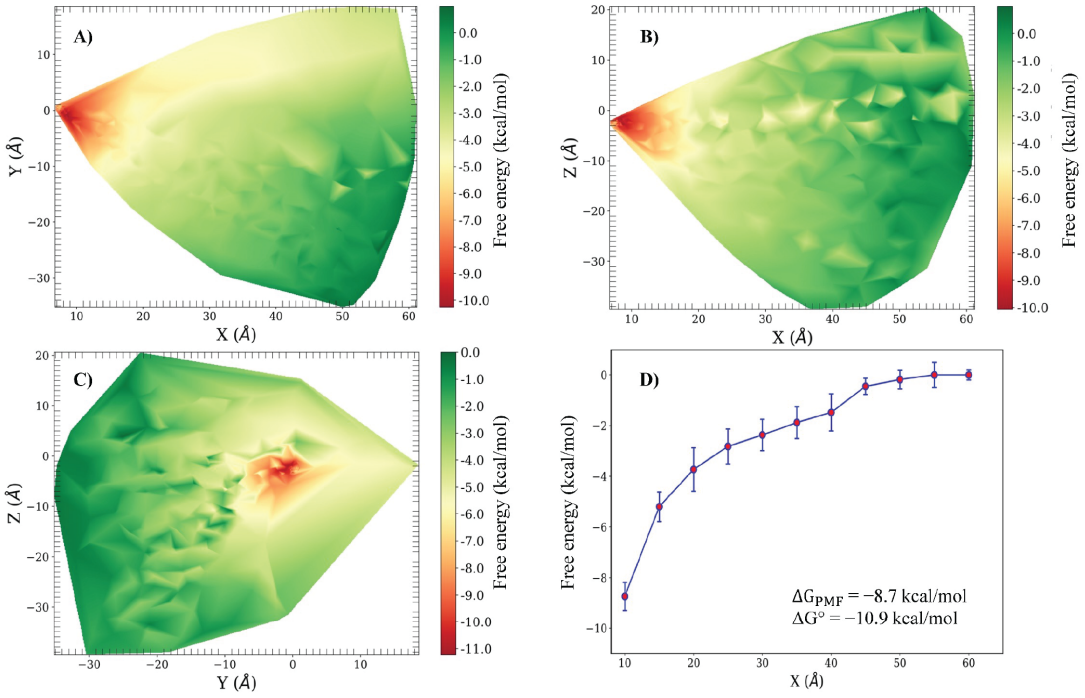

FIGURE 3.12: Free energy landscape of p53-DBD dissociation from DNA mapped onto the A) XY-, B) XZ-, and C) YZ-planes. D) The potential of mean force F as a function of X obtained by averaging microstate probabilities only in the -Y area. The obtained values of ΔG_{PMF} and ΔG^{o} are also shown.

To further investigate the reason for this asymmetry I examined the transition probabilities between microstates of all cluster centers and the free energy values of all microstates, as shown in Fig.3.13 and 3.14. Figure 3.14 shows all the microstates and transition probabilities between them along the dissociation directions. In Figs.3.15A and 3.14, the thickness of the line between microstates shows the magnitude of the probability. These figures clearly indicate that point C is a major critical point to going to the +Y directions. The points indicated by C in these panels show the microstate that can reach to the +Y directions upon transition to point D, and D indicates the microstate just after the critical transition toward the +Y directions. Figures 3.15 C and D show representative snapshots of these critical microstates. In Fig.3.15A, the thickness of the line between microstates shows the magnitude of the probability. The transition probability from C to D is significantly less than those going to the -Y directions. From C, the total probability going to the +Y directions is 0.0091 whereas that toward the -Y directions is 0.0549, which means that ratio between the +Y and -Y directions is 1:6.5. If the other minor pathways are considered, this ratio is 1:4.5. Of 75 trials, 19 reached point C, indicating that point C was visited many times but the transition to the +Y directions only occur with significantly lower probability. This is consistent with a larger free energy gap from C to D (1.1 kcal/mol) toward the +Y directions compared to those

toward the -Y directions ($\leq$ 0.5 kcal/mol). These results indicate that, although the +Y directions show overall shallower free energy changes after the transition to the partially bound state, most dissociations occur along the -Y directions at the branch point where the dissociation along the -Y directions is energetically preferable. I also examined the free energy difference between the +Y and -Y directions by comparing PMFs and confirmed that the difference is significant. This is consistent with our group previous results, showing that the PMF convergence can be reached by averaging 3–5 trials[51, 52]. As is clear from comparison of Fig.3.15 C and D, the minor groove contacts were lost upon the transition from C to D, indicating greater importance of p53-DBD binding to the minor groove. This also indicates larger free energy contributions of minor groove binding compared to the contributions of major groove. As mentioned in section 3.3.4, R248 plays major roles in minor groove binding by interacting with seven nucleotide residues, including the middle of the consensus sequence, and by bridging the two DNA strands, and R248 was the last residue to detach from the DNA in 75% of the PaCS-MD trials. Also, R248Q and R248W are ranked in the top five of the top 50 missense p53 mutations, with a frequency of around 7.9%. I conclude that R248 is essential in recognizing the consensus sequence and in stabilizing p53-DBD binding with DNA.

To examine the validity of the calculated FEL, I estimated the standard binding free energy ΔG^o and compared it to the experimental value. As already shown previously in Eq. 2.51, ΔG^o is obtained from the probabilities of the bound and unbound states with volume correction. The unbound state can be any state in which the FEL reaches a plateau, as mentioned in section 3.2.5. As shown in Fig.3.12A, the FEL converged to flat values along the -Y directions. To illustrate this convergence, the PMF as a function of X obtained by averaging microstate probabilities only in the -Y area is shown in Fig.3.12D. The distribution of these microstates in the COM space is shown in Fig.3.16. During the dissociation process, the PMF increased at a higher rate in the bound state (d < 15 Å), and continued increasing with a lower rate in the partially bound state (15 $\leq$ d < 35 Å) and became flat in the unbound state, especially at around 45 - 60 Å. Regardless of the direction of dissociation, the PMF converges to a certain value when two molecules are separated sufficiently and the interactions between these molecules are negligible, as shown previously[44, 50]. Thus, the unbound state was defined as the flat region (d $\geq$ 45 Å). After calculating the unbound volume of the trajectories, I obtained ΔG^o = -10.9 ± 0.4 kcal/mol (ΔG_{PMF} = -8.7 ± 0.4 and the correction value = -2.2 kcal/mol). This value is very close to the binding free energy of p53-DBD (residues

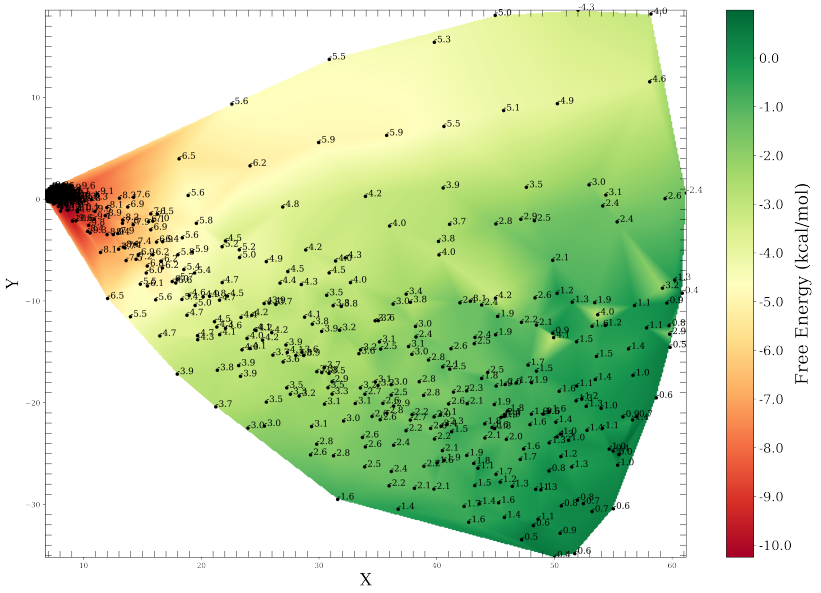

FIGURE 3.13: The distribution of all cluster centers and the predicted PMF of each cluster center

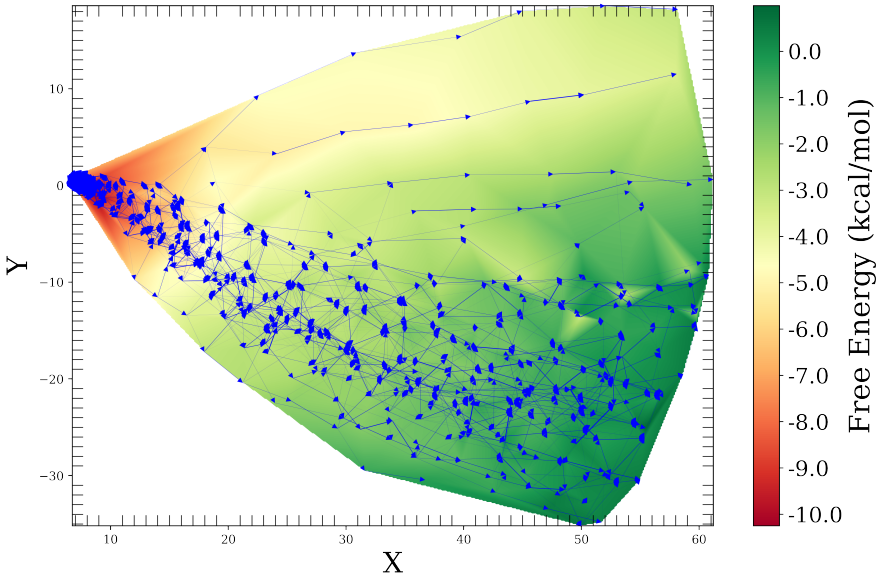

FIGURE 3.14: Transition probabilities between all microstates (800 microstates) along the dissociation directions visualized by the thickness of the lines on the FEL

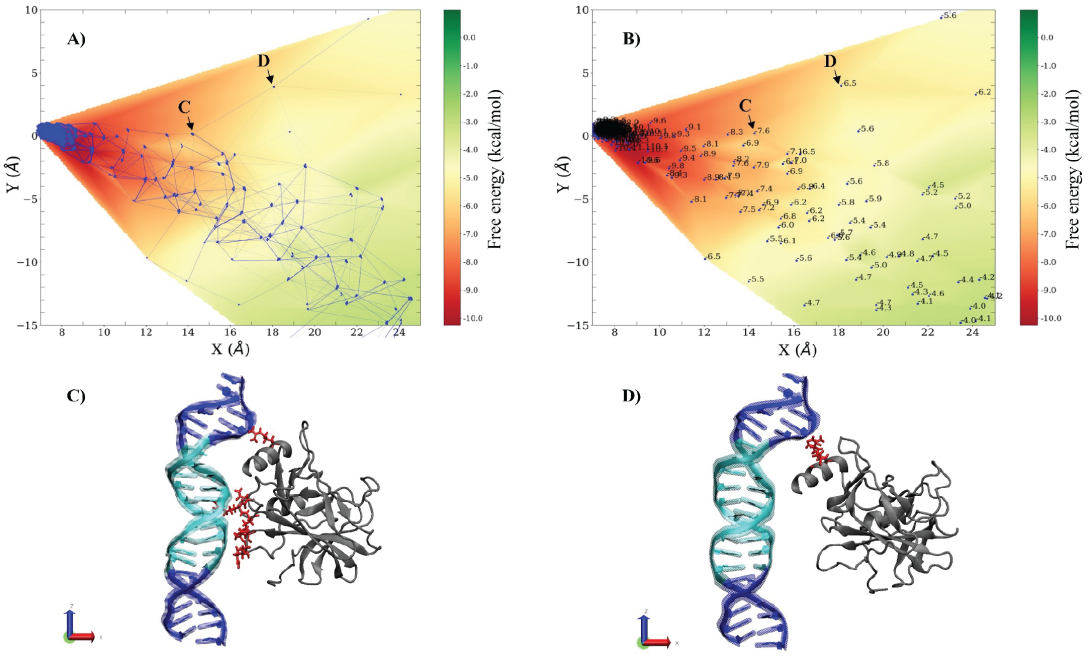

FIGURE 3.15: A) Transition probabilities between microstates along the dissociation directions visualized by the thickness of the lines in a close-up view of the FEL up to X $\leq$ 25 Å. When the probabilities are lower than 0.001, no lines are shown. B) Free energy values are shown for the corresponding microstates. C) A representative snapshot of the microstate before the critical transition toward dissociation to the +Y directions and D) a snapshot just after the critical transition. The positions of C and D are indicated in panels A and B. p53-DBD residues contacting DNA are shown in red.

94–312) with the consensus DNA sequence of –11.1 kcal/mol measured by isothermal titration calorimetry (ITC)[151], suggesting that the calculated FEL is reasonable. Since this p53 fragment did not include the tetramerization domain (325–356), this experimental value was considered as for the monomer[151]. It was also reported that the 80–320 fragments of p53 mostly exist as monomers at low protein concentrations and bind consensus DNA as four monomers and only as four monomer as already mentioned in Section 3.2.1[116].

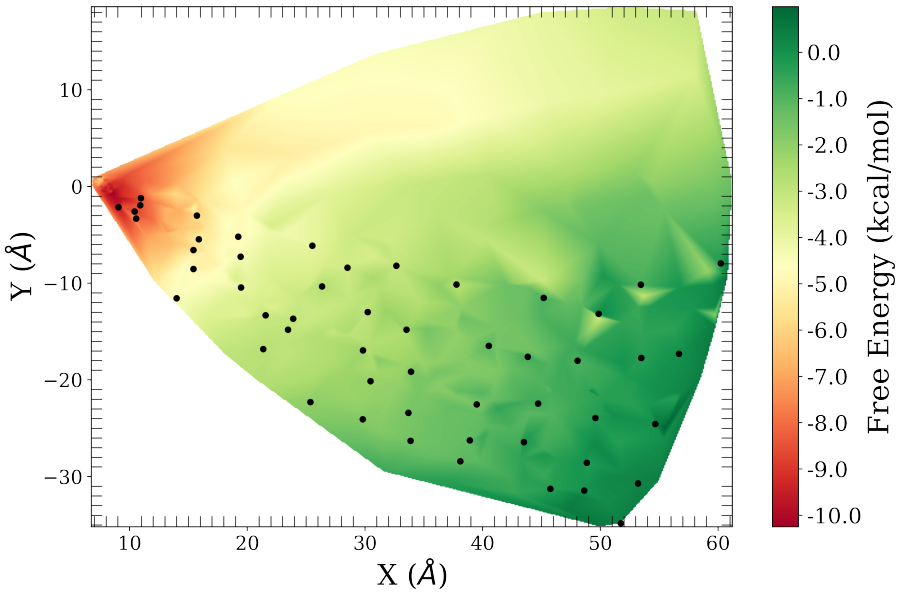

FIGURE 3.16: The microstates used for the calculation of PMF shown in Fig. 3.12D are shown by black dots on the FEL of the XY-plane.

Chapter 4

Conclusions and Perspectives

In this chapter, I present the conclusions and the scientific contribution of my work in section 4.1 . After that, I introduce identified points for the future work of this research in section 4.2.

4.1 Conclusions

In this work, I observed the dissociation processes of p53-DBD from DNA that contains the consensus sequence in the middle of the sequence by dissociation PaCS-MD (dPaCS-MD) simulations with an all-atom model including explicit solvent. Seventy-five trials of dPaCS-MD were conducted with an average simulation time of 11.2 ± 2.2 ns. During the dissociation process, 93% of the trials dissociated along the +X and -Y directions (-Y directions), while 7% moved along the +X and +Y directions (+Y directions). Along the -Y directions, p53-DBD dissociated from the major groove first and then detached from the minor groove, while unbinding from the minor groove occurred first along the +Y directions, followed by dissociation from the major groove. Since the loss of minor groove interaction with p53-DBD has a relatively high free energy cost (1.1 kcal/mol) upon the critical transition toward the +Y direction, major groove detachment occurs more frequently with a lower free energy cost (< 0.5 kcal/mol) as the initial step of dissociation. The standard binding free energy calculated from the free energy landscape was ΔG^o = - 10.9 ± 0.4 kcal/mol, which agrees with the value obtained by ITC, –11.1 kcal/mol. These results indicate that the dPaCS-MD/MSM approach can be a powerful tool to investigate the dissociation mechanisms of two large molecules, such as a protein and a DNA molecule.

The minor groove binding is stabilized mainly by R248 and R249 and sometimes also by N247. Among them, R248 is the most important residue that tightly packs deep inside the minor groove. This residue contacts with 7 nucleotide residues including the middle of the consensus sequence, bridging the two DNA strands that form the minor groove. Also, R248 was the last residue detached from the DNA in 75 % of the dPaCS-MD trials. These results suggest that R248 mutations can significantly affect the binding of p53 to DNA and interfere p53 functions. R248 is one of the most frequently mutated residues in the top 50 list of p53 missense mutations. Two major mutations, R248Q and R248W, loose the charged side chain that interact with 7 nucleotide residues. Previously, atomistic MD simulations of p53-DBD/DNA also showed that a few crucial residues, specifically R248, S241, and N239, contribute to binding energy by using the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA)[58]. The binding of p53-DBD to the major groove is stabilized by R280 and R283. R280 was the last residue dissociated from the DNA in 20 % of trials while R283 tended to dissociate faster. R280 mutations are found in the top 50 list but situated below top 25 and the total low percentage is only around 0.8 %. These results show that the p53 key residues for the DNA binding are known as the cancer-related mutations, which implies that impairments of interactions between p53-DBD and DNA can be frequently related to cancer.

Upon binding to consensus DNA, p53-DBD can form a tetramer, stabilized by protein-DNA and protein-protein interactions. The crystal structures of p53-DBD-tetramer (2ADY and 3KMD)[118, 120] indicate that the tetramer is stabilized by two distinct types of protein-protein interactions. One occurs within each core dimer (usually referred to as dimer interface), while the other occurs at the dimer-dimer interface. One of the core dimers is stabilized by networks of water-mediated protein-DNA interactions formed by the SS241, N239, and R248 side chains[118, 120]. As limitations of my work, I was unable to analyze the important residues in dimers and to determine if the dimerization has significant effects on the binding free energy because I only focused on the monomer as a fundamental step of the binding. Therefore, future studies should focus on the dimer or tetramer binding of p53 to obtain significant insight into how protein-protein interactions affect the process of dissociation and the binding free energy during dissociation.

In summary, the promising combination of dPaCS-MD/MSM can be used not only to investigate different pathways during dissociations of two large molecules but also to identify key residues for major dissociation pathways and to quantitatively calculate the binding free energy of the complex, which should also be useful in elucidating the effects of mutations. The presence of allosteric

roles and inactivating effects of the p53-DBD mutations located distant from the DNA binding surface that were recently revealed[59, 60, 152] may also be investigated by dPaCS-MD/MSM to quantitatively analyze mutational effects on binding free energy and binding mechanisms in the future. I conclude that this combination sheds light on underlying mechanisms, which are highly necessitated for developing small molecules as anti-tumor drugs that can reactivate functions of p53 mutants.

4.2 Future Works

Although the dPaCS-MD/MSM combination overcomes the difficulties associated with simulating the dissociation of large complex and facilitated the calculation of the p53-DBD binding free energy to DNA, there are still some improvements, which may be achieved to apply this technique to investigate more biological processes. The research conducted in this study leaves a number of unanswered issues and several research topics that may be addressed in the future, including the following:

1. **Mutations Effects**:

   As p53 is a potential target for cancer treatment, drugs designed to attach to mutants and restore their stability or wild-type conformation are being developed. Using the dPaCS-MD/MSM to study the Structural conformations of p53 mutants and the effect of mutations of key residues during dissociation on the binding free energy and the dissociation pathways is very important. In particular, I would study the effect of mutation of the key residues. I showed in our study that these residues, R248 and R249, stabilize the minor groove binding. I would expect that R248 affects not only in the binding free energy but in the dissociation pathway as well. The primary explanation is that R248 stabilizes the minor groove by attaching its long side chain to its deep part through a number of non-bonded interactions with the DNA. Additionally, the influence of mutation R280, which indicated a crucial function in maintaining binding to the major groove bases and is likely to affect the dissociation pathway, despite the fact that it is not among the top 25 mutations.

2. **association/dissociation simulation**:

   As I explored the dissociation pathways of the p53-DBD/DNA complex using a dPaCS-MD/MSM combination, I can now expand this work to analyze the association/dissociation process utilizing a/dPaCS-MD/MSM. By using this approach, I might obtain important insight into the association process, examine favorable association pathways, and answer the question if the p53-DBD binds to the DNA's minor or major groove first.

3. **p53 anti tumor drug discovery**:

   Based on the predication of binding free energy, this combination of PaCS-MD/MSM can improve the most promising therapeutic strategies in cancer treatment; the reactivation of the mutant p53,

   (a) Evaluate the efficacy of currently suggested reactivators that restore p53's wild-type activity.

   (b) Parallel to docking approaches, PaCS-MD/MSM can suggest additional lead compounds that may lead to the development of a novel anti-tumor medicine.

Bibliography

[1] *Protein Physics*. Elsevier, 2016. doi: 10.1016/c2015-0-04816-x. URL https://doi.org/10.1016%2Fc2015-0-04816-x.

[2] Daniel I.H. Linzer and Arnold J. Levine. Characterization of a 54k dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell*, 17(1):43–52, may 1979. doi: 10.1016/0092-8674(79)90293-9. URL https://doi.org/10.1016%2F0092-8674%2879%2990293-9.

[3] Moshe Oren. p53: not just a tumor suppressor. *Journal of Molecular Cell Biology*, 11(7): 539–543, jul 2019. doi: 10.1093/jmcb/mjz070. URL https://doi.org/10.1093%2Fjmcb%2Fmjz070.

[4] Kathryn T. Bieging, Stephano Spano Mello, and Laura D. Attardi. Unravelling mechanisms of p53-mediated tumour suppression. *Nature Reviews Cancer*, 14(5):359–370, apr 2014. doi: 10.1038/nrc3711. URL https://doi.org/10.1038%2Fnrc3711.

[5] Margot E Bowen and Laura D Attardi. The role of p53 in developmental syndromes. *Journal of Molecular Cell Biology*, 11(3):200–211, jan 2019. doi: 10.1093/jmcb/mjy087. URL https://doi.org/10.1093%2Fjmcb%2Fmjy087.

[6] I Goldstein, V Marcel, M Olivier, M Oren, V Rotter, and P Hainaut. Understanding wild-type and mutant p53 activities in human cancer: new landmarks on the way to targeted therapies. *Cancer Gene Therapy*, 18(1):2–11, oct 2010. doi: 10.1038/cgt.2010.63. URL https://doi.org/10.1038%2Fcgt.2010.63.

[7] Pratik Vyas, Itai Beno, Zhiqun Xi, Yan Stein, Dmitrij Golovenko, Naama Kessler, Varda Rotter, Zippora Shakked, and Tali E. Haran. Diverse p53/DNA binding modes expand the repertoire of p53 response elements. *Proceedings of the National Academy of Sciences*, 114

(40):10624–10629, sep 2017. doi: 10.1073/pnas.1618005114. URL https://doi.org/10.1073%2Fpnas.1618005114.

[8] Vladimir J. N. Bykov, Sofi E. Eriksson, Julie Bianchi, and Klas G. Wiman. Targeting mutant p53 for efficient cancer therapy. *Nature Reviews Cancer*, 18(2):89–102, dec 2017. doi: 10.1038/nrc.2017.109. URL https://doi.org/10.1038%2Fnrc.2017.109.

[9] Wei-Sheng Wu, Jer-Wei Chang, Hung-Jiun Liaw, Yu-Han Chu, and Yu-Xuan Jiang. p53 binding loci database (p53bld): a repository for the genome-wide binding loci of human TP53. *Clinical Microbiology and Research*, pages 1–10, jul 2019. doi: 10.31487/j.cmr.2018.01.01. URL https://doi.org/10.31487%2Fj.cmr.2018.01.01.

[10] Yunje Cho, Svetlana Gorina, Philip D. Jeffrey, and Nikola P. Pavletich. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science*, 265(5170):346–355, jul 1994. doi: 10.1126/science.8023157. URL https://doi.org/10.1126%2Fscience.8023157.

[11] D. P. Lane. p53, guardian of the genome. *Nature*, 358(6381):15–16, jul 1992. doi: 10.1038/358015a0. URL https://doi.org/10.1038%2F358015a0.

[12] Dirk Dittmer, Sibani Pati, Gerard Zambetti, Shelley Chu, Angelika K. Teresky, Mary Moore, Cathy Finlay, and Arnold J. Levine. Gain of function mutations in p53. *Nature Genetics*, 4 (1):42–46, may 1993. doi: 10.1038/ng0593-42. URL https://doi.org/10.1038%2Fng0593-42.

[13] Liacine Bouaoun, Dmitriy Sonkin, Maude Ardin, Monica Hollstein, Graham Byrnes, Jiri Zavadil, and Magali Olivier. TP53variations in human cancers: New lessons from the IARC TP53 database and genomics data. *Human Mutation*, 37(9):865–876, jul 2016. doi: 10.1002/humu.23035. URL https://doi.org/10.1002%2Fhumu.23035.

[14] Andrew C.R. Martin, Angelo M. Facchiano, Alison L. Cuff, Tina Hernandez-Boussard, Magali Olivier, Pierre Hainaut, and Janet M. Thornton. Integrating mutation data and structural analysis of the TP53 tumor-suppressor protein. *Human Mutation*, 19(2):149–164, jan 2002. doi: 10.1002/humu.10032. URL https://doi.org/10.1002%2Fhumu.10032.

[15] A Petitjean, M I W Achatz, A L Borresen-Dale, P Hainaut, and M Olivier. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*, 26(15):2157–2165, apr 2007. doi: 10.1038/sj.onc.1210302. URL https://doi.org/10.1038%2Fsj.onc.1210302.

[16] Qiang Lu, Yu-Hong Tan, and Ray Luo. Molecular dynamics simulations of p53 DNA-binding domain. *The Journal of Physical Chemistry B*, 111(39):11538–11545, sep 2007. doi: 10.1021/jp0742261. URL https://doi.org/10.1021%2Fjp0742261.

[17] Ran Brosh and Varda Rotter. When mutants gain new powers: news from the mutant p53 field. *Nature Reviews Cancer*, 9(10):701–713, aug 2009. doi: 10.1038/nrc2693. URL https://doi.org/10.1038%2Fnrc2693.

[18] Gili Ben-Nissan and Michal Sharon. Capturing protein structural kinetics by mass spectrometry. *Chemical Society Reviews*, 40(7):3627, 2011. doi: 10.1039/c1cs15052a. URL https://doi.org/10.1039%2Fc1cs15052a.

[19] Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry*, 59(9): 4035–4061, feb 2016. doi: 10.1021/acs.jmedchem.5b01684. URL https://doi.org/10.1021%2Facs.jmedchem.5b01684.

[20] Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9(9):646–652, 2002.

[21] Maria Arnittali, Anastassia N Rissanou, and Vagelis Harmandaris. Structure of biomolecules through molecular dynamics simulations. *Procedia Computer Science*, 156:69–78, 2019.

[22] George Hedger, David Shorthouse, Heidi Koldsø, and Mark SP Sansom. Free energy landscape of lipid interactions with regulatory binding sites on the transmembrane domain of the egf receptor. *The Journal of Physical Chemistry B*, 120(33):8154–8163, 2016.

[23] Ron O Dror, Daniel H Arlow, David W Borhani, Morten Ø Jensen, Stefano Piana, and David E Shaw. Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. *Proceedings of the National Academy of Sciences*, 106(12):4689–4694, 2009.

[24] Ron O Dror, Albert C Pan, Daniel H Arlow, David W Borhani, Paul Maragakis, Yibing Shan, Huafeng Xu, and David E Shaw. Pathway and mechanism of drug binding to g-protein-coupled receptors. *Proceedings of the National Academy of Sciences*, 108(32):13118–13123, 2011.

[25] Efrem Braun, Justin Gilmer, Heather B. Mayes, David L. Mobley, Jacob I. Monroe, Samarjeet Prasad, and Daniel M. Zuckerman. Best practices for foundations in molecular simulations [article v1.0]. *Living Journal of Computational Molecular Science*, 1(1), 2019. doi: 10.33011/livecoms.1.1.5957. URL https://doi.org/10.33011%2Flivecoms.1.1.5957.

[26] Yibing Shan, Eric T Kim, Michael P Eastwood, Ron O Dror, Markus A Seeliger, and David E Shaw. How does a drug molecule find its target binding site? *Journal of the American Chemical Society*, 133(24):9181–9183, 2011.

[27] Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

[28] Peter L. Freddolino, Feng Liu, Martin Gruebele, and Klaus Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophysical Journal*, 94(10): L75–L77, may 2008. doi: 10.1529/biophysj.108.131565. URL https://doi.org/10.1529%2Fbiophysj.108.131565.

[29] Stewart A. Adcock and J. Andrew McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *ChemInform*, 37(30), jul 2006. doi: 10.1002/chin.200630297. URL https://doi.org/10.1002%2Fchin.200630297.

[30] Kira A Armacost, Sereina Riniker, and Zoe Cournia. Novel directions in free energy methods and applications, 2020.

[31] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, oct 2011. doi: 10.1126/science.1208351. URL https://doi.org/10.1126%2Fscience.1208351.

[32] Anna S. Kamenik, Stephanie M. Linker, and Sereina Riniker. Enhanced sampling without borders: on global biasing functions and how to reweight them. *Physical Chemistry Chemical Physics*, 2022. doi: 10.1039/d1cp04809k. URL https://doi.org/10.1039%2Fd1cp04809k.

[33] William L. Jorgensen and C. Ravimohan. Monte carlo simulation of differences in free energies of hydration. *The Journal of Chemical Physics*, 83(6):3050–3054, sep 1985. doi: 10.1063/1.449208. URL https://doi.org/10.1063%2F1.449208.

[34] Glenn M Torrie and John P Valleau. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chemical Physics Letters*, 28 (4):578–581, 1974.

[35] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, feb 1977. doi: 10.1016/0021-9991(77)90121-8. URL https://doi.org/10.1016%2F0021-9991%2877%2990121-8.

[36] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *The Journal of Chemical Physics*, 113(15):6042–6051, oct 2000. doi: 10.1063/1.1308516. URL https://doi.org/10.1063%2F1.1308516.

[37] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[38] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *WIREs Computational Molecular Science*, 1(5):826–843, feb 2011. doi: 10.1002/wcms.31. URL https://doi.org/10.1002%2Fwcms.31.

[39] Barry Isralewitz, Mu Gao, and Klaus Schulten. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, 11(2):224–230, apr 2001. doi: 10.1016/s0959-440x(00)00194-9. URL https://doi.org/10.1016%2Fs0959-440x%2800%2900194-9.

[40] Donald Hamelberg, John Mongan, and J. Andrew McCammon. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal*

of Chemical Physics, 120(24):11919–11929, jun 2004. doi: 10.1063/1.1755656. URL https://doi.org/10.1063%2F1.1755656.

[41] Arthur F. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78(20):3908–3911, may 1997. doi: 10.1103/physrevlett.78.3908. URL https://doi.org/10.1103%2Fphysrevlett.78.3908.

[42] Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of Chemical Physics*, 128(14): 144120, apr 2008. doi: 10.1063/1.2829861. URL https://doi.org/10.1063%2F1.2829861.

[43] Ryuhei Harada and Akio Kitao. Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *The Journal of Chemical Physics*, 139(3): 035103, jul 2013. doi: 10.1063/1.4813023. URL https://doi.org/10.1063%2F1.4813023.

[44] Duy Phuoc Tran, Kazuhiro Takemura, Kazuo Kuwata, and Akio Kitao. Protein–ligand dissociation simulated by parallel cascade selection molecular dynamics. *Journal of Chemical Theory and Computation*, 14(1):404–417, dec 2017. doi: 10.1021/acs.jctc.7b00504. URL https://doi.org/10.1021%2Facs.jctc.7b00504.

[45] Eric Vanden-Eijnden and Maddalena Venturoli. Markovian milestoning with voronoi tessellations. *The Journal of Chemical Physics*, 130(19):194101, may 2009. doi: 10.1063/1.3129843. URL https://doi.org/10.1063%2F1.3129843.

[46] Eric Vanden-Eijnden, Maddalena Venturoli, Giovanni Ciccotti, and Ron Elber. On the assumptions underlying milestoning. *The Journal of Chemical Physics*, 129(17):174102, nov 2008. doi: 10.1063/1.2996509. URL https://doi.org/10.1063%2F1.2996509.

[47] G.A. Huber and S. Kim. Weighted-ensemble brownian dynamics simulations for protein association reactions. *Biophysical Journal*, 70(1):97–110, jan 1996. doi: 10.1016/s0006-3495(96)79552-8. URL https://doi.org/10.1016%2Fs0006-3495%2896%2979552-8.

[48] Rosalind J Allen, Chantal Valeriani, and Pieter Rein ten Wolde. Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter*, 21(46):463102, oct

2009. doi: 10.1088/0953-8984/21/46/463102. URL https://doi.org/10.1088%2F0953-8984%2F21%2F46%2F463102.

[49] Matthew Carter Childers and Valerie Daggett. Insights from molecular dynamics simulations for computational protein design. *Molecular Systems Design & Engineering*, 2 (1):9–33, 2017. doi: 10.1039/c6me00083e. URL https://doi.org/10.1039%2Fc6me00083e.

[50] Duy Phuoc Tran and Akio Kitao. Dissociation process of a MDM2/p53 complex investigated by parallel cascade selection molecular dynamics and the markov state model. *The Journal of Physical Chemistry B*, 123(11):2469–2478, jan 2019. doi: 10.1021/acs.jpcb.8b10309. URL https://doi.org/10.1021%2Facs.jpcb.8b10309.

[51] Hiroaki Hata, Yasutaka Nishihara, Masayoshi Nishiyama, Yoshiyuki Sowa, Ikuro Kawagishi, and Akio Kitao. High pressure inhibits signaling protein binding to the flagellar motor and bacterial chemotaxis through enhanced hydration. *Scientific Reports*, 10(1), feb 2020. doi: 10.1038/s41598-020-59172-3. URL https://doi.org/10.1038%2Fs41598-020-59172-3.

[52] Hiroaki Hata, Duy Tran, Mohamed Marzouk, and Akio Kitao. Binding free energy of protein/ligand complexes calculated using dissociation parallel cascade selection molecular dynamics and markov state model. aug 2021. doi: 10.33774/chemrxiv-2021-51bwf. URL https://doi.org/10.33774%2Fchemrxiv-2021-51bwf.

[53] Ryuhei Harada and Yasuteru Shigeta. Parallel cascade selection molecular dynamics simulations for transition pathway sampling of biomolecules. In *Quantum Systems in Physics, Chemistry and Biology - Theory, Interpretation, and Results*, pages 129–147. Elsevier, 2019. doi: 10.1016/bs.aiq.2018.05.002. URL https://doi.org/10.1016%2Fbs.aiq.2018.05.002.

[54] Ryuhei Harada and Akio Kitao. Nontargeted parallel cascade selection molecular dynamics for enhancing the conformational sampling of proteins. *Journal of Chemical Theory and Computation*, 11(11):5493–5502, oct 2015. doi: 10.1021/acs.jctc.5b00723. URL https://doi.org/10.1021%2Facs.jctc.5b00723.

[55] Daniel Nagel, Anna Weber, and Gerhard Stock. MSMPathfinder: Identification of pathways in markov state models. *Journal of Chemical Theory and Computation*, 16(12):7874–7882, nov 2020. doi: 10.1021/acs.jctc.0c00774. URL https://doi.org/10.1021%2Facs.jctc.0c00774.

[56] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D. Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics*, 134(17):174105, may 2011. doi: 10.1063/1.3565032. URL https://doi.org/10.1063%2F1.3565032.

[57] Frank Noé. Markov models of molecular kinetics. In *Encyclopedia of Biophysics*, pages 1385–1394. Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-16712-6_726. URL https://doi.org/10.1007%2F978-3-642-16712-6_726.

[58] Khaled Barakat, Bilkiss B. Issack, Maria Stepanova, and Jack Tuszynski. Effects of temperature on the p53-DNA binding interactions and their dynamical behavior: Comparing the wild type to the r248q mutant. *PLoS ONE*, 6(11):e27651, nov 2011. doi: 10.1371/journal.pone.0027651. URL https://doi.org/10.1371%2Fjournal.pone.0027651.

[59] Emilia P. Barros, Özlem Demir, Jenaro Soto, Melanie J. Cocco, and Rommie E. Amaro. Markov state models and NMR uncover an overlooked allosteric loop in p53. *Chemical Science*, 12(5):1891–1900, 2021. doi: 10.1039/d0sc05053a. URL https://doi.org/10.1039%2Fd0sc05053a.

[60] Mohan R Pradhan, Jia Wei Siau, Srinivasaraghavan Kannan, Minh N Nguyen, Zohra Ouaray, Chee Keong Kwoh, David P Lane, Farid Ghadessy, and Chandra S Verma. Simulations of mutant p53 DNA binding domains reveal a novel druggable pocket. *Nucleic Acids Research*, 47(4):1637–1652, jan 2019. doi: 10.1093/nar/gky1314. URL https://doi.org/10.1093%2Fnar%2Fgky1314.

[61] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, 27(5):1208–1209, nov 1957. doi: 10.1063/1.1743957. URL https://doi.org/10.1063%2F1.1743957.

[62] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, jun 1977. doi: 10.1038/267585a0. URL https://doi.org/10.1038%2F267585a0.

[63] David E Shaw, Peter J Adams, Asaph Azaria, Joseph A Bank, Brannon Batson, Alistair Bell, Michael Bergdorf, Jhanvi Bhatt, J Adam Butts, Timothy Correia, et al. Anton 3: twenty microseconds of molecular dynamics simulation before lunch. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–11, 2021.

[64] Stephen K Burley, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Li Chen, Gregg V Crichlow, Cole H Christie, Kenneth Dalenberg, Luigi Di Costanzo, Jose M Duarte, Shuchismita Dutta, Zukang Feng, Sai Ganesan, David S Goodsell, Sutapa Ghosh, Rachel Kramer Green, Vladimir Guranović, Dmytro Guzenko, Brian P Hudson, Catherine L Lawson, Yuhe Liang, Robert Lowe, Harry Namkoong, Ezra Peisach, Irina Persikova, Chris Randle, Alexander Rose, Yana Rose, Andrej Sali, Joan Segura, Monica Sekharan, Chenghua Shao, Yi-Ping Tao, Maria Voigt, John D Westbrook, Jasmine Y Young, Christine Zardecki, and Marina Zhuravleva. RCSB protein data bank: powerful new tools for exploring 3d structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*, 49(D1):D437–D451, nov 2020. doi: 10.1093/nar/gkaa1038. URL https://doi.org/10.1093%2Fnar%2Fgkaa1038.

[65] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8): 3696–3713, jul 2015. doi: 10.1021/acs.jctc.5b00255. URL https://doi.org/10.1021%2Facs.jctc.5b00255.

[66] Xiao Zhu, Pedro E. M. Lopes, and Alexander D. MacKerell. Recent developments and applications of the CHARMM force fields. *WIREs Computational Molecular Science*, 2 (1):167–185, jun 2011. doi: 10.1002/wcms.74. URL https://doi.org/10.1002%2Fwcms.74.

[67] Michael J. Robertson, Julian Tirado-Rives, and William L. Jorgensen. Improved peptide and protein torsional energetics with the OPLS-AA force field. *Journal of Chemical Theory and Computation*, 11(7):3499–3509, jun 2015. doi: 10.1021/acs.jctc.5b00356. URL https://doi.org/10.1021%2Facs.jctc.5b00356.

[68] Nathan Schmid, Andreas P. Eichenberger, Alexandra Choutko, Sereina Riniker, Moritz Winger, Alan E. Mark, and Wilfred F. van Gunsteren. Definition and testing of the GROMOS force-field versions 54a7 and 54b7. *European Biophysics Journal*, 40(7):843–856, apr 2011. doi: 10.1007/s00249-011-0700-9. URL https://doi.org/10.1007%2Fs00249-011-0700-9.

[69] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, may 1995. doi: 10.1021/ja00124a002. URL https://doi.org/10.1021%2Fja00124a002.

[70] Jacob D Durrant and J Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1), oct 2011. doi: 10.1186/1741-7007-9-71. URL https://doi.org/10.1186%2F1741-7007-9-71.

[71] David A. Case, Thomas E. Cheatham, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J. Woods. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688, 2005. doi: 10.1002/jcc.20290. URL https://doi.org/10.1002%2Fjcc.20290.

[72] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, jan 1982. doi: 10.1063/1.442716. URL https://doi.org/10.1063%2F1.442716.

[73] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, 159(1):98–103, jul 1967. doi: 10.1103/physrev.159.98. URL https://doi.org/10.1103%2Fphysrev.159.98.

[74] Michel A. Cuendet and Wilfred F. van Gunsteren. On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of Chemical Physics*, 127(18):184102, nov 2007. doi: 10.1063/1.2779878. URL https://doi.org/10.1063%2F1.2779878.

[75] Seppo Mikkola and Sverre J. Aarseth. A time-transformed leapfrog scheme. *Celestial Mechanics and Dynamical Astronomy*, 84:343–354, 2002. doi: 10.1023/A:1021149313347.

[76] Andrew R Leach and Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[77] Philippe H. Hünenberger. Thermostat algorithms for molecular dynamics simulations. In *Advanced Computer Simulation*, pages 105–149. Springer Berlin Heidelberg, jan 2005. doi: 10.1007/b99427. URL https://doi.org/10.1007%2Fb99427.

[78] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8): 3684–3690, oct 1984. doi: 10.1063/1.448118. URL https://doi.org/10.1063%2F1.448118.

[79] Jesús A. Izaguirre, Daniel P. Catarello, Justin M. Wozniak, and Robert D. Skeel. Langevin stabilization of molecular dynamics. *The Journal of Chemical Physics*, 114(5):2090–2098, feb 2001. doi: 10.1063/1.1332996. URL https://doi.org/10.1063%2F1.1332996.

[80] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, feb 1980. doi: 10.1063/1.439486. URL https://doi.org/10.1063%2F1.439486.

[81] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, dec 1981. doi: 10.1063/1.328693. URL https://doi.org/10.1063%2F1.328693.

[82] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, sep 1994. doi: 10.1063/1.467468. URL https://doi.org/10.1063%2F1.467468.

[83] John L Klepeis, Kresten Lindorff-Larsen, Ron O Dror, and David E Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*, 19(2):120–127, apr 2009. doi: 10.1016/j.sbi.2009.03.004. URL https://doi.org/10.1016%2Fj.sbi.2009.03.004.

[84] Vojtěch Spiwok and Igor Tvaroška. Conformational free energy surface of -n-acetylneuraminic acid: An interplay between hydrogen bonding and solvation. *The Journal of Physical Chemistry B*, 113(28):9589–9594, apr 2009. doi: 10.1021/jp8113495. URL https://doi.org/10.1021%2Fjp8113495.

[85] Volodymyr Babin, Christopher Roland, Thomas A. Darden, and Celeste Sagui. The free energy landscape of small peptides as obtained from metadynamics with umbrella sampling corrections. *The Journal of Chemical Physics*, 125(20):204909, nov 2006. doi: 10.1063/1.2393236. URL https://doi.org/10.1063%2F1.2393236.

[86] Stefano Piana, Alessandro Laio, Fabrizio Marinelli, Marleen Van Troys, David Bourry, Christophe Ampe, and José C. Martins. Predicting the effect of a point mutation on a protein fold: The villin and advillin headpieces and their pro62ala mutants. *Journal of Molecular Biology*, 375(2):460–470, jan 2008. doi: 10.1016/j.jmb.2007.10.020. URL https://doi.org/10.1016%2Fj.jmb.2007.10.020.

[87] J. Sun, D. D. Klug, R. Martonak, J. A. Montoya, M.-S. Lee, S. Scandolo, and E. Tosatti. High-pressure polymeric phases of carbon dioxide. *Proceedings of the National Academy of Sciences*, 106(15):6077–6081, mar 2009. doi: 10.1073/pnas.0812624106. URL https://doi.org/10.1073%2Fpnas.0812624106.

[88] Mercedes Alfonso-Prieto, Xevi Biarnés, Pietro Vidossich, and Carme Rovira. The molecular mechanism of the catalase reaction. *Journal of the American Chemical Society*, 131(33):11751–11761, aug 2009. doi: 10.1021/ja9018572. URL https://doi.org/10.1021%2Fja9018572.

[89] Sergei Izrailev, Sergey Stepaniants, Barry Isralewitz, Dorina Kosztin, Hui Lu, Ferenc Molnar, Willy Wriggers, and Klaus Schulten. Steered molecular dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*, pages 39–65. Springer Berlin Heidelberg, 1999. doi: 10.1007/978-3-642-58360-5_2. URL https://doi.org/10.1007%2F978-3-642-58360-5_2.

[90] Outi M. H. Salo-Ahen, Ida Alanko, Rajendra Bhadane, Alexandre M. J. J. Bonvin, Rodrigo Vargas Honorato, Shakhawath Hossain, André H. Juffer, Aleksei Kabedev, Maija Lahtela-Kakkonen, Anders Støttrup Larsen, Eveline Lescrinier, Parthiban Marimuthu, Muhammad Usman Mirza, Ghulam Mustafa, Ariane Nunes-Alves, Tatu Pantsar, Atefeh Saadabadi, Kalaimathy Singaravelu, and Michiel Vanmeert. Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes*, 9(1):71, dec 2020. doi: 10.3390/pr9010071. URL https://doi.org/10.3390%2Fpr9010071.

[91] Helmut Grubmüller, Berthold Heymann, and Paul Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999, feb 1996. doi: 10.1126/science.271.5251.997. URL https://doi.org/10.1126%2Fscience.271.5251.997.

[92] G. Binnig, C. F. Quate, and Ch. Gerber. Atomic force microscope. *Physical Review Letters*, 56(9):930–933, mar 1986. doi: 10.1103/physrevlett.56.930. URL https://doi.org/10.1103%2Fphysrevlett.56.930.

[93] K Svoboda and S M Block. Biological applications of optical forces. *Annual Review of Biophysics and Biomolecular Structure*, 23(1):247–285, jun 1994. doi: 10.1146/annurev.bb.23.060194.001335. URL https://doi.org/10.1146%2Fannurev.bb.23.060194.001335.

[94] E. Evans, K. Ritchie, and R. Merkel. Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. *Biophysical Journal*, 68(6):2580–2587, jun 1995. doi: 10.1016/s0006-3495(95)80441-8. URL https://doi.org/10.1016%2Fs0006-3495%2895%2980441-8.

[95] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690–2693, apr 1997. doi: 10.1103/physrevlett.78.2690. URL https://doi.org/10.1103%2Fphysrevlett.78.2690.

[96] Hui Lu, Barry Isralewitz, André Krammer, Viola Vogel, and Klaus Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*, 75(2):662–671, aug 1998. doi: 10.1016/s0006-3495(98)77556-3. URL https://doi.org/10.1016%2Fs0006-3495%2898%2977556-3.

[97] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical Journal*, 72(4):1568–1581, apr 1997. doi: 10.1016/s0006-3495(97)78804-0. URL https://doi.org/10.1016%2Fs0006-3495%2897%2978804-0.

[98] Toni Giorgino and Gianni De Fabritiis. A high-throughput steered molecular dynamics study on the free energy profile of ion permeation through gramicidin a. *Journal of Chemical Theory and Computation*, 7(6):1943–1950, may 2011. doi: 10.1021/ct100707s. URL https://doi.org/10.1021%2Fct100707s.

[99] Yechun Xu, Jianhua Shen, Xiaomin Luo, Israel Silman, Joel L. Sussman, Kaixian Chen, and Hualiang Jiang. How does huperzine a enter and leave the binding gorge of acetylcholinesterase? steered molecular dynamics simulations. *Journal of the American Chemical Society*, 125(37):11340–11349, aug 2003. doi: 10.1021/ja029775t. URL https://doi.org/10.1021%2Fja029775t.

[100] Yinglong Miao and J. Andrew McCammon. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Molecular Simulation*, 42(13):1046–1055, jul 2016. doi: 10.1080/08927022.2015.1121541. URL https://doi.org/10.1080%2F08927022.2015.1121541.

[101] I. Buch, M. J. Harvey, T. Giorgino, D. P. Anderson, and G. De Fabritiis. High-throughput all-atom molecular dynamics simulations using distributed computing. *Journal of Chemical Information and Modeling*, 50(3):397–403, mar 2010. doi: 10.1021/ci900455r. URL https://doi.org/10.1021%2Fci900455r.

[102] Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*, 116(20):9058–9067, may 2002. doi: 10.1063/1.1472510. URL https://doi.org/10.1063%2F1.1472510.

[103] Nitin Rathore, Manan Chopra, and Juan J. de Pablo. Optimal allocation of replicas in parallel tempering simulations. *The Journal of Chemical Physics*, 122(2):024111, jan 2005. doi: 10.1063/1.1831273. URL https://doi.org/10.1063%2F1.1831273.

[104] David A. Kofke. On the acceptance probability of replica-exchange monte carlo trials. *The Journal of Chemical Physics*, 117(15):6911–6914, oct 2002. doi: 10.1063/1.1507776. URL https://doi.org/10.1063%2F1.1507776.

[105] David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910, 2005. doi: 10.1039/b509983h. URL https://doi.org/10.1039%2Fb509983h.

[106] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, nov 1999. doi: 10.1016/s0009-2614(99)01123-9. URL https://doi.org/10.1016%2Fs0009-2614%2899%2901123-9.

[107] Antonija Kuzmanic and Bojan Zagrovic. Determination of ensemble-average pairwise root mean-square deviation from experimental b-factors. *Biophysical Journal*, 98(5):861–871, mar 2010. doi: 10.1016/j.bpj.2009.11.011. URL https://doi.org/10.1016%2Fj.bpj.2009.11.011.

[108] Samuel Genheden, Anna Reymer, Patricia Saenz-Méndez, and Leif A. Eriksson. Chapter 1. computational chemistry and molecular modelling basics. In *Computational Tools for Chemical Biology*, pages 1–38. Royal Society of Chemistry. doi: 10.1039/9781788010139-00001. URL https://doi.org/10.1039%2F9781788010139-00001.

[109] Jun hui Peng, Wei Wang, Ye qing Yu, Han lin Gu, and Xuhui Huang. Clustering algorithms to analyze molecular dynamics simulation trajectories for complex chemical and biological systems. *Chinese Journal of Chemical Physics*, 31(4):404–420, aug 2018. doi: 10.1063/1674-0068/31/cjcp1806147. URL https://doi.org/10.1063%2F1674-0068%2F31%2Fcjcp1806147.

[110] Jianyin Shao, Stephen W. Tanner, Nephi Thompson, and Thomas E. Cheatham. Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. *Journal of Chemical Theory and Computation*, 3(6):2312–2334, oct 2007. doi: 10.1021/ct700119m. URL https://doi.org/10.1021%2Fct700119m.

[111] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2):129–137, mar 1982. doi: 10.1109/tit.1982.1056489. URL https://doi.org/10.1109%2Ftit.1982.1056489.

[112] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006. URL https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf.

[113] Slimane Doudou, Neil A. Burton, and Richard H. Henchman. Standard free energy of binding from a one-dimensional potential of mean force. *Journal of Chemical Theory and Computation*, 5(4):909–918, mar 2009. doi: 10.1021/ct8002354. URL https://doi.org/10.1021%2Fct8002354.

[114] I. Buch, T. Giorgino, and G. De Fabritiis. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 108(25):10184–10189, jun 2011. doi: 10.1073/pnas.1103547108. URL https://doi.org/10.1073%2Fpnas.1103547108.

[115] Mohamed Sobeh and Akio kitao. Dissociation pathways of the p53 DNA binding domain from DNA and critical roles of key residues elucidated by dPaCS-MD/MSM. dec 2021. doi: 10.26434/chemrxiv-2021-z08zx. URL https://doi.org/10.26434%2Fchemrxiv-2021-z08zx.

[116] Y Wang, J F Schwedes, D Parks, K Mann, and P Tegtmeyer. Interaction of p53 with its consensus DNA-binding site. *Molecular and Cellular Biology*, 15(4):2157–2165, apr 1995. doi: 10.1128/mcb.15.4.2157. URL https://doi.org/10.1128%2Fmcb.15.4.2157.

[117] William C. Ho, Mary X. Fitzgerald, and Ronen Marmorstein. Structure of the p53 core domain dimer bound to DNA. *Journal of Biological Chemistry*, 281(29):20494–20502, jul 2006. doi: 10.1074/jbc.m603634200. URL https://doi.org/10.1074%2Fjbc.m603634200.

[118] Malka Kitayner, Haim Rozenberg, Naama Kessler, Dov Rabinovich, Lihi Shaulov, Tali E. Haran, and Zippora Shakked. Structural basis of DNA recognition by p53 tetramers. *Molecular Cell*, 22(6):741–753, jun 2006. doi: 10.1016/j.molcel.2006.05.015. URL https://doi.org/10.1016%2Fj.molcel.2006.05.015.

[119] K.A. Malecka. Crystal structure of a p53 core tetramer bound to DNA, dec 2008. URL https://doi.org/10.2210%2Fpdb3exl%2Fpdb.

[120] Yongheng Chen, Raja Dey, and Lin Chen. Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure*, 18(2):246–256, feb 2010. doi: 10.1016/j.str.2009.11.011. URL https://doi.org/10.1016%2Fj.str.2009.11.011.

[121] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, feb 1996. doi: 10.1016/0263-7855(96)00018-5. URL https://doi.org/10.1016%2F0263-7855%2896%2900018-5.

[122] Jumin Lee, Manuel Hitzenberger, Manuel Rieger, Nathan R. Kern, Martin Zacharias, and Wonpil Im. CHARMM-GUI supports the amber force fields. *The Journal of Chemical Physics*, 153(3):035103, jul 2020. doi: 10.1063/5.0012280. URL https://doi.org/10.1063%2F5.0012280.

[123] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, jul 1983. doi: 10.1063/1.445869. URL https://doi.org/10.1063%2F1.445869.

[124] Marie Zgarbová, Jiří Šponer, Michal Otyepka, Thomas E. Cheatham, Rodrigo Galindo-Murillo, and Petr Jurečka. Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of z- and b-DNA. *Journal of Chemical Theory and Computation*, 11(12):5723–5736, nov 2015. doi: 10.1021/acs.jctc.5b00716. URL https://doi.org/10.1021%2Facs.jctc.5b00716.

[125] Martin B. Peters, Yue Yang, Bing Wang, László Füsti-Molnár, Michael N. Weaver, and Kenneth M. Merz. Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *Journal of Chemical Theory and Computation*, 6(9):2935–2947, aug 2010. doi: 10.1021/ct1002626. URL https://doi.org/10.1021%2Fct1002626.

[126] Catherine Méplan, Marie-Jeanne Richard, and Pierre Hainaut. Metalloregulation of the tumor suppressor protein p53: zinc mediates the renaturation of p53 after exposure to

metal chelators in vitro and in intact cells. *Oncogene*, 19(46):5227–5236, nov 2000. doi: 10.1038/sj.onc.1203907. URL https://doi.org/10.1038%2Fsj.onc.1203907.

[127] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, mar 1977. doi: 10.1016/0021-9991(77)90098-5. URL https://doi.org/10.1016%2F0021-9991%2877%2990098-5.

[128] Shuichi Miyamoto and Peter A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, oct 1992. doi: 10.1002/jcc.540130805. URL https://doi.org/10.1002%2Fjcc.540130805.

[129] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, nov 1995. doi: 10.1063/1.470117. URL https://doi.org/10.1063%2F1.470117.

[130] Melissa F Adasme, Katja L Linnemann, Sarah Naomi Bolz, Florian Kaiser, Sebastian Salentin, V Joachim Haupt, and Michael Schroeder. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Research*, 49(W1):W530–W534, may 2021. doi: 10.1093/nar/gkab294. URL https://doi.org/10.1093%2Fnar%2Fgkab294.

[131] Evan H Baugh, Hua Ke, Arnold J Levine, Richard A Bonneau, and Chang S Chan. Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death & Differentiation*, 25(1):154–160, nov 2017. doi: 10.1038/cdd.2017.180. URL https://doi.org/10.1038%2Fcdd.2017.180.

[132] Gregory R. Bowman, Vijay S. Pande, and Frank Noé, editors. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. Springer Netherlands, 2014. doi: 10.1007/978-94-007-7606-7. URL https://doi.org/10.1007%2F978-94-007-7606-7.

[133] Martin K. Scherer, Benjamin Trendelkamp-Schroer, Fabian Paul, Guillermo Pérez-Hernández, Moritz Hoffmann, Nuria Plattner, Christoph Wehmeyer, Jan-Hendrik Prinz, and Frank Noé. PyEMMA 2: A software package for estimation, validation, and analysis of markov models. *Journal of Chemical Theory and Computation*, 11(11):5525–5542, oct 2015. doi: 10.1021/acs.jctc.5b00743. URL https://doi.org/10.1021%2Facs.jctc.5b00743.

[134] Slimane Doudou, Raman Sharma, Richard H. Henchman, David W. Sheppard, and Neil A. Burton. Inhibitors of PIM-1 kinase: A computational analysis of the binding free energies of a range of imidazo [1,2-b] pyridazines. *Journal of Chemical Information and Modeling*, 50 (3):368–379, feb 2010. doi: 10.1021/ci9003514. URL https://doi.org/10.1021%2Fci9003514.

[135] C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, dec 1996. doi: 10.1145/235815.235821. URL https://doi.org/10.1145%2F235815.235821.

[136] Sergey Yu Noskov, Jon D. Wright, and Carmay Lim. Long-range effects of mutating r248 to q/w in the p53 core domain. *The Journal of Physical Chemistry B*, 106(50):13047–13057, nov 2002. doi: 10.1021/jp022140w. URL https://doi.org/10.1021%2Fjp022140w.

[137] Shah Md. Abdur Rauf, Mohamed Ismael, Kamlesh Kumar Sahu, Ai Suzuki, Michihisa Koyama, Hideyuki Tsuboi, Nozomu Hatakeyama, Akira Endou, Hiromitsu Takaba, Carlos A. Del Carpio, Momoji Kubo, and Akira Miyamoto. The effect of r249s carcinogenic and h168r–r249s suppressor mutations on p53–DNA interaction, a multi scale computational study. *Computers in Biology and Medicine*, 40(5):498–508, may 2010. doi: 10.1016/j.compbiomed.2010.03.004. URL https://doi.org/10.1016%2Fj.compbiomed.2010.03.004.

[138] Andreas C. Joerger, Hwee Ching Ang, Dmitry B. Veprintsev, Caroline M. Blair, and Alan R. Fersht. Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations. *Journal of Biological Chemistry*, 280(16):16030–16037, apr 2005. doi: 10.1074/jbc.m500179200. URL https://doi.org/10.1074%2Fjbc.m500179200.

[139] Tsuyoshi Terakawa, Hiroo Kenzaki, and Shoji Takada. p53 searches on DNA by rotation-uncoupled sliding at c-terminal tails and restricted hopping of core domains. *Journal of the American Chemical Society*, 134(35):14555–14562, aug 2012. doi: 10.1021/ja305369u. URL https://doi.org/10.1021%2Fja305369u.

[140] Lavi S. Bigman, Harry M. Greenblatt, and Yaakov Levy. What are the molecular requirements for protein sliding along DNA? *The Journal of Physical Chemistry B*, 125(12):3119–3131, mar 2021. doi: 10.1021/acs.jpcb.1c00757. URL https://doi.org/10.1021%2Facs.jpcb.1c00757.

[141] Dwiky Rendra Graha Subekti, Agato Murata, Yuji Itoh, Satoshi Takahashi, and Kiyoto Kamagata. Transient binding and jumping dynamics of p53 along DNA revealed by sub-millisecond resolved single-molecule fluorescence tracking. *Scientific Reports*, 10(1), aug 2020. doi: 10.1038/s41598-020-70763-y. URL https://doi.org/10.1038%2Fs41598-020-70763-y.

[142] Andreas C. Joerger and Alan R. Fersht. The p53 pathway: Origins, inactivation in cancer, and emerging therapeutic approaches. *Annual Review of Biochemistry*, 85(1):375–404, jun 2016. doi: 10.1146/annurev-biochem-060815-014710. URL https://doi.org/10.1146%2Fannurev-biochem-060815-014710.

[143] K.-B. Wong, B. S. DeDecker, S. M. V. Freund, M. R. Proctor, M. Bycroft, and A. R. Fersht. Hot-spot mutants of p53 core domain evince characteristic local structural changes. *Proceedings of the National Academy of Sciences*, 96(15):8438–8442, jul 1999. doi: 10.1073/pnas.96.15.8438. URL https://doi.org/10.1073%2Fpnas.96.15.8438.

[144] James S. Butler and Stewart N. Loh. Structure, function, and aggregation of the zinc-free form of the p53 DNA binding domain. *Biochemistry*, 42(8):2396–2403, feb 2003. doi: 10.1021/bi026635n. URL https://doi.org/10.1021%2Fbi026635n.

[145] A C Joerger and A R Fersht. Structure–function–rescue: the diverse nature of common p53 cancer mutants. *Oncogene*, 26(15):2226–2242, apr 2007. doi: 10.1038/sj.onc.1210291. URL https://doi.org/10.1038%2Fsj.onc.1210291.

[146] J. D. Wright. Factors governing loss and rescue of DNA binding upon single and double mutations in the p53 core domain. *Nucleic Acids Research*, 30(7):1563–1574, apr 2002. doi: 10.1093/nar/30.7.1563. URL https://doi.org/10.1093%2Fnar%2F30.7.1563.

[147] Oded Suad, Haim Rozenberg, Ran Brosh, Yael Diskin-Posner, Naama Kessler, Linda J.W. Shimon, Felix Frolow, Atar Liran, Varda Rotter, and Zippora Shakked. Structural basis of restoring sequence-specific DNA binding and transactivation to mutant p53 by suppressor mutations. *Journal of Molecular Biology*, 385(1):249–265, jan 2009. doi: 10.1016/j.jmb.2008.10.063. URL https://doi.org/10.1016%2Fj.jmb.2008.10.063.

[148] Ana Gomes, Filipa Trovão, Benedita Andrade Pinheiro, Filipe Freire, Sara Gomes, Carla Oliveira, Lucília Domingues, Maria Romão, Lucília Saraiva, and Ana Carvalho. The crystal structure of the r280k mutant of human p53 explains the loss of DNA binding. *International Journal of Molecular Sciences*, 19(4):1184, apr 2018. doi: 10.3390/ijms19041184. URL https://doi.org/10.3390%2Fijms19041184.

[149] A. C. Joerger and A. R. Fersht. The tumor suppressor p53: From structures to drug discovery. *Cold Spring Harbor Perspectives in Biology*, 2(6):a000919–a000919, feb 2010. doi: 10.1101/cshperspect.a000919. URL https://doi.org/10.1101%2Fcshperspect.a000919.

[150] Qiang Zhang, Vladimir J. N. Bykov, Klas G. Wiman, and Joanna Zawacka-Pankau. APR-246 reactivates mutant p53 by targeting cysteines 124 and 277. *Cell Death & Disease*, 9 (5), apr 2018. doi: 10.1038/s41419-018-0463-7. URL https://doi.org/10.1038%2Fs41419-018-0463-7.

[151] Fan He, Wade Borcherds, Tanjing Song, Xi Wei, Mousumi Das, Lihong Chen, Gary W. Daughdrill, and Jiandong Chen. Interaction between p53 n terminus and core domain regulates specific and nonspecific DNA binding. *Proceedings of the National Academy of Sciences*, 116(18):8859–8868, apr 2019. doi: 10.1073/pnas.1903077116. URL https://doi.org/10.1073%2Fpnas.1903077116.

[152] Özlem Demir, Emilia P Barros, Tavina L Offutt, Mia Rosenfeld, and Rommie E Amaro. An integrated view of p53 dynamics, function, and reactivation. *Current Opinion in Structural Biology*, 67:187–194, apr 2021. doi: 10.1016/j.sbi.2020.11.005. URL https://doi.org/10.1016%2Fj.sbi.2020.11.005.