論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Prediction of Protein Stability and Stability Changes Using Dynamic Information from Molecular Dynamics Simulation
著者(和文)	KurniawanJason
Author(English)	Jason Kurniawan
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12781号, 授与年月日:2024年3月26日, 学位の種別:課程博士, 審査員:石田 貴士,秋山 泰,瀧ノ上 正浩,関嶋 政和,大上 雅史
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12781号, Conferred date:2024/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Prediction of Protein Stability and Stability Changes Using Dynamic Information from Molecular Dynamics Simulation

Jason Kurniawan
Department of Computer Science
School of Computing
Tokyo Institute of Technology

Advisor: Takashi Ishida

A thesis submitted for the degree of
Doctor of Artificial Intelligence
2024

Tokyo Tech

Contents

List of Figures

Abstract

The prediction of protein stability and its changes is fundamental to understanding protein function in health and disease, guiding drug design, and engineering proteins with desired properties. To facilitate these predictions, various computational methods have been developed. Current state-of-the-art methods for predicting protein stability, predominantly based on machine learning (ML), face critical limitations due to the inherent nature of supervised learning. These include data set biases, the black-box nature of algorithms, and a reliance on static protein information, which overlooks the intrinsic dynamic behavior of proteins. Such limitations can lead to inaccuracies, especially when predicting the stability of proteins under various conditions, including point mutations. To address these gaps and leverage the dynamic nature of proteins, this research adopts Molecular Dynamics (MD) simulations, renowned for their success in capturing the complex, real-time interactions within proteins.

The first phase of the research proposes the use of MD simulations to evaluate the accuracy of predicted three-dimensional protein structures. This approach provides a deeper understanding of protein stability, moving beyond static models to incorporate the dynamic aspects that are crucial for an accurate representation of protein behavior. By incorporating the protein stability information obtained from MD simulations, the proposed method introduced new useful information for protein model quality estimation while overcoming the limitation of ML-based methods.

In the second phase, the research extends the application field from protein stability to stability changes, particularly in the domain of point mutations on protein stability by employing alchemical molecular dynamics simulations to predict the resulting stability changes. We introduced a novel protocol to enhance the performance by specifically handling charge-changing mutation cases. Our proposed approach achieves significantly better performance in comparison with state-of-the-art machine learning-based methods. The research overcomes the limitations of static ML-based methods, providing a more accurate prediction.

Finally, the methodologies developed and tested in this study not only enhance the prediction accuracy for both protein stability and stability changes but also pave the way for

more comprehensive and highly applicable predictions in protein research. We explored the potential to extend the method to more complex scenarios, such as the integration with ML-based methods and multiple mutation cases. By integrating with machine learning-based methods, it can tackle computational limitations which is the main disadvantage of the simulation methods, enhancing the speed without compromising the accuracy of stability predictions. The initial results in multiple mutations also represent a positive outcome. These open new alternatives for more reliable and insightful predictions in protein research and its applications, particularly in the fields of drug design, therapeutic interventions, and protein engineering.

Acknowledgements

This section is dedicated to expressing my deepest appreciation to those who have been integral in my journey through Ph.D. studies, making it an enriching and fulfilling endeavor.

First and foremost, I am profoundly thankful to Professor Takashi Ishida for his relentless guidance and encouragement throughout my Ph.D. journey. His steadfast support has been a cornerstone in my academic pursuit, inspiring me to overcome challenges and achieve my goals. The lessons learned under his mentorship have been invaluable, and this thesis owes its completion to his guidance over the past five years.

My sincere thanks also go to Professor Yutaka Akiyama, Masakazu Sekijima, Masahiro Takinoue, and Masahito Ohue as the dissertation committee members. Their constructive feedback and discussions have greatly contributed to the refinement of this thesis.

I cannot express enough gratitude to my family for their boundless love and support. My deepest indebtedness goes to my parents, whose nurturing has shaped who I am today. Their unwavering encouragement, despite the physical distance, has been a pillar of strength and motivation, enabling me to venture into new realms in my life.

A heartfelt thank you to my colleague Hasic Haris, whose valuable insights and engaging discussions have greatly enhanced my comprehension of my research. Special recognition is given to Kuniko Nakayama and all members of the Takashi Ishida Lab, whose assistance and camaraderie have been a significant part of my academic life.

Chapter 1

Introduction

1.1 Protein structure

A protein is a large, complex molecule composed of amino acids, essential for the structure, function, and regulation of the body's cells, tissues, and organs. Proteins are essential in all biological processes, dictating the functional dynamics of living organisms at the molecular level. Protein structure is a fundamental concept in molecular biology, encompassing a hierarchy of organizational levels that dictate the function and interaction of proteins. [1] The primary structure, consisting of a linear sequence of amino acids, is determined by the genetic code and sets the stage for higher-order structures. This sequence undergoes folding to form secondary structures, such as alpha-helices and beta-sheets, driven by patterns of hydrogen bonding and the chemical properties of the constituent amino acids. These secondary structures further fold into a unique three-dimensional configuration known as the tertiary structure, which is stabilized by various interactions including hydrophobic interactions, hydrogen bonds, ionic bonds, and disulfide bridges. [2] The highest level of protein structure, the quaternary structure, emerges when multiple protein sub-units assemble into a functional complex.

One of essential characteristics of proteins is their dynamic nature. [3, 4] While often depicted as static entities, proteins in reality exhibit a high degree of flexibility and undergo various conformational changes (Figure 1.1). [5] This dynamic behavior is crucial for their biological function, allowing proteins to adapt to different functional states, interact with diverse biomolecules, and participate in complex cellular processes. The dynamic nature of proteins is intrinsically linked to their stability. [6] Environmental factors such as temperature, pH, and interactions with other molecules can induce these conformational changes, highlighting the importance of understanding protein dynamics for accurate functional prediction. The interplay between protein dynamics and stability is a key area of study, as changes in conformation can significantly impact a protein's stability. This relationship is particularly evident in cases where alterations in protein dynamics lead to stability changes, affecting protein function and potentially leading to disease states.

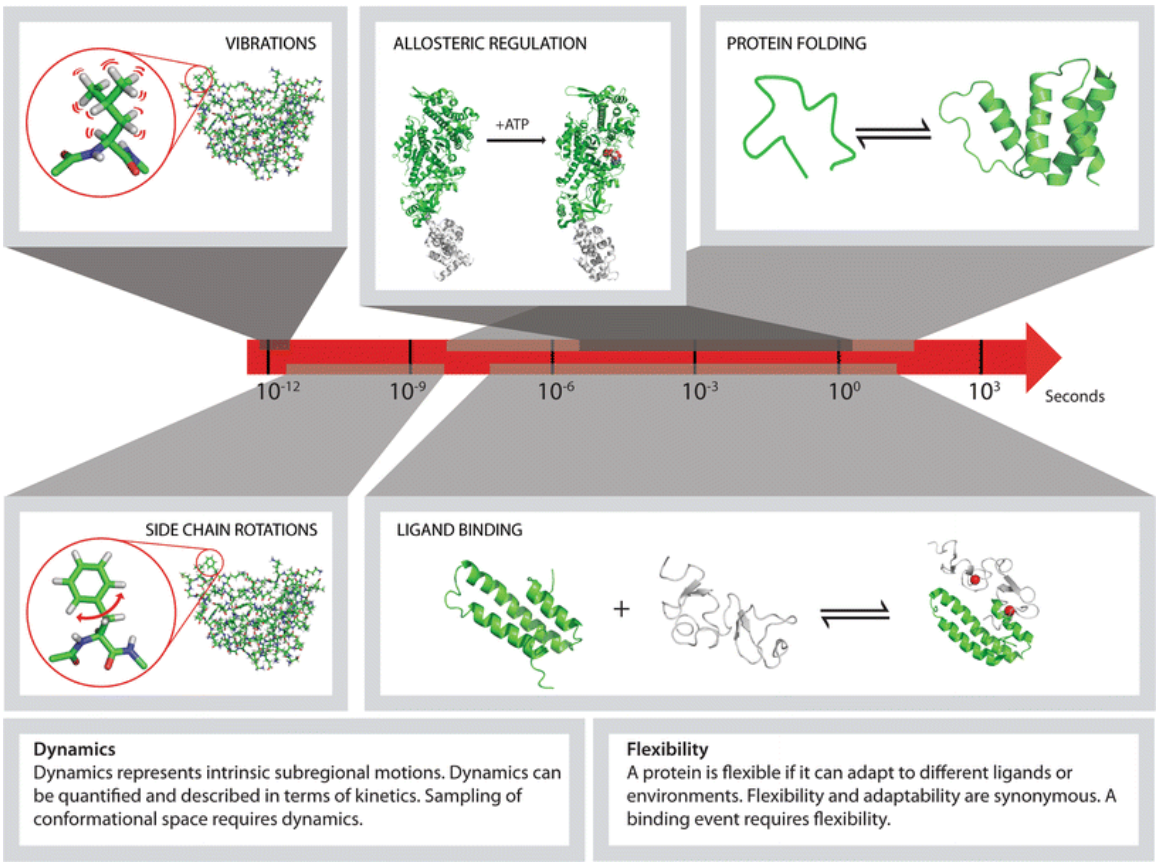


Figure 1.1: Illustration of dynamic conformational events in proteins. Reproduced with permission from Springer Nature. [5] This schematic depicts the range of motions altering protein conformations, spanning time scales from nanoseconds to hours. It highlights the diverse temporal dynamics inherent in protein structural changes, encompassing rapid fluctuations to slower, more gradual transformations.

1.2 Protein structure stability

Protein stability is a fundamental aspect of molecular biology and biophysics that intricately ties into every domain of life processes. It represents the balance between the unfolded and folded states of a protein in specific environmental conditions as depicted by the example shown in Figure 1.2. This balance is underpinned by the thermodynamic measures of a protein's propensity to remain in its native, folded state as opposed to a denatured or unfolded state. [7] In biological systems, the stability of a protein is important for its function, as every protein's role is deeply intertwined with its three-dimensional structure. [8] This spatial arrangement of amino acids is what bestows proteins with their unique abilities, whether it's enzymatic catalysis, structural support, or cellular signaling. Protein stability also dictates the longevity of a protein within cellular systems. Proteins with higher stability are more resistant to degradation, thereby maintaining their functional presence for the necessary duration. On the other hand, instability can lead to protein aggregation, which is an aberration in protein behavior and has been linked to numerous diseases, including neurodegenerative conditions like Alzheimer's and Parkinson's. [9] The balance of protein stability is not just internally determined by the amino acid sequence and structural configuration of the protein. It's also influenced by an array of external factors. These range from environmental aspects like pH, temperature, and ionic strength, to cellular interactions with ligands, other proteins, or even molecular chaperones that assist in protein folding.

1.2.1 Measuring protein stability

To measure the stability of a protein, one must consider both its thermodynamic and kinetic characteristics. Thermodynamically, the free energy difference, represented as ΔG, stands out as a primary indicator. ΔG provides profound insights by quantifying the energy difference between a protein's folded and unfolded states. A negative ΔG indicates a spontaneous process, suggesting that the protein is inherently stable in its native conformation. Another important measure is the melting temperature, or Tm, which elucidates the exact thermal point at which a protein transitions from its native, folded state to an unfolded conformation. [11, 12, 13] While these measures provide quantitative insights, the behavior of proteins under denaturing conditions can offer qualitative understandings. For instance, the way a protein responds to chemical agents that disrupt its structure can reveal not only its structural robustness but also pinpoint specific regions that may be vulnerable to destabilization.

1.2.2 Experimental Determination of Protein Stability

The experimental determination of protein stability is a crucial aspect of molecular biology, providing essential insights into the behavior of proteins under various conditions. This understanding is pivotal for applications ranging from drug design to the elucidation of disease mechanisms. Over

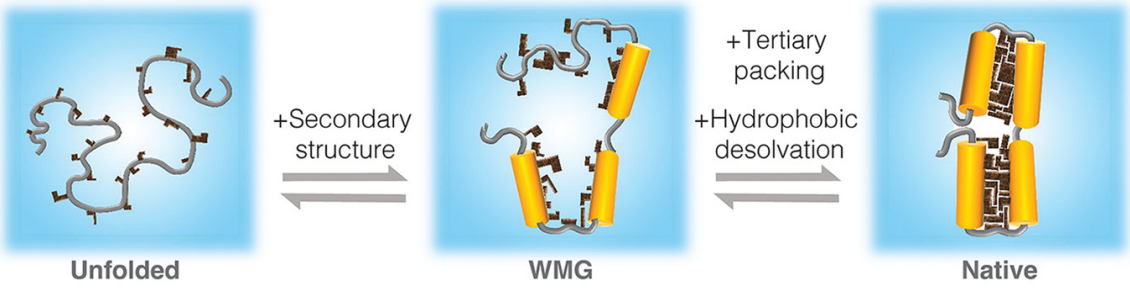


Figure 1.2: This schematic illustrates the stepwise folding process of a protein, depicting its transition from an unfolded state to a stabilized native state. Reprinted with permission from [10]. Copyright 2022 American Chemical Society. From left to right, the protein progressively stabilizes towards its native conformation. Conversely, moving from right to left, the protein destabilizes, transitioning towards an unfolded state. A key feature of this process is the WMG (Warm Molten Globule) intermediate state, represented as a partially folded stage with a hydrated core. The transition from the WMG intermediate to the native state involves a critical step of hydrophobic desolvation and the establishment of tertiary packing, culminating in the protein achieving its fully folded, native structure.

the decades, the methodologies employed in determining protein stability have evolved significantly, transitioning from basic techniques to advanced, sophisticated methodologies. The field of experimental protein stability determination has undergone significant evolution, transitioning from basic techniques to advanced, sophisticated methodologies. [14]

Central to these developments are calorimetric techniques, which have been instrumental in elucidating the thermodynamic aspects of protein behavior. [15] These techniques, by precisely measuring the heat exchange during protein folding and unfolding processes, offer direct insights into the thermodynamics governing protein stability. In conjunction with calorimetry, spectroscopic methods have emerged as indispensable tools in protein stability analysis. [16] Techniques such as absorbance spectrometry, fluorescence resonance, and circular dichroism are employed to monitor subtle structural alterations in proteins under various conditions. [17, 18] These methods are particularly adept at detecting changes such as the exposure of hydrophobic cores or alterations in secondary structural elements like alpha helices and beta sheets.

Recent decades have witnessed the advent of more advanced biophysical methods, integrating cutting-edge technologies like nuclear magnetic resonance (NMR) and X-ray crystallography. [19, 20] These methodologies have significantly broadened the scope of protein stability studies, providing a more detailed and comprehensive understanding of the factors influencing protein stability. Collectively, these diverse techniques have greatly enhanced our ability to characterize protein stability, contributing profoundly to the field of molecular biology. [21] The insights gained from these advanced methods have not only deepened the fundamental understanding of protein behavior but have also opened new fields for various applications.

1.2.3 Application

The experimentally determined protein stability information finds extensive applications across various scientific and medical fields. In drug design and development, understanding protein stability is crucial for identifying stable protein targets essential for effective drug binding and action. [22] This knowledge aids in the development of drugs with optimal binding affinity and efficacy. Similarly, in the pharmaceutical industry, the stability of therapeutic proteins such as enzymes and antibodies is a key consideration for their efficacy and shelf life. Stable therapeutic proteins are more effective in treatment regimes and less prone to degradation, making them more reliable in clinical applications. [23]

In the field of disease research, protein stability data is instrumental in understanding the molecular basis of various disorders, particularly those involving protein misfolding or aggregation, such as Alzheimer's and Parkinson's diseases. By analyzing how mutations or environmental factors affect protein stability, researchers can uncover the underlying mechanisms of these diseases and develop targeted treatments. [24] Additionally, in the field of biotechnology, protein stability information is vital for the design of enzymes used in industrial processes. [25] Enzymes with enhanced stability are more efficient and can withstand harsh industrial conditions, making them more suitable for applications in biofuels, waste management, and food processing. [26]

Furthermore, in structural biology, the stability data of proteins assists in the accurate modeling of protein structures, which is crucial for understanding their function and interactions. This information is particularly valuable in the design of biomimetic materials and in synthetic biology, where stable protein structures are engineered for specific applications. [27] In proteomics, stability data helps in predicting the behavior of proteins under different conditions, facilitating the discovery of new biomarkers for disease diagnosis and prognosis. [28] Lastly, in the emerging field of personalized medicine, protein stability information can be used to tailor treatments based on individual genetic profiles. [29] Understanding how specific genetic variations influence protein stability can guide the development of personalized therapeutic strategies, ensuring more effective and targeted treatments for patients. [30]

Additionally, accurate modeling of protein structures informed by stability data plays a pivotal role in protein model quality assessment. [31] This application is particularly relevant in the field of computational biology, where the assessment of predicted protein models against experimental data is crucial. The ability to accurately predict and assess protein structures using stability data enhances the reliability of computational models, making them indispensable tools in drug discovery and protein engineering. [32] By improving the accuracy of protein models, researchers can better understand protein behavior in various conditions, leading to more effective drug designs and novel therapeutic approaches.

Lastly, in the emerging field of personalized medicine, protein stability information can be used to tailor treatments based on individual genetic profiles. Understanding how specific genetic variations influence protein stability can guide the development of personalized therapeutic strategies, ensuring more effective and targeted treatments for patients. [30, 33]

1.3 Protein stability prediction

The attempt to predict protein stability is a critical aspect of contemporary molecular biology, primarily driven by the importance of addressing the limitations of wet laboratory experimental methods. Experimental methods while foundational in elucidating protein behavior, are often constrained by their inherent limitations in scalability, resource intensity, and temporal efficiency. [34, 35] These constraints are particularly essential in the context of the expansive and complex data sets characteristic of modern biological research.

Computational methods have emerged as a pivotal response to these challenges, offering an alternative approach by performing prediction of protein stability through theoretical and algorithmic frameworks. These methods encompass a broad spectrum of computational strategies, each tailored to specific aspects of protein analysis. For instance, when examining protein structures, computational models can predict how alterations in amino acid sequences or environmental conditions might affect the overall stability of the protein. Similarly, in the context of protein-ligand interactions, these methods can forecast how binding events may influence the structural integrity of the protein. [36, 37]

The concept paradigm of computational predictions in protein stability involves comprehensive analysis of various protein attributes. Inputs such as primary sequence data, tertiary structure information, or environmental parameters are processed through computational models to yield predictions about the protein's stability. [38, 39, 40] By providing predictions on protein stability under a variety of conditions and scenarios, computational approaches extend the reach of experimental techniques. They enable researchers to hypothesize and test a broader range of biological phenomena in a virtual environment, which can then guide and inform subsequent experimental investigations. [41] They address some of the limitations of experimental approaches by offering a more scalable and efficient means to explore protein behavior.

1.3.1 Prediction target

Protein stability supports a wide array of research applications. In the domain of computational prediction, the targets for prediction extend from direct thermodynamic measures to include various derivatives of stability information. These targets encompass a range of parameters, from thermodynamic measures like the free energy difference (ΔG) and the melting temperature (Tm), as well as

their derivatives (ΔTm and $\Delta\Delta G$), each with multiple applications across various fields of protein research. [42, 18]

The prediction of ΔG is crucial for understanding the inherent stability of proteins in their native conformation. Computational methods that accurately predict ΔG provide insights into how proteins maintain their structural integrity under varying environmental conditions. [43] Such predictions are fundamental for accurate modeling of protein structures, especially in designing stable protein structures for specific applications like biomimetic materials. Similarly, the prediction of Tm is vital for determining the thermal stability of proteins, a key factor in both drug design and the development of therapeutic proteins. The ability to predict Tm aids in identifying proteins that can maintain functionality under thermal stress, which is crucial in pharmaceutical applications.

Furthermore, the changes in melting temperature (ΔTm) and free energy difference ($\Delta\Delta G$) are particularly significant in the context of protein-protein interactions, protein-ligand binding, and the effects of mutations. Computational predictions of ΔTm and $\Delta\Delta G$ assist in evaluating how structural alterations, such as those induced by point mutations or binding events, influence the overall stability of proteins. [44, 45, 46] This information is critical for unraveling the molecular mechanisms underlying various diseases and for the development of targeted drug therapies, as it provides a deeper understanding of how stability changes can impact protein function and interactions.

In the context of prediction targets that are the derivatives of stability information, understanding the stability of a protein's native state is crucial. According to Anfinsen's dogma, the native structure of a protein, determined solely by its amino acid sequence, is the most stable conformation under physiological conditions. [47] This principle underscores the significance of the native state's stability in protein research, as it is indicative of the protein's functional form. To assess the stability of the native state indirectly, one effective approach is through model quality assessment. [48] This involves generating a score that reflects the accuracy and reliability of a predicted protein model. Although the score is not a direct physical measure akin to ΔG, it is intrinsically linked to the stability information of the protein, serving as an indicator of the protein's structural integrity and stability. [49, 50]

1.3.2 Scope of this research

In this research, the application of protein stability information is initially explored within the field of protein model quality assessment. This area was chosen due to its critical role in computational biology, where accurate prediction and assessment of protein structures are essential for understanding protein function and interactions. Accurately predicted structure theoretically is the structure with the most stable conformation. Reliable quality assessment is pivotal in determining the usability of predicted models in further biological and biomedical research, making it a foundational step in protein analysis. The discussion then progresses to address protein stability changes, especially in the context of point mutations. The focus on point mutations was selected because of their profound impact on protein stability, leading to significant structural and functional alterations. Such changes are at the heart of many genetic diseases and are crucial in the development of targeted therapies. Understanding these stability changes through the prediction of $\Delta\Delta G$ is essential for unraveling the molecular mechanisms underlying various diseases and for the development of effective therapeutic strategies. By exploring both protein model quality assessment and stability changes due to point mutations, this research aims to contribute to a more comprehensive understanding of protein dynamics, offering insights that are vital for both basic scientific inquiry and practical applications in medicine and drug design.

1.3.3 Methods of protein stability prediction

The field of protein stability prediction has evolved significantly over time, reflecting advancements in computational methods and a deeper understanding of protein dynamics. Initially, statistical methods were employed, leveraging sequence and structural data to infer stability changes, particularly in the context of mutations. [51, 52, 53] These methods were foundational in the field, providing initial insights into how protein stability could be predicted from available data. However, their effectiveness was often limited by the availability and quality of empirical data. These early methods laid the groundwork for future advancements, setting the stage for more sophisticated approaches. [54]

As the field progressed, physics-based methods were introduced, offering a more theoretical approach to understanding protein behavior. [55] Grounded in thermodynamics and molecular interactions, these methods aimed to provide a deeper understanding of the forces governing protein stability. They represented a significant step forward, moving beyond empirical data to theoretical modeling. However, these methods often required simplifications in their models due to computational limitations at the time, which sometimes limited their accuracy and applicability. [56]

With the advent of machine learning, a significant shift occurred in the field. These algorithms, capable of processing large data sets, began to enhance prediction accuracy by incorporating a variety of features, including physicochemical properties and empirical energy functions. Machine

learning methods marked an improvement over earlier techniques, offering more data-driven and sophisticated analyses. However, they sometimes faced challenges with interpretability and data dependency, highlighting the need for large, high-quality data sets for training. [57, 58]

Deep learning, a subset of machine learning, further advanced the field of protein stability prediction. These models, adept at handling complex, high-dimensional data, offered more nuanced and detailed predictions. Deep learning's ability to discern intricate patterns in protein sequences and structures led to more reliable predictions. [59, 60] However, these models often required substantial computational resources and were sometimes criticized for their 'black box' nature, which made it challenging to understand the basis of their predictions. [61]

The latest advancement in this field is the integration of molecular dynamics simulation. Molecular dynamics simulations provide a dynamic perspective, capturing real-time interactions within proteins. This approach is particularly effective in understanding stability changes due to mutations and environmental factors. [62, 63] It represents a comprehensive method, considering both the static structure and the dynamic nature of proteins. However, molecular dynamics simulations are computationally intensive and require a high level of expertise for accurate interpretation, making them a sophisticated but resource-demanding tool in protein stability prediction. [64]

1.4 Problem of current protein stability prediction methods

The field of protein stability prediction, while having made significant strides, faces several challenges. A primary issue is the heavy reliance on extensive, high-quality data sets, particularly in machine learning and deep learning approaches. [65] These methods require a significant amount of data for training, and the accuracy of their predictions is heavily dependent on the quality and diversity of the data used. This reliance becomes a major hurdle when dealing with rare or novel proteins, for which extensive data may not be available. Additionally, the collection and preparation of such large datasets can be resource-intensive and time-consuming. [66]

Another critical issue is the black-box nature of many machine learning and deep learning models. [67] While these models can provide accurate predictions, understanding the underlying reasons for these predictions can be challenging. This lack of interpretability is a significant drawback in research settings where understanding the basis of predictions is crucial for further scientific inquiry and validation. [68] Moreover, the complexity of these models often makes them less transparent, hindering the ability to improve or troubleshoot them based on specific needs or observations.

Generalizability also poses a significant challenge. Many existing prediction methods are developed and tested on specific types of proteins or conditions, limiting their broader applicability. This specialization can limit their applicability to a broader range of proteins or different environmental conditions, reducing their utility in diverse research or practical applications. [69] The challenge lies

in developing methods that are both accurate and versatile, capable of handling a wide range of proteins and conditions with equal efficacy.

Physics-based methods, though less reliant on large data sets, often require simplifications in their models due to computational limitations. These simplifications can sometimes lead to inaccuracies, especially when dealing with complex proteins or interactions. Additionally, both machine learning and physic-based methods may not fully capture the dynamic nature of proteins, focusing more on static structural aspects. As a result, they might miss critical insights into how proteins behave under various physiological conditions. [54]

Another significant challenge is the computational intensity of advanced methods like molecular dynamics simulations. While molecular dynamics simulations provide detailed and dynamic insights into protein behavior, they require substantial computational resources, making them less accessible for widespread use. This limitation is particularly pronounced in resource-constrained environments or for researchers who do not have access to high-performance computing facilities. Furthermore, the complexity of these simulations necessitates a high level of expertise for accurate setup and interpretation, which can be a barrier for many researchers. [70]

1.5 Research purpose

The purpose of this thesis is to introduce new useful information to improve the accuracy of protein stability and stability change prediction by introducing a novel approach that integrates dynamic information from molecular dynamics simulations. It also aimed to address and overcome the inherent limitations of current machine learning-based methods, which predominantly rely on static structural data and require complex feature extraction and extensive training. This was accomplished on two research applications where first, by incorporating dynamic protein stability features derived from molecular dynamics simulation trajectories, we developed a new method for protein model quality estimation. In the further extension to stability changes research, the developed method aims to automate and improve the prediction process for protein stability changes due to point mutations, specifically targeting the challenges posed by charge-changing mutations.

1.6 Contributions

The main contribution of this thesis is the development of a novel approach that integrates protein dynamics information, obtained from molecular dynamics simulations, into the prediction of protein stability and stability changes in single protein structures. This novel approach addresses and overcomes the inherent limitations found in machine learning-based methods. A key advantage of the proposed method is it does not need complex feature extraction and extensive training processes typically associated with machine learning models. By incorporating dynamic information from the

molecular dynamics simulations, this approach not only introduces additional useful information but also enhances the accuracy of protein stability predictions. This offers a more comprehensive and dynamically informed perspective in understanding protein behavior and functionality.

The first part of the research detailed in chapter 3 introduces a new approach for protein model quality estimation. The method leverages explicit dynamic information, integrating protein stability features extracted from molecular dynamics simulation trajectories. Existing methods in this domain have primarily relied on static structure information. Model inaccuracies can significantly impact protein stability due to its structural changes over time in an in-silico system [63]. By incorporating protein structural stability information that has not been employed by any existing methods, this research offers a new useful information in model quality assessment. The use of short molecular dynamics simulations with simple parameters and short time frames further enhances the method's efficiency and practical applicability.

The second part of the research described in chapter 4 introduces a novel workflow that not only automates the prediction process but also specifically targets the challenges in handling charge-changing mutations. It addresses the challenges faced by existing machine learning-based methods in predicting protein stability changes due to point mutations. These methods often struggle with biased training sets predominantly consisting of neutral mutations and exhibit poor performance in highly stabilizing or destabilizing cases. Furthermore, their reliance on local, static information leads to a neglect of global and dynamic aspects of protein structures. The proposed method employs alchemical molecular dynamics simulations for free energy calculations, this approach provides a more accurate and comprehensive analysis compared to existing state-of-the-art methods.

1.7 Content of This Thesis

The chapter 2 delves into the broader field of protein stability prediction, providing an overview of existing research and its diverse applications in computational biology and protein engineering. This chapter begins by exploring the current landscape of protein stability prediction, highlighting the key challenges and advancements in the field. It then focuses on two specific applications of research related to the prediction of protein stability and stability changes. The chapter discusses the existing problems in these applications, particularly emphasizing the limitations of current methods that do not adequately incorporate dynamic information about proteins. Towards the end of the chapter, we further explain about molecular dynamics simulations as the main methodology that used in this research to encapsulate the dynamic aspects of proteins.

In chapter 3 we introduce a novel approach for estimating the quality of protein models by harnessing structural stability information gleaned from molecular dynamics simulations. This chapter details how MD simulations are utilized to obtain accurate protein stability information, which is

essential for evaluating the quality of protein models. The focus here is on demonstrating the importance of reliable stability assessment in ensuring the accuracy and usefulness of protein models in subsequent analyses and engineering efforts.

In chapter 4 we propose a novel workflow is proposed to enhance the alchemical method for predicting protein stability changes due to point mutations. Here we shift the focus to predicting stability changes in proteins upon point mutations. This chapter extends the discussion from inherent protein stability to the dynamic changes in stability resulting from point mutations. This chapter also includes a comparative analysis of the performance between supervised learning-based methods and our rigorous alchemical approach.

Chapter 5 discusses the potential extensions of the research, particularly focusing on the integration of machine learning methods with MD simulations to achieve optimal predictions of protein stability. This chapter explores how machine learning can enhance prediction computation time. It also discusses the further extension of the application that involves multiple mutations and mutation pathway reliability which is very important in the actual application.

Chapter 2

Prediction of protein stability and stability changes

2.1 Research overview

The study of protein stability and stability changes encompasses a broad spectrum of biological phenomena, from the macroscopic interactions within protein networks to the microscopic behavior of individual proteins, encompassing various aspects from the stability of protein complexes to the intricacies of individual protein folding and mutations. This diversity underscores the importance of protein stability in molecular biology, driving advancements in our understanding of biological systems and informing the development of new therapeutic strategies. The stability of protein complexes, including protein-protein and protein-ligand interactions, is fundamental to cellular function. These interactions are crucial for the formation and regulation of complex biological systems, and understanding their stability is key to deciphering cellular mechanisms. [71, 72, 73].

At the level of individual proteins, stability plays a critical role in maintaining structural integrity and functional efficacy. This is particularly evident in enzymatic activity and cellular signaling pathways, where the proper folding and stability of proteins are essential for their activity. Protein folding dynamics are central to this discussion, as correct folding is vital for protein function. Misfolding or instability can lead to aggregation and dysfunction, often associated with diseases. The impact of mutations on protein stability is another significant area of research. Point mutations can drastically alter protein stability, leading to changes in structure and function that are crucial in the development of genetic diseases and in therapeutic interventions. Research in this area has revealed how even minor alterations in amino acid sequences can have profound effects on protein stability and function. In addition to genetic factors, environmental conditions such as temperature, pH, and ionic strength play a significant role in protein stability. The response of proteins to these external factors is a key area of study, particularly in understanding how proteins maintain their stability under varying conditions. This has implications for biotechnological applications, where protein stability under different environmental conditions is critical. [74, 75]

Furthermore, the study of intrinsically disordered proteins (IDPs) has added a new dimension to our understanding of protein stability. IDPs lack a fixed three-dimensional structure but are functionally active, challenging the traditional view of the structure-function relationship in proteins [76]. Recent advancements in computational techniques, including machine learning algorithms and molecular dynamics simulations, have revolutionized the study of protein stability. These computational approaches, particularly machine learning and deep learning, have become state-of-the-art in predicting protein stability changes, complementing traditional methods and providing a more comprehensive understanding of protein dynamics.

In this research, the application of protein stability information is initially explored within the field of protein model quality assessment. The discussion then progresses to address protein stability changes, especially in the context of point mutations. The primary goal of this chapter is to provide an overview of applications in each research field, reviewing the current state-of-the-art methods in protein model quality assessment and the prediction of stability changes due to point mutations. The subsequent sections will delve into a detailed discussion of molecular dynamics simulation, the principal methodology used in this research, exploring its potential to provide a more comprehensive understanding of protein dynamics. This chapter also highlights the limitations of existing methods. A significant observation is the lack of utilization of dynamic information in these fields, despite its crucial role in protein stability. The utilization of dynamic information can offer valuable insights and enhance the performance of existing methodologies. Additionally, this chapter will discuss how recent advancements in computational techniques can be leveraged to better understand and predict protein stability changes.

2.2 Molecular dynamics simulation

In this section, we provide an overview of molecular dynamics simulation, as it is the primary methodology employed in this research. Molecular dynamics simulation is rooted in the principles of classical mechanics, offering a computational lens to scrutinize the dynamic behavior inherent to molecular systems. By employing Newton's equations of motion for each atom or particle, MD simulations present a real-time, temporal depiction of atomic and molecular trajectories. This dynamic view bridges the theoretical constructs with the tangible world, giving us a granular look into molecular activities as they unfold. Further, as computational power has grown, so has the resolution and depth of MD simulations. Not only do they visualize molecular activities, but they also enable predictions and provide explanations for phenomena otherwise elusive to direct observation. From fundamental research to applied sciences, the role of MD simulations in reshaping our molecular understanding cannot be understated.

2.2.1 Background

MD simulations have revolutionized our understanding of biomolecular dynamics, offering a window into the atomic-level interactions within proteins, nucleic acids, and other biological macromolecules. The historical development of MD simulations marks a journey from rudimentary models of simple molecular systems to sophisticated simulations of complex biological processes. This evolution has been propelled by significant advancements in computational power and algorithmic efficiency, enabling researchers to explore the dynamic nature of biomolecules with increasing accuracy and detail. At the heart of MD simulations lies the application of classical mechanics principles, particularly Newton's laws of motion. These simulations involve the iterative solution of motion equations for each atom in a system, allowing for the prediction of molecular behavior over time. This approach provides critical insights into the structural dynamics, interactions, and functional mechanisms of biomolecules under various conditions. [77] The ability to simulate these intricate processes has profound implications for understanding biological systems and developing therapeutic interventions.

2.2.2 Algorithm

The choice of algorithms in MD simulations is crucial for accurately integrating the equations of motion, which are fundamental to predicting the behavior of molecular systems over time. The Verlet algorithm, a cornerstone in MD simulations, exemplifies the balance between computational efficiency and numerical stability. This algorithm calculates the positions of particles at each time step based on their current positions and velocities, along with the forces acting upon them. Its simplicity and lack of velocity dependence make it highly efficient, especially for large systems. [78]

Variations of the Verlet algorithm, such as the Velocity Verlet and Leapfrog methods, offer additional benefits. The Velocity Verlet algorithm, for instance, is particularly advantageous for its ability to provide velocity information at each time step, which is essential for calculating kinetic properties and thermodynamic quantities. This algorithm is often preferred in simulations where energy conservation and detailed analysis of particle velocities are crucial. [79] Another important aspect of these algorithms is their impact on the stability of MD simulations. The choice of time step is a critical factor; too large a time step can lead to inaccuracies and instability, while too small a time step significantly increases computational demand. The algorithms must strike a balance, ensuring that the simulation remains stable over the desired time scale while maintaining computational feasibility. This balance is particularly challenging when simulating systems with a wide range of motion timescales, from fast-moving electrons to slower-moving larger molecules. Furthermore, the integration algorithms in MD simulations are continually evolving. Recent advancements aim to enhance the accuracy of simulations under extreme conditions, such as high pressure or temperature, and for complex systems involving multiple phases or interfaces.

2.2.3 Force Fields and Potential Energy Functions

Force fields play a pivotal role in MD simulations, dictating the accuracy with which molecular interactions are modeled. A force field is a collection of equations and parameters that describe the potential energy of a system as a function of the positions of its atoms. These equations typically include terms for bond stretching, angle bending, torsional interactions, and non-bonded interactions such as van der Waals forces and electrostatics. The choice of a force field depends on the system being studied and the specific interactions that need to be accurately represented. For instance, the CHARMM and AMBER force fields are widely used in protein simulations for their detailed treatment of biomolecular interactions. The development of more sophisticated force fields, incorporating polarizability and quantum mechanical effects, continues to be an area of active research, aiming to enhance the realism and predictive power of MD simulations. [80]

2.2.4 Time integration, stability, and computational considerations

Time integration is a critical aspect of MD simulations, involving the numerical integration of Newton's equations of motion to update the positions and velocities of atoms over time. The stability and accuracy of these simulations depend on the choice of the time step and the integration algorithm. A smaller time step leads to more accurate simulations but requires more computational resources. The stability of the simulation is also influenced by factors such as the stiffness of the system and the accuracy of the force field. Computational considerations, therefore, play a significant role in the design and execution of MD simulations. With advancements in high-performance computing and parallel processing, it has become feasible to conduct longer and more complex simulations, opening new avenues in biomolecular research.

2.2.5 General protocol of molecular dynamics simulation

MD simulations follow a structured protocol to accurately model the behavior of biomolecules. This process involves several key steps, each contributing to the overall fidelity and relevance of the simulation as shown in Figure 2.1.

The protocol of MD simulation involves several key steps, each contributing to the overall fidelity and relevance of the simulation. The following numbered list outlines these steps:

1. **Preparing protein structures for simulation:** The process begins with obtaining high-quality structural data, typically from experimental methods like X-ray crystallography or NMR spectroscopy. The protein structure is then optimized and energy-minimized to remove any steric clashes or unrealistic bond lengths and angles, ensuring the starting conformation is as close to its natural state as possible.

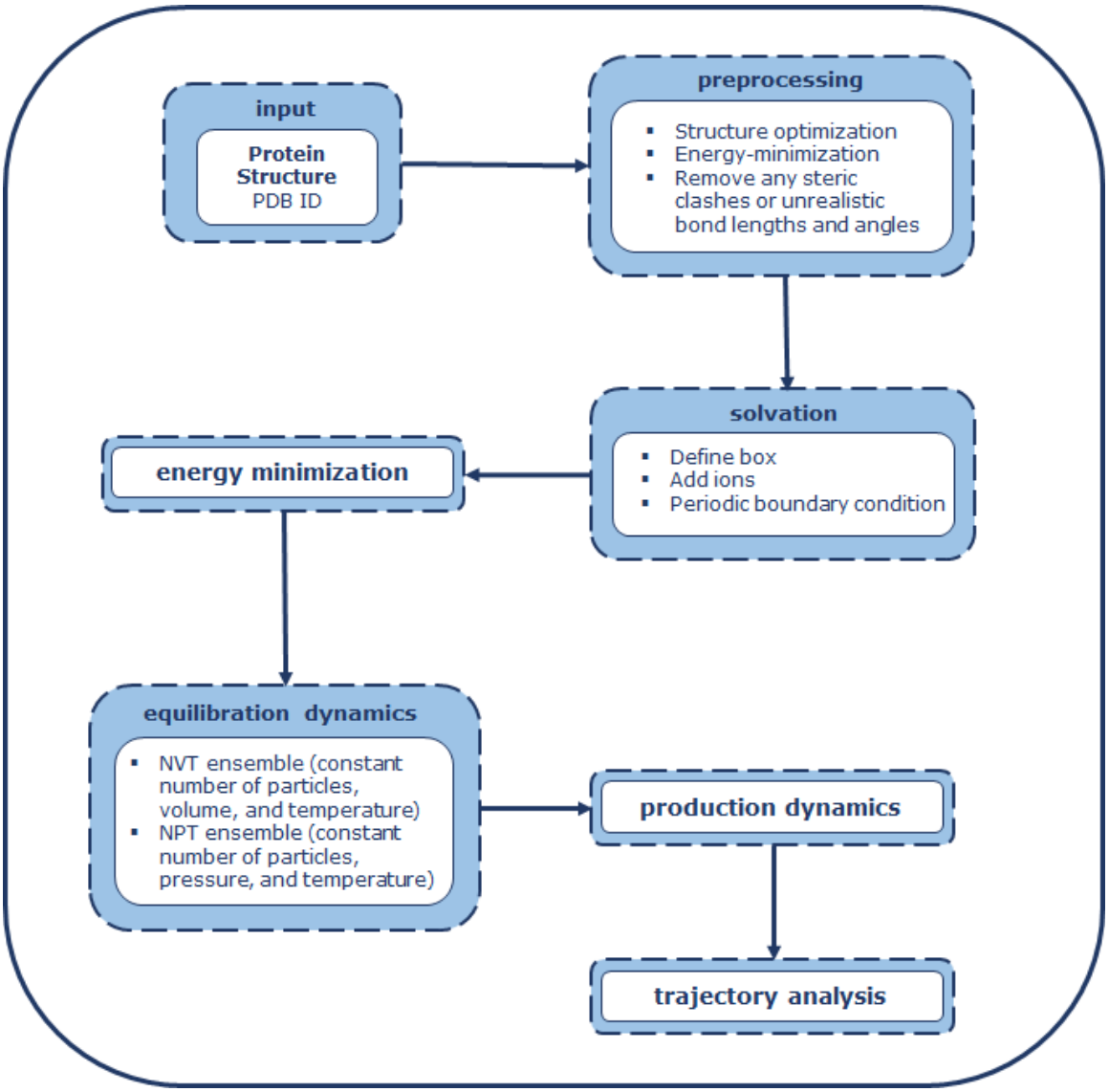


Figure 2.1: Flow chart depicting steps involved in MD simulation of a protein

2. **Solvation models, ionic conditions, and periodic boundary conditions:** The prepared protein structure is placed in a simulated environment that mimics its natural surroundings. This includes adding a solvation model, such as a water box, around the protein. Ionic conditions are replicated by adding counterions and salt ions to neutralize the system and establish physiological ionic strength. Periodic boundary conditions are applied to simulate an infinite system and avoid edge effects.

3. **Equilibration dynamics (NVT and NPT ensembles):** Before the production phase of the simulation, the system undergoes equilibration in both the NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) ensembles. This step ensures that the system reaches a stable state with respect to temperature and pressure, respectively.

4. **Production dynamics:** In this phase, the actual MD simulation is carried out, where the system evolves over time under the defined conditions. This step generates the trajectory data that will be analyzed in subsequent steps.

5. **Trajectory analysis:** The final step involves analyzing the trajectory data obtained from the production phase. This analysis can include studying various properties such as root-mean-square deviation, radius of gyration, hydrogen bonding patterns, and more, to understand the behavior and interactions of the biomolecules.

2.2.6 Molecular dynamics software and tools

Several software packages and tools are available for conducting MD simulations, each with its unique features and capabilities. Popular MD software includes GROMACS, AMBER, and NAMD, which offer a range of functionalities from basic simulations to advanced modeling techniques. These tools are continually updated to incorporate the latest advancements in computational biology, making them indispensable in the field of molecular dynamics. Choosing the right MD software depends on the specific requirements of the simulation. Factors such as system size, complexity, and the desired level of detail play a crucial role in this decision. The continuous development of these software tools, including improvements in algorithms and the introduction of new features, ensures that they remain at the forefront of molecular dynamics research.

GROMACS is renowned for its high performance and efficiency, especially in simulating large systems. It is optimized for speed and can handle a vast number of particles, making it a go-to choice for large-scale simulations. [81] AMBER, on the other hand, offers a comprehensive suite of tools for biomolecular simulations, with a particular focus on nucleic acids. Its force fields and simulation protocols are widely used in the study of DNA and RNA. [82] NAMD is designed for high-performance parallel computing, making it suitable for very large systems. It excels in scalability

and can be used effectively on both CPU- and GPU-based architectures. NAMD's versatility allows it to perform a wide range of simulations, from standard MD to more complex enhanced-sampling techniques. [83]

2.2.7 Data analysis and interpretation

Data analysis and interpretation in MD simulations are crucial for extracting meaningful insights from the complex simulation data. This process involves a variety of techniques and tools, each contributing to a deeper understanding of the biomolecular systems under study. The analysis transforms raw simulation data into a coherent narrative, revealing the dynamic behavior and interactions of molecules. It involves a meticulous examination of the trajectories and a careful interpretation of their significance in the context of biological function and molecular stability. This step is vital for bridging the gap between theoretical predictions and experimental observations.

Trajectory analysis is a fundamental aspect of MD simulations, where the trajectories representing the time evolution of molecular systems are examined. Techniques such as clustering, principal component analysis, and free-energy landscape mapping are employed to analyze these trajectories. The use of inherent structure analysis to discretize trajectories and construct free-energy surfaces can provide insights into the dynamic organization of biomolecular systems. [84] This approach helps in identifying key conformations and transition states, offering a deeper understanding of molecular mechanisms and interactions.

Analyzing key properties like root-mean-square deviation (RMSD) and energy profiles is crucial for understanding the stability and conformational changes of biomolecules. RMSD measures the deviation of a structure from a reference conformation over time, revealing information about the structural stability and flexibility of the molecule. Energy profile analysis is vital for understanding protein folding and ligand-binding processes. These analyses provide quantitative measures of the simulation's accuracy and the reliability of the predicted molecular behavior, making them indispensable in the validation of simulation results.

Visualization tools and techniques also play an integral role in MD data analysis. Software such as VMD and PyMOL offer capabilities for visualizing molecular structures, trajectories, and interactions. [85, 86] These tools facilitate the exploration of molecular dynamics, enabling researchers to gain insights into the mechanisms underlying biomolecular functions. Visualization not only aids in the interpretation of complex data but also allows for the communication of findings in a more accessible and understandable manner.

2.2.8 Applications of molecular simulation in protein stability research

MD simulations have become a fundamental tool in protein research, offering insights into various aspects of protein structure and function. These simulations enable the study of protein folding, dynamics, ligand-protein interactions, and much more, providing a detailed understanding of the molecular mechanisms underlying biological processes. The versatility of MD simulations allows researchers to explore a wide range of biological phenomena, from the behavior of individual proteins to complex interactions within larger biomolecular systems. This has led to significant advancements in our understanding of protein behavior at the molecular level, contributing to fields ranging from biochemistry to drug design.

Protein folding and dynamics are key areas where MD simulations have made significant contributions. By simulating the folding pathways of proteins, researchers can gain insights into the factors that influence protein stability and function. This understanding is crucial for the design of proteins with desired properties and for unraveling the mechanisms of protein-related diseases. MD simulations provide a dynamic view of the folding process, offering a detailed picture of how proteins attain their functional conformations. This knowledge is vital for understanding the fundamental principles of protein biochemistry and for developing strategies to manipulate protein behavior for therapeutic purposes. [87]

Ligand-protein interactions are another critical area explored through MD simulations. These interactions are central to understanding the function of proteins and are particularly important in drug discovery. MD simulations help in predicting the binding modes and affinities of ligands to proteins, thereby aiding in the identification of potential drug candidates. The ability to simulate these interactions with high accuracy has made MD an invaluable tool in the pharmaceutical industry, where it is used to screen and optimize drug candidates before they are tested in experimental settings.

MD simulations also play a vital role in the stability prediction of proteins. By simulating different environmental conditions and mutations, researchers can predict how these changes affect protein stability. This information is invaluable in protein engineering and in understanding the molecular basis of diseases caused by protein instability. The ability to accurately predict the effects of mutations on protein stability has implications for understanding genetic diseases and for designing proteins with enhanced stability or novel functions. [88]

Furthermore, the integration of MD simulations with experimental data has enhanced our understanding of proteins. This combination allows for the validation of simulation results and provides a more comprehensive picture of protein behavior. The synergy between computational and experimental approaches is key to advancing our knowledge in the field of protein research. It enables researchers to test hypotheses generated from simulations with experimental data, leading to more robust conclusions and a deeper understanding of complex biological systems.

2.2.9 Challenges and limitations

MD simulations have become a cornerstone in protein research, but they face significant challenges related to computational resources, accuracy, and system size. One of the primary challenges is the computational intensity required for large-scale simulations. Recent advancements have aimed to reduce the memory footprint and computational time, allowing for simulations of unprecedented system sizes. However, these improvements often require sophisticated computational infrastructure, which may not be accessible to all research groups. The need for high-performance computing facilities and the associated costs remain a significant barrier in the field. [89]

Real-time rendering of large MD simulation data sets is another challenge, especially when the data size is too large for conventional memory management. Advanced techniques are being developed to optimize rendering frame rates and handle large data sets effectively. However, these techniques require specialized knowledge in computer graphics and programming, adding another layer of complexity to MD simulations. The challenge lies in making these advanced rendering techniques more user-friendly and accessible to researchers who may not have a background in computational graphics.

The accuracy of MD simulations is also a critical concern. The use of multiple-time-step algorithms has been explored to save computational effort, but this approach involves a trade-off between computational efficiency and the accuracy of the simulations. Ensuring the reliability of simulation results while managing computational resources effectively remains a delicate balance. Furthermore, accurately parameterizing force fields to reflect real-world molecular interactions continues to be a challenging aspect of MD simulations. Additionally, the application of constraints to increase the time step in simulations can lead to potential side effects, limiting the applicability of this method. While constraints can help in managing the computational load, they may introduce artifacts or inaccuracies in the simulation results. Researchers must carefully consider these limitations when designing their simulations and interpreting the results. [90]

Future directions in MD simulation for protein research involve leveraging advancements in computational technology to overcome existing challenges. The development of parallel computing techniques and the utilization of high-performance computing power are key to managing the computational demands of large-scale simulations. These approaches enable the distribution of computational tasks across multiple processors, significantly reducing simulation times and allowing for more complex and larger system analyses. Additionally, the emergence of distributed cloud computing systems offers a scalable and accessible solution, providing researchers with the computational resources needed without the requirement of maintaining their own high-end computing infrastructure. As computational technology continues to evolve, it is expected that these advancements will greatly enhance the accuracy, efficiency, and applicability of MD simulations in protein research, opening new avenues for exploration and discovery.

2.2.10 Alchemical simulation

Alchemical simulations represent a sophisticated and versatile computational approach within the field of molecular dynamics, offering profound insights into the study of molecular interactions and transformations. The term of 'alchemical' refers to the theoretical transformation of one molecular into another within a computational framework. Alchemical simulations in molecular dynamics represent a sophisticated computational technique where virtual, stepwise transformations of molecular systems are performed. These transformations could involve changes in atomic types, bonding patterns, or electronic states, mirroring the metaphorical essence of ancient alchemy. The use of the term alchemical thus signifies the transformative aspect of these simulations, where one molecular entity is methodically and theoretically converted into another. This process is achieved through a series of intermediate states, allowing for a detailed exploration of the transition pathway and its energetic profile. The ability to perform such transformations in silico offers a powerful tool for probing molecular mechanisms that are challenging to study through direct experimental approaches.

The applications of alchemical simulations are extensive and diverse, encompassing a range from drug-receptor interactions, such as protein-protein interactions and the intricacies of ligand binding, to the complexities of protein folding and enzyme catalysis. One of the primary applications is in the field of protein folding and stability studies. By applying this theorem in alchemical simulations, researchers can accurately calculate the free energy changes associated with protein folding, unfolding, and mutations. This is particularly important in understanding the stability of proteins and their variants, which has implications in drug design and understanding disease mechanisms. The simulations can be broadly categorized into two types: equilibrium and non-equilibrium simulations, each with distinct methodologies and applications.

Equilibrium simulation

Equilibrium simulations allow the system to attain a steady state where macroscopic properties remain constant over time. These simulations ensure that a system reaches a steady state, where its macroscopic properties, like temperature and pressure, remain constant over time. This approach is crucial for accurately determining the free energy changes associated with molecular transformations. One prominent method within equilibrium simulations is Free Energy Perturbation (FEP). FEP involves the calculation of free energy differences between two states of a molecular system under equilibrium conditions, typically used for studying protein folding, unfolding, and mutations. FEP is based on the principle of equilibrium simulations. It calculates the free energy differences between a reference state and a target state by perturbing the system slightly and measuring the response in terms of free energy change. This method is particularly effective for detailed studies where precision in free energy calculations is important, though computationally intensive.

The FEP method operates on the principle of gradually perturbing a system from a reference state to a target state and measuring the resultant changes in free energy. This gradual perturbation allows for a detailed examination of the transition between states, making FEP a preferred method for studies requiring high precision in free energy calculations. For instance, the implementation of FEP in a molecular dynamics study typically follows a structured approach:

1. *State Selection:* The process begins with the selection of the initial and final states of the molecular system. These states could represent different conformational states, binding states, or compositions, such as the folded and unfolded forms of a protein, a protein with and without a bound ligand, or wild-type and mutant forms of a molecule.

2. *Intermediate State Generation:* A series of intermediate states are created to model the transition between the initial and final states. These intermediate states, often defined by varying the alchemical coupling parameter lambda (λ), form a bridge, allowing a gradual and controlled transformation of the system. The parameter λ smoothly interpolates between the initial and final states, enabling the system to transition through a series of intermediate configurations.

3. *Equilibration and Sampling:* Each intermediate state, defined by a specific λ value, is equilibrated, allowing the system to adequately explore its conformational space. Extensive sampling at these states is crucial for capturing accurate statistical data, which forms the backbone of reliable free energy calculations.

4. *Free Energy Calculation:* The essence of FEP lies in the calculation of free energy differences between successive intermediate states, each characterized by different λ values. By computing these differences and summing them up, the total free energy change associated with the system's transformation is determined. This calculation often involves statistical reweighting techniques to accurately estimate the free energy differences from the sampled distributions at each λ state.

FEP's reliance on equilibrium conditions renders it an invaluable tool in molecular dynamics. Its versatility allows it to be applied to a wide range of scenarios, providing detailed insights into the energetics of various molecular transformations. Whether investigating protein folding, ligand binding dynamics, or the effects of mutations, FEP offers a rigorous approach to understanding the changes in free energy that accompany these processes. FEP's reliance on prolonged equilibrium conditions makes it highly accurate for free energy calculations, but also computationally intensive, especially in large systems where extensive sampling over prolonged periods is necessary. This is particularly true when observing detailed protein behavior during state changes, as FEP requires extensive sampling over extended periods. In contrast, non-equilibrium simulations provide a different approach, particularly useful in scenarios where observing real-time conformational changes is less critical than rapidly obtaining free energy estimates.

Non-equilibrium simulation

Non-equilibrium simulations in molecular dynamics provide an efficient alternative to equilibrium methods. This efficiency is particularly beneficial when rapid molecular transformations are studied, and the detailed observation of conformational changes is not the primary focus. These simulations offer a time-efficient means to estimate free energy differences, especially in scenarios where molecular systems are large. This approach is particularly useful for studying molecular systems undergoing transformations that are either too fast to reach equilibrium or where the interest lies in the immediate response of the system to perturbations, rather than detailed state changes. Non-equilibrium simulations, therefore, focus on capturing snapshots of the system at various stages of transformation and quickly assessing the energetic implications. The methodology typically involves:

1. *Establishing Initial Conditions:* The process starts with a shorter equilibrium phase to establish initial conditions for the system. This phase sets the necessary baseline before any perturbation and typically involves equilibrating the system at an initial λ value.

2. *Capturing Snapshots for Non-Equilibrium Simulations:* The system's state is rapidly changed to capture snapshots at various stages of transformation. These snapshots, taken at different λ values, serve as starting points for subsequent non-equilibrium simulations. By varying λ, the system is transitioned between different states, capturing a range of configurations.

3. *Performing Non-Equilibrium Simulations:* Non-equilibrium simulations are then performed from these snapshots. In this phase, the system is rapidly perturbed to new states by altering λ over time, thus capturing the dynamic response of the system to these changes.

4. *Calculating Free Energy Differences:* The final step involves calculating the free energy differences based on the data gathered from these non-equilibrium simulations. The changes in λ across the simulations facilitate the computation of free energy differences, offering a quicker yet effective assessment of the system's energetics. This approach contrasts with the extensive sampling periods required in equilibrium methods like FEP.

Among the various methodologies employed, the Crooks Fluctuation Theorem stands out as a particularly revolutionary principle. It provides a quantitative relationship between the probability distributions of work done on a system during forward and reverse non-equilibrium processes. This theorem is particularly significant in alchemical simulations where it allows for the accurate calculation of free energy changes in molecular systems undergoing transformations. The theorem states that the ratio of the probability of observing a certain amount of work W in the forward process to the probability of observing the same amount of work in the reverse process is exponentially related to the work done and the difference in free energy between the initial and final states. Mathematically, it is expressed as:

$$\frac{P_{\text{Forward}}(W)}{P_{\text{Reverse}}(-W)} = e^{\beta(W-\Delta G)} \tag{2.1}$$

where $P_{\text{Forward}}(W)$ and $P_{\text{Reverse}}(-W)$ are the normalized probability distributions of work values obtained from the forward and reverse transformation paths and ΔG is the free energy difference.

For an actual example, in the context of mutations, the study typically involves both equilibrium and non-equilibrium simulations as depicted in Figure 2.2. In equilibrium simulations, the system is allowed to attain a steady state where macroscopic properties such as temperature and pressure remain constant. This equilibrium state serves as a critical reference point for assessing the impact of mutations in a controlled environment. On the other hand, non-equilibrium simulations involve inducing rapid changes to the system. These simulations are crucial for observing the system's immediate and dynamic response to sudden molecular transformations, such as those induced by mutations.

Compared with equilibrium methods, non-equilibrium simulations in molecular dynamics often require less computational cost but entail greater computational complexity. This complexity stems from several factors that impact both the accuracy of predictions and the computational burden. Firstly, thorough sampling of conformations at the initial and final states is essential for accurate predictions. Secondly, work values derived from non-equilibrium runs typically exhibit non-normal distributions, which affects the overlap of work distributions in forward and reverse directions; insufficient overlap can lead to inaccuracies in free energy calculations. Thirdly, the criteria for sufficient overlap, crucial for the reliability of the method, necessitate the use of ensembles. Moreover, when differences between the two physical endpoints are substantial, extensive sampling at these states and adequately long non-equilibrium transitions are required to ensure convergence of the predicted free energy changes. While a flexible protocol, adapting various parameters to the system under study, is recommended for both equilibrium and non-equilibrium approaches, a significant limitation of the latter is the inability to simply extend the length of non-equilibrium transitions. Increasing transition duration necessitates rerunning the entire set of transitions, which can hamper flexibility and prove inconvenient in large-scale applications. [91]

The choice between equilibrium and non-equilibrium approaches depends on the specific goals of the study, considering factors such as the desired level of detail, computational resources, system size, and the nature of the molecular transformations being investigated. For instance, in the domain of protein stability prediction upon mutation, especially when confronted with the necessity of conducting large mutational scan analyses, non-equilibrium simulations emerge as the preferred approach over equilibrium simulations. The inherent advantages of non-equilibrium methods are their computational efficiency and the ability to rapidly process extensive datasets—align closely with the demands of high-throughput screening endeavors. These methods enable the swift prediction

of changes in free energy ($\Delta\Delta G$), a critical parameter in assessing the impact of mutations on protein stability. This efficiency is indispensable in contexts where the primary goal is to identify mutations with significant effects from a vast pool of possibilities, rather than to delve into the intricate molecular dynamics characterizing each mutation. While it is recognized that equilibrium simulations offer more accuracy, the pragmatic benefits of non-equilibrium simulations are their speed and reduced computational cost which make them particularly advantageous for studies aiming to balance between mutational coverage within large systems. Thus, in scenarios prioritizing the rapid assessment of protein stability across numerous mutations, non-equilibrium simulations is more preferable as the method of choice.

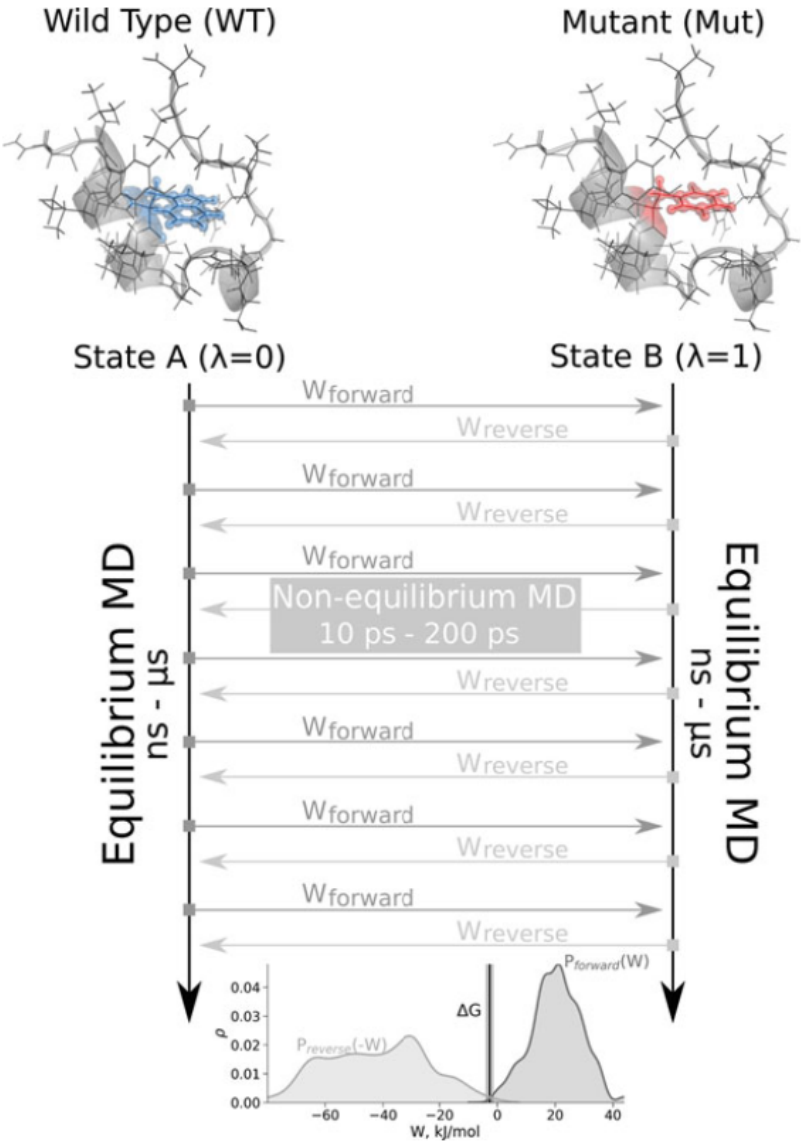


Figure 2.2: Illustration of the alchemical simulation in the case of mutation. Reproduced with permission from Springer Nature [92] The procedure involves two distinct equilibrium simulations, one representing the wild-type (WT) protein at $\lambda = 0$ and the other depicting the mutant form at $\lambda = 1$. These simulations are conducted over a range from nanoseconds to microseconds to ensure comprehensive sampling of the end state ensembles, which is critical for the accuracy of the free energy estimation. From these simulations, specific snapshots are chosen to initiate rapid transition simulations (spanning 10–200 ps) that drive the system in both forward ($\lambda : 0 \rightarrow 1$) and reverse ($\lambda : 1 \rightarrow 0$) directions. The work values associated with these transitions are recorded, and the Crooks Fluctuation Theorem is subsequently applied to determine the free energy difference between the wild-type and mutant states

2.3 Protein stability prediction

Historically, the endeavor to understand protein stability was primarily confined to experimental methodologies, ranging from X-ray crystallography to nuclear magnetic resonance spectroscopy. [93, 94] These techniques, though revolutionary, were limited by their scalability, high resource demand, and the slow pace at which data could be generated and analyzed. The advent of computational biology in the late 20th century marked a significant paradigm shift, offering new avenues to predict protein stability through theoretical models and simulations. [1]

The protein stability prediction embedded within the broader tapestry of molecular biology and biochemistry, mirrors the scientific community's quest to unravel the complexities of protein function and its determinants. This narrative begins in the mid-20th century, with the pioneering work of Christian Anfinsen laying the conceptual foundation through his seminal experiments on ribonuclease A. [47] Anfinsen's dogma, positing that the primary sequence of a protein dictates its three-dimensional structure and by extension, its stability has catalyzed decades of research aimed at deciphering the intricate mechanisms governing protein folding and stability. [95, 9]

The foundation rests on Anfinsen's dogma, positing that the native structure of a protein, characterized by its most stable conformation, corresponds to the global minimum on the energy landscape. Proteins fold through multiple intermediate states, navigating through a funnel-shaped energy landscape toward the native structure. This funnel represents not a single, deterministic pathway but a multitude of routes that guide the protein to its lowest energy state. Historically, the quest to understand protein folding and predict stability involved molecular simulations, an approach that, while groundbreaking, was significantly hampered by the computational limitations of the era. [96, 97] These simulations sought to explore the vast conformational space to identify the most stable structures, a task that is computationally intensive due to the sheer number of possible configurations and interactions to consider. [98, 99]

The development of computational approaches to protein stability prediction can be seen as a response to the growing need for scalable, efficient methods to complement traditional experimental techniques. Early computational models were primarily focused on understanding the thermodynamic properties of proteins, using physical principles to predict how changes in sequence or environment might influence protein stability. These efforts were greatly enhanced by the increasing availability of protein structure databases in the 1990s, such as the Protein Data Bank (PDB), which provided a wealth of data for computational analysis. [100]

The advent of sophisticated computational techniques, notably those leveraging machine learning, has further revolutionized the field. These methods can now predict protein structures with high accuracy, identifying candidates for the most stable conformations without the need for exhaustive simulation of the folding process. Following prediction, model quality assessment tools are employed to select the most stable structures from among these candidates, ensuring that predictions align

with biophysical realities. This evolution in approach has been facilitated by Critical Assessment of protein Structure Prediction (CASP) [69] and technological advancements, allowing for rapid progress in computational biology.

As a result of these advancements, the focus of protein stability prediction has broadened significantly. Researchers are no longer confined to studying the folding of individual proteins in isolation. The field focus progressed and expanded to encompass a wider understanding of the factors influencing protein stability, including but not limited to, the effects of mutations and protein-ligand interactions. [42, 101] Researchers began exploring how post-translational modifications, intramolecular forces, protein folding pathways, and the cellular environment contribute to the stability landscape of proteins. [102] This broader perspective recognized the multifaceted nature of protein stability, influenced by a complex interplay of biochemical and biophysical factors.

Advancements in algorithmic development, particularly in the area of machine learning and artificial intelligence, have been instrumental in facilitating this expanded focus. These computational tools, capable of handling vast data sets and extracting meaningful patterns, have significantly enhanced our ability to predict stability across a diverse range of proteins and conditions. [103] Machine learning models, trained on data from both experimental observations and theoretical predictions, have become adept at forecasting stability changes, thereby broadening the scope of protein stability research beyond traditional confines. [104]

The integration of computational predictions with experimental data has indeed become a defining characteristic of modern protein stability research. This interdisciplinary approach not only enhances the predictive power of computational models but also provides a feedback mechanism to refine experimental designs based on computational insights. The symbiotic relationship between computational and experimental methodologies has propelled the field forward, enabling a more nuanced understanding of protein stability and its implications for biotechnology, pharmaceutical research, and the elucidation of disease mechanisms.

2.3.1 Prediction target

The field now encompasses a wide range of applications predicting the thermodynamic properties of proteins such as changes in Gibbs free energy (ΔG) and melting temperatures (T_m). These developments have opened new avenues for exploring the complex interplay of forces that govern protein stability and function, marking a new era in the study of proteins and their roles in biological systems. Reflecting on the historical development of protein stability prediction, it is evident that the field has grown to encompass a diverse range of approaches, each contributing to our comprehensive understanding of protein behavior. This journey, from Anfinsen's foundational work to the sophisticated computational models of today, illustrates the continuous evolution and integration of multiple research domains within protein stability prediction.

In the domain of drug discovery and development, computational approaches play a pivotal role in identifying and optimizing drug-protein interactions. Central to this endeavor is the quantitative assessment of stability changes induced by potential drug compounds, specifically, the change in Gibbs free energy, represented as $\Delta\Delta G_{binding}$. [105] This metric is crucial for designing molecules with optimal binding properties that modulate protein function effectively. By leveraging computational models that integrate structural data and dynamic simulations, researchers can predict how binding events affect the stability of target proteins. [106] The output, quantified as changes in binding free energy, guides the selection and optimization of lead compounds, ensuring their efficacy and specificity.

In the study of protein-protein interaction networks, understanding the stability dynamics within these complexes is essential for elucidating cellular processes and signal transduction pathways. [107] Computational analyses aim to quantify the energetics of protein complexes $\Delta G_{binding}$, providing insights into the stability contributions of individual interactions. This quantitative approach allows researchers to map the stability landscape of protein assemblies, offering a deeper understanding of cellular machinery and identifying potential therapeutic targets within these networks.

In the realm of thermal stability prediction, the focus is on estimating the melting temperature (T_m) of proteins. This parameter is essential for understanding how proteins behave under various temperature conditions, a factor that is crucial for biotechnological processes and the formulation of biopharmaceuticals. [108] Accurate predictions of T_m allow researchers to identify proteins that maintain their functionality at different temperatures, enabling the development of stable therapeutic proteins and industrially relevant enzymes.

Allosteric regulation prediction aims to understand how modifications or ligand binding at sites distant from the active site influence protein stability and function. [109] This analysis is pivotal for the design of allosteric modulators, offering a strategy to regulate protein activity in a nuanced manner. Predictive models assess the impact of allosteric interactions on protein stability, quantifying changes in stability ($\Delta\Delta G_{binding}$) or activity that results from these interactions. [110] Such predictions inform on the potential of allosteric sites as therapeutic targets, guiding the design of modulators that can either enhance or inhibit protein function with high specificity.

Mutation impact analysis focuses on assessing the effects of natural or induced mutations on protein stability. This analysis is vital for unraveling the molecular underpinnings of diseases, guiding therapeutic interventions, and shaping genetic engineering efforts. [111] By quantifying the stability change due to mutations ($\Delta\Delta G$), computational models provide insights into how alterations in amino acid sequences influence the structural integrity and functionality of proteins. This quantitative metric serves as a cornerstone for predicting the phenotypic consequences of mutations, facilitating the identification of potential targets for drug development and the design of proteins with enhanced properties. [112]

Biophysical property prediction extends the scope of stability predictions by integrating them with assessments of other critical biophysical properties, such as disorder regions and binding affinities. [113, 114] This holistic approach enables a more comprehensive understanding of protein behavior, correlating stability with functional attributes that influence protein interactions and activity. By leveraging computational models to predict these properties, researchers can gain insights into the complex relationship between protein stability, solubility, and interaction dynamics, informing on the protein's role in cellular processes and its suitability for therapeutic applications.

The field of structural bioinformatics plays a pivotal role in advancing protein stability prediction by developing and refining algorithms that predict stability based on three-dimensional structures. Incorporating factors like folding energy landscapes and conformational dynamics, these algorithms enhance our ability to model the energetics of protein folding and stability. [115] The development of such algorithms not only improves the accuracy of stability predictions but also contributes to our understanding of the principles governing protein structure and function, facilitating the design of proteins with desired characteristics.

Model quality assessment (MQA) for stability prediction underscores the importance of ensuring the reliability of computational models used in stability predictions. Developing metrics and validation tools to assess the quality of these models is essential for guaranteeing that predictions are based on accurate, the most stable, and reliable structural information [48]. This aspect of protein stability research is critical for bridging the gap between computational predictions and experimental validation, ensuring that the insights derived from computational models are reflective of true protein behavior and can be confidently applied in scientific research and therapeutic development [116].

The significance of MQA extends beyond the foundational assessment of model accuracy; it serves as a crucial prerequisite for all subsequent computational analyses aimed at understanding protein behavior. Reliable protein structures, validated through MQA, form the basis for exploring a wide range of biological phenomena, for instance in the context of the effects of mutations, the dynamics of protein-protein interactions, and the protein-ligand binding. These analyses hinge on the initial confidence in the protein model's fidelity, as inaccuracies in the structural representation can lead to misleading conclusions about a protein's stability and function.

As what existing works have shown for instance, a comprehensive QMEAN Z-score analysis of all experimental structures in the PDB, which found that membrane proteins accumulate on one side of the score spectrum and thermostable proteins on the other, underlines the significance of the QMEAN Z-score as an estimate of protein stability. Proteins from the thermophilic organism *Thermatoga maritima* received significantly higher QMEAN Z-scores in a pairwise comparison with their homologous mesophilic counterparts, highlighting the utility of QMEAN scores not only in assessing model quality but also in indirectly estimating protein stability [49, 50].

By ensuring that computational models accurately reflect the protein's native state, MQA enables researchers to confidently extend their investigations into these critical areas, thereby enriching our comprehension of protein mechanisms and facilitating the development of targeted therapeutic strategies. This layered importance of MQA underscores its pivotal role in the broader context of protein research, emphasizing that reliable structure prediction is not merely the first step but a continuous necessity for meaningful computational biology and biochemistry studies.

The integration of model quality assessment into protein stability prediction is a testament to the field's commitment to accuracy and reliability. Given Anfinsen's dogma that the native structure is the most stable, MQA's role in selecting the best model closest to the native structure is crucial. It ensures that the selected model likely represents the most stable conformation, providing a reliable basis for predicting protein stability. This connection between MQA and stability prediction is not just methodological but foundational, reinforcing the need for precise and accurate computational models to advance our understanding of protein stability and function.

2.4 Protein stability: application in protein model quality assessment

Protein stability is a critical factor in determining the accuracy and reliability of protein structure models. This sub-chapter delves into the essential role of protein stability information in the field of protein model quality assessment (MQA) research. We begin with an overview of MQA, highlighting its significance in computational biology and structural bioinformatics. This overview lays the groundwork for a detailed examination of existing MQA methods, focusing on their strengths, limitations, and particularly the challenges they face due to the lack of utilization of dynamic information related to protein stability.

2.4.1 Protein model quality assessment

The three-dimensional (3D) structure of proteins is pivotal for understanding their functional mechanisms and plays a crucial role in drug discovery. [117, 118] Typically determined through experimental methods like X-ray crystallography, NMR spectroscopy, and electron microscopy, these techniques, while effective, are often costly and time-intensive. [119] To address these challenges, computational approaches have been developed to predict protein 3D structures from primary sequence data. Comparative modeling, for instance, identifies homologous known protein structures and aligns the query sequence to these templates. In cases where no homologous sequences are found, de novo modeling is employed, utilizing general principles of folding and energetics. Despite advancements, current computational schemes often generate multiple structure models, necessitating a method to identify the most accurate model that closely resembles the native structure. This process is known as model quality assessment. The importance of MQA lies in its ability to determine the reliability of

protein models, which are essential for understanding protein function, interactions, and for guiding further experimental or computational studies. The evolution of MQA has been marked by a significant shift from traditional comparative methods to the incorporation of sophisticated computational techniques.

Key metrics used in MQA include the Global Distance Test - Total Score (GDT_TS) and the Local Distance Difference Test (lDDT). GDT_TS measures the similarity between the predicted model and the native structure by calculating the average percentage of residues within a certain cutoff distance. lDDT, on the other hand, evaluates local conformational accuracy by comparing distances between atoms in the model and the native structure. Both GDT_TS and lDDT have become standard benchmarks in MQA, providing comprehensive assessments of model accuracy. These metrics are crucial in evaluating the performance of computational methods in accurately predicting protein structures, and they play a significant role in guiding the development of more advanced and reliable modeling algorithms.

2.4.2 Related work

Initially, MQA relied on basic structural comparisons, where predicted protein models were evaluated against known structures using fundamental physicochemical principles. Traditional MQA methods relied on statistical potential scoring functions like DFIRE, [41] DOPE, [120] GOAP, [121] and RWplus. [122] These early methods laid the groundwork for understanding protein structure and function. As computational biology advanced, it introduced more sophisticated methodologies, enhancing the accuracy and efficiency of model assessment. This transition was driven by the increasing complexity of protein structures and the growing volume of protein models generated by various prediction techniques. Concurrently, the development and utilization of benchmark data sets, notably those from the Critical Assessment of Protein Structure Prediction (CASP) competitions (`https://predictioncenter.org/index.cgi`), have been pivotal in the evolution of MQA methods. CASP data sets have provided a platform for systematic evaluation and comparison of various assessment techniques, fostering an environment conducive to both competitive and collaborative advancements in the field. The CASP competitions have underscored the importance of accuracy in model predictions, pushing the boundaries of computational methods in protein structure prediction. Over the years, the field has seen a trend towards machine learning and data-driven approaches, leveraging the power of extensive data sets to learn complex patterns and relationships within protein structures. This shift towards computational and data-centric methods reflects a broader movement within bioinformatics, harnessing the capabilities of modern computational tools to navigate the complexities of biological data. In the latest CASP14 competition, the top five machine learning-based methods attained an average top 1 GDT_TS score of 8.36 and an lDDT score of 5.59, respectively. [123]

Machine learning-based methods typically extract features from protein structures and sequences, employing supervised learning to predict model quality. Examples include ProQ2, [124] which combines structural and evolutionary information in a support vector machine, and ProQ3, [125] which adds Rosetta energy terms to a similar approach. The advent of deep learning has further elevated MQA methods. ProQ3D [126] and DeepQA [127] utilize high-level features from other methods, feeding them into a multilayer perceptron. Variations of convolutional neural networks (CNNs), such as 3DCNN [128, 129, 130, 131] and graph CNN, [132, 133] have been employed to learn low-level structural information. DeepAccNet [134] and QDeep [135] represent further innovations, estimating per-residue accuracy and combining inter-residue distance with multiple sequence alignment information, respectively.

QDeep

Typically, machine learning-based methods begin by extracting features, either from a sequence usually through multiple sequence alignment and/or from structural features. These extracted features are then fed into the machine-learning model. The final output is generally a predicted MQA score, which is used to select the best model. The workflow for the prediction process used in machine learning methods is illustrated in Figure 2.3. Here. we use the method used by QDeep as an example. QDeep is amongst of the top 5 machine learning-based method in CASP14 competition.

QDeep starts with generating a multiple sequence alignment (MSA), essential for predicting inter-residue distances and other sequence-based features like secondary structure and solvent accessibility. This framework then derives 23 features for each residue in a model, including distance-based weighted histogram alignment, sequence versus structure consistency, and ROSETTA centroid energy terms. The feature later used as the input to be fed to the deep residual network (ResNet). The architecture features stacked deep ResNet classifiers tailored to 1, 2, 4, and 8Å error thresholds. Each classifier within this network consists of 13 residual blocks, each containing three 1-dimensional convolutional layers, with the output channels for these blocks set at 128, 256, and 512, respectively. Following the residual blocks, an average pooling layer is employed to reduce the number of parameters, thus facilitating faster computation. This layer is followed by a flatten layer that transforms the pooled feature map into a 1-dimensional vector, which is then input into the fully connected (dense) layer. The final stage involves residue-level ensemble classifications and their combination for model quality estimation. In QDeep, each of the four deep ResNet models acts as an independent residue-specific error classifier. Their ensemble collectively assesses the structural quality of a model. At each classifier's output layer, the sigmoid activation function predicts the likelihood of residue-level errors at 1, 2, 4, and 8Å error thresholds, combining the ensemble of residue-level classifiers using the QDeep score to evaluate the model's quality.

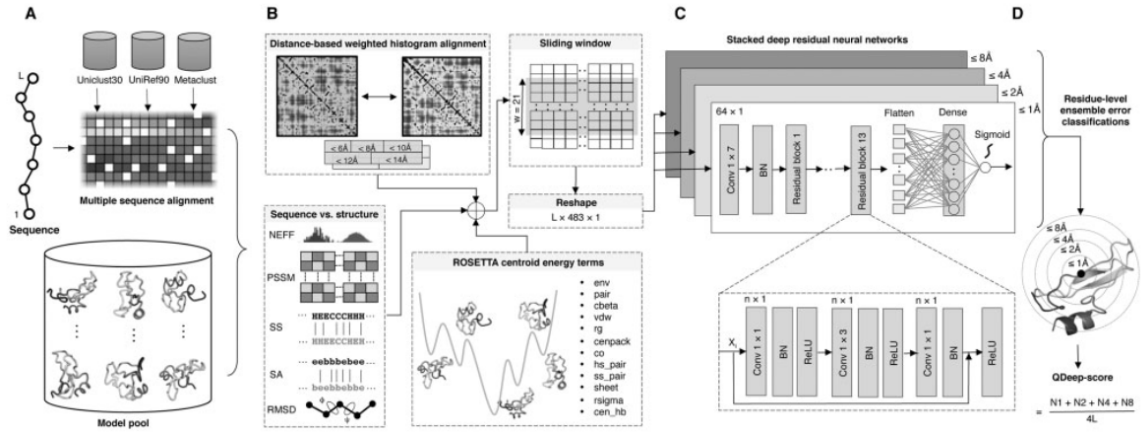


Figure 2.3: The flowchart of QDeep. Reproduced with permission from [135]. Copyright 2020 Oxford University Press. (A) Multiple sequence alignment generation (B) Feature extraction (C) deep learning model (D) Residue-level ensemble error classifications and prediction of MQA score

2.4.3 Problem of existing methods

The advancements in MQA have predominantly focused on extracting static structure information, such as geometric, sequence, and energetic features. [136] This approach, while providing valuable insights, overlooks a fundamental aspect of proteins: their dynamic and flexible nature. Proteins are not static entities; they undergo conformational changes influenced by environmental factors, which are crucial for understanding their stability. These changes, occurring over various scales, significantly impact the protein's functional state. High-quality structures or accurate models are often characterized by their ability to maintain stability, preserving their equilibrium shape under varying conditions. Conversely, low-quality structures are more likely to exhibit significant deviations and instability. [63] This correlation between structural deviation and model quality loss highlights a critical gap in current MQA methodologies: the lack of consideration for the dynamic nature of proteins. The static models used in these methods fail to capture the fluctuating conformational states of proteins, leading to potential inaccuracies in stability predictions.

To address the limitations, a more dynamic approach is required. MD simulation techniques present a promising solution. MD simulations can provide a detailed view of protein movements and conformational changes over time, offering a more comprehensive understanding of protein stability. The later sub-chapter 2.2 will delve into the intricacies of MD simulations, discussing how they can be effectively applied to enhance MQA. Furthermore, we introduce a novel method in Chapter 3 that incorporates explicit dynamic information derived from protein stability studies. The proposed method aims to bridge the gap in current MQA approaches by integrating dynamic stability data, thus providing new useful information and realistic representation of protein behavior.

2.5 Protein Stability Changes: prediction upon point mutation

Transitioning from the protein stability field, we further extends our research to the protein stability changes, particularly those resulting from point mutations. Point mutations significantly influence protein behavior, affecting both structure and function. The section begins by contextualizing point mutations within the broader landscape of protein stability changes. It then critically examines the current state-of-the-art prediction methods, highlighting the challenges in accurately predicting the effects of mutations.

2.5.1 Point mutation

Point mutations, involving single amino acid substitutions, are critical modifiers of protein stability and function. These mutations can lead to significant changes in the protein's structure, often resulting in altered biological activity. The implications of point mutations are far-reaching, as they are associated with a wide array of diseases, including cancer, neurodegenerative disorders, and various genetic conditions. [137, 138, 139, 140] The study of these mutations is crucial for understanding the molecular basis of these diseases and for developing targeted therapeutic interventions.

Experimentally, protein stability in the context of point mutations is often assessed by measuring the folding free energy change ($\Delta\Delta G$). This measurement provides insights into how a mutation affects the protein's tendency to maintain its native conformation. However, experimental approaches, while accurate, face several challenges. They are typically time-consuming, costly, and require substantial amounts of pure protein, which may not always be feasible, especially for rare or difficult-to-express proteins.

In response to these challenges, computational methods have gained prominence as a viable alternative for predicting the $\Delta\Delta G$ value of point mutations. These methods offer a faster, more cost-effective means of assessing mutation impacts on protein stability. They leverage various computational techniques, ranging from empirical and statistical models to machine learning algorithms, to predict how a single amino acid change might alter the protein's stability. [141, 142]

2.5.2 Related Work

Computational prediction methods for protein stability have evolved from mechanistic models to advanced supervised learning techniques. [143] These methods often use simplified models treating protein structures as rigid, relying on statistical potentials or empirical energy functions for direct prediction [144, 145, 146] or as inputs for machine learning algorithms to estimate $\Delta\Delta G$ values. Machine learning-based methods typically involve training algorithms on large data sets of known protein structures and their stability characteristics. These methods use a variety of features extracted

from protein sequences, such as amino acid composition, physicochemical properties, and evolutionary information derived from multiple sequence alignments. [147, 148, 149, 150, 151, 152, 153, 154] Additionally, structural features like solvent accessibility, secondary structure elements, and inter-residue contacts are also utilized. [155, 156, 157, 158, 53, 159, 160, 161]

Advanced techniques like deep learning have enabled the integration of these diverse data types, leading to more accurate and robust predictions. Deep learning models, such as convolutional neural networks and recurrent neural networks, have been particularly effective in capturing complex patterns in protein sequences and structures. These models can learn hierarchical feature representations, which are crucial for understanding the intricate relationship between protein sequence, structure, and stability. Ensemble methods, which combine predictions from multiple models, further enhance the accuracy by leveraging the strengths of different algorithms and reducing the likelihood of overfitting to specific data set characteristics. [162, 163] However, these ensemble methods often rely on the performance of machine learning-based methods or use supervised learning for the ensembling. Thus, they can also be included as machine learning-based methods.

Another approach, the rigorous approach, involves explicit calculations of protein structure and energetics, such as free energy calculations based on molecular dynamics simulations or Monte Carlo simulations. This approach provides a detailed understanding of the underlying mechanisms of protein stability and can capture the effects of nonlocal interactions and context-dependent effects of mutations. [164, 165] Alchemical free energy calculations based on molecular dynamics simulations are a rigorous method that is widely used to calculate the relative free energy differences between related molecular systems. [92] This method is particularly useful for studying small localized perturbations in molecular systems, such as point mutations or different protonation states of a ligand. It can provide valuable insights into the energetic and structural changes that occur during a transformation, allowing for the identification of key interactions and residues that contribute to the overall stability. Alchemical methods involve transforming one amino acid residue into another (i.e., from wild to mutant variants) while keeping the rest of the system unchanged. The free energy difference between the two states then can be computed by using MD simulations and statistical mechanics, providing a quantitative measure of the effect of a mutation on protein stability. Numerous studies have demonstrated the consistency between the predicted effects of mutations obtained from alchemical methods and experimental results in various fields. [166, 167, 168]. One of examples is free energy perturbation plus (FEP+), a commercial software initially designed for protein-ligand interaction predictions, has also been adapted for protein stability prediction upon point mutation. [169] They used increasingly multiple longer peptides with native sequences and conformations as the unfolded mutant structure representation which caused inconsistency in their performance.

pmx

De Groot et al. were one of the first to investigate the effect of mutations on protein folding $\Delta\Delta G$ using alchemical methods [92, 170, 171]. They developed the *pmx* package, which automates the generation of hybrid protein structures and topologies by utilizing force field-specific pregenerated mutation libraries. The group also suggested a double-system/single-box approach that involves two protein systems within a single simulation box to preserve the overall charge of the system while undergoing alchemical transformation [170, 172]. For the free energy difference calculation, they applied a fast-growth thermodynamic approach with Crooks Fluctuation Theorem [173] using the maximum likelihood estimator. While their performance validations revealed larger fluctuations around the expected outcome value for charge-changing mutations compared to charge-conserving mutations [170]. The group later also performed a large-scale mutation scan of 143 unique mutations in barnase and S. nuclease [174]. Despite showing good mean unsigned error for conserving mutation, the results were significantly worse for charge-mutation cases.

2.5.3 Problem of Existing Methods

Despite majorly being utilized and becoming state-of-the-art, machine learning-based methods have not shown significant improvement over the past decade. Recent reviews suggest that the upper limit of prediction accuracy, around 1 kcal/mol, has nearly been reached. [175, 75] Even with additional experimental data, substantial improvements in performance are unlikely. Contrary to this theoretical upper limit, current machine learning methods still demonstrate relatively low performance, especially when tested on challenging data sets like frataxin and p53, with high prediction errors and no notable improvements. [154] This may stem from a heavy reliance on training data sets biased towards neutral mutations and principally were trained using local, static protein information. Sequence-based predictors fail to capture structural change effects, while structure-based predictors often overlook mutant structure information, focusing only on local aspects like neighboring residues of wild-type structure information.

Rigorous approaches, such as molecular dynamics simulations-based methods, can address these limitations by considering global and dynamic effects of mutations. However, these methods are computationally intensive and requiring both wild-type and mutant structures. Nevertheless, advances in computing power and GPU parallelization have significantly improved the efficiency and accuracy of these rigorous free energy calculation methods. [176, 177, 178] The advent of accurate 3D structure prediction models like AlphaFold2 [104] has enabled the generation of high-quality protein structure models, comparable to experimental data, useful when structural information is unavailable.

Current alchemical methods often use simplified models like capped n-peptides, for instance, the *pmx* package employs glycine-capped tripeptides to represent mutant structures. [179, 180] While

it showed good performance in conserving mutations, their effectiveness in charge-changing mutation cases is limited. This is because charge-changing mutations can significantly impact protein stability by altering electrostatic interactions, hydrogen bonding networks, and solvation effects. These effects may be challenging for simplified peptide models to accurately capture, as they do not fully represent the native environment of full-length proteins. Additionally, the determination of the peptide number can often be arbitrary. To address this issue, it might be necessary to incorporate the whole structure information on the mutant to enhance the prediction of charge-changing mutations. While AlphaFold2 offers the potential to predict mutant structures, it typically outputs these structures in a folded state. Conversely, alchemical methods necessitate the mutant structure to be presented in an unfolded state. Some studies also suggest that AlphaFold2 may not accurately predict the impact of point mutations on the mutant structure because it relies on existing structures in the PDB database rather than the underlying principles governing protein folding. [181, 182] To overcome this limitation, Buel and Walters suggest integrating AlphaFold2 with current and future developments in molecular dynamics simulations for protein structure. [181] Recent findings also suggest that AlphaFold2-predicted structures provide valuable information about protein dynamics, comparable to results from molecular dynamics simulations. [183] Thus, further extension might reveal the hidden potential of the AlphaFold2 structure. One possibility is unfolding the predicted structure through MD simulation.

To address the limitations, we propose a novel workflow of an alchemical free-energy calculation method based on *pmx* and utilizing AlphaFold2-predicted structure. The proposed method can handle charge-changing mutation cases and improve the accuracy. We directly compare the performance of this method with existing state-of-the-art methods on challenging data sets: frataxin and p53. The comprehensive details of this approach, including its methodology and comparative analysis, are elaborated in Chapter 4.

Chapter 3

Protein stability information for protein model quality assessment

Protein model quality assessment (MQA) is the process of evaluating the accuracy and reliability of computationally predicted protein structures, ensuring they align with actual biological structures and functional characteristics. Existing methods in MQA research predominantly focus on static structural features, such as geometric, sequence, and energetic attributes. However, this approach overlooks the dynamic nature of proteins, which is crucial for a comprehensive understanding of their stability and function. Proteins are inherently dynamic entities, undergoing conformational changes influenced by various environmental factors. These dynamic changes play a critical role in determining the quality of protein models. Recognizing this gap in existing MQA methods, this chapter introduces a novel approach for estimating the quality of protein models by leveraging structural stability information derived from MD simulations.

Central to this approach are three key protein stability features: the root-mean-square deviation (RMSD), the fraction of secondary structure, and the fraction of native contacts relative to the initial structure. The main hypothesis of this method is that the quality of a predicted protein structure is intrinsically linked to its stability in silico systems, as evidenced by its deviation from the original structure. Insights gleaned from structural stability are proposed to introduce new and useful information and enhance the accuracy of MQA. This approach is particularly advantageous in scenarios where comprehensive training data or sequence homology is lacking, offering a more accurate and realistic representation of protein behavior. By employing a basic feature set extracted from standardized MD simulations, this methodology demonstrates efficacy comparable to more advanced methods that rely on extensive CASP data sets and complex deep learning architectures.

The novel contribution of this chapter lies in its pioneering integration of protein structural stability information into the field of MQA tasks. This integration represents a significant advancement in the field, offering a new perspective and methodology for assessing the quality of protein models. It paves the way for more accurate predictions in protein structure analysis, particularly in cases

where existing methods fall short, and is designed to be accessible to researchers who may not have extensive experience with MD simulations.

3.1 Materials and methods

The proposed method consists of three steps: protein trajectory generation through MD simulation, feature extraction, and best model selection (Figure 3.1).

3.1.1 Data sets

CASP data sets are one of the standard benchmark data sets broadly used to evaluate MQA method performance. [184] The data sets are updated every 2 years and can be accessed through the CASP website `https://predictioncenter.org/download_area/`. A single CASP data set contains numerous protein pools. Each pool consists of multiple predicted structure models from different structure prediction methods (Figure 3.1A). MQA methods generally train and evaluate using multiple CASP data sets, including ten to a hundred thousand predicted structures. However, due to the different approaches of our method and the computational cost limitation of MD simulation, we only chose sample pools from a single CASP. In this work, we selected the test data from QDeep, which consisted of 20 protein pools with various protein sizes from CASP13 stage 2. Each pool here contains 150 predicted models, except for T0951, with 149 models.

3.1.2 Molecular dynamics simulation

MD simulation requires protein structure information as the input. This information is provided by the predicted structures in the CASP data set. On the other hand, knowledge about the protein environment is rarely available. [129] To compensate for such limitations, we make the simulation parameters and environment uniform by following the relatively simple setup described by Lemkul, J., which solvates the proteins in a cubic box filled with water molecules and added ions. [185] Protein folding simulations typically require long simulation times that range from tens to hundreds of nanoseconds (ns) for the conformational space searches. [186] Here, we perform a relatively short 1 ns simulation. MD simulation is commonly performed at room temperature of 300 K. A previous study shows that helical proteins in explicit water tend to destabilize faster within the same timeframe while being simulated at higher temperatures. [187] Thus, we use a higher temperature of 500 K to speed up the stabilization effect, allowing us to get such information in short simulations (Table 3.1). We then perform post-processing in the final step by removing the periodic boundary condition after centering the protein molecule to avoid boundary effect problems. The MD simulation is implemented using GROMACS version 2019.4. [176] The final output of the simulation is protein trajectories containing raw information, including structural and positional changes of the protein over time. To validate the simulation results, we calculate the RMSD value of the final to the initial

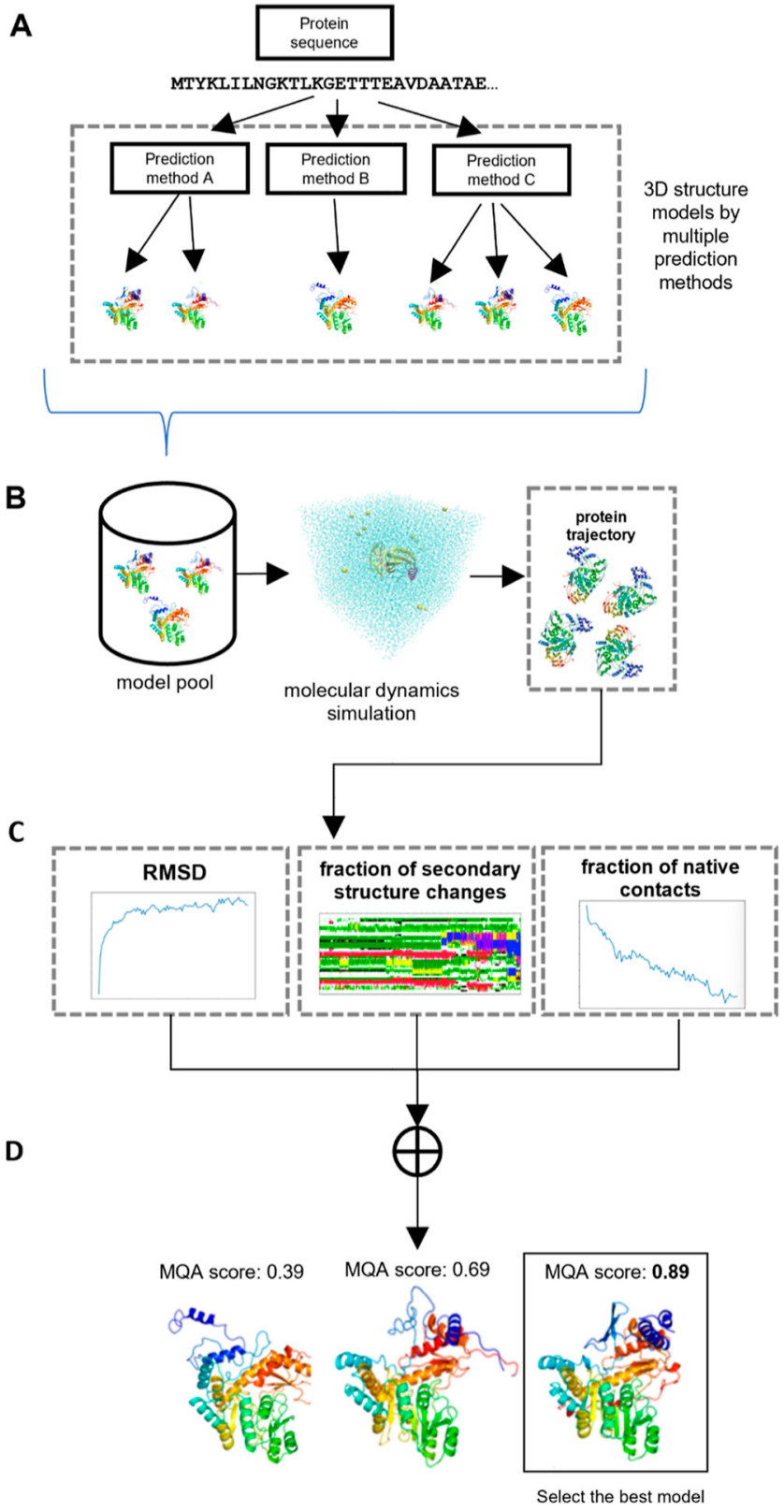


Figure 3.1: Schematic diagram of the proposed method: (A) model pool generation by structure prediction methods, (B) protein trajectory generation through MD simulation, (C) feature extraction, and (D) best model selection with GDT_TS values as an example.

Table 3.1: Chosen parameters for the MD Simulation

Parameter	Value
force field	OPLS/LA ([189])
water model	SPC/E ([190])
ions	Na (+), CL (-)
temperature	500 K
energy minimization	steepest descent

structure with the threshold of 2.5 Å. [188] A simulation is marked as invalid if the RMSD value of the final trajectory to the initial structure is larger than 2.5 Å. This is to ensure that there are no unrealistic movements during the simulation as a result of drastic structure changes or clashes. The final output of the MD simulation is the protein trajectory required for the feature extraction step (Figure 3.1B).

3.1.3 Feature extraction

Our method suggests that protein structure stability in an in silico system might indicate structural quality. To incorporate such structure stability information, we define features extracted from the protein trajectories. The results of previous work [191, 192] reveal that the RMSD and fraction of native contacts to the initial structure are useful to monitor structure stability. We also define new stability information: the fraction of the secondary structure to the initial structure. These three types of structural change information are extracted from the MD trajectories and defined as features (Figure 3.1C). The feature extraction step is implemented using the MDTraj library. [193] These features are later used as the input for the best model selection method. In the pilot phase of this work, we also calculated other potential features from the MD trajectories, such as the radius of gyration and solvent accessibility surface area. However, these features did not significantly correlate to the structure quality.

3.1.3.1 Root-mean-square deviation

RMSD is an evaluation metric to measure structural similarity by calculating the average distance of the selected atoms between a trajectory of structures to one reference state. The reference is often defined as the initial structure of the trajectory. This can provide insight into the overall structure movement from the initial structure. Stable RMSD values can measure structural stability and conformational convergence. Low-quality models hypothetically have unstable structures with more significant atom position deviations than the initial structure. In contrast, high-quality predicted models ideally have lower RMSD values. For the featurization, we calculate the RMSD value of the last trajectory to the initial structure for heavy atoms and then apply a cutoff at 1 Å. Any RMSD value above 1 Å is set to 1 Å. As the final step, we invert the RMSD values as $1 - \text{RMSD}$. Thus,

from this featurization, the best model is selected based on the highest feature value. The RMSD for heavy atoms is computed as follows:

$$\text{RMSD} = \min\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_{heavy,i} - d_{heavy,\text{ref},i})^2}, 1\ \text{Å}\right) \tag{3.1}$$

where N is the number of heavy atoms, $d_{heavy,i}$ is the position of the i-th heavy atom in the last trajectory frame, and $d_{heavy,\text{ref},i}$ is its position in the reference (initial) structure. The inverted RMSD value, accounting for the 1 Å cutoff, is calculated as:

$$\text{Inverted RMSD} = 1 - \text{RMSD} \tag{3.2}$$

3.1.3.2 Fraction of secondary structure changes

The protein stability can be examined through the secondary structure-type changes among conformational-state transitions. For example, unstable structures might have numerous secondary-type changes between a trajectory state and the initial structure. Conversely, stable structures tend to maintain their secondary structure type. To represent this information as a feature, we define the fraction of secondary structures as follows:

$$SS(X) = \frac{\sum(c^X - c^o)}{n} \tag{3.3}$$

where X is a conformation state, n is the total number of residues, and c is a $1 \times n$ binary matrix, where the value of 1 represents a residue whose secondary structure type remains unchanged compared to the initial structure. Each residue is assessed independently for secondary structure type changes. The secondary structure types are determined using the DSSP program, adhering to its standard classification criteria for eight structure types. [194] The binary matrix c^X is computed for each conformation state X in the trajectory, and c^o represents the initial structure. A higher fraction value of $SS(X)$ indicates a greater stability in secondary structure, with high-quality models typically exhibiting minimal changes from the initial structure. This fraction is calculated for the comparison between the last trajectory frame and the initial structure.

3.1.3.3 Fraction of native contacts

The results of previous work have revealed that the fraction of native contacts is useful for monitoring structure stability alongside RMSD. [191] Hence, we apply this information as an additional feature. Native contacts, formed during the transition between two conformational states, reflect the stability of proteins in the presence of denaturants. This feature is computed according to the following definition: [195]

$$N(X) = \frac{1}{|S|} \sum_{(i,j) \in S} \frac{1}{1 + \exp[\beta(r_{ij}(X) - r_{ij}^{o})]} \tag{3.4}$$

where X is a conformational state, $r_{ij}(X)$ represents the distance between heavy atoms i and j in conformation X, and r_{ij}^{0} is the corresponding distance in the native-state conformation. The set S includes all pairs of heavy atoms (i, j) from residues θ_i and θ_j such that $|\theta_i - \theta_j| > 3$ and the distance in the native state $r_{ij}^{0} < 4.5$ Å. The parameter β is set to 5 Å^{-1}, indicating the steepness of the contact function. The final structural conformation X of the last trajectory is used for this calculation. A higher value of $N(X)$ indicates greater structural stability. Similar to previous features, the best model is selected as the one with the highest value of this feature.

3.1.4 Best model selection

As mentioned in each feature definition, high-quality models hypothetically have a stable structure. Stable structures ideally have low atom position deviation, less secondary structure-type changes, and high native heavy atom contacts to their initial structure. These are represented by the inverted RMSD, a fraction of secondary structure changes, and a fraction of native contacts, respectively. Thus, the best model in a prediction pool is the model with the highest feature value, defined as:

$$\text{Best model}(T) = \arg\max_{i}^{n} \left(x(M_i)\right) \tag{3.5}$$

where T is the selected pool of models, n is the total number of models in the pool, x is the selected feature of a model M, and i represents the index of a model within the pool. We later also investigate the model selection results from the combination of the features in chapter 3.2.2. The features are combined by simple addition, and the best model is selected based on the highest value (Figure 3.1D).

3.1.5 Evaluation method

The quality of the predicted structure model is quantified using global distance test total scores (GDT_TS). GDT_TS is an accuracy-like score that indicates the structure similarity between the predicted models and the native structure [196]. We evaluate the performance of MQA methods by calculating the GDT_TS loss, that is, the difference between true GDT_TS of the model selected by the MQA methods and GDT_TS of the best/most accurate model in a protein pool (Figure 3.2). A lower loss indicates better performance. This evaluation method measures the ability of the MQA methods to find the best model in protein pools.

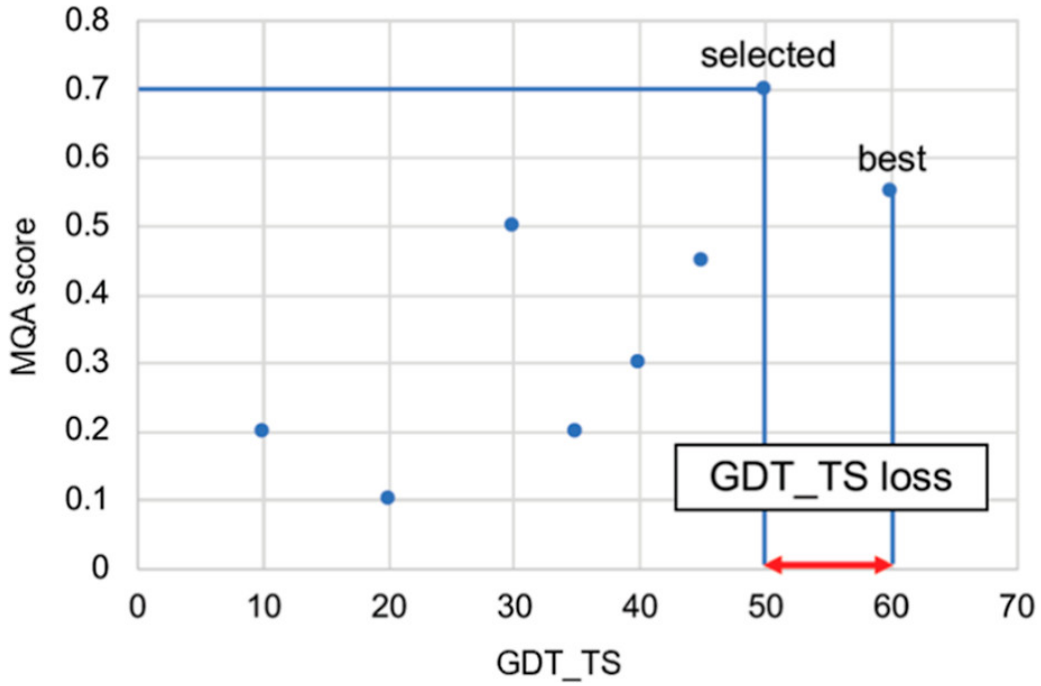


Figure 3.2: Illustration for GDT_TS loss. MQA methods generally select the best model using certain scoring methods.

3.2 Results

3.2.1 Simulation results

During the energy minimization step, not all prediction models from the HMSCraper-refiner group could be simulated due to the incompleteness/missing atoms. Thus, we excluded models from this group in all pools except for T1016 since it does not contain prediction models from the group. Additionally, a few different prediction models from other groups could not be simulated due to the occurrence of overlapping atoms. Only a small percentage of unsuccessful simulations in each pool (excluding prediction from the HMSCraper-refiner group) were found. Theoretically, a larger protein size tends to have higher difficulty in the simulation. This is found in the increasing percentage of unsuccessful simulations on larger proteins, particularly when the number of residues is > 300 (Table 3.2). The results also show that our proposed MD simulation setup is effective for all pools with an average simulation success rate larger than 90%. Since the ratio of unsuccessful simulations is less than 10%, we omit these data from the simulation results. In addition, the simulation validation results show that 7 of 20 pools contain invalid simulations (RMSD > 2.5 Å) with the ratio relative to the number of successful simulations that is smaller than 10% (Table 3.3). Since the ratio of invalid simulations is less than 10%, we also exclude them from the simulation results.

Table 3.2: The percentage of successful simulation in each protein pool excluding prediction models from HMSCraper-refiner group

Pool Name	Number of Residues	Successful Simulation (%)
T0955	41	99.33
T0953s1	72	100.00
T0958	77	100.00
T1008	77	100.00
T0968s2	115	97.24
T0968s1	118	97.96
T0957s2	158	99.32
T0957s1	162	100.00
T1016	202	100.00
T0953s2	248	95.17
T0951	266	100.00
T1005	326	95.21
T0950	342	91.78
T0954	342	97.24
T0963	364	89.66
T0960	374	93.79
T1003	434	97.95
T1011	444	93.10
T0966	492	91.10
T1009	718	93.10
Average		96.60

Table 3.3: The percentage of invalid simulation in each corresponding pool

Pool Name	Number of Residues	Invalid Simulation (%)
T0968s1	118	4.47
T0957s1	162	2.89
T0953s2	248	1.36
T0950	342	9.62
T0963	364	8.46
T0960	374	2.77
T1011	444	2.27
Average		4.55

Table 3.4: Performance of each singular feature and feature combination measured by Top 1 model GDT_TS loss and number of actual best model successfully selected. The successfully selected actual best model in a pool is represented by zero value

Pool Name	RMSD	SS	Native Contacts	RMSD and SS	RMSD and Native Contacts	SS and Native Contacts	All Combined	GDT_TS of Actual Best Model
T0950	0.053	0.129	0.165	0.053	0.053	0	0.053	0.385
T0951	0.032	0.032	0.015	0.032	0.032	0.032	0.032	0.943
T0953s1	0.205	0.067	0.078	0.067	0.078	0.067	0.067	0.489
T0953s2	0.261	0.290	0.443	0.261	0.261	0.443	0.261	0.631
T0954	0.025	0.125	0.022	0.025	0	0.022	0	0.699
T0955	0.043	0.122	0.031	0.122	0.031	0.122	0.031	0.951
T0957s1	0.051	0.109	0.118	0.051	0.051	0.118	0.051	0.544
T0957s2	0.174	0.310	0.019	0.310	0.108	0.310	0.310	0.610
T0958	0.133	0.045	0.325	0.325	0.325	0.325	0.325	0.740
T0960	0.109	0.102	0.023	0.102	0.109	0.023	0.102	0.484
T0963	0.113	0.040	0.226	0.186	0.113	0.137	0.186	0.516
T0966	0.049	0.054	0.007	0.001	0.049	0.007	0.001	0.611
T0968s1	0.305	0.222	0.047	0.222	0.047	0.047	0.047	0.667
T0968s2	0.180	0	0.122	0	0.122	0	0	0.713
T1003	0.054	0.047	0.001	0.047	0.047	0.001	0.047	0.895
T1005	0.063	0.137	0	0.063	0.063	0.059	0.063	0.558
T1008	0.179	0.484	0.179	0.179	0.179	0.179	0.179	0.870
T1009	0.016	0.017	0.022	0.053	0.053	0.017	0.053	0.673
T1011	0.055	0.105	0.043	0.055	0.055	0.043	0.055	0.686
T1016	0.030	0.031	0.022	0.030	0.030	0.014	0.014	0.816
Average	0.107	0.123	0.095	0.109	0.090	0.098	0.094	0.674
Actual Best Model Selected:	0	1	1	1	1	2	2	

3.2.2 Feature combination

The best model in each pool is selected based on the highest feature values. In the experiment, we evaluate the performance of each singular feature and all possible feature combinations. The results show that combining all three proposed features led to the best performance according to the average top 1 GDT_TS loss and the number of actual best models in each pool (Table 3.4). Even though the combination between RMSD and native contacts features results in a slightly lower average GDT_TS loss, the difference is only 0.004, and it only successfully selected the actual best model in one pool, while the combination of all the three features selected two actual best models in two pools. Thus, the combination of all of the three features is selected as the main proposed method.

3.2.3 Performance at different simulation times and temperatures

Performing simulation at unusually high temperatures like 500 K might be harmful to the protein structure stability even for high-quality models. However, this also might accelerate the destabilization effect in shorter simulation length and could reduce the computational cost. To confirm this effect, we perform additional MD simulations at lower temperatures of 300K and 400 K. We then compare the performance results at different temperatures and simulation lengths. The results show that simulation at unusually high temperature and shorter length fastened the destabilization effect as the 500 K and 0.5 ns simulation achieved the best performance with the lowest average

Table 3.5: The top 1 GDT_TS loss performance in all pools on different temperatures and simulation length pairs

Pool Name	300K (0.5ns)	300K (1ns)	400K (0.5ns)	400K (1ns)	500K (0.5ns)	500K (1ns)	GDT_TS of Actual Best Model
T0950	0.174	0.053	0.243	0.236	0.246	0.053	0.385
T0951	0.010	0.007	0.013	0.015	0.008	0.032	0.943
T0953s1	0.052	0.205	0.239	0.168	0.067	0.067	0.489
T0953s2	0.165	0.068	0.358	0.216	0.358	0.261	0.631
T0954	0.031	0	0.112	0.025	0	0	0.699
T0955	0.024	0.024	0.006	0.031	0.043	0.031	0.951
T0957s1	0.051	0.051	0.225	0.192	0	0.051	0.544
T0957s2	0.261	0.261	0.316	0.316	0.019	0.310	0.610
T0958	0.325	0.325	0.338	0.338	0.127	0.325	0.740
T0960	0.078	0.102	0.109	0.109	0.102	0.102	0.484
T0963	0.186	0.186	0.121	0.129	0.121	0.186	0.516
T0966	0.056	0	0	0.050	0.098	0.001	0.611
T0968s1	0.186	0.273	0.057	0.057	0.057	0.047	0.667
T0968s2	0	0	0.091	0.091	0	0	0.713
T1003	0.076	0.005	0.047	0.047	0.047	0.047	0.895
T1005	0.044	0.087	0.048	0.048	0.063	0.063	0.558
T1008	0.471	0.110	0.068	0.188	0.179	0.179	0.870
T1009	0.046	0.017	0.046	0	0.016	0.053	0.673
T1011	0.046	0.098	0.098	0.098	0.105	0.055	0.686
T1016	0	0.021	0.048	0.011	0.005	0.014	0.816
Average	0.114	0.095	0.129	0.118	0.083	0.094	0.674
Actual Best Model Selected:	2	3	1	1	3	2	

top 1 GDT_TS loss and the highest number of the actual selected best model (Table 3.5). However, both results of 400 K simulations show worse performance than 300K simulation within the same simulation length. This might be because the temperature difference is not sufficiently high enough and thus the performance did not significantly change. The best performance results from 500 K and 0.5 ns simulation then are taken as the main results for the proposed method.

3.2.4 Performance evaluation

For this scenario, we compare the main results of the proposed method with the results from QDeep. The proposed method shows comparable performance to QDeep, where it achieved a lower average top 1 GDT_TS loss with 0.008 difference (Table 3.6). The individual pool results also show comparable performance, where the method achieves eight successes, four draws, and eight fails. In several pools, our method significantly outperforms QDeep. For instance, in T1008, our method attains a GDT_TS loss of 0.110, while QDeep achieves a poor performance of 0.455. This comes from the disadvantage of using the MSA feature-based method, including QDeep, where the alignment depth of the MSA forT1008 is zero with no identifiable homologous sequences [135]. Our method does not rely on such information since the features are acquired solely from the protein structural information. The results also show that our proposed method successfully selected the actual best model

Table 3.6: Top 1 Model GDT_TS loss comparison between our proposed method and QDeep on the CASP13 Stage 2 data set[a]

Pool Name	Baseline	Proposed	QDeep	GDT_TS of Actual Best Model
T0950	0.215	0.246	0.030	0.385
T0951	0.167	**0.008**	0.057	0.943
T0953s1	0.171	0.067	0.041	0.489
T0953s2	0.267	0.358	0.028	0.631
T0954	0.239	**0**	0	0.699
T0955	0.308	**0.043**	0.171	0.951
T0957s1	0.259	**0**	0.151	0.544
T0957s2	0.267	**0.019**	0.261	0.610
T0958	0.253	**0.127**	0.133	0.740
T0960	0.101	0.102	0.078	0.484
T0963	0.124	**0.121**	0.121	0.516
T0966	0.171	0.098	0.006	0.611
T0968s1	0.323	**0.057**	0.057	0.667
T0968s2	0.387	**0**	0.130	0.713
T1003	0.110	**0.047**	0.047	0.895
T1005	0.154	0.063	0	0.558
T1008	0.449	**0.179**	0.455	0.870
T1009	0.140	0.016	0.003	0.673
T1011	0.171	0.105	0.043	0.686
T1016	0.055	**0.005**	0.014	0.816
Average	0.217	**0.083**	0.091	0.674

[a] The underlined marks indicate that the baseline method performs better than the proposed method. The bold marks indicate that our proposed method achieves better/draw performance than QDeep.

with zero GDT_TS loss in three pools (T0954, T0957s1, and T0968s2) while QDeep was successfully selected in two pools (T0954 and T1005). To compare the performance of our method to random selection, we add the GDT_TS loss of the baseline method, which is the average GDT_TS of each pool. It is shown that our method achieves significantly superior overall performance. However, in T0950, T0953s1, and T0960, our method shows worse performance than the baseline method. In addition, we computed the Wilcoxon signed-rank test with $\alpha = 0.05$ between the proposed versus baseline method and the proposed method versus QDeep results. The statistical test results between the proposed versus baseline method reject the null hypothesis with p-value = 0.0001, which means that the proposed method is significantly different from random selection. On the other hand, the test results of the proposed method versus QDeep fail to reject the null hypothesis with p-value = 0.87. This means that the proposed method achieves comparable performance with QDeep.

3.3 Discussion

This work shows the possibility and potential application of using protein stability information to estimate the quality of a protein model. This information is derived from the structural change information over time obtained through MD simulation. We propose three features representing the protein stability information: rmsd, a fraction of the secondary structure, and a fraction of

native contact information to the initial structure. Thus far, no previous MQA method has utilized the stability information explicitly. Our approach does not use any additional predictive features or evolutionary information, such as the predicted secondary structure or sequence profiles from multiple sequence alignment homologues. Furthermore, our method does not rely heavily on machine learning methods that require training on tens to hundreds of thousands of models, that is, training on multiple CASP data sets.

3.3.1 Quality of unsuccessful and invalid simulated structures

Our method requires protein trajectory information in the first step by conducting MD simulation. A small percentage of the models could not be simulated successfully and was omitted from the simulation results. However, the omission might discard the top models in each pool related to the selection of the best model in each pool. Hypothetically, poor-quality models and/or models with structure incompleteness are the main causes of unsuccessful simulations; thus, the omission will not discard good-quality models. To prove this hypothesis, we compare the model quality distribution between successful and unsuccessful data (Figure 3.3). It is shown that the omission of unsuccessful simulation data "filters" the low-quality models from each pool, especially for low-size proteins. This is plausible because low-quality models typically have structural problems as mentioned above and are found in the omitted models.

Running MD simulation under extreme physical conditions like high temperature might cause drastic structural deviation even in a short simulation. We further investigate whether invalid simulations are coming from low-quality models in the pools. The results show that all these invalid simulations are found in pools with no high-quality models (GDT_TS 80) and models with lower quality relative to other models in the same pool, except for T0960 and T0963 (Figure 3.4). They also have a larger percentage of invalid simulations compared to the other five pools. This is because these two models did not contain any outstanding prediction models, as the GDT_TS score ranged between 30 and 50 s, unlike other pools with larger ranges. Nevertheless, none of the actual highest quality models in these two pools were marked as invalid simulations.

3.3.2 Case study

The poor performance in the T1008 pool induced the huge disadvantage of QDeep when there were no identifiable homologous sequences. Correspondingly, our method achieves significantly poor performance compared to QDeep with the difference of GDT_TS loss larger than 0.1 in T0950 and T0953s2. Our method suggests that the structural stability might indicate the quality of the model, which is represented by the feature values. When there are no high-quality models in a pool, it is more difficult for our method to select and distinguish between bad- and good-quality models since there is no significant feature value difference between them. This is found in two pools where

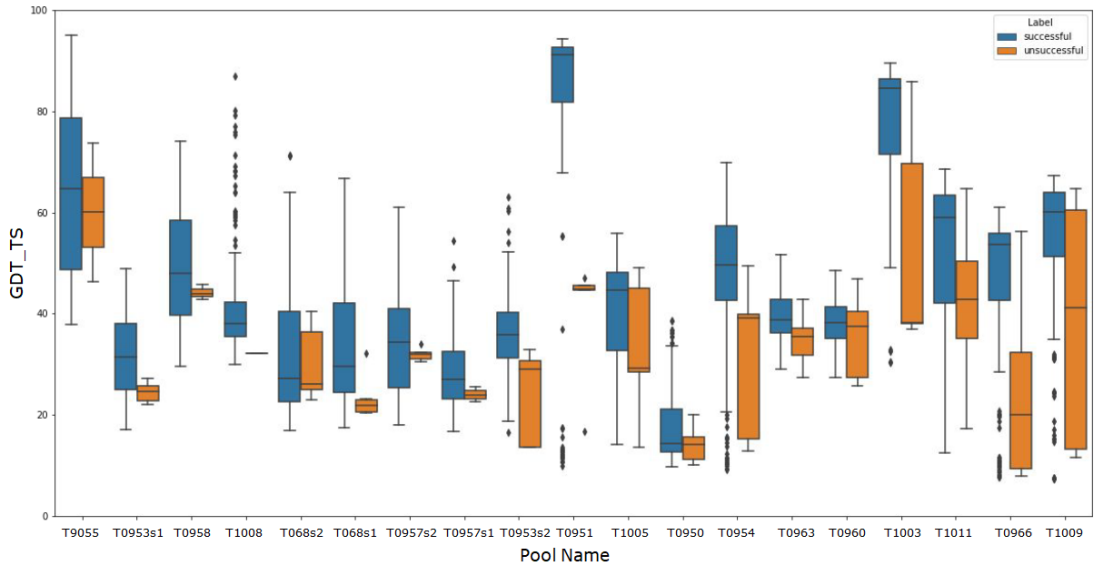


Figure 3.3: GDT_TS distribution between successful (blue) and unsuccessful (orange) simulation of all pools sorted from smallest to largest protein size (left to right). The prediction models from HMSCraper-refiner group are included. T1016 pool is excluded since all of prediction models are successfully simulated

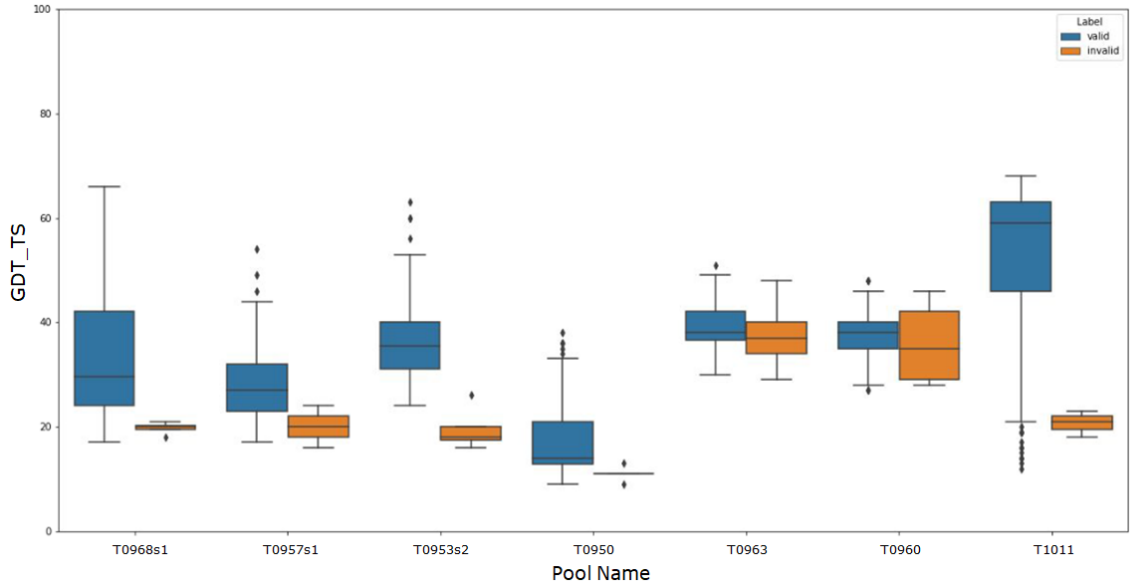


Figure 3.4: GDT_TS distribution between successful (blue) and invalid (orange) simulation of pools containing invalid simulations sorted from smallest to largest protein size (left to right)

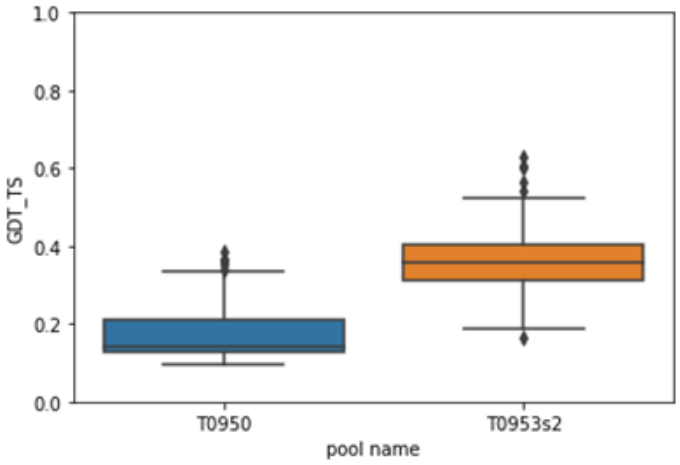


Figure 3.5: GDT_TS distribution pools where the proposed method shows significant worse performance than QDeep

the proposed method shows significantly worse performance than QDeep, and none of the pools has high-quality models (Figure 3.5). On further inspection, we find consistent results with our hypothesis, where the proposed feature values can select the best model in the successful cases if high-quality models (GDT_TS 70) are available in pools (Figure 3.6). This is also found in the comparison between the proposed feature value and the GDT_TS of the actual best model. Thus, our method has a major advantage when there are high-quality models in the prediction pools as found in T0951, T0955, T1003, T1008, and T1016. This is more useful and applicable for real applications, where selecting the best, high-quality models is the primary goal. Interestingly, the successful cases are also found in the pools where high-quality models are not available such as in T0957s1 and T0957s2. We then investigate further by comparing the best model between the two successful cases with the failed cases whose GDT_TS is larger than 50. In protein MQA, the GDT_TS value of larger than 50 often indicates that the majority of secondary structure composition is correctly predicted. We observed that the actual best model's structure in the two successful pools exhibits fewer random and extensive terminal coils compared to those in the five losing cases (Figure 3.7). All structures in the losing cases have a loop/coil region percentage exceeding one-third (33.33%) of their total structure, whereas in the successful cases, the loop region percentage remains below one-quarter (25%). This discrepancy likely impacts the method's performance in losing cases, as the proposed features – rmsd and the fraction of secondary structure – are sensitive to the fluctuation bias originating from these coil regions.

Additionally, we also investigate each feature value between low- and high-quality models. A high-quality model should have lower and stable fluctuations over time. We took the T0951 pool as an example since this pool contains a large number of high-quality models and also low-quality models. Each singular feature value between the actual best and worst model in the T0951 pool

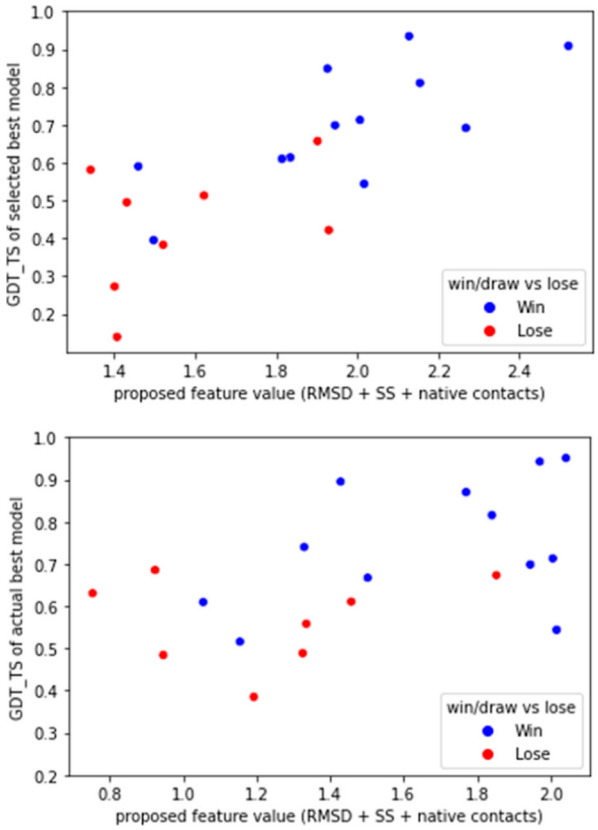


Figure 3.6: Top: proposed feature value versus GDT_TS of the best model selected by the proposed method. Bottom: proposed feature value versus GDT_TS of the actual best model

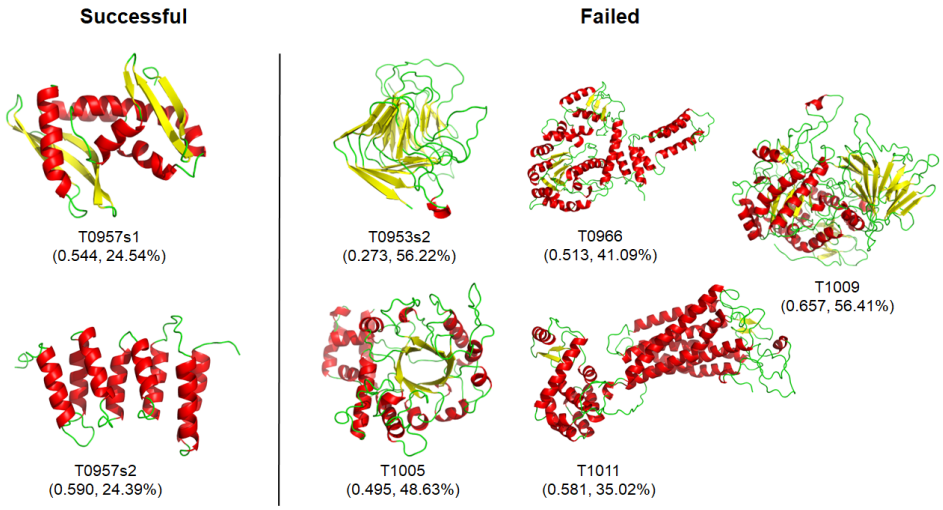


Figure 3.7: The 3D structure and GDT_TS of the selected best models by the proposed method in successful (left) and failed (right) cases where the actual highest GDT_TS is higher than 50. The alpha helix and beta sheet structures are depicted in red and yellow, respectively, while loop structures are illustrated in green. The values presented within the brackets represent the normalized GDT_TS and the corresponding percentage of loop/coil region

Table 3.7: The average top 1 GDT_TS loss performance in all pools using different weight ratios

	RMSD : SS : Native Contacts							
Ratios	**1:1:2**	**1:1:3**	**1:2:1**	**1:2:2**	**1:2:3**	**1:3:1**	**1:3:2**	**1:3:3**
Average	0.110	0.110	0.093	0.083	0.105	0.096	0.084	0.084

	RMSD : SS : Native Contacts							
Ratios	**2:1:1**	**2:1:2**	**2:1:3**	**2:2:1**	**2:2:3**	**2:3:1**	**2:3:2**	**2:3:3**
Average	0.101	0.122	0.130	0.090	0.106	0.090	0.096	0.089

	RMSD : SS : Native Contacts							
Ratios	**3:1:1**	**3:1:2**	**3:1:3**	**3:2:1**	**3:2:2**	**3:2:3**	**3:3:1**	**3:3:2**
Average	0.109	0.129	0.129	0.101	0.099	0.120	0.093	0.090

	RMSD : SS : Native Contacts
Ratios	**1:1:1 (proposed)**
Average	0.083

shows a significant feature value change over time (Figure 3.8).

3.3.3 Feature weight ratio

Each feature might have more contributions than the others. To examine this, we define the weights for the features as follows:

$$\text{Best model}(T) = \arg\max_{i}^{n} \left(\frac{\sum_{j=1}^{m} w_j \cdot x_j(M_i)}{\sum_{j=1}^{m} w_j} \right) \tag{3.6}$$

where T is the selected pool of models, n is the total number of models in the pool, m is the number of features, $x_j(M_i)$ is the value of the j-th feature for model M_i, w_j is the weight of the j-th feature, and the sums run over all m features. Using various weight ratio combinations, the currently proposed method with balanced weight ratios achieves the best performance with the lowest average top 1 GDT_TS loss (Table 3.7). Our proposed method feature combination is already appropriate for the best model selection. The weight optimization itself might be contrary to the advantages of the proposed method that does not require any training or optimization steps, making the method data set-dependent.

3.3.4 Simulation at higher temperature

We also further explored simulations under more extreme conditions at a significantly higher temperature of 600K. This exploration was motivated by observations from previous simulations, which indicated that operating at higher temperatures, such as 500K, could significantly accelerate the destabilization effect within a shorter simulation time frame, thereby offering potential reductions in

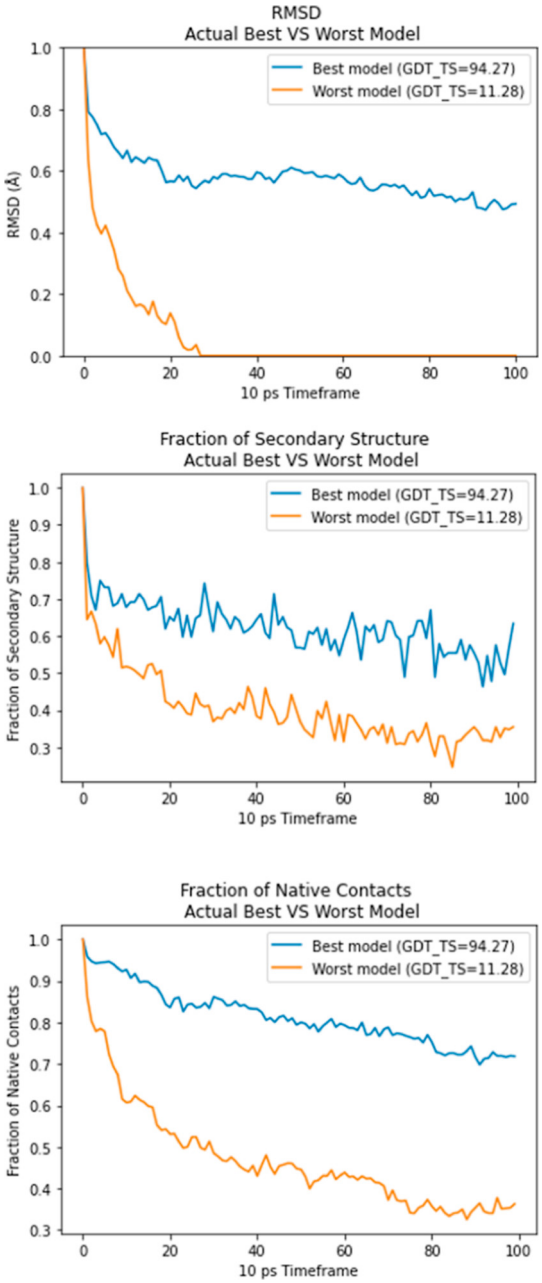


Figure 3.8: Each singular feature value between the actual best versus the worst model in the T0951 pool

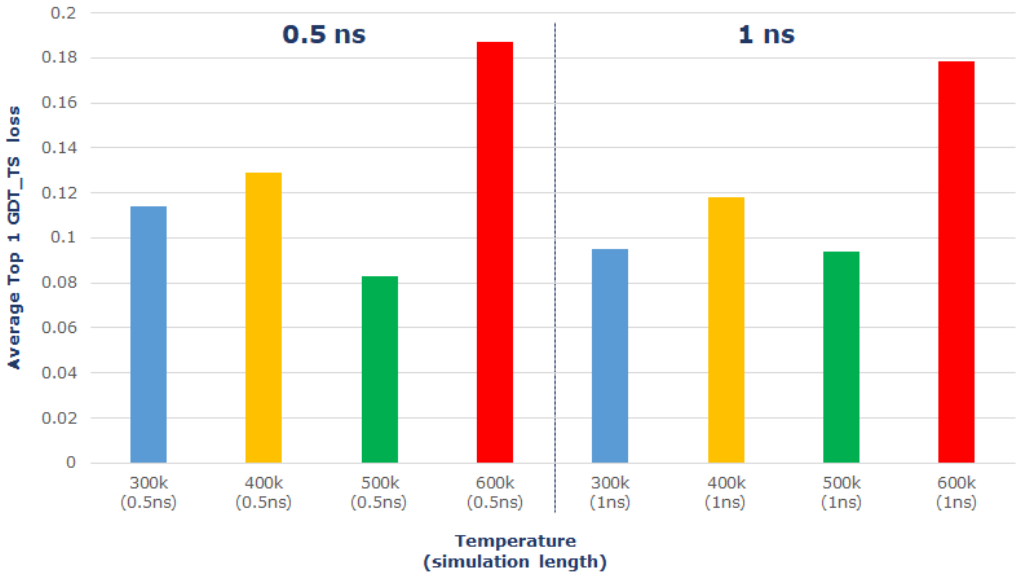


Figure 3.9: The performance comparison between 500 K simulation and 600 K simulation

computational costs and further validating whether an extended simulation time influences performance. Given the main results at 500K, we hypothesized that increasing the simulation temperature to 600K might enhance these effects, potentially leading to even more rapid and pronounced destabilization phenomena. Such conditions are not typically employed due to concerns over the structural integrity of proteins. However, they may offer unique insights into the dynamics of protein stability under stress conditions.

As demonstrated in Figure 3.9, the average top 1 GDT_TS loss performance is worse at 600K for both tested simulation times of 0.5 and 1 ns compared with the best results of 500K at 0.5 ns. This deterioration in performance may be attributable to the excessively high temperature causing structures consisting mostly of coil regions to have larger fluctuations in stability features, and subsequently flattening the RMSD feature, making it less informative. The absence of informative RMSD feature poses greater challenges for our proposed method in selecting the best model. While higher temperatures can destabilize structures more rapidly, and well-constructed models are expected to be more resistant, excessively high temperatures might be too excessive. We observed that simulations at 0.5 and 1 ns deliver similar performance, suggesting that shorter simulation times are not only more cost-effective but also do not significantly affect performance.

3.3.5 Performance at each time frames

In an effort to optimize computational efficiency without compromising the accuracy of protein stability predictions, we explored the performance of molecular dynamics simulations conducted over different time frames. Specifically, simulations were executed for lengths of 100 ps, 200 ps, 300 ps, 400 ps, 500 ps (0.5 ns), 600 ps, 700 ps, 800 ps, 900 ps, and 1000 ps (1 ns) at varying

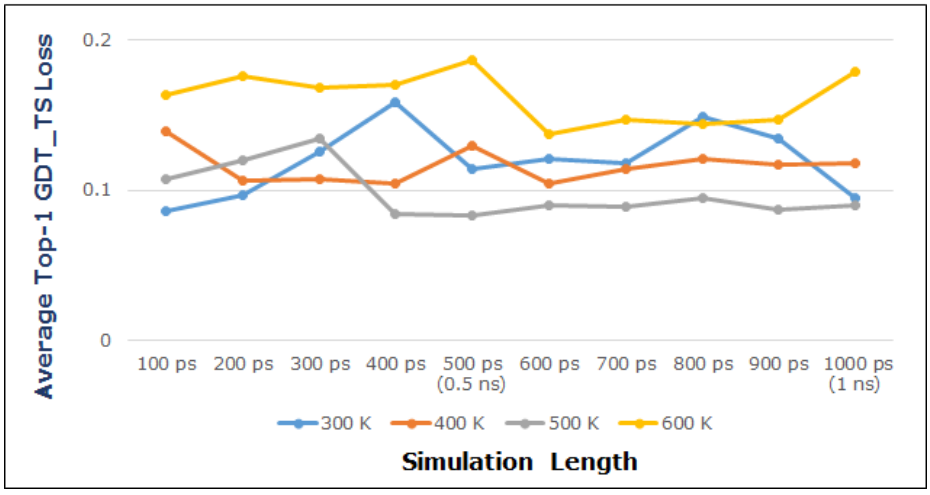


Figure 3.10: The performance at shorter simulation length and different temperatures

temperatures of 300 K, 400 K, 500 K, and 600 K. This investigation aimed to determine whether shorter simulations could yield comparable or superior results to longer simulations, thereby reducing the computational resources required for accurate model quality assessments.

The comparative analysis focused on simulations conducted at 0.5 ns and 500 K versus those extended to 1 ns, with the anticipation that shorter simulations might exhibit similar or enhanced performance metrics. The results, as illustrated in Figure 3.10, indicate that the optimal outcomes were consistently achieved with the 0.5 ns simulations at 500 K. This observation suggests a critical balance between simulation length and temperature in capturing the dynamic stability features of proteins where performance after 0.5 ns at all temperatures did not fluctuate significantly.

A notable observation from this study was the fluctuation of stability features within the initial 200/300 ps of the simulations, as depicted in Figure 3.8. These fluctuations suggest that the early phase of the simulation is characterized by significant conformational exploration, which may impact the reliability of stability predictions derived from shorter simulations. Despite the initial instability and larger features fluctuation, the 0.5 ns simulations at 500 K eventually converge towards a more stable representation of the protein's native state, underscoring the importance of allowing sufficient time for the system to equilibrate.

The findings underscore the complexity of balancing simulation length and temperature to achieve accurate protein stability predictions. While shorter simulations offer the allure of reduced computational demand, our results affirm that a minimum simulation length of 0.5 ns at elevated temperatures is critical for capturing the nuanced dynamics of protein folding and stability. This insight is instrumental in guiding the design of future molecular dynamics studies aimed at efficient and accurate prediction of protein behavior.

3.3.6 Computing time for model quality assessment

The proposed method uses an MD simulation, and thus, it needs more computing resources than machine learning-based MQA methods, whose running time is generally less than a minute. In this research, we used 1 NVIDIA Tesla V100 SMX2 GPU for the MD simulation, and the running times of the proposed method generally ranged between 1000 to 3000 s (Figure 3.11). The running time slightly depends on the protein size, but each pool has a different running time although the proposed method employs uniform MD simulation parameters for all models. Especially, there were some outliers represented by extremely long running time. We found that extremely long running times were mainly caused by the unusual structures. For instance, the prediction model MUFold_server_TS4 in T1003, whose running time was the longest, had long coil terminals (Figure 3.12). The long coil terminals caused larger energy minimization, and the simulation system became much larger than the others. Thus, to reduce the computing cost, we may need to remove such regions before applying the proposed method.

Furthermore, the previous work shows that a systematic MD simulation study of temperature dependency requires numerous temperature parameters that run on ten to hundreds of nanosecond simulations. [197] This becomes the limitation of our proposed method since such experiments demand huge computational resources and time. Despite the fact that the current temperature and simulation length parameters have shown promising results, further systematic studies using different simulation conditions might be necessary to re-evaluate and improve the performance of the methodology.

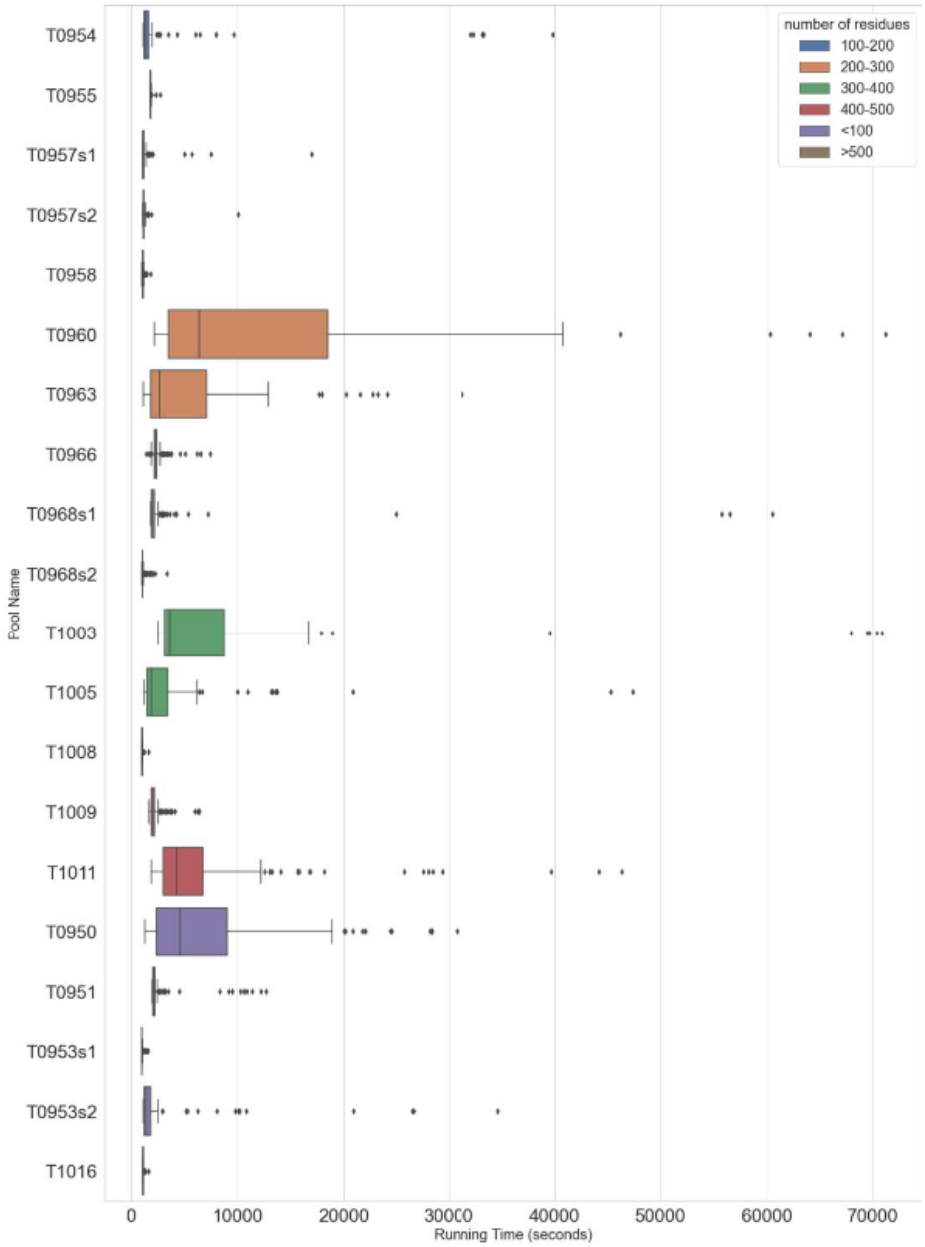


Figure 3.11: MD simulation running time for each pool

Figure 3.12: 3D structure of prediction models that has longest MD simulation running time. Top: MUFold_server_TS4 prediction model in T1003 pool. Bottom: Seok-assembly_TS2 prediction model in T0960 pool

3.4 Summary

In this chapter, we introduced a novel approach for model quality estimation, utilizing explicit protein structure stability information derived from Molecular Dynamics (MD) simulations. Our method employs relatively simple, uniform parameters and a short MD simulation time to extract key stability features. These features are then combined to effectively select the best prediction model. Despite relying on simple feature combinations and short MD simulation durations, our proposed method demonstrates performance comparable to existing state-of-the-art deep learning-based methods, which are typically trained on extensive, multiple CASP data sets. The integration of explicit protein stability information into the model quality assessment process represents a significant advancement, potentially offering a valuable addition to existing MQA methodologies.

Chapter 4

Predicting protein stability change upon point mutation

Building upon the foundational research in protein stability as applied to model quality assessment, this chapter transitions to a focused exploration of protein stability changes, particularly emphasizing the effects of point mutations. This progression from studying overall protein stability to investigating the specific nuances of stability changes due to point mutations represents an extension of the research scope. It highlights the critical importance of understanding how subtle genetic variations, such as point mutations, can significantly alter the stability landscape of proteins.

Machine learning-based methods, despite becoming state-of-the-art in predicting protein stability changes, have reached a performance barrier in prediction accuracy over the past decade. The reliance on biased training data sets and the focus on static protein information have limited their effectiveness, particularly in challenging blind test sets. This has necessitated a shift towards more rigorous approaches, such as molecular dynamics simulations-based methods, which consider the global and dynamic effects of mutations. Advances in computing power have made these rigorous free-energy calculation methods more feasible and accurate.

Furthermore, the existing alchemical methods often employ simplified models, such as capped n-peptides, which may not fully capture the complex effects of charge-changing mutations on protein stability. These mutations can significantly impact protein stability by altering electrostatic interactions, hydrogen bonding networks, and solvation effects, which are challenging for simplified models to accurately represent. The advent of high-quality protein structure prediction methods like AlphaFold2 offers a new possibility to enhance the prediction of charge-changing mutations using whole mutant structure information. In this context, we particularly underscore the utility of non-equilibrium simulation techniques. These methods, leveraging the latest advancements in computational efficiency, offer a promising alternative to equilibrium simulations. Non-equilibrium simulations enable the faster assessment of protein stability changes across large data sets of mutations, providing a critical advantage in high-throughput analyses. Their ability to deliver swift

predictions of changes in free energy ($\Delta\Delta G$) without the more extensive computational demands of equilibrium methods.

In this chapter, a novel workflow is proposed to enhance the alchemical method for predicting protein stability changes due to point mutations. This workflow employs the *pmx* double-system/single-box approach, specifically to address the challenges posed by charge-changing mutation cases. The refinement of this protocol enables accurate prediction of folding free energy changes for both conserving and charge-changing mutations, representing a significant advancement over existing methods. A key aspect of this chapter is the comparative analysis of the performance between supervised learning-based methods and our rigorous alchemical approach. This comparison is conducted using two challenging targets: frataxin and p53, which are known for their difficulty. The results demonstrate that our proposed rigorous alchemical method significantly outperforms state-of-the-art techniques, showcasing superior accuracy as evidenced by lower root-mean-squared error and higher Pearson correlation coefficient in these blind test sets.

The advancements presented in this chapter contribute substantially to the field of protein stability prediction. By offering a more robust and accurate approach to understanding the implications of point mutations, this research addresses the limitations of the machine learning-based approach and charge-changing mutation case of the rigorous approach. The novel integration of computational techniques and advanced protein structure prediction models marks a significant step forward in the predictive capabilities of protein stability changes upon point mutations.

4.1 Materials and methods

Figure 4.1 illustrates the general workflow of the proposed method. Alchemical free energy calculation upon point mutation requires both a folded wild-type structure and an unfolded mutant structure. In the conserving mutation case, the folded wild-type structure can be either obtained from the crystal structure or AlphaFold2 prediction. For the unfolded mutant structure, we use glycine-capped tripeptide (GxG) as suggested by Gapsys et al. [170] In the charge-changing mutation case, AlphaFold2 predictions are used for both the folded wild-type and unfolded mutant structure. Since AlphaFold2 will output the mutant structure in the folded state, we apply high-temperature simulation to unfold the predicted mutant structure (AF2HT). We use *pmx* to generate the hybrid structure for state A ($\lambda = 0$) and state B ($\lambda = 1$). The equilibrium and non-equilibrium simulation is then performed using the double-system/single-box (DSSB) simulation protocol. The final $\Delta\Delta G$ value is calculated using the Bennett acceptance ratio (BAR) method.

4.1.1 Data sets and wild-type structures

The proposed method does not require any training process, allowing us to evaluate its performance directly on two commonly used blind test data sets: p53 and frataxin data sets obtained from

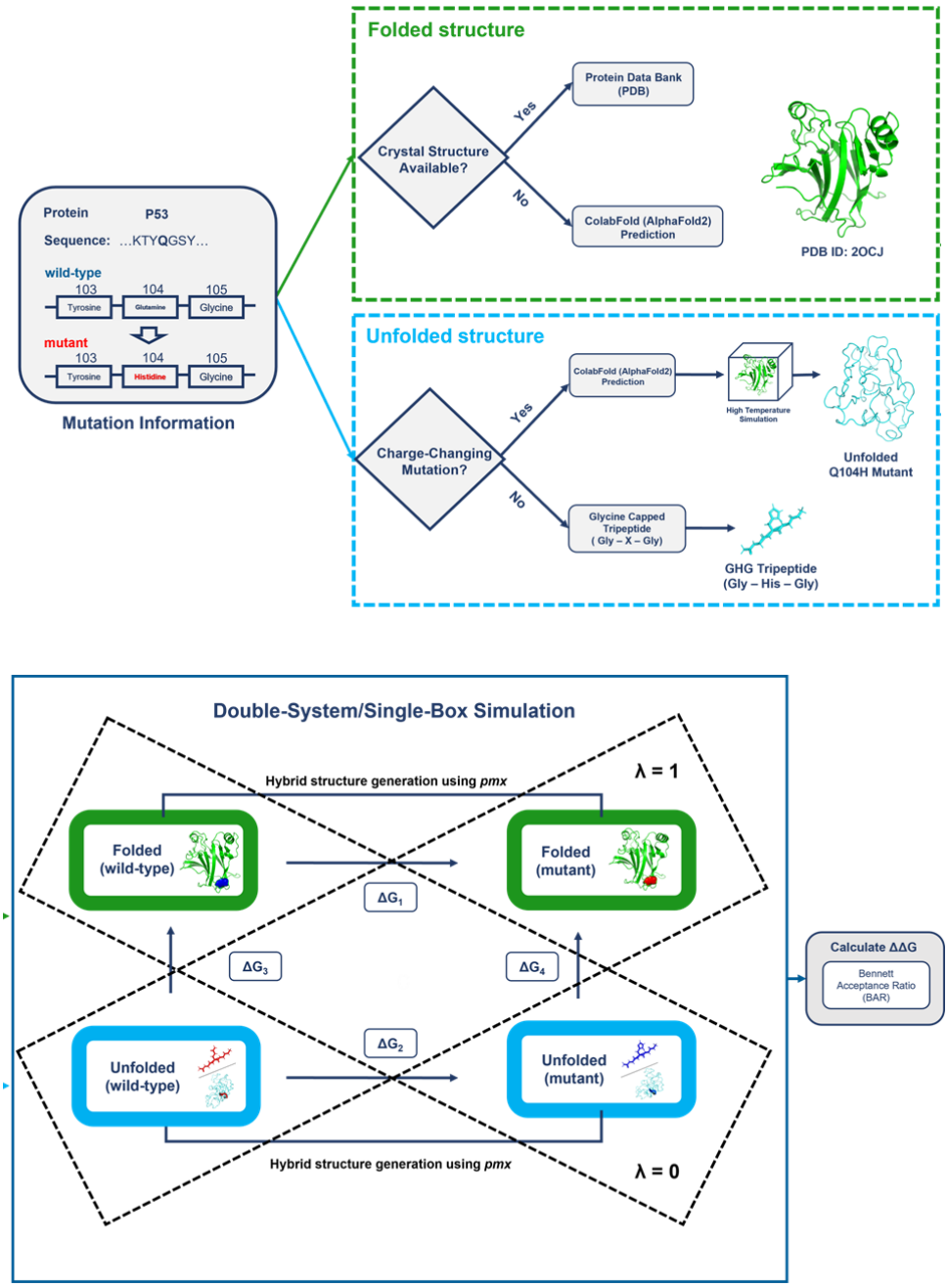


Figure 4.1: General overview of the proposed method using the p53 Q104H mutation as an example. The double-system, single-box protocol requires both folded wild-type and unfolded mutant structures. For the folded wild-type structure, if an existing structure is available in the Protein Data Bank (PDB), it will be used. If not, then the AlphaFold2-generated structure will be used. For the unfolded mutant structure, two approaches are employed, depending on the nature of the mutation. If the mutation results in a change in charge, then the AlphaFold2 structure will be unfolded through high-temperature simulation. If the mutation is charge-conserving, then a GxG will be used instead. For the DSSB protocol, hybrid structures are required for both state A and state B, which are generated through alchemical mutation using the *pmx* library.

Table 4.1: Input parameters used for AlphaFold2 run using ColabFold.

Parameter Name	Value
num_relax	5
template_mode	pdb70
msa_mode	Mmseqs2_uniref_env
pair_mode	unpaired_paired
model_type	auto
recycle_early_stop_tolerance	auto
max_msa	auto
num_seeds	1
use_dropout	no

SAAFEC-SEQ. [147] The p53 data set comprises 42 mutations in the DNA binding domain of the p53 tumor suppressor protein, as previously assembled by Pires et al. [157] The frataxin data set, consisting of eight single-point mutations in the Frataxin protein, was obtained from the Frataxin challenge in CAGI 5. [198] We utilized the protein data bank (PDB) to acquire the published crystal structures and sequences of p53 (PDB ID: 2OCJ, apo form) and frataxin (PDB ID: 2EKG). For AlphaFold2 structure prediction, we employed ColabFold v1.5.2 (https://github.com/sokrypton/ColabFold, accessed between 16-23 February 2023). [199] The detailed inputs and parameters used for the structure prediction are listed in Table 4.1. AlphaFold2, as expected, is able to accurately predict both p53 and frataxin wild-type structures since it is trained using PDB data Figure 4.2. In the p53 sequence, unmodeled/disordered regions (sequence number: 197-219) were trimmed, as AlphaFold2 predicted these regions as long coils, which would have significantly increased the system size and thus the computing time.

4.1.2 Unfolded mutant and hybrid structure generation

Alchemical free-energy calculation upon point mutation requires both folded wild-type and unfolded mutant structures. In the conserving mutation case, we generate the GxG structure using the Amber LEaP program with the Amber99SB-ILDN force field [200, 201]. For charge-changing mutations, we performed short 1 ns equilibrium MD simulations at high temperatures to unfold the AlphaFold2-predicted mutant structures. Mutant structures are initially solvated using dodecahedral water boxes and then neutralized with Na^+ and Cl^- ions. Energy minimization is carried out using the steepest descent algorithm, followed by 100 ps of NVT and NPT ensemble simulations. In the final MD simulation, we applied a 1000 K temperature and used the last snapshot as the unfolded mutant structure. An example of the unfolded mutant structure from the Q104H mutation in the p53 protein is provided in Figure 4.3.

Calculating alchemical free energy requires an intermediate pathway connecting the wild-type residue ($\lambda = 0$) and its mutant form ($\lambda = 1$), which is not physically feasible. The *pmx* package automates the generation of these intermediate states by producing hybrid amino acid states that

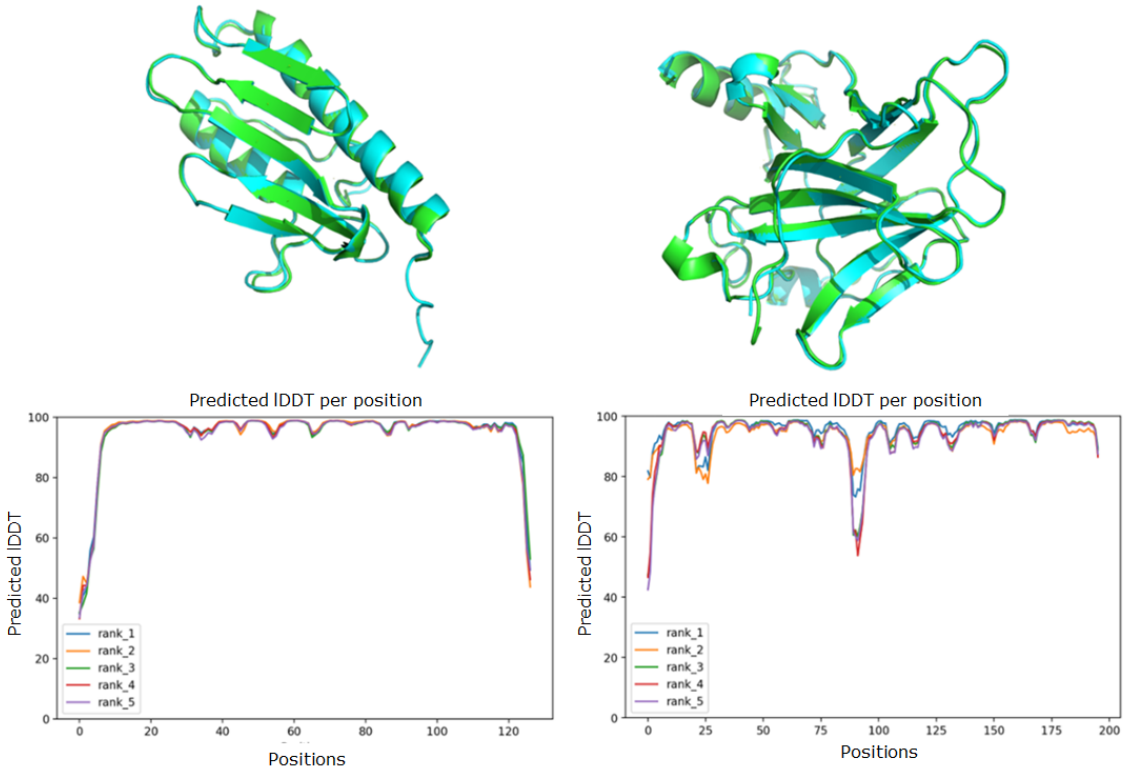


Figure 4.2: AlphaFold2 predicted structure and pLDDT scores for frataxin (left) and p53 (right). The crystal structure for frataxin (PDB ID: 2EKG) and p53 (PDB ID: 2OCJ) are highlighted in green. The relaxed rank_1 structure is chosen as the representative structure for each protein

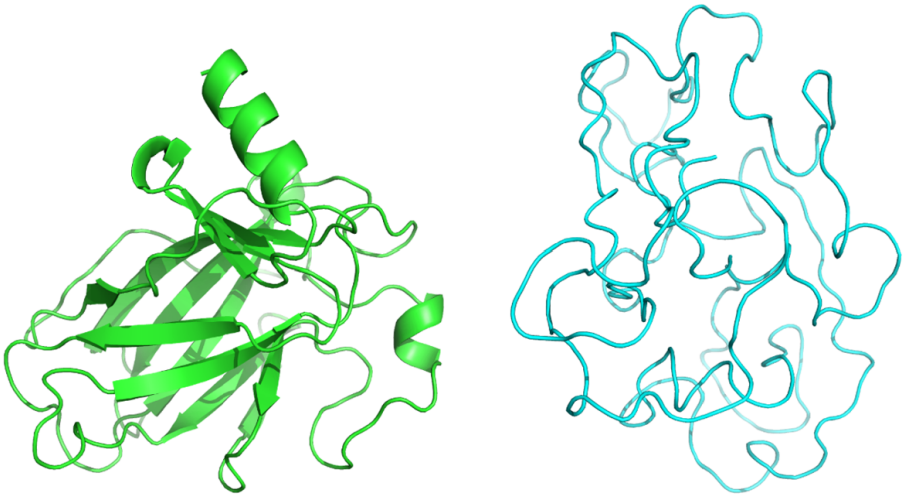


Figure 4.3: AlphaFold2-predicted wild-type structure (left) and the unfolded structure of p53 Q104H mutant (right) after applying high tem-perature MD simulation

represent a mixture of the wild-type and mutant forms. We use the *pmx* "develop" branch for this purpose (`https://github.com/deGrootLab/pmx` b0b5684 commit). The generated hybrid structure files are then processed with the GROMACS *pdb2gmx* command to add hydrogen atoms by using the Amber99SB-ILDN modified force field. This results in hybrid structure and topology files that are compatible with GROMACS for the alchemical MD simulations.

4.1.3 Thermodynamic cycle and DSSB protocol

The folding free energy change ($\Delta\Delta G$) upon point mutation can be calculated by using the thermodynamic cycle principle. In the cycle shown in Figure 4.1 (DSSB part), the difference $\Delta\Delta G = \Delta G_3 - \Delta G_4$ (vertical arrows) represents the change in folding free energy between the wild-type and mutant proteins. Directly computing the folding free energies ΔG_3 and ΔG_4 from simulations can be computationally challenging, particularly for larger peptides and proteins, due to the long folding times involved [92]. Nevertheless, we can more efficiently calculate the mutation-free energies through the indirect path of the "alchemical" mutations in the folded and unfolded states ($\Delta\Delta G = \Delta G_1 - \Delta G_2$), which involve morphing the wild-type amino acids to their mutated and then compute the corresponding free energy change. By performing this process in both folded and unfolded states (two horizontal arrows), we can leverage the fact that the total free energy change of the thermodynamic cycle is zero ($\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$).

In the implementation, we applied the DSSB simulation protocol (Figure 4.1) to maintain charge neutrality during alchemical transformations involving charge changes between the wild-type and mutant states. The first system consists of the $\text{folded}_{\text{wild-type}}$ and $\text{unfolded}_{\text{mutant}}$ proteins in a simulation box ($\lambda = 0$), while the second system contains the $\text{folded}_{\text{mutant}}$ protein and $\text{unfolded}_{\text{wild-type}}$ proteins in a simulation box ($\lambda = 1$). In each box, the two proteins are kept 3 nm apart by applying positional restraints on a single backbone atom near the center of mass, preventing them from diffusing closer together during simulations. The alchemical transformation from $\lambda = 0$ to 1, known as the "forward" process, transforms the $\text{folded}_{\text{wild-type}}$ into the $\text{folded}_{\text{mutant}}$, and simultaneously, the $\text{unfolded}_{\text{mutant}}$ into the $\text{unfolded}_{\text{wild-type}}$, which is referred to as the "reverse" process ($\lambda = 1$ to 0). This protocol conducts two independent simulations (forward and reverse), generating work distributions that can be used later to calculate the $\Delta\Delta G$ value.

4.1.4 MD simulations and free-energy calculations

All MD simulations were executed using the GROMACS-2022 [176] (single-precision) with the Amber99SB-ILDN [200] force field and TIP3P [202] water model. The hybrid structures with modified force field generated by *pmx* were used as the input. The MD simulation protocol is based on previous work by Patel et al., [168] but with shorter simulation times as we focused on single, smaller

proteins rather than large protein complexes. For equilibrium MD simulations, we set up two simulation boxes for conducting forward and backward transition states for each mutation. Both states were solvated with dodecahedron water boxes, and Na^+ and Cl^- ions were added at a concentration of 0.15 M to balance the net charge. The simulation boxes underwent energy minimization for 10,000 steps utilizing the steepest descent algorithm. NVT ensemble simulations, followed by NPT ensemble simulations, were then executed for 100 ps in each box. Throughout the MD simulation, the pressure and temperature were consistently maintained at 1 atm and 300 K. A time step of 2 fs was employed with snapshots saved every 10 ps. Lastly, final production MD simulations were conducted for 12 ns. Following equilibrium MD simulations, fast-growth nonequilibrium alchemical simulations were conducted to approximate $\Delta\Delta G$. From each equilibrated MD simulation, the initial 2 ns of the trajectory was discarded, while the remaining 10 ns was utilized to generate 100 snapshots (one every 100 ps). These snapshots initiated nonequilibrium simulations with a 100 ps transition time, during which λ shifted continuously from 0 to 1 or from 1 to 0. The λ value change speed was set at 2×10^{-5}/fs for all forward and backward transitions. For the soft-core interactions, we used the soft-core potential proposed by Gapsys et al. [203] The $\Delta\Delta G$ was estimated by determining the overlap of forward and backward work distributions by using the BAR method provided by the *pmx* script. This estimator yields a more precise approximation without presuming the work distributions' shape. Aldeghi et al. [92] suggest that the most reliable way to assess the precision of free energy estimates is to repeat the entire procedure, including both equilibrium and nonequilibrium simulations, multiple times. In their example, they repeated the entire calculation five times and took the average free energy values. In this work, we opted to repeat the whole process three times to reduce computational costs while still maintaining accuracy, as the calculations were performed on a large data set.

4.2 Results

To assess the performance of the proposed approach, we utilized two common metrics: root-mean-squared error (RMSE) and Pearson correlation coefficient (PCC). For the frataxin and p53 data sets, the proposed method achieved RMSE values of 1.67 kcal/mol and 1.11 kcal/mol, respectively, and exhibited notably high correlation values (PCC = 0.90 and PCC = 0.81) for both data sets. This suggests that the proposed approach can accurately predict whether a mutation leads to destabilization or stabilization (Figure 4.4). The predicted $\Delta\Delta G$ values for each mutation in the corresponding data sets are provided in Tables 4.2 and 4.3. These results were obtained while applying the proposed protocol, which uses a GxG as the unfolded mutant structure for conserving mutations and an AlphaFold2-predicted mutant structure unfolded in high-temperature simulation (AF2HT) for charge-changing mutations. Additionally, we also assessed the performance using GxG-only and AF2HT-only protocols on all mutations to gauge the effectiveness of the proposed protocol. The

Table 4.2: Predicted $\Delta\Delta G$ values for frataxin data set by the proposed method.

Mutation	**Predicted $\Delta\Delta G$ (kcal/mol)**	**Experimental $\Delta\Delta G$ (kcal/mol)**
D104G	0.32	0.40
A107V	-0.94	0.00
F109L	-2.12	-2.80
Y123S	-4.23	-5.10
S161I	1.13	-3.10
W173C	-10.95	-9.50
S181F	-2.98	-3.00
S202F	0.16	-0.20

frataxin data set comprises 1 charge-changing and 7 conserving mutations, while the p53 data set includes 16 charge-changing and 26 conserving mutations. Both protocols demonstrated inferior performance on both data sets, as shown in Table 4.4. While the GxG-only protocol achieved better performance on conserving mutations, it performed poorly on charge-changing mutations. This aligns with *pmx* validations, where charge-changing mutations exhibited larger fluctuations. On the other hand, the AF2HT-only protocol showed the opposite results, with better accuracy on charge-changing mutations but low accuracy on conserving mutations. The predicted $\Delta\Delta G$ values for by each protocol in the corresponding data sets are provided in Tables A.1, A.2, A.3, and A.4 (see appendix A).

The GxG structure is effective for conserving mutations due to its simplicity and stability. However, this simplicity falls short when it comes to charge-changing mutations, as it does not adequately capture the more complex electrostatic interactions. Charge-changing mutations can significantly alter the local and global electrostatic environments in the protein structure. Conversely, the AF2HT structure is adept at handling charge-changing mutations as it considers both the global effects and intricate electrostatic interactions arising from alterations in amino acid charges.

While the AF2HT structure accounts for global mutation effects, it may struggle to capture the subtle local effects of conserving mutations, which do not involve significant charge changes or substantial local environmental alterations. Consequently, the AF2HT method might overestimate the impact of conserving mutations on the protein stability. Thus, combining both methods provided a more accurate approach to addressing different mutation types.

4.2.1 Performance comparison with other predictors

Considering the hard limit of $\sim$1 kcal/mol numerical accuracy [75], recent benchmarks on the frataxin and p53 data sets indicate that existing methods demonstrate relatively low performance, with high average prediction errors, as evidenced by RMSE values of 3.27 kcal/mol and 1.53 kcal/mol, respectively [154]. In comparison with existing methods (Table 4.5), the proposed method achieves superior performance, obtaining the lowest RMSE values (1.67 kcal/mol and 1.11 kcal/mol, respectively). The proposed method also attains exceptional PCC values (0.90 and 0.81, respectively),

Table 4.3: Predicted $\Delta\Delta G$ values for p53 data set by the proposed method.

Mutation	**Predicted $\Delta\Delta G$ (kcal/mol)**	**Experimental $\Delta\Delta G$ (kcal/mol)**
Q104H	0.40	0.24
Q104P	-0.18	0.11
T123A	-1.09	-0.13
A129D	-2.26	-0.70
A129E	-1.40	-0.38
A129S	0.21	-0.19
M133L	0.87	0.30
F134L	-4.17	-4.78
V143A	-3.43	-3.50
L145Q	-2.56	-2.98
D148E	-0.05	-0.43
D148S	0.09	0.22
T150P	0.02	-0.08
P151S	-4.78	-4.49
V157F	-2.48	-3.88
Q165K	-0.40	-1.27
Q167E	-0.70	-0.43
H168R	-4.35	-2.75
R174K	0.04	-0.22
R175A	-1.22	-0.73
R175H	-3.61	-3.52
C182S	0.07	0.16
I195T	-4.74	-4.12
L201P	-0.61	0.35
V203A	-1.25	0.49
L206S	0.18	-0.10
Y220C	-2.71	-3.98
D228E	-0.39	0.05
I232T	-2.92	-3.19
Y236F	0.23	0.27
M237I	0.08	-3.18
N239Y	-0.05	1.49
C242S	-0.23	-3.07
G245S	0.56	-1.21
R248Q	0.69	-1.87
R249S	-2.98	-1.92
I255F	-2.45	-3.29
S260P	0.37	-0.32
N268D	1.37	1.21
F270C	-3.75	-4.54
R273H	-0.18	-0.45
R282W	-3.76	-3.30

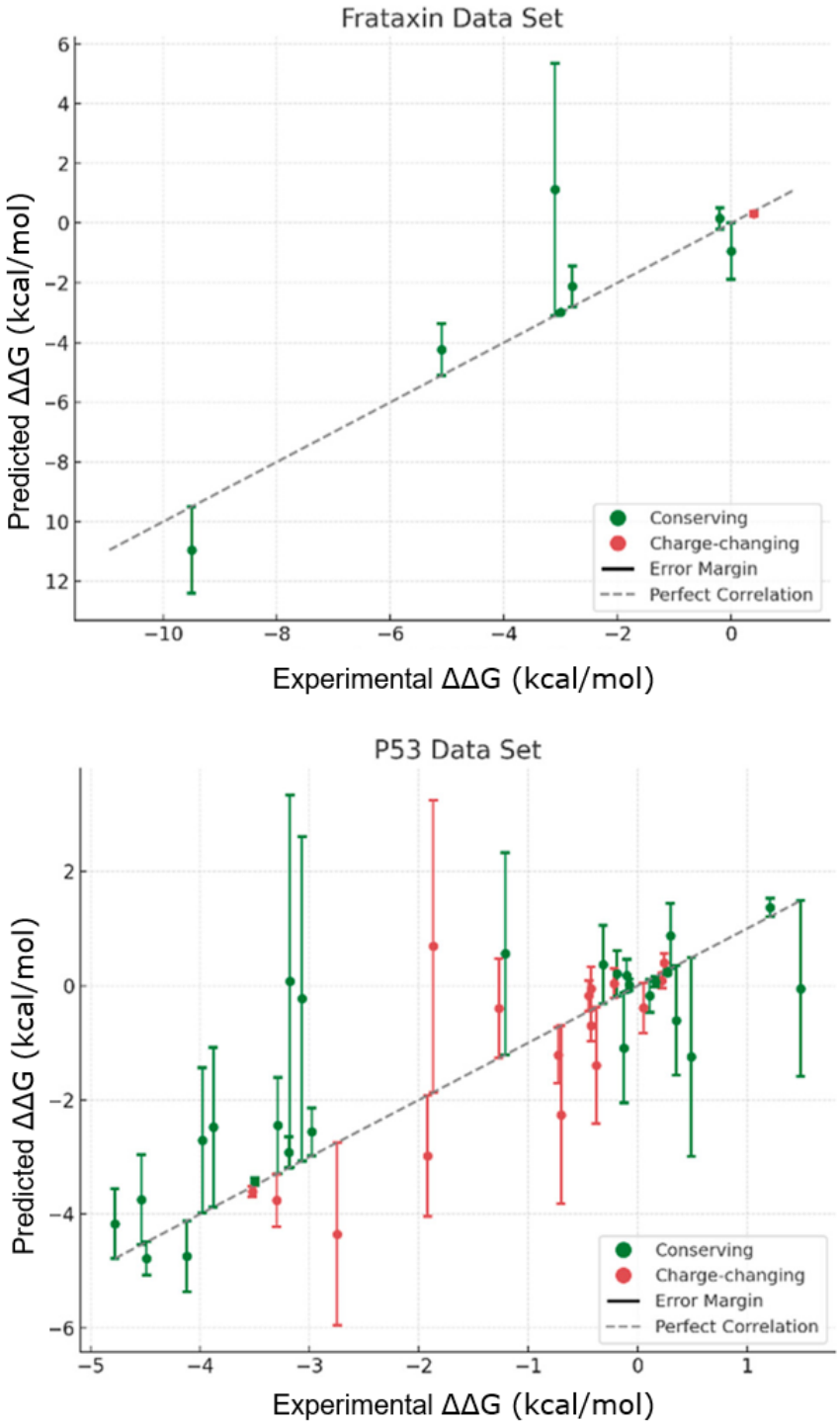


Figure 4.4: Correlation plot between predicted and experimental for the frataxin (top) and p53 (bottom) data set. The green and red colors represent conserving and charge-changing mutations, respectively. The error margin is the difference between the experimental and predicted $\Delta\Delta G$ values.

Table 4.4: Performance Comparison between the Three Protocols for Frataxin and p53 Data Sets[a]. The RMSE values are in kcal/mol

	Frataxin (All)		**P53 (All)**	
	RMSE	PCC	RMSE	PCC
Proposed	**1.67**	**0.90**	**1.11**	**0.81**
GxG	**1.67**	**0.90**	2.08	0.60
AF2HT	2.29	0.76	1.62	0.71

	Frataxin (Conserving)		**P53 (Conserving)**	
	RMSE	PCC	RMSE	PCC
GxG	**1.78**	**0.89**	**1.18**	**0.84**
AF2HT	2.29	0.76	1.90	0.69

	Frataxin (Charge-changing)		**P53 (Charge-changing)**	
	RMSE	PCC	RMSE	PCC
GxG	-	-	3.01	0.27
AF2HT (charge changing)	-	-	**0.98**	**0.79**

[a]Frataxin data set only contains one charge-changing mutation.

which are the best for the frataxin data sets and comparable to DDGun and PROST (0.89 and 0.91, respectively) for the p53 data sets. It is important to note that some predictions generated by other methods were derived from existing literature [147, 154, 204]. In the frataxin data set, the proposed method achieves the lowest squared error for five of the eight mutations (Table A.5, see appendix). Interestingly, the proposed approach obtains a significantly lower squared error for the W173C mutation, whereas all existing methods exhibit a high squared error. Despite being significantly lower than existing methods, RMSE of 1.67 kcal/mol, it is still relatively high when compared to the 1 kcal/mol barrier. This can be largely attributed to the small sample size in the frataxin data set, which contains only eight mutations. Within this small sample, a single substantial error in prediction, as observed in the S161I mutation, can considerably skew the overall error measure. When we exclude this mutation, the average error falls significantly to 0.79 kcal/mol. This mutation also presents a significant challenge for all existing methods tested, as none have succeeded in predicting it accurately. Moreover, supervised learning-based methods, including PROST, typically exhibit high PCC values but low accuracy, as they are primarily trained using data sets heavily skewed toward neutral mutations. This training bias enabled them to successfully predict whether a mutation is stabilizing or destabilizing (positive or negative), thus achieving high correlation values. However, it underestimated the effect of highly stabilizing or destabilizing mutations, resulting in low prediction accuracy. For the p53 data set, the proposed method achieves a squared error of less than 1 kcal/mol for 30 of the 42 mutations, whereas the average squared error for existing methods is attained for only 21 of the 42 mutations (Table A.6, see appendix). This can be attributed to highly stabilizing/destabilizing cases ($|\Delta\Delta G| \geq 1$ kcal/mol), where existing methods yield significantly higher squared error.

Table 4.5: Performance Comparison on p53 and Frataxin Data Sets[b]. The RMSE values are in kcal/mol

Predictor	**Frataxin**		**P53**	
	RMSE	PCC	RMSE	PCC
SDM	3.73	0.42	1.48	0.67
mCSM	3.29	0.53	1.40	0.67
DUET	3.33	0.53	1.39	0.68
PoPMuSic	-	-	1.58	0.56
SDM2	3.48	0.54	1.56	0.68
iStable	3.43	0.70	1.69	0.36
INPS	3.30	0.64	1.51	0.69
EASE-MM	2.96	0.85	1.64	0.59
BoostDDG	-	-	1.89	0.49
SAAFEC-SEQ	3.41	0.72	1.49	0.64
ACDC-NN-Seq	2.83	0.88	1.62	0.62
DynaMut2	3.31[a]	0.56[a]	1.28	0.75
DDGun	2.24[a]	0.89[a]	1.45	0.69
MU3DSP	-	-	1.41	0.72
PROST	3.02	**0.91**	1.5	0.67
Proposed	**1.67**	0.90	**1.11**	**0.81**

[a]Results obtained from the Web server.
[b]Best performance is marked with bold letters.

4.2.2 Convergence analysis

In this study, we employed the BAR method as an estimator for free energy differences and associated uncertainties in the alchemical MD simulations [205, 206, 207]. To evaluate the convergence and reliability of the free energy calculations, we performed a thorough convergence analysis using both analytical error and bootstrap error. The analytical error associated with the BAR estimator can escalate quickly, even with a small overlap between the work distributions. A substantial value for the analytical estimator may act as a reliable indicator of insufficient convergence during nonequilibrium transitions. If the two error estimates are consistent with each other, it suggests that the simulations have provided sufficient sampling and the calculated free energy differences are reliable.

The analytical and bootstrap errors for the frataxin and p53 data sets indicate that the current protocol is suitable for these systems (Figure 4.5). Most errors have values close to zero, demonstrating agreement between analytical and bootstrap methods, with a few exceptions showing high analytical error in one or all three independent simulations (Tables A.7 and A.8, see appendix). In such cases, we reperform the equilibrium and nonequilibrium simulations and recalculate the $\Delta\Delta G$ values. Notably, the analytical errors remain exceptionally high in all three simulations for some charge-changing mutations, even though they do not impact the predicted $\Delta\Delta G$ accuracy as found in H168R, R175A, R175H, and R282W mutations in the p53 data set.

In cases of substantial analytical error, the uncertainty in ΔG estimates may be attributed to insufficient overlap between the work distributions, as pointed out by Aldeghi et al. [92]. To maintain

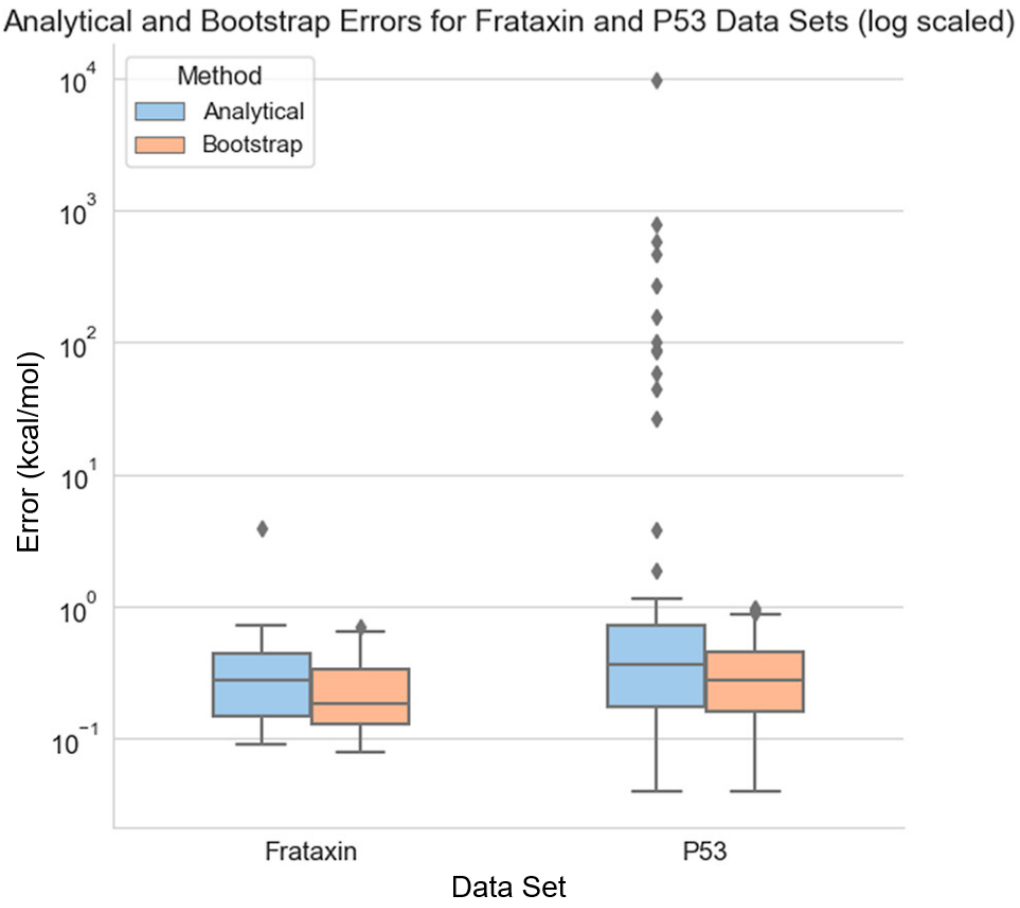
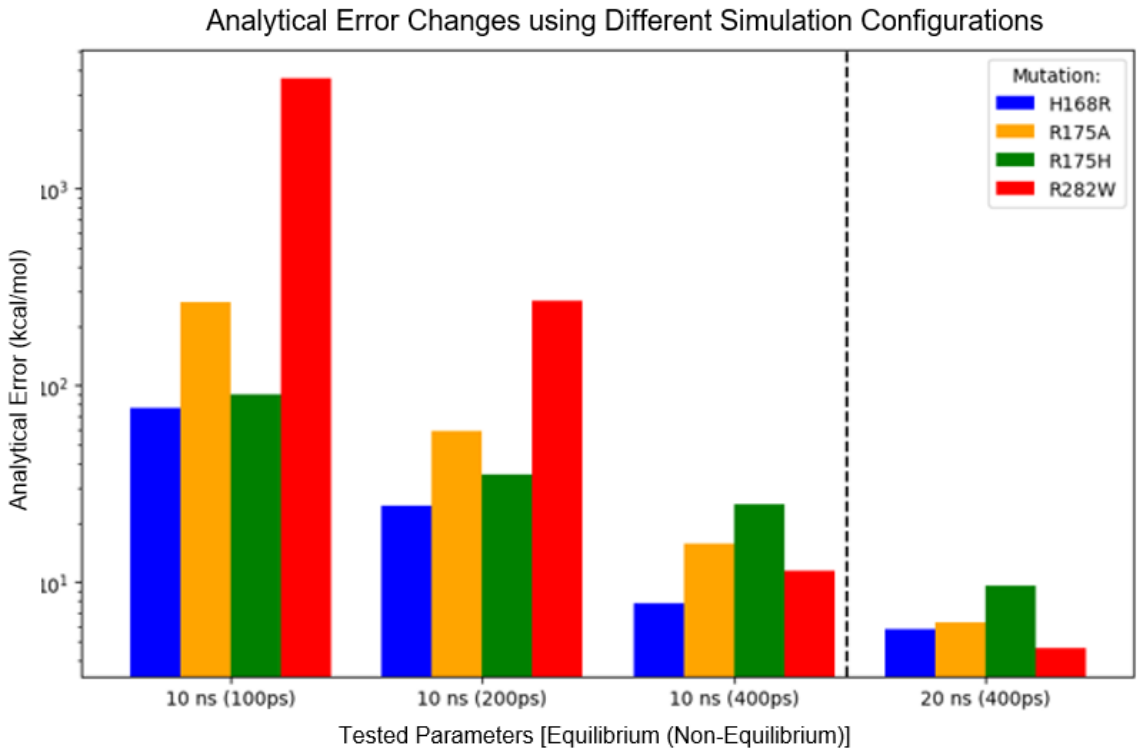


Figure 4.5: Log-scaled analytical (blue) and bootstrap (orange) error on frataxin (left) and p53 (right) data sets.

equilibrium and thus increase overlap, it may be beneficial to execute slower transitions, while conducting additional nonequilibrium transitions can enhance the likelihood of observing low-dissipation work values. In the efforts to rectify this issue in instances of high analytical error, we conducted multiple longer equilibrium and nonequilibrium simulations across varied configurations. As demonstrated in Figure 4.6, the extension of simulation time in the equilibrium and nonequilibrium states effectively minimized analytical errors. It was noted that, although the squared errors augmented with an increase in nonequilibrium simulation time, they significantly decreased when the equilibrium simulations were prolonged. We anticipate that simultaneously extending the simulation time for both equilibrium and nonequilibrium conditions could reduce analytical errors while preserving prediction accuracy.



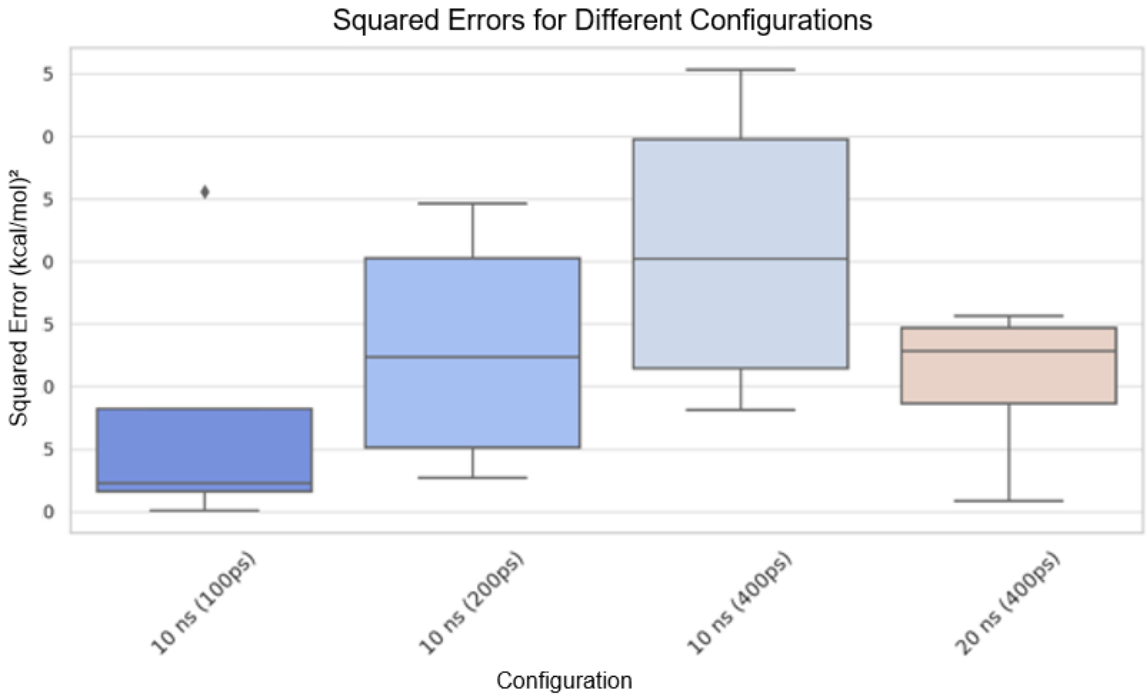


Figure 4.6: Top: Log-scaled changes in analytical errors average across mutation cases with unusually high analytical errors, assessed by using extended equilibrium (and non-equilibrium) simulation configurations. Bottom: Distribution of squared errors across vari-ous simulation configurations, illustrating the effects of different adjustments on error magnitudes.

4.3 Discussion

In this study, we applied an improved alchemical MD simulation-based method for predicting protein stability upon point mutations. The alchemical method is based on the *pmx* hybrid topologies and double-system/single-box developed by Gapsys et al. [170, 172]. The proposed approach showed robust performance on both conserved and charge-changing mutation cases, significantly outperforming existing state-of-the-art methods in overall performance. The approach does not require any training process, making it robust against limitations associated with supervised learning-based methods. Supervised learning-based methods often rely on complex features at the expense of explainability and are highly dependent on the quality of the training data. This can be observed in the results for frataxin and p53 data sets, where existing supervised learning-based methods exhibit low accuracy on these two challenging blind test sets. The proposed rigorous approach eliminates the need for complex feature extraction steps as it solely requires protein structure information, which can be acquired from PDB or AlphaFold2 predictions if crystal structures are unavailable.

4.3.1 Reverse mutations

To further validate the proposed method and assess the reliability of mutant structures predicted by AlphaFold2, reverse mutations were performed on the p53 data set. This step was crucial in highlighting the symmetry inherent in our approach, as reverse mutations necessitate the use of the mutant structure itself. This also probes the reliability of the predicted mutant structures, in this case, those derived from AlphaFold2. This additional validation bolsters the confidence in the proposed approach and the practicality of the predicted mutant structure, especially those obtained from AlphaFold2, for studying protein mutations. As demonstrated in Figure 4.7, there is a notable internal consistency between the predicted forward and reverse mutations in most cases. The predicted $\Delta\Delta G$ values for each mutation are provided in Table A.9 (see appendix A).

Some outliers were observed, primarily resulting from charge-changing mutations. Initially, we suspected that these discrepancies arose from cases where either the forward or reverse mutations had unusually high analytical errors. However, even after attempting to reduce the high analytical errors through simulations with longer configurations, the prediction accuracy did not show substantial improvement. Upon further investigation, we found that the annotated mutation cases were known as tumorigenic mutations in p53 that are located within the DNA-binding substructure (with the exception of D148S). In this region, mutations may exert a structural effect that destabilizes the local conformation or even induces global denaturation [208]. In direct mutation, the proposed method can accurately predict $\Delta\Delta G$ values where the AlphaFold2-predicted mutant structure is used for the unfolded state and is subsequently unfolded through high-temperature simulation. Conversely, in the reverse mutation scenario, the AlphaFold2-predicted mutant structure is employed in the folded state, and this appears to be the primary issue. This observation aligns with findings from previous

studies [181] indicating that AlphaFold2 may not accurately predict the structural destabilization in these particular mutation regions. As AlphaFold2 predicts these regions in a folded conformation, it consequently affects the accuracy of the reverse mutation prediction.

We conducted a further evaluation of the internal consistency performance using the Ssym data set. The Ssym data set comprises 634 variants derived from 357 structures, corresponding to 13 distinct clusters. [209] This data set includes 342 mutations along with their reverse. It is commonly utilized to assess the anti-symmetric bias of prediction methods, as it provides crystal structures for both direct and reverse mutations. Given the computational demands of simulating the entire Ssym data set, we selected a substantial and representative subset. This subset ensured coverage across all proteins, with a minimum of three samples per protein when available, and included mutations spanning a comprehensive range of $\Delta\Delta G$ values (destabilizing-neutral-stabilizing). The subset comprising 39 pairs of direct and reverse mutations (30 conserving and 9 charge-changing) leveraged the Ssym data set crystal structures for both wild-type and mutant forms (Table A.10, see appendix). Thus, the performance of the proposed protocol, especially for charge-changing mutations, could be evaluated. The proposed method demonstrated internal consistency, as shown in Figure 4.8. Since the crystal mutant structure is used, the proposed method does not seem to be affected by the limitations associated with the predicted mutant structure, such as those found with AlphaFold2 in the p53 DNA-binding region, signifying that the proposed charge-changing mutation protocol is reliable for both direct and reverse mutations. We also evaluated the prediction performance for both directions with DDGun and DDGun3D, unsupervised methods that achieved good performance in the frataxin data set. Most of the supervised learning-based approaches were trained on the S2647 data set, where many of the proteins also exist in the Ssym data set. Thus, a good performance in the Ssym data set might come from overfitting. The performance comparison showed that the proposed method showed comparable performance with DDGun and DDGun3D in terms of RMSE (direct: 1.61, 1.63, 1.59 kcal/mol; reverse: 1.73, 1.63, 1.62 kcal/mol, respectively) but superior Pearson correlation (direct: 0.79, 0.62, 0.68; reverse: 0.75, 0.63, 0.67) (Table 4.6). It is worth noting that the D10N and its inverse N10D mutations caused significant prediction errors in the proposed method. When these data are excluded, the proposed method achieves the best performance in both RMSE and Pearson correlation for both direct and reverse mutations. We were unable to identify a clear reason for the failure of the proposed method in predicting for both D10N and N10D mutations. These specific mutations present challenges for the proposed method, which will be further explored in a subsequent subsection. When focusing only on charge-changing mutations, the proposed method showed worse RMSE but significantly better Pearson correlation in both directions. Excluding the problematic D10N and N10D mutations, the proposed method exhibited significantly improved RMSE and Pearson performance.

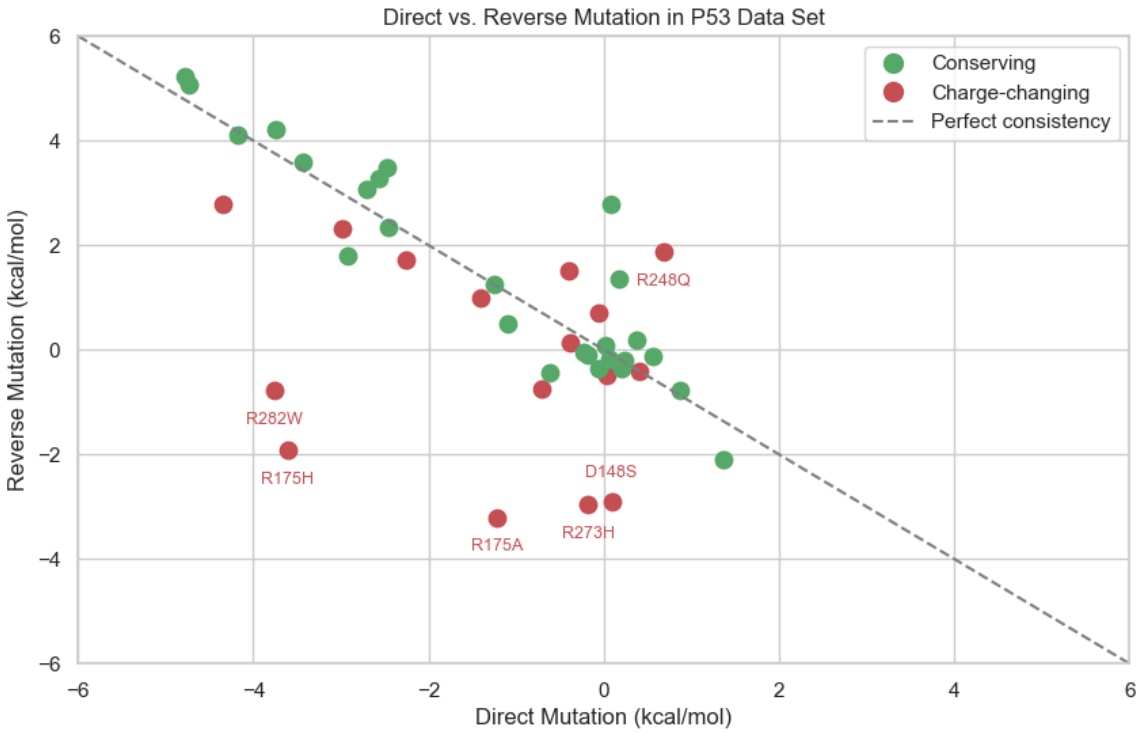


Figure 4.7: Internal consistency of direct and reverse mutations of the p53 data set. The green and red colors represent conserving and charge-changing mutations, respectively. The annotated points are the mutation cases where high inconsistencies are found.

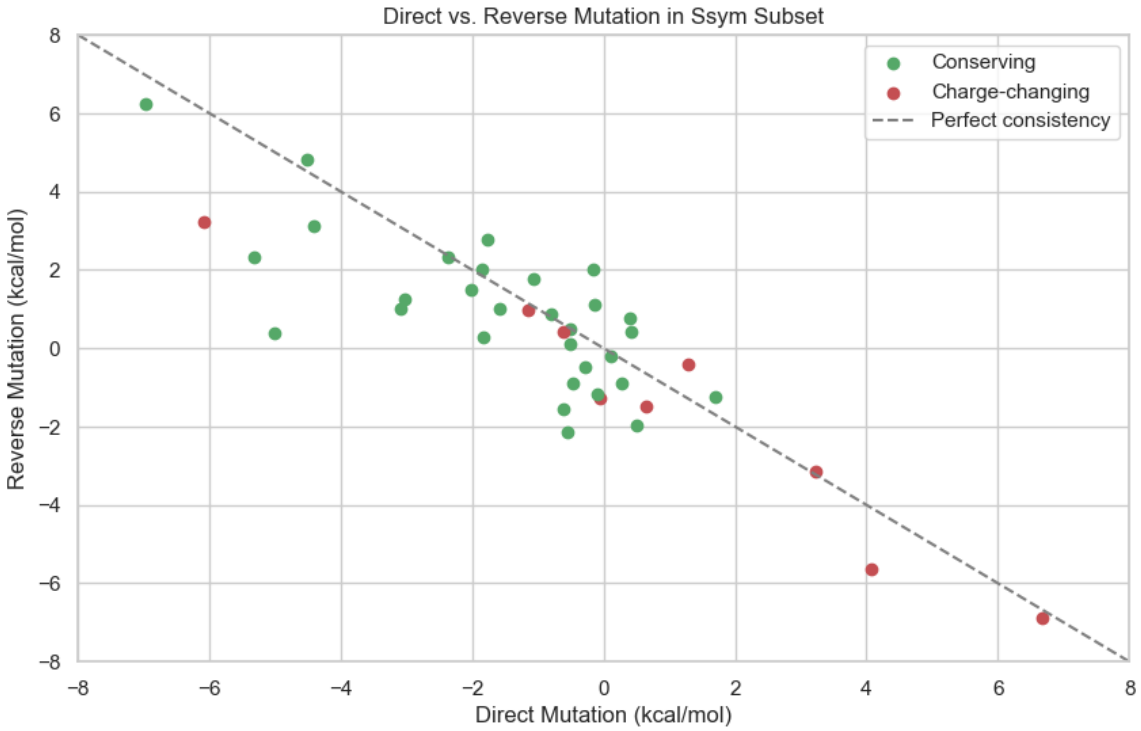


Figure 4.8: Internal consistency of direct and reverse mutations of the Ssym subset. The green and red colors represent conserving and charge-changing mutations, respectively.

Table 4.6: Comparative performance analysis of our proposed method with DDGun and DDGun3D in the Ssym subset for both direct and inverse mutations. The left section includes the entire subset, while the right excludes the D10N and its inverse N10D mutations. The top table represents all conserving and charge-changing mutation cases, while the bottom table focuses on charge-changing mutations only. Best performances are highlighted with bold letters. All RMSE values are in kcal/mol

All data (conserving and charge-changing mutations)

Method	Direct		Reverse	
	RMSE	Pearson	RMSE	Pearson
Proposed	1.61	0.79	1.73	0.75
DDGun	1.63	0.62	1.63	0.63
DDGun3D	1.59	0.68	1.62	0.67

Charge-changing mutations only

Method	Direct		Reverse	
	RMSE	Pearson	RMSE	Pearson
Proposed	2.42	0.71	2.49	0.66
DDGun	2.31	0.33	2.31	0.30
DDGun3D	2.18	0.40	2.13	0.48

D10N and N10D excluded (conserving and charge-changing mutations)

Method	Direct		Reverse	
	RMSE	Pearson	RMSE	Pearson
Proposed	1.33	0.83	1.45	0.79
DDGun	1.64	0.63	1.64	0.64
DDGun3D	1.60	0.68	1.63	0.68

Charge-changing mutations only (D10N and N10D excluded)

Method	Direct		Reverse	
	RMSE	Pearson	RMSE	Pearson
Proposed	1.56	0.85	1.58	0.81
DDGun	2.40	0.34	2.40	0.31
DDGun3D	2.27	0.43	2.22	0.50

Overall, while the proposed method exhibits internal consistency especially in direct mutation scenarios, caution is advised in reverse mutation cases. This consideration is particularly pertinent when relying on predicted folded mutant structures such as those from AlphaFold2, which may not accurately represent structural destabilization in certain mutation regions. Future work will aim to address these challenges and further refine the method for broader applicability.

4.3.2 Mutation types

Considering the high errors observed in the predictions of D10N and N10D mutations, we sought to investigate whether specific mutation types consistently pose challenges for the alchemical method. To this end, we examined the impact of different mutation types on the prediction squared error, taking into account their inherent characteristics and the potential limitations of the alchemical approach. Our analysis drew on results from frataxin (direct), p53 (direct), and the Ssym subset

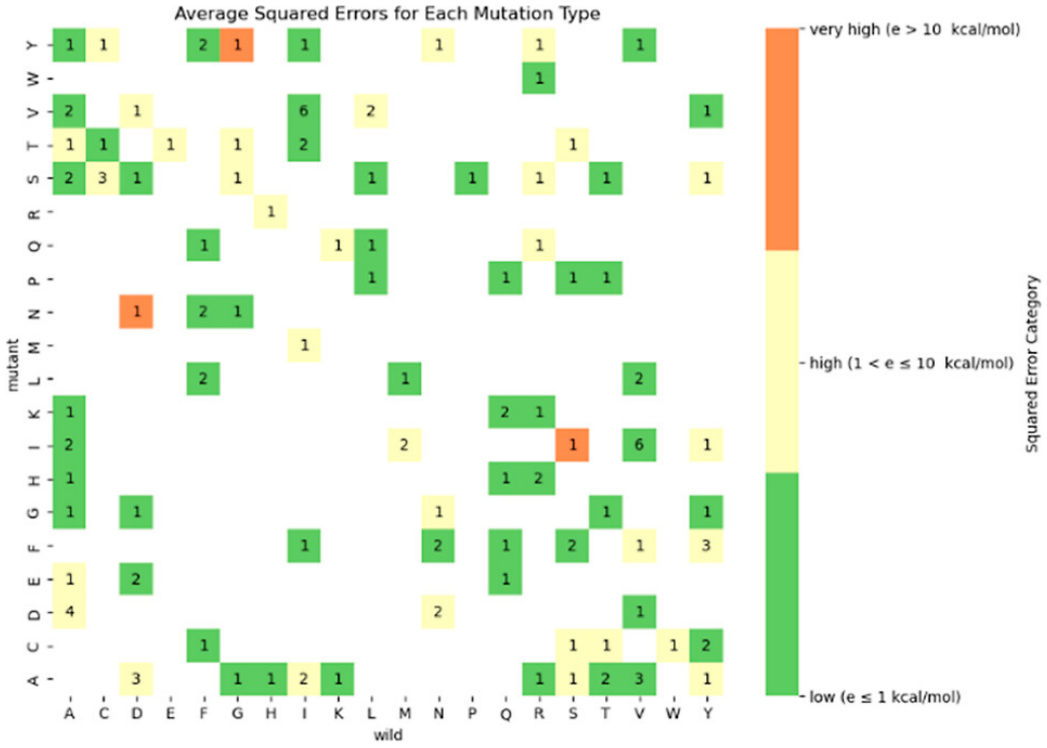


Figure 4.9: Average squared errors, categorized by magnitude, in the frataxin (direct), p53 (direct), and Ssym subset (direct and reverse) data sets.

(both direct and reverse). As depicted in Figure 4.9, this investigation revealed distinctive patterns of prediction error across various mutation types. Notably, mutations involving cysteine yielded high errors, likely attributed to the difficulties in capturing the breaking and reforming of disulfide bonds, which introduce significant free energy barriers into the simulations. Similarly, transitions from aspartic acid (D) to alanine (A) and vice versa, as well as transitions from D to asparagine (N) and N to D, have posed challenges. We suspect that these issues may stem from inherent limitations in the *pmx* library, as indicated in the original work [170], where $\Delta\Delta G$ values for charge-changing mutations involving aspartate tend to fluctuate significantly. Further refinement, including adjustments to the simulation parameters and possibly extending the simulation time for better convergence, might be necessary to address these challenges. It is important to emphasize, however, that our analysis was constrained by the available data. The limited number of data points precluded the establishment of clear statistical tendencies among the various mutation types.

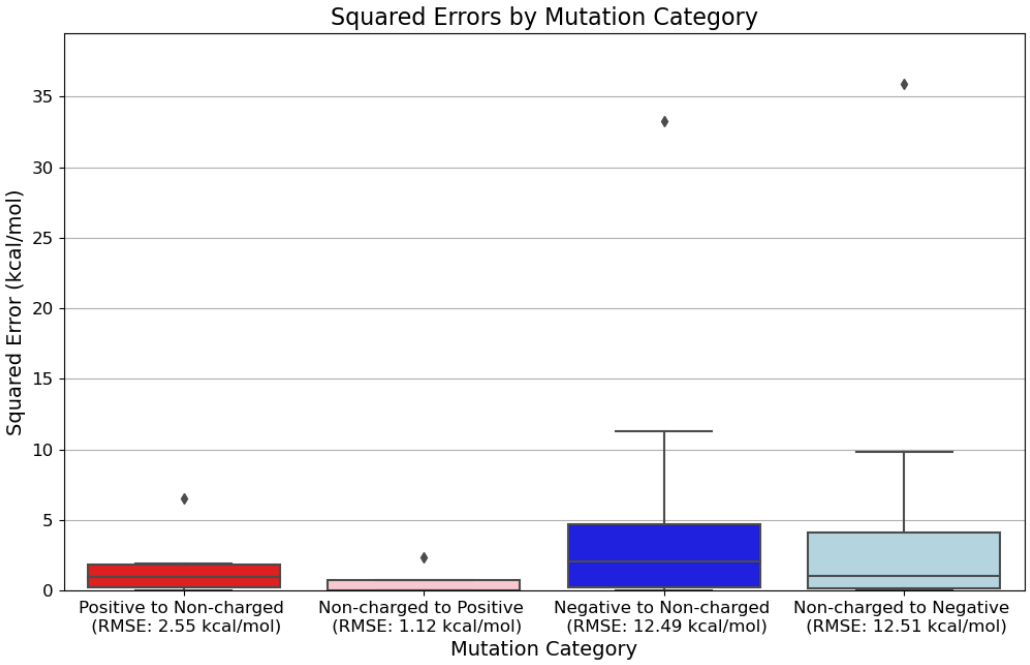


Figure 4.10: Squared errors for charge-changing mutation categorized by mutation type

Charge-changing mutation

Additionally, we conducted an investigation focused specifically on charge-changing mutations to assess our proposed workflow for handling these cases. Our analysis categorized these mutations into four types:

- Positive to Negative and vice versa (no cases observed)
- Positive to Non-charged and vice versa
- Negative to Non-charged and vice versa
- Non-charged to Positive and vice versa

No instances of positive to negative mutations (or vice versa) were observed in the tested sets. This may be due to their rare occurrence or exclusion in large mutational scan analyses. Figure 4.10 presents the squared errors for the other categories of charge-changing mutations. The small sample size for each category (less than 10 instances) limits our ability to draw definitive conclusions. However, most squared errors were under 1 kcal/mol, except for specific cases like D to N, D to A, and their vice versa. As mentioned earlier, these trends align with the challenges associated with transitions involving aspartic acid, potentially stemming from limitations in the *pmx* library. This analysis highlights the need for more extensive data to establish clearer statistical trends among various types of charge-changing mutations and to enhance the reliability of our conclusions.

4.3.3 Case study

The proposed method demonstrated exceptional performance on two blind test sets, frataxin (RMSE = 1.67 kcal/mol and PCC = 0.90) and p53 (RMSE = 1.11 kcal/mol and PCC = 0.81), outperforming existing methods. The notable advantage of the proposed method is evident in the frataxin test set, where the RMSE is significantly lower compared to other predictors where they attain the average RMSE >3 kcal/mol. This superior performance can be attributed to the ability of the proposed method to predict $\Delta\Delta G$ values outside the typical range found in training data for machine learning-based predictors, which is $-5 \leq \Delta\Delta G \leq 5$ kcal/mol. The proposed method is not limited by any specific range and can predict higher $\Delta\Delta G$ values, as seen in the Y123S and W173C mutations (experimental $\Delta\Delta G = -5.1$ kcal/mol and -9.5 kcal/mol, respectively). This observation is also true for the p53 test set, where the proposed method accurately predicts highly destabilizing cases (Figure 4.11) with squared experimental $\Delta\Delta G$ values around -3.5 kcal/mol or above, unlike existing methods that yield significant squared errors in these cases (Table A.11, see appendix). This can be attributed to the fact that training sets for existing methods are often dominated by neutral mutations with an $\Delta\Delta G$ range of -1 to 1 kcal/mol.

For instance, consider the R282W mutation as the case study, where the experimental $\Delta\Delta G$ = -3.30 kcal/mol. Our proposed method accurately predicted this as a highly destabilizing mutation with a $\Delta\Delta G$ = -3.76 kcal/mol. In contrast, existing methods incorrectly predicted it as a neutral mutation, resulting in an average squared error of 10.89 kcal/mol. Even the most accurate among them, DynaMut2, achieved a squared error of 3.13 kcal/mol, which is significantly higher than the 0.21 squared error of our proposed method. This discrepancy is likely due to the R282W mutation being a tumorigenic mutation in p53, located within the DNA-binding substructure. The mutation induces structural alterations, particularly the loss of interactions between the H2 helix and L1 Loop as shown in Figure 4.12. [88] Our proposed method successfully captures these dynamic changes from the protein structre, leading to a more accurate prediction.

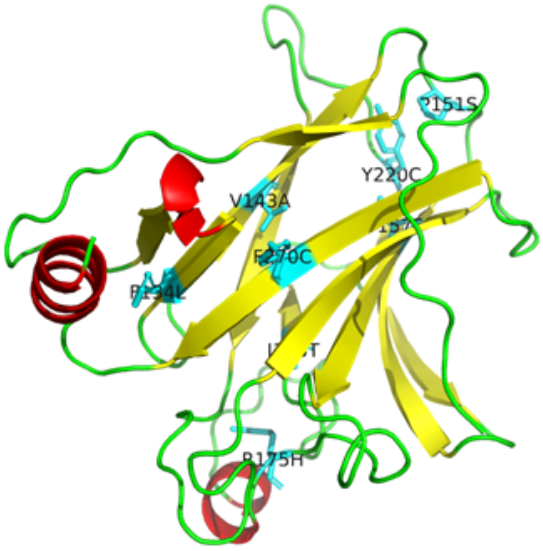


Figure 4.11: The highly destabilizing mutation position in p53 test set with squared experimental $\Delta\Delta G$ values around -3.5 kcal/mol or larger

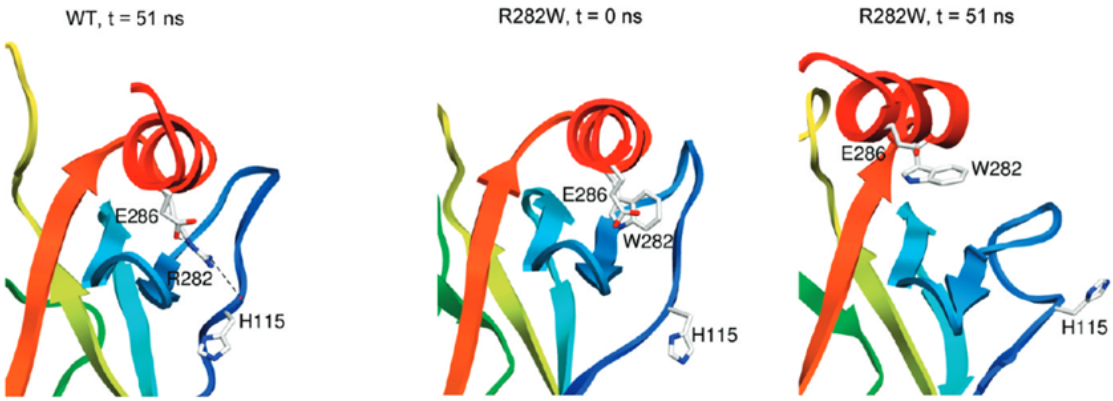


Figure 4.12: Impact of the R282W mutation on the local structure: detailed view of the H2 Helix and L1 loop in silico system. Reprinted with permission from [88]. Copyright 2011 American Chemical Society. The side chains of residue 282 and adjacent critical side chains are depicted as sticks, with coloring based on atom type.

Table 4.7: Run Time Comparison between Top-Performing Methods on Frataxin and p53 Data Sets[b]

Predictor	**Run Time**
DDGun (webserver)	<2 min
Dynamut2 (webserver)	<1 min
PROST[a]	11–13 min
GxG	2(11–13 h)
AF2HT	2(17–20 h)

[a]Results are taken from the original literature.
[b]Note: the running time in the parentheses is for a single direction (forward/reverse).

A notable limitation was identified when AlphaFold2 was employed for predicting the structural effect of mutations in the p53 DNA-binding region. AlphaFold2's predicted mutant structure works effectively when used for the unfolded state, where the mutant structure is unfolded using high-temperature simulation. This aligns with previous findings where AlphaFold2 conveys protein dynamics information obtainable from MD simulation [183]. However, it appears to inaccurately predict structural destabilization in the folded state, possibly due to AlphaFold2 predicting these specific regions in a folded conformation. Such predictions could inherently bias toward stability, affecting the accuracy of reverse mutation predictions. These findings highlight both the potential of AlphaFold2's structure dynamics and limitations that caution the use of predicted mutant structures, especially for reverse mutations. The observations suggest that further refinement of prediction methodologies or exploration of alternative structure prediction techniques may be warranted.

The major limitation of the rigorous approach, including the proposed method, is their substantial computational requirements. Thus, it is essential to compare the runtime of the proposed method with existing state-of-the-art methods to understand the trade-off between accuracy and runtime. We evaluated the runtime of the GxG and AF2HT protocols on the frataxin data sets on the workstation (AMD EPYC 7302 16-Core Processor, two NVIDIA A100-PCIE-40GB). In the test cases, frataxin and p53 structures have $\leq$200 residues (127 and 196, respectively). Larger proteins might increase the system size, resulting in longer execution times. Table 4.7 displays the computational runtimes of the existing methods and the proposed method. The results show, as expected, that the proposed rigorous approach-based method yielded significantly longer running time compared to existing methods, which only took minutes. Given the potential of a rigorous approach to deliver higher accuracy, the execution time, though longer, is still within feasible limits, making this trade-off a reasonable one in our view. Nonetheless, advancements in computing power and GPU parallelization facilitate faster and more practical implementation of rigorous methods. We anticipate that future developments in those fields and distributed cloud computing [210] will also help address computational resource limitations.

4.4 Summary

This chapter presented a novel workflow for rigorous alchemical method and comprehensive comparison between existing supervised learning-based state-of-the-art methods in predicting protein stability changes due to point mutations, particularly focusing on the challenging frataxin and p53 data sets. The enhanced alchemical approach, utilizing *pmx* hybrid topologies within a double-system/single-box setup, demonstrated significant advancements. For conserving mutation cases, the method employed capped tripeptides, while for charge-changing mutations, we introduced a new protocol that utilized predicted mutant structures derived from AlphaFold2.

A key finding of this study is the superior performance of the proposed alchemical method in accurately predicting $\Delta\Delta G$ values for the frataxin and p53 blind test sets, outperforming existing state-of-the-art techniques. This success is particularly notable given that the method does not rely on a training process and is therefore not dependent on the quality of training data. This independence was evident in scenarios involving highly destabilizing mutations, where supervised learning-based methods typically struggled to predict higher $\Delta\Delta G$ ranges.

Additionally, convergence analysis results underscored the reliability of the proposed protocol, especially for proteins with sizes of $\leq$200 residues. These results were achieved even with relatively short simulation times. However, in cases of charge-changing mutations with high analytical errors, the study found that longer time configurations for both equilibrium and non-equilibrium states were necessary to minimize these errors. Overall, the findings in this chapter contribute significantly to the field of protein stability prediction, particularly in the context of point mutations, offering a robust and reliable alternative to existing methods.

Chapter 5

Discussion

5.1 Integration with machine learning-based methods

The advent of highly sophisticated protein structure prediction models such as AlphaFold2 has significantly minimized the margin for improvement in the prediction of individual protein structures. These models have set a new benchmark for accuracy, leaving limited scope for further advancements in this particular area of structural biology. The current research trajectory, as a consequence, is now pivoting toward the analysis of protein complexes, where the interactions and stability within these systems remain less understood and more challenging to predict. While the computational prediction of single protein structures has reached a point of near perfection with existing tools, protein complexes, which involve multiple protein interactions, still present a significant challenge. MD simulations provide a dynamic perspective that is indispensable for investigating the fluctuating nature of these complexes. This method offers a window into the transient interactions and the conformational flexibility of proteins within complexes, which are critical to understanding their stability and biological function.

On the other hand, protein stability upon mutation research still has a broader area for improvement. The proposed leverage the robustness of alchemical MD simulations, has provided detailed insights into protein dynamics. However, despite the accuracy of this approach, the computational intensity presents a significant bottleneck, limiting the throughput of such simulations and presenting challenges for large-scale applications. Notably, in the cases of neutral mutations—mutations that do not significantly alter the energy landscape of the protein, machine learning-based methods have shown comparable accuracy with much faster computing time. Thus, by integrating the proposed alchemical methods with machine learning-based methods, we believe that accurate and faster prediction can be obtained. To achieve the goal, we conduct additional experiments where the most recent state-of-the-art machine learning models DDGun, and Dynamut2 are selected as the machine learning models as both methods have a webserver available.

Table 5.1: Neutral mutation pairs based on biochemical similarity and amino acid characteristics. The two direction arrows represent the corresponding wild type to mutant and its reverse mutations.

Mutation Pairs
L ↔ I
V ↔ A
M ↔ L
V ↔ M
S ↔ T
N ↔ Q
G ↔ A
S ↔ A
C ↔ S
F ↔ Y
Y ↔ W
Q ↔ H
K ↔ H
K ↔ R
D ↔ E
F ↔ W

The integration strategy involves utilizing the average predicted values from these state-of-the-art machine-learning models for neutral mutation cases. This approach is based on a set of predefined rules that identify neutral mutations, which include amino acid substitutions that are biochemically similar and are less likely to cause significant changes in protein stability. The neutral mutation cases are identified according to the pairs of substitutions as described in Table 5.1. [211, 212] The integration of machine learning-based methods with the base proposed method in the analysis of the p53 dataset, which comprises 42 mutations (26 conserving and 16 charge-changing mutations), demonstrates a notable improvement in computational efficiency. This efficiency is particularly evident when analyzing neutral mutations, identified based on defined pairs. In the proposed method, conserving mutations require approximately 2x(11 to 13 hours) each, while charge-changing mutations take about 2x(17 to 20 hours) each. In contrast, DDGun and DynaMut2 webserver can process mutations in just 1 to 2 minutes each.

Upon analyzing the p53 dataset with the defined neutral mutation pairs, 11 mutations are identified as neutral. These include pairs such as M ↔ L, V ↔ A, and others. Notably, the D ↔ E mutation is categorized as a charge-changing mutation. The integration of machine learning methods for these neutral mutations significantly reduces the computational time. For the base proposed method, the total estimated running time for the p53 dataset is calculated as follows:

- Conserving Mutations: 26 x 2x(11 to 13 hours) = 572 to 676 hours
- Charge-Changing Mutations: 16 x 2x(17 to 20 hours) = 544 to 640 hours
- Total: 1116 to 1316 hours

In comparison, the integrated method with machine learning for neutral mutations yields:

- Neutral Mutations (Machine Learning): 11 x (1 to 2 minutes) = 11 to 22 minutes
- Remaining Conserving Mutations: 16 x 2x(11 to 13 hours) = 352 to 416 hours
- Charge-Changing Mutations: 15 x 2x(17 to 20 hours) = 510 to 600 hours
- Total: 862 to 1016 hours and 11 to 22 minutes

The integration of machine learning methods for neutral mutations in the p53 data set demonstrates a significant reduction in computational time, with the integrated method taking approximately 22.75% to 26.88% less time compared to the base proposed method. Additionally, in terms of accuracy, the integrated method achieves an improved root-mean-square error (RMSE) of 1.03 kcal/mol compared to 1.11 kcal/mol for the base proposed method. It's important to note that these improvements are based on the characteristics of data set and the proportion of neutral mutations. The results indicate the potential benefits of combining computational and data-driven methods in protein stability research.

5.2 Multiple mutation

The major focus of prediction methods in protein stability has been on single point mutations, largely due to the complexities involved in analyzing the cumulative effects of multiple mutations. However, biological systems often exhibit polymorphisms and concurrent mutations, necessitating an understanding of the compounded effects of multiple mutations on protein stability and function. Extending our method to multiple mutations presents a challenge due to the exponential increase in computational demand. Each additional mutation adds complexity, leading to a greater number of possible mutation combinations and, consequently, an increased number of simulations required for accurate stability assessment.

The current alchemical MD simulation protocol, based on *pmx*, can be applied to cases of multiple mutations, either simultaneously or sequentially. When mutations are applied at once, a caveat is the slower convergence of the free energy estimate. More mutations impose a larger perturbation to the system, leading to more work dissipation along the path and reduced accuracy of the free energy estimate. In such cases, performing non-equilibrium transitions more slowly may be necessary. [92] Alternatively, the effect of multiple mutations can be calculated sequentially. For instance, the free energy difference of introducing mutations X and Y simultaneously is equivalent to the combined $\Delta\Delta G$ of first introducing mutation X, followed by calculating $\Delta\Delta G$ for mutation Y in a system where mutation X is already present. Since free energy is a state function, the sequence of introducing mutations does not influence the final $\Delta\Delta G$ estimate. Therefore, mutation Y can be

Table 5.2: Comparison of predicted values: proposed method versus DDGun and DynaMut2.

mutation	experimental $\Delta\Delta G$ (kcal/mol)	mutant PDB ID
L99A, E108V	-3.1	(L99A: 1l90), (L99A, E108V: 1quo)
L99G, E108V	-5.6	(L99G: 1qud), (L99G, E108V: 1quh)
L99A, F153A	-6.9	(L99A: 1l90), (L99A, F153: 1l89)
E128A, V131A	0.44	-
E128A, V131A, N132A	0.9	1l36

performed first, followed by mutation X. The free energy differences calculated in all three scenarios (simultaneous X and Y, first X then Y, first Y then X) should yield the same estimate. [92]

Applying all mutations at once might require more careful sampling and longer simulation time. Additionally, using the current proposed protocol can be problematic for mixed cases (conserving and charge-changing mutations occurring simultaneously) due to the unfolded structure representation. In this research, we prefer to apply mutations sequentially. Although applied sequentially, simulations can run in parallel; for example, the simulation of mutation X to Y and Y to Z can be conducted simultaneously, where $\Delta\Delta G_{X\_Y\_Z} = \Delta\Delta G_{X\_Y} + \Delta\Delta G_{Y\_Z}$. The main drawback of this approach is the requirement for accurate mutant structures in their folded state.

To evaluate the performance of our proposed method, we utilized sample data from the PTmul dataset [160]. We specifically chose 5 mutation cases in T4 lysozyme (PDB id: 2lzm), for which experimental mutant structures are available, except for the E128A and V131A mutations where only experimental value was used. The details of these selected samples are provided in Table 5.2. Currently, only a few existing methods support predictions for multiple mutations. We compared our method's performance with predictions from DDGun and DynaMut2, obtained via their respective web servers. As shown in Table 5.3, our proposed method demonstrated superior performance in the test sample, achieving an RMSE of 1.12 kcal/mol. This is significantly lower compared to DDGun and DynaMut2, which recorded RMSE of 3.46 kcal/mol and 2.89 kcal/mol, respectively. This superiority aligns with our findings in the case of single-point mutations in p53, particularly in predicting highly destabilizing mutations, and highlights the limitations of existing methods.

The flow hierarchy, depicted in Figure 5.1, illustrates how the final $\Delta\Delta G$ of multiple mutations can be calculated by accumulating the free energy changes of each single mutation state, in either forward or reverse direction. For instance, in the case of the L99A, F153A, and L99A, E108V mutations, we can calculate the free energy from the L99A mutant along two paths: towards F153A and E108V in the forward direction. Similarly, for the E128A, V131A, N132A mutant, we calculate the free energy in the reverse direction, starting from the N132A mutant and reversing to E128A and V131A. The predicted final $\Delta\Delta G$ values accurately reflect these changes, underscoring that free energy is a state function and validating the effectiveness of our rigorous approach for multiple mutation cases. Future work should extend this methodology to other mutation scenarios and proteins to further validate its performance.

Table 5.3: Predicted value by proposed methods in comparison with DDGun and DynaMut2

mutation	experimental ddg (kcal/mol)	proposed (kcal/mol)	DDGun (kcal/mol)	DynaMut2 (kcal/mol)
L99A, E108V	-3.1	-5.03	-0.68	-2.31
L99G, E108V	-5.6	-6.73	-1.35	-2.07
L99A, F153A	-6.9	-6.78	-2.11	-2.42
E128A, V131A, N132A	0.9	1.02	-0.15	0.31
RMSE		1.12	3.46	2.89

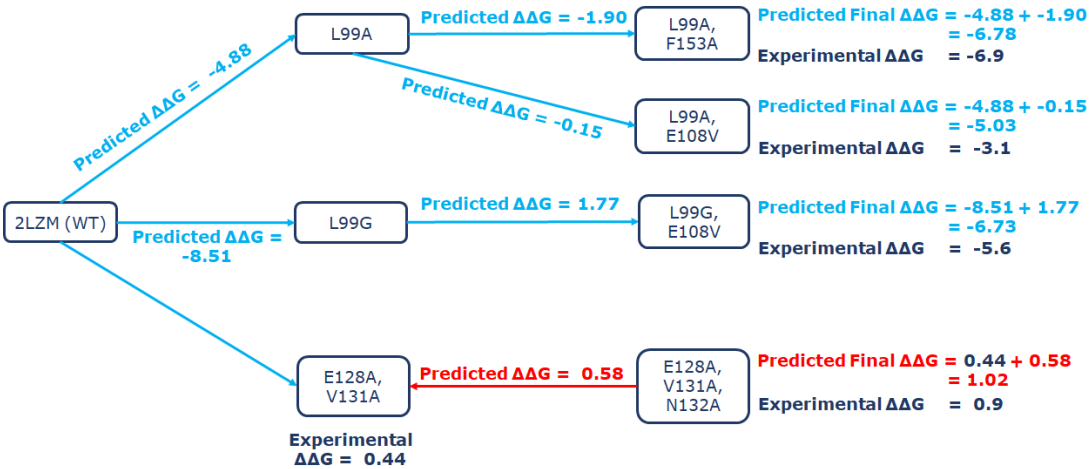


Figure 5.1: Flow hierarchy of multiple mutation states and the corresponding $\Delta\Delta G$ (in kcal/mol), with forward mutations indicated by light blue arrows and reverse mutations by red arrows

5.3 Mutation Pathway Reliability

The motivation behind this exploration arises from observed inconsistencies in the calculated free energy changes ($\Delta\Delta G$) for direct and reverse mutations within the proposed method. These inconsistencies challenge the fundamental principle of microscopic reversibility, which posits that the $\Delta\Delta G$ values of direct mutations should be equal in magnitude but opposite in sign to those of their reverse counterparts. However, empirical data often display non-anti-symmetric results, underscoring a significant gap in our current understanding or application of simulation methodologies.

To address these challenges, this section delves into the systematic evaluation of mutation pathways, leveraging cyclic pathway analysis combined with advanced error propagation strategies. This approach is rooted in the construction of thermodynamic cycles that incorporate both direct and reverse mutations, potentially traversing through intermediate states, to assess their conformity to thermodynamic laws. Cyclic pathway analysis enables the identification of the most reliable pathways by pinpointing cycles with the lowest absolute cyclic $\Delta\Delta G$ sum and propagated error. This dual criterion ensures that selected pathways not only align with theoretical expectations but also demonstrate the highest degree of computational accuracy. Analyzing three outliers (T41C, V66L, S91A mutations, and their reverse counterparts) from the S_{sym} subset, characterized by their non-anti-symmetric $\Delta\Delta G$ values, underscores the challenge to the principle of microscopic reversibility. Here, we select the alternative pathways where mutation at the same position and the crystal structures for the mutant type are available.

For each $\Delta\Delta G$ calculation, since each simulation is repeated three times, the error is quantified through the standard deviation (σ) of $\Delta\Delta G$ values. In contrast, for pathways involving intermediates, a more complex error propagation strategy is employed. Suppose for cyclic mutation pathway of A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ A, the propagated error can be calculated as follow:

$$\sigma_{\text{prop}} = \sqrt{\sigma^2_{\text{A}\rightarrow\text{B}} + \sigma^2_{\text{B}\rightarrow\text{C}} + \sigma^2_{\text{C}\rightarrow\text{A}}}$$

This equation aggregates the variances (σ^2) from each step within the pathway, offering a comprehensive metric of uncertainty for the complete pathway. The investigation of the predicted $\Delta\Delta G$ with the error for each pathway is documented in the Table 5.4.

As illustrated in Figure 5.2, direct and reverse mutation pairs are selected based on pathways with the lowest absolute cyclic $\Delta\Delta G$ sum and propagated error. Table 5.5 enumerates the selected pathways with the lowest cyclic $\Delta\Delta G$ sum and propagated error for each mutation case, demonstrating a rigorous approach in reconciling observed discrepancies with theoretical expectations via alternative pathways. The antisymmetry metric δ is further employed to evaluate the predicted direct and reverse $\Delta\Delta G$ pairs, providing a quantitative measure of bias in predicted free energy

Table 5.4: Predicted $\Delta\Delta G$ and the standard deviation for each mutational pathway

Group	**Mutational Pathway**	**Predicted $\Delta\Delta G$ (kcal/mol)**	**Standard Deviation (kcal/mol)**
T41C ↔ C41T	T → C	-0.61	0.25
	C → T	-1.55	0.88
	T → I	-0.81	1.05
	I → C	1.25	0.39
	C → I	-0.50	0.98
	I → T	0.53	0.83
	T → S	-0.87	0.55
	S → C	0.47	0.72
	C → S	0.36	0.45
	S → T	-0.87	0.16
	T → V	0.10	0.23
	V → C	0.85	0.78
	C → V	-1.15	0.52
	V → T	0.19	0.39
V66L ↔ L66V	V → L	-0.55	2.10
	L → V	-2.14	0.58
	V → I	-0.91	0.15
	I → L	4.77	0.80
	L → I	-0.51	0.28
	I → V	-0.48	0.29
S91A ↔ A91S	S → A	-3.10	0.22
	A → S	1.00	0.06
	S → T	-0.15	0.63
	T → A	-0.77	0.21
	A → T	1.11	0.78
	T → S	-0.48	0.29

Table 5.5: Comparison of $\Delta\Delta G$ pairs for each mutation case using the δ metric. The δ metric quantifies the anti-symmetry between direct and reverse mutations. Cases where δ is closer to zero indicating better performance are highlighted in bold

				Predicted DDG Pair	
Mutation	Pathway	Cyclic Sum	Propagated Error	Direct	Reverse
T41C ↔ C41T	T → V → C → T	-0.60	1.20	0.95	-0.6
V66L ↔ L66V	V → I → L → V	1.72	0.99	3.87	-2.14
S91A ↔ A91S	S → T → A → S	0.08	0.86	-0.92	1

Table 5.6: Selected $\Delta\Delta G$ pairs for each mutation case according to the lowest cycle sum and propagated error

	Direct Pathway Predicted $\Delta\Delta G$ (kcal/mol)			Cyclic Pathway Predicted $\Delta\Delta G$ (kcal/mol)		
Mutation	Direct	Reverse	δ	Direct	Reverse	δ
T41C	-0.61	-1.55	-2.61	0.95	-0.6	**0.35**
V66L	-0.55	-2.14	-2.69	3.87	-2.14	**1.67**
S91A	-3.10	1	-2.1	-0.92	1	**0.08**

changes for mutations and their reversals. [209] The δ metric is defined as

$$\delta = \Delta\Delta G_{\text{inv}} + \Delta\Delta G_{\text{dir}}, \tag{5.1}$$

where $\Delta\Delta G_{\text{dir}}$ and $\Delta\Delta G_{\text{inv}}$ represent the predicted free energy changes for the direct and inverse mutations, respectively. A δ value of 0 denotes perfect anti-symmetry, suggesting a non-biased prediction where the energetic effects of a mutation are precisely counterbalanced upon mutation reversal. This metric is utilized to compare the direct and reverse predicted $\Delta\Delta G$ values, both prior to and subsequent to the application of the cyclic pathway, with the mutation pairs presented in Table 5.6. The analyses reveal that mutation pairs selected via the cyclic pathway exhibit enhanced performance, evidenced by an improvement in the δ value. This meticulous analysis not only augments our comprehension of the anomalies within the S_{sym} subset but also refines the methodologies deployed in predicting the effects of mutations, particularly in the context of anti-symmetry.

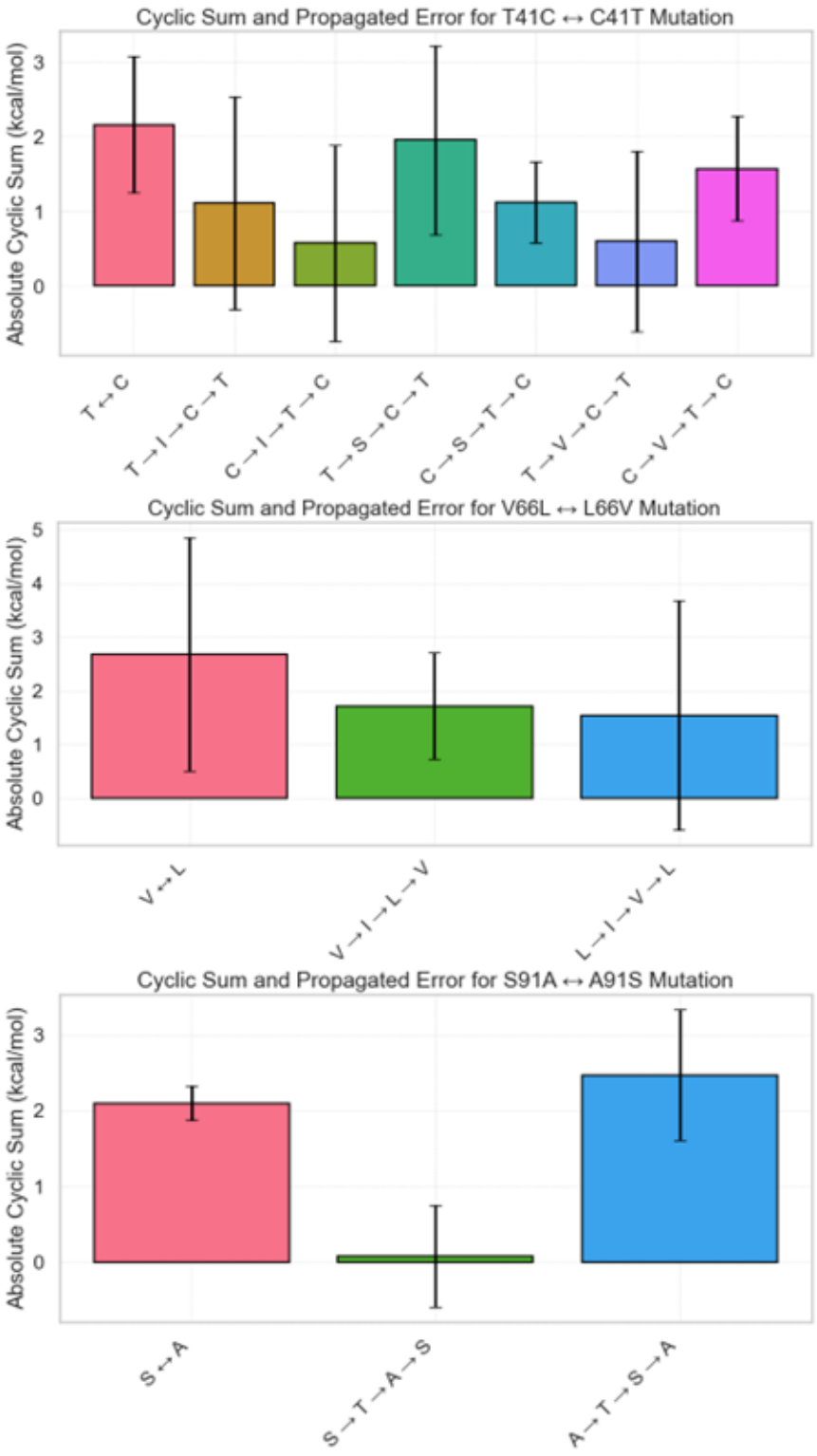


Figure 5.2: The absolute cyclic $\Delta\Delta G$ sum and propagated error of each mutational pathway for corresponding mutation case

Chapter 6

Conclusion

This thesis has explored and expanded the frontiers of protein stability prediction through rigorous computational methods. A key observation throughout this research is the limitations of existing methods, particularly their lack of utilization of dynamic information, despite its crucial role in protein stability. The integration of dynamic information can offer valuable insights and enhance the performance of existing methodologies.

In the first phase of our research, we demonstrated the ability of molecular dynamics simulations to assess the quality of predicted single protein 3D structures. This has laid a solid foundation for understanding inherent protein stability, which is critical for accurate structural prediction and the subsequent rational design of proteins, but still has not been utilized by existing methods in the field. Building upon this, the second phase extends the field from protein stability to stability changes, particularly in the field of point mutations on protein stability, employing alchemical molecular dynamics simulations to predict the resulting stability changes. We introduced a novel protocol to enhance the performance by specifically handling charge-changing mutation cases. Our proposed approach achieves significantly better performance in comparison with state-of-the-art machine learning-based methods.

The main contribution of this thesis is the development of a novel approach that integrates protein dynamics information, obtained from molecular dynamics simulations, into the prediction of protein stability and stability changes in single protein structures. This novel approach addresses and overcomes the inherent limitations found in machine learning-based methods. A key advantage of the proposed method is that it does not require complex feature extraction and extensive training processes typically associated with machine learning models. By incorporating dynamic information from molecular dynamics simulations, this approach not only introduces additional useful information but also enhances the accuracy of protein stability predictions. This offers a more comprehensive and dynamically informed perspective in understanding protein behavior and functionality.

The additional experiments and discussions presented in Chapter 5 have opened up new directions for the application of the proposed methods. The possibility of integrating with machine learning-

based methods to tackle computational limitations, enhancing the speed without compromising the accuracy of stability predictions. Further extension of the proposed approach to the field of multiple mutations also represents a positive outcome in the field.

In conclusion, the research embodied in this thesis has achieved its objectives. The methodologies developed and tested here have proven successful in predicting protein stability and its changes. They pave the way for more reliable and insightful predictions in protein research and its applications, particularly in the fields of drug design, therapeutic interventions, and protein engineering. The potential to extend these methods to more complex scenarios, such as multiple mutations, opens new possibilities for comprehensive and insightful protein stability analysis. This thesis not only contributes to the scientific community by enhancing our understanding of protein dynamics and stability but also sets a new benchmark for future research in this vital area of molecular biology.

Future directions

Moving forward, we identify the following areas as promising directions for further exploration and development, building upon the insights and discoveries gained from this research. The following areas represent promising directions for further exploration and development:

- **Development of accurate structure prediction methods for mutant structure**

  While AlphaFold2 has demonstrated high accuracy in predicting wild-type protein structures, it often falls short in accurately predicting the effects of missense mutations on three-dimensional protein structures. [181, 182] In our research, we utilized structures predicted by AlphaFold2, exploiting the conveyed information on the residue flexibility and dynamics. [183] Specifically, we employed AlphaFold2-predicted mutant structures in an unfolded state within our workflow. However, this approach encounters challenges in reverse mutation cases where experimentally determined mutant structures are unavailable, and AF2-predicted mutant structures are used as the folded state. Future work should therefore concentrate on developing novel methods that can more accurately predict the impact of missense mutations on protein structures. Such methods would need to address the limitations in scenarios where understanding the mutation's effect is critical and relies on the accurate representation of the mutant structure.

- **Addressing multiple mutations**

  In real-world applications, the occurrence of multiple mutations is a critical factor that significantly influences protein behavior and functionality. Accurately calculating the free energy changes associated with multiple mutations, whether performed simultaneously or sequentially, is essential for understanding their cumulative impact on protein stability. This task becomes

particularly challenging due to the need for precise mutant structures. The development of accurate structure prediction methods for mutants as discussed previously, is therefore not only crucial for single mutations but also becomes imperative when dealing with multiple mutations. Future research should focus on methodologies that can reliably predict the structural effects of multiple mutations on proteins.

- **Creation of data sets providing protein structure dynamics**

  Currently, obtaining detailed information on protein dynamics typically requires performing molecular simulations. Developing data sets that offer insights into protein structure dynamics would be invaluable. Such data sets could serve as a new benchmarking tool and also be utilized to train machine learning models, particularly deep learning models that are capable of learning spatio-temporal features of protein dynamics. This advancement would significantly enhance the predictive accuracy of these models in understanding protein behavior and stability changes. The creation of these datasets would provide a crucial resource for researchers, allowing for a more efficient and comprehensive analysis of protein dynamics without the sole reliance on computationally intensive molecular simulations.

- **Advancements in computing power**

  Future research should leverage advancements in computing power, distributed cloud computing, software optimization, and graphics processing unit parallelization. These advancements are expected to make molecular simulations more efficient and faster, thereby overcoming current computational intensity limitations. As a result, molecular simulations could become more widely accessible and applicable for extensive research, allowing for a deeper exploration of protein dynamics and interactions. For instance, we can explore larger systems, such as triple or quarter systems within a single simulation box, especially when applying multiple mutations simultaneously.

These future directions not only build upon the findings of this thesis but also open new pathways for advancing our understanding of protein stability and dynamics. The integration of computational power, advanced modeling techniques, and innovative data sets will likely lead to significant breakthroughs in protein research, with wide-ranging implications in molecular biology, drug design, and therapeutic development.

Appendix A

Supporting tables

Table A.1: Predicted $\Delta\Delta G$ values for frataxin data set using GxG-only protocol. Charge changing mutation is marked with red font.

Mutation	Predicted $\Delta\Delta G$ (kcal/mol)	Experimental $\Delta\Delta G$ (kcal/mol)
D104G	0.40	0.40
A107V	-0.94	0.00
F109L	-2.12	-2.80
Y123S	-4.23	-5.10
S161I	1.13	-3.10
W173C	-10.95	-9.50
S181F	-2.98	-3.00
S202F	0.16	-0.20

Table A.2: Predicted $\Delta\Delta G$ values for p53 data set using GxG-only protocol. Charge changing mutations are marked with red font.

Mutation	Predicted $\Delta\Delta G$ (kcal/mol)	Experimental $\Delta\Delta G$ (kcal/mol)
Q104H	-0.23	0.24
Q104P	-0.18	0.11
T123A	-1.09	-0.13
A129D	0.13	-0.70
A129E	0.14	-0.38
A129S	0.21	-0.19
M133L	0.87	0.30
F134L	-4.17	-4.78
V143A	-3.43	-3.50
L145Q	-2.56	-2.98
D148E	2.03	-0.43
D148S	0.66	0.22
T150P	0.02	-0.08
P151S	-4.78	-4.49
V157F	-2.48	-3.88
Q165K	-4.42	-1.27
Q167E	0.59	-0.43
H168R	-3.23	-2.75
R174K	0.06	-0.22
R175A	2.65	-0.73
R175H	-1.72	-3.52
C182S	0.07	0.16
I195T	-4.74	-4.12
L201P	-0.61	0.35
V203A	-1.25	0.49
L206S	0.18	-0.10
Y220C	-2.71	-3.98
D228E	2.05	0.05
I232T	-2.92	-3.19
Y236F	0.23	0.27
M237I	0.08	-3.18
N239Y	-0.05	1.49
C242S	-0.23	-3.07
G245S	0.56	-1.21
R248Q	3.75	-1.87
R249S	6.06	-1.92
I255F	-2.45	-3.29
S260P	0.37	-0.32
N268D	1.37	1.21
F270C	-3.75	-4.54
R273H	2.45	-0.45
R282W	-1.35	-3.30

Table A.3: Predicted $\Delta\Delta G$ values for frataxin data set using AF2HT-only protocol. Charge changing mutation is marked with red font.

Mutation	Predicted $\Delta\Delta G$ (kcal/mol)	Experimental $\Delta\Delta G$ (kcal/mol)
D104G	-0.32	0.40
A107V	-2.38	0.00
F109L	-2.24	-2.80
Y123S	-1.90	-5.10
S161I	0.83	-3.10
W173C	-7.08	-9.50
S181F	-3.93	-3.00
S202F	1.71	-0.20

Table A.4: Predicted $\Delta\Delta G$ values for p53 data set using AF2HT-only protocol. Charge changing mutations are marked with red font.

Mutation	Predicted $\Delta\Delta G$ (kcal/mol)	Experimental $\Delta\Delta G$ (kcal/mol)
Q104H	0.40	0.24
Q104P	-1.84	0.11
T123A	-1.05	-0.13
A129D	-2.26	-0.7
A129E	-1.40	-0.38
A129S	0.11	-0.19
M133L	1.65	0.3
F134L	-3.01	-4.78
V143A	-2.23	-3.5
L145Q	-2.56	-2.98
D148E	-0.05	-0.43
D148S	0.09	0.22
T150P	0.65	-0.08
P151S	-8.62	-4.49
V157F	-5.74	-3.88
Q165K	-0.40	-1.27
Q167E	-0.70	-0.43
H168R	-4.35	-2.75
R174K	0.04	-0.22
R175A	-1.22	-0.73
R175H	-3.61	-3.52
C182S	-0.24	0.16
I195T	-4.81	-4.12
L201P	-1.51	0.35
V203A	-1.06	0.49
L206S	-0.44	-0.1
Y220C	-4.43	-3.98
D228E	-0.39	0.05
I232T	-5.08	-3.19
Y236F	-0.19	0.27
M237I	0.17	-3.18
N239Y	-0.30	1.49
C242S	-0.20	-3.07
G245S	0.05	-1.21
R248Q	0.69	-1.87
R249S	-2.98	-1.92
I255F	-3.39	-3.29
S260P	3.29	-0.32
N268D	-2.37	1.21
F270C	-5.49	-4.54
R273H	-0.18	-0.45
R282W	-3.76	-3.3

Table A.5: Prediction squared error (in kcal/mol) by predictors for frataxin data set. The lowest squared error for each mutation is marked with green color.

Mutation	Proposed	iStable	mCSM	SDM	DUET	DeepDDG	iDeepDDG	SDM2	DynaMut2	DDGun3D	DDGun Seq	PROST
D104G	0.01	0.86	1.98	0.37	1.48	2.91	2.74	0.31	1.12	0.81	1.96	1.04
A107V	0.72	0.00	0.08	2.13	0.13	0.57	0.13	1.64	0.19	1.44	0.81	0.22
F109L	0.27	2.50	4.12	12.25	5.50	3.49	4.41	11.83	9.67	6.76	6.76	4.33
Y123S	1.10	10.82	4.15	6.86	3.08	8.82	5.22	8.70	2.66	6.25	7.84	11.64
S161I	32.72	12.39	8.50	21.07	11.84	9.35	9.39	11.42	6.66	9.00	5.29	8.01
W173C	1.80	56.10	61.15	62.73	62.03	47.13	53.76	59.91	60.06	16.81	7.84	42.46
S181F	0.18	11.16	5.88	4.75	4.27	4.06	2.86	2.59	6.86	17.64	9.61	5.17
S202F	0.52	0.20	0.68	1.25	0.28	0.00	0.11	0.52	0.40	0.09	0.25	0.28

Table A.6: Squared error (in kcal/mol) analysis for p53 data set. Squared errors greater than 1 kcal/mol are highlighted in yellow for both the proposed method and the average of existing methods.

Mutation	Proposed	Average	DynaMut2	MUpro	EASE-MM-web	PoPMuSiC	I-MutantA	I-MutantB	INPS	MU3-DSP	DDGun	MAESTRO	SAAFEC-SEQ	m-CSM	MU3DSP-unbalance	PROST
Q104H	0.02	0.06	2.40	0.87	0.00	0.01	0.59	2.10	0.14	0.03	0.03	0.00	2.22	1.54	0.04	0.03
Q104P	0.08	0.01	0.26	0.71	0.22	0.10	1.23	1.88	0.02	0.03	1.19	1.24	1.12	0.23	0.02	0.75
T123A	0.93	0.02	0.09	0.37	0.60	0.94	2.56	0.83	0.44	0.30	0.45	0.67	1.37	0.56	0.29	0.37
A129D	2.44	0.49	0.06	0.04	1.48	0.15	0.01	0.34	0.44	0.47	0.49	1.91	0.00	0.00	0.45	0.44
A129E	1.03	0.14	0.04	0.02	0.37	0.03	0.12	0.14	0.31	0.11	0.34	0.47	0.12	0.19	0.10	0.17
A129S	0.16	0.04	0.02	0.48	0.16	0.04	0.30	0.00	0.03	0.02	0.24	0.03	0.25	0.40	0.02	0.15
M133L	0.33	0.09	0.08	1.69	0.06	0.86	4.80	0.10	0.04	0.37	1.44	0.01	1.59	0.01	0.36	0.14
F134L	0.37	22.85	3.10	20.16	12.27	10.18	20.16	7.18	5.17	9.27	4.75	30.06	9.80	2.89	9.94	8.48
V143A	0.01	12.25	2.02	8.75	1.46	0.15	1.21	0.01	1.71	0.74	0.25	29.32	4.54	2.50	0.77	3.00
L145Q	0.18	8.88	0.01	1.41	0.68	0.52	1.51	1.72	1.28	0.20	0.01	32.26	2.53	0.05	0.21	0.79
D148E	0.14	0.18	0.01	0.02	0.00	0.09	1.08	0.06	0.21	0.06	0.40	0.32	0.04	0.02	0.05	0.33
D148S	0.02	0.05	0.02	1.26	0.20	0.08	0.29	2.22	0.04	0.18	0.01	0.01	0.49	0.01	0.14	0.04
T150P	0.01	0.01	0.02	1.71	0.17	0.71	0.12	0.05	0.17	0.00	0.08	0.68	0.64	0.09	0.00	0.39
P151S	0.08	20.16	3.84	11.00	9.52	6.66	6.97	11.22	11.54	6.85	6.71	34.06	9.86	4.04	6.87	7.62
V157F	1.96	15.05	5.34	8.64	6.35	7.24	2.62	0.66	6.79	7.19	6.66	16.79	8.01	6.50	7.14	4.89
Q165K	0.76	1.61	0.52	0.01	0.39	0.42	0.41	0.85	2.26	0.92	0.45	6.91	0.00	0.21	1.00	3.96
Q167E	0.07	0.18	0.14	0.06	0.67	0.59	0.01	0.53	0.36	0.06	0.02	0.01	0.15	0.09	0.05	1.38
H168R	2.55	7.56	4.37	3.60	6.49	2.69	8.58	5.76	4.61	2.79	2.40	10.37	2.46	2.62	2.94	3.62
R174K	0.07	0.05	1.06	0.57	0.00	0.21	1.54	0.08	0.00	0.06	1.17	0.76	1.37	0.22	0.06	0.31
R175A	0.24	0.53	0.06	2.29	0.04	0.16	0.35	2.25	0.00	0.74	0.94	6.46	0.42	0.01	0.68	1.83
R175H	0.01	12.39	1.93	0.37	8.66	10.11	0.71	9.06	5.85	1.98	7.95	19.62	1.37	6.15	2.08	0.27
C182S	0.01	0.03	0.98	3.28	0.53	0.36	3.17	2.46	0.26	0.25	0.19	0.00	0.85	1.54	0.27	0.00
I195T	0.38	16.97	0.44	1.89	4.00	4.71	0.01	0.00	1.61	2.56	1.25	40.17	3.84	1.51	2.93	0.79
L201P	0.93	0.12	0.00	7.70	0.91	0.12	4.12	3.03	0.81	0.35	1.10	1.02	1.69	0.44	0.35	0.07
V203A	3.04	0.24	4.28	2.94	2.42	5.52	7.67	14.75	0.12	2.43	0.24	0.48	3.69	4.49	2.44	0.09
L206S	0.08	0.01	0.08	1.85	4.90	0.66	15.52	3.69	2.05	2.34	1.44	2.06	1.51	1.14	2.17	0.43
Y220C	1.62	15.84	3.42	13.96	6.31	2.50	21.81	6.00	5.97	5.11	1.17	27.23	5.90	4.33	5.20	7.17
D228E	0.20	0.00	0.00	0.38	0.07	0.14	0.06	0.55	0.01	0.04	0.00	0.04	0.42	0.10	0.06	0.15
I232T	0.07	10.18	0.34	2.09	2.35	1.06	0.19	0.62	0.66	1.66	0.79	21.89	1.08	0.67	1.66	0.61
Y236F	0.00	0.07	3.28	1.39	0.18	1.44	1.06	1.74	0.06	0.55	2.99	0.03	0.31	3.53	0.57	1.17
M237I	10.63	10.11	4.08	5.93	8.14	6.86	1.96	1.19	2.58	5.09	5.20	11.02	7.78	1.80	4.51	7.35
N239Y	2.36	2.22	3.03	3.53	3.31	4.54	1.82	9.49	0.45	7.20	0.08	2.02	6.10	3.06	7.21	0.51
C242S	8.07	9.42	5.52	0.05	1.28	2.79	0.29	0.22	0.27	1.69	2.66	14.42	0.90	4.45	1.78	0.00
G245S	3.14	1.46	0.18	0.16	0.32	0.12	1.35	0.00	0.19	0.00	0.08	2.60	0.31	0.00	0.00	0.00
R248Q	6.55	3.50	1.61	0.00	4.53	3.61	0.10	0.01	0.35	0.80	0.59	2.29	0.08	2.89	0.89	0.00
R249S	1.12	3.69	0.20	0.00	2.38	0.27	0.00	0.05	0.52	0.18	0.01	15.18	0.41	0.06	0.24	0.03
I255F	0.71	10.82	1.21	2.97	8.57	5.76	2.56	0.59	20.26	5.86	22.00	14.31	4.54	2.96	5.89	20.72
S260P	0.48	0.10	0.01	0.84	0.10	0.11	0.88	0.02	0.00	0.01	0.01	0.12	0.00	0.02	0.01	0.45
N268D	0.02	1.46	6.05	4.46	5.02	5.38	8.35	9.30	1.58	2.89	2.28	0.32	4.00	7.95	3.06	1.45
F270C	0.63	20.61	4.93	9.19	6.08	5.81	6.76	3.57	7.80	5.05	4.16	34.02	6.00	7.02	5.49	9.66
R273H	0.07	0.20	0.45	0.00	0.20	0.00	0.35	0.01	0.35	0.45	0.20	0.93	0.00	2.02	0.46	0.05
R282W	0.21	10.89	3.13	6.41	6.11	12.25	8.35	8.82	11.07	6.99	5.76	18.91	5.57	4.33	7.09	6.41

Table A.7: Analytical and Bootstrap Error in Frataxin Dataset

Mutation	Analytical Error (kcal/mol)	Bootstrap Error (kcal/mol)
D104G	0.630	0.710
	0.640	0.650
A107V	3.970	0.590
	0.090	0.090
	0.110	0.110
F109L	0.090	0.110
	0.120	0.080
	0.110	0.110
Y123S	0.120	0.130
	0.730	0.370
	0.570	0.250
S161I	0.460	0.230
	0.170	0.130
	0.430	0.190
W173C	0.170	0.180
	0.380	0.390
	0.420	0.490
S181F	0.330	0.290
	0.300	0.330
	0.260	0.180
S202F	0.320	0.330
	0.160	0.130
	0.230	0.170

Table A.8: Analytical and bootstrap error (in kcal/mol) in p53 dataset. Error larger than 1 kcal/mol is marked with yellow color.

Mutation	Analytical Error	Bootstrap Error	Mutation	Analytical Error	Bootstrap Error	Mutation	Analytical Error	Bootstrap Error
Q104H	0.29	0.25	V157F	0.3	0.32	I232T	0.13	0.18
	0.48	0.32		0.73	0.25		0.16	0.18
	0.26	0.19		0.34	0.33		0.10	0.15
Q104P	0.87	0.76	Q165K	0.37	0.42	Y236F	0.04	0.05
	0.73	0.85		0.33	0.38		0.09	0.07
	0.67	0.79		0.44	0.42		0.04	0.04
T123A	0.07	0.09	Q167E	0.63	0.65	M237I	0.18	0.11
	0.08	0.08		0.51	0.61		0.14	0.11
	0.08	0.10		0.85	0.50		0.21	0.19
A129D	0.83	0.95	H168R	101.07	0.55	N239Y	0.41	0.20
	0.63	0.59		44.11	0.57		0.24	0.30
	0.56	0.23		85.72	0.35		0.40	0.31
A129E	0.57	0.24	R174K	0.49	0.48	C242S	0.10	0.07
	0.50	0.38		0.81	0.27		0.50	0.05
	0.96	0.90		0.77	0.39		0.07	0.07
A129S	0.06	0.06	R175A	462.91	0.94	G245S	0.26	0.24
	0.09	0.09		271.45	0.93		0.26	0.24
	0.09	0.09		58.39	0.63		0.21	0.19
M133L	0.17	0.17	R175H	26.71	0.45	R248Q	0.27	0.27
	0.15	0.12		154.14	0.56		0.33	0.31
	0.17	0.11		88.53	0.88		0.19	0.20
F134L	0.21	0.18	C182S	0.05	0.05	R249S	0.73	0.59
	0.22	0.16		0.06	0.05		0.96	0.46
	0.20	0.18		0.07	0.06		3.80	0.19
V143A	0.08	0.09	I195T	0.15	0.13	I255F	0.23	0.22
	0.50	0.10		0.19	0.25		0.37	0.34
	0.09	0.09		0.26	0.16		0.23	0.28
L145Q	0.93	0.50	L201P	0.45	0.32	S260P	0.73	0.50
	0.41	0.29		0.97	0.25		-0.74	0.29
	0.44	0.21		0.62	0.43		0.61	0.38
D148E	0.19	0.20	V203A	0.11	0.13	N268D	0.41	0.36
	0.15	0.16		0.16	0.17		0.52	0.54
	0.27	0.38		0.07	0.08		0.29	0.23
D148S	0.17	0.21	L206S	0.21	0.18	F270C	0.13	0.13
	0.39	0.16		0.34	0.36		0.35	0.31
	0.28	0.25		0.34	0.16		0.15	0.14
T150P	0.52	0.56	Y220C	0.38	0.32	R273H	0.64	0.29
	0.83	0.26		0.46	0.43		0.97	0.47
	0.58	0.32		0.32	0.38		0.77	0.68
P151S	1.13	0.58	D228E	0.67	0.31	R282W	9610.18	0.51
	1.89	0.53		0.49	0.43		772.91	0.61
	1.15	0.53		0.73	0.97		580.26	0.45

Table A.9: The predicted $\Delta\Delta G$ values of the direct and reverse mutations of p53 data sets

Mutation	Direct $\Delta\Delta G$ (kcal/mol)	Reverse $\Delta\Delta G$ (kcal/mol)
Q104H	0.40	-0.40
Q104P	-0.18	-0.10
T123A	-1.09	0.50
A129D	-2.26	1.72
A129E	-1.40	0.99
A129S	0.21	-0.36
M133L	0.87	-0.78
F134L	-4.17	4.10
V143A	-3.43	3.60
L145Q	-2.56	3.29
D148E	-0.05	0.70
D148S	0.09	-2.90
T150P	0.02	0.07
P151S	-4.78	5.22
V157F	-2.48	3.49
Q165K	-0.40	1.50
Q167E	-0.70	-0.75
H168R	-4.35	2.79
R174K	0.04	-0.49
R175A	-1.22	-3.21
R175H	-3.61	-1.92
C182S	0.07	-0.19
I195T	-4.74	5.08
L201P	-0.61	-0.44
V203A	-1.25	1.25
L206S	0.18	1.35
Y220C	-2.71	3.07
D228E	-0.39	0.13
I232T	-2.92	1.79
Y236F	0.23	-0.21
M237I	0.08	2.78
N239Y	-0.05	-0.37
C242S	-0.23	-0.05
G245S	0.56	-0.12
R248Q	0.69	1.87
R249S	-2.98	2.31
I255F	-2.45	2.35
S260P	0.37	0.18
N268D	1.37	-2.09
F270C	-3.75	4.21
R273H	-0.18	-2.96
R282W	-3.76	-0.78

Table A.10: The selected Ssym subset and predicted $\Delta\Delta G$ (in kcal/mol). The sub set consisted of 39 pairs of direct and reverse mutations, including 30 conserving and 9 charge-changing mutations. Charge changing mutations are marked with red letter.

Direct				Reverse			
Mutation	Experimental $\Delta\Delta G$	Predicted $\Delta\Delta G$	Squared Error	Mutation	Experimental $\Delta\Delta G$	Predicted $\Delta\Delta G$	Squared Error
D10N	0.9	6.67	33.29	D10N	-0.9	-6.89	35.88
D12A	2.5	4.08	2.50	D12A	-2.5	-5.64	9.86
D13A	2.7	3.23	0.28	D13A	-2.7	-3.13	0.18
Q25K	0.93	-0.61	2.37	Q25K	-0.93	0.41	1.80
K43A	-1	-1.16	0.03	K43A	1	0.98	0.00
D57A	3.3	-0.06	11.29	D57A	-3.3	-1.28	4.08
H62A	0.4	1.29	0.79	H62A	-0.4	-0.41	0.00
R96Y	-4.7	-6.09	1.93	R96Y	4.7	3.23	2.16
V131D	0.1	0.65	0.30	V131D	-0.1	-1.50	1.96
V2Y	-0.4	0.11	0.26	V2Y	0.4	-0.19	0.35
V2L	-0.05	-1.07	1.04	V2L	0.05	1.78	2.99
I3Y	-2.3	-3.02	0.52	I3Y	2.3	1.26	1.08
T9G	-2.6	-2.02	0.34	T9G	2.6	1.50	1.21
Y23A	-5.9	-6.97	1.14	Y23A	5.9	6.24	0.12
T26A	-1.7	-1.84	0.02	T26A	1.7	0.29	1.99
Y32F	0.2	1.69	2.22	Y32F	-0.2	-1.25	1.10
V35I	-0.6	0.28	0.77	V35I	0.6	-0.91	2.28
Y35G	-5	-5.02	0.00	Y35G	5	0.37	21.44
T41C	0.6	-0.61	1.46	T41C	-0.6	-1.55	0.90
N43G	-5.7	-4.52	1.39	N43G	5.7	4.82	0.77
I47M	-1.7	-0.13	2.46	I47M	1.7	1.11	0.35
I47V	-2	-1.59	0.17	I47V	2	1.01	0.98
I50A	-2	-0.51	2.22	I50A	2	0.50	2.25
I51V	-1.1	-1.77	0.45	I51V	1.1	2.76	2.76
A52V	1.7	0.51	1.42	A52V	-1.7	-1.96	0.07
I55V	-0.9	-0.16	0.55	I55V	0.9	2.00	1.21
I56V	-1.2	-2.36	1.35	I56V	1.2	2.32	1.25
Y57F	-1.2	0.40	2.56	Y57F	1.2	0.78	0.18
T59S	-0.2	-0.46	0.07	T59S	0.2	-0.89	1.19
V66I	-1	-0.09	0.83	V66I	1	-1.16	4.67
V66L	-0.3	-0.55	0.06	V66L	0.3	-2.14	5.95
G67A	-0.5	-0.79	0.08	G67A	0.5	0.86	0.13
F71N	0.7	-0.51	1.46	F71N	-0.7	0.12	0.67
F82Q	-0.4	-0.28	0.01	F82Q	0.4	-0.47	0.76
S91A	-0.2	-3.10	8.41	S91A	0.2	1.00	0.64
F92N	-1.1	-1.86	0.58	F92N	1.1	2.00	0.81
I96A	-3.2	-4.42	1.49	I96A	3.2	3.14	0.00
C191S	-1.9	0.42	5.38	C191S	1.9	0.42	2.19
C191Y	-2.3	-5.33	9.18	C191Y	2.3	2.32	0.00

Table A.11: Prediction squared error (in kcal/mol) by predictors for p53 data set where the absolute experimental $\Delta\Delta G$ values ≥ 3.5 kcal/mol. The squared error greater than 1 kcal/mol between proposed method and average of existing methods for each mutation is marked with yellow color

Mutation	Experimental	Proposed	DynaMut2	MUpro	EASE-MM-web	PoPMuSiC	I-MutantA	I-MutantB	INPS	MU3-DSP	DDGun	MAESTRO	SAAFEC-SEQ	mCSM	MU-3DSP-unbalance	PROST	Average
F134L	-4.78	0.37	3.10	20.16	12.27	10.18	20.16	7.18	5.17	9.27	4.75	30.06	9.80	2.89	9.94	8.48	22.85
V143A	-3.50	0.01	2.02	8.75	1.46	0.15	1.21	0.01	1.71	0.74	0.25	29.32	4.54	2.50	0.77	3.00	12.25
P151S	-4.49	0.08	3.84	11.00	9.52	6.66	6.97	11.22	11.54	6.85	6.71	34.06	9.86	4.04	6.87	7.62	20.16
V157F	-3.88	1.96	5.34	8.64	6.35	7.24	2.62	0.66	6.79	7.19	6.66	16.79	8.01	6.50	7.14	4.89	15.05
R175H	-3.52	0.01	1.93	0.37	8.66	10.11	0.71	9.06	5.85	1.98	7.95	19.62	1.37	6.15	2.08	0.27	12.39
I195T	-4.12	0.38	0.44	1.89	4.00	4.71	0.01	0.00	1.61	2.56	1.25	40.17	3.84	1.51	2.93	0.79	16.97
Y220C	-3.98	1.62	3.42	13.96	6.31	2.50	21.81	6.00	5.97	5.11	1.17	27.23	5.90	4.33	5.20	7.17	15.84
F270C	-4.54	0.63	4.93	9.19	6.08	5.81	6.76	3.57	7.80	5.05	4.16	34.02	6.00	7.02	5.49	9.66	20.61

Bibliography

[1] Ken A Dill and Justin L MacCallum. The protein-folding problem, 50 years on. *Science*, 338 (6110):1042–1046, 2012. DOI: 10.1126/science.1219021.

[2] Elliott J Stollar and David P Smith. Uncovering protein structure. *Essays in Biochemistry*, 64(4):649–680, 2020. DOI: 10.1042/EBC20190042.

[3] Ivano Tavernelli, Simona Cotesta, and Ernesto E Di Iorio. Protein dynamics, thermal stability, and free-energy landscapes: a molecular dynamics investigation. *Biophysical Journal*, 85(4): 2641–2649, 2003. DOI: 10.1016/S0006-3495(03)74687-6.

[4] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007. DOI: 10.1038/nature06522.

[5] Kaare Teilum, Johan G Olsen, and Birthe B Kragelund. Functional aspects of protein flexibility. *Cellular and Molecular Life Sciences*, 66:2231–2247, 2009. DOI: 10.1007/s00018-009-0014-6.

[6] Natali A Gonzalez, Brigitte A Li, and Michelle E McCully. The stability and dynamics of computationally designed proteins. *Protein Engineering, Design and Selection*, 35:gzac001, 2022. DOI: 10.1093/protein/gzac001.

[7] Jose M Sanchez-Ruiz. Protein kinetic stability. *Biophysical Chemistry*, 148(1-3):1–15, 2010. DOI: 10.1016/j.bpc.2010.02.004.

[8] Marc C Deller, Leopold Kong, and Bernhard Rupp. Protein stability: a crystallographer's perspective. *Acta Crystallographica Section F: Structural Biology Communications*, 72(2): 72–95, 2016. DOI: 10.1107/S2053230X15024619.

[9] Christopher M Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, 2003. DOI: 10.1038/nature02261.

[10] Nirbhik Acharya and Santosh Kumar Jha. Dry molten globule-like intermediates in protein folding, function, and disease. *The Journal of Physical Chemistry B*, 126(43):8614–8622, 2022. DOI: 10.1021/acs.jpcb.2c04991.

[11] C Nick Pace. Measuring and increasing protein stability. *Trends in Biotechnology*, 8:93–98, 1990. DOI: 10.1016/0167-7799(90)90146-o.

[12] C Nick Pace. Conformational stability of globular proteins. *Trends in Biochemical Sciences*, 15(1):14–17, 1990. DOI: 10.1016/0968-0004(90)90124-T.

[13] Peter L Privalov. Cold denaturation of protein. *Critical Reviews in Biochemistry and Molecular Biology*, 25(4):281–306, 1990. DOI: 10.3109/10409239009090612.

[14] Valerie Daggett and Alan R Fersht. Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences*, 28(1):18–25, 2003. DOI: 10.1016/S0968-0004(02)00012-9.

[15] PL Privalov and NN Khechinashvili. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *Journal of Molecular Biology*, 86(3): 665–684, 1974. DOI: 10.1016/0022-2836(74)90188-0.

[16] Daniel D Vallejo, Carolina Rojas Ramiérez, Kristine F Parson, Yilin Han, Varun V Gadkari, and Brandon T Ruotolo. Mass spectrometry methods for measuring protein stability. *Chemical Reviews*, 122(8):7690–7719, 2022. DOI: 10.1021/acs.chemrev.1c00857.

[17] Sharon M Kelly, Thomas J Jess, and Nicholas C Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - roteins and Proteomics*, 1751(2):119–139, 2005. DOI: 10.1016/j.bbapap.2005.06.005.

[18] C Nick Pace, Felix Vajdos, Lanette Fee, Gerald Grimsley, and Theronica Gray. How to measure and predict the molar absorption coefficient of a protein. *Protein Science*, 4(11):2411–2423, 1995. DOI: 10.1002/pro.5560041120.

[19] Adelinda A Yee, Alexei Savchenko, Alexandr Ignachenko, Jonathan Lukin, Xiaohui Xu, Tatiana Skarina, Elena Evdokimova, Cheng Song Liu, Anthony Semesi, Valerie Guido, et al. Nmr and x-ray crystallography, complementary tools in structural proteomics of small proteins. *Journal of the American Chemical Society*, 127(47):16512–16517, 2005. DOI: 10.1021/ja053565+.

[20] Lewis E Kay. Nmr studies of protein structure and dynamics. *Journal of Magnetic Resonance*, 213(2):477–491, 2011. DOI: 10.1016/j.jmr.2011.09.009.

[21] Brian W Matthews. Studies on protein stability with t4 lysozyme. *Advances in Protein Chemistry*, 46:249–278, 1995. DOI: 10.1016/S0065-3233(08)60337-X.

[22] Mark C Manning, Kamlesh Patel, and Ronald T Borchardt. Stability of protein pharmaceuticals. *Pharmaceutical Research*, 6:903–918, 1989. DOI: 10.1023/A:1015929109894.

[23] Maciej Majewski, Sergio Ruiz-Carmona, and Xavier Barril. An investigation of structural stability in protein-ligand complexes reveals the balance between order and disorder. *Communications Chemistry*, 2(1):110, 2019. DOI: 10.1038/s42004-019-0205-5.

[24] Evan T Powers, Richard I Morimoto, Andrew Dillin, Jeffery W Kelly, and William E Balch. Biological and chemical approaches to diseases of proteostasis deficiency. *Annual Review of Biochemistry*, 78:959–991, 2009. DOI: 10.1146/annurev.biochem.052308.114844.

[25] Andreas S Bommarius and Mariétou F Paye. Stabilizing biocatalysts. *Chemical Society Reviews*, 42(15):6534–6565, 2013. DOI: 10.1039/C3CS60137D.

[26] Luke F Hartje and Christopher D Snow. Protein crystal based materials for nanoscale applications in medicine and biotechnology. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 11(4):e1547, 2019. DOI: 10.1002/wnan.1547.

[27] Heungsoo Shin, Seongbong Jo, and Antonios G Mikos. Biomimetic materials for tissue engineering. *Biomaterials*, 24(24):4353–4364, 2003. DOI: 10.1016/S0142-9612(03)00339-9.

[28] Ruedi Aebersold and Matthias Mann. Mass spectrometry-based proteomics. *Nature*, 422 (6928):198–207, 2003. DOI: 10.1038/nature01511.

[29] Margaret A Hamburg and Francis S Collins. The path to personalized medicine. *New England Journal of Medicine*, 363(4):301–304, 2010. DOI: 10.1056/NEJMp1006304.

[30] Ikjin Kim, Christina R Miller, David L Young, and Stanley Fields. High-throughput analysis of in vivo protein stability. *Molecular & Cellular Proteomics*, 12(11):3370–3378, 2013. DOI: 10.1074/mcp.O113.031708.

[31] Yang Zhang. Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology*, 18(3):342–348, 2008. DOI: 10.1016/j.sbi.2008.02.004.

[32] William L Jorgensen. The many roles of computation in drug discovery. *Science*, 303(5665): 1813–1818, 2004. DOI: 10.1126/science.1096361.

[33] Francis S Collins and Harold Varmus. A new initiative on precision medicine. *New England Journal of Medicine*, 372(9):793–795, 2015. DOI: 10.1056/NEJMp1500523.

[34] Adi Goldenzweig and Sarel J Fleishman. Principles of protein stability and their application in computational design. *Annual Review of Biochemistry*, 87:105–129, 2018. DOI: 10.1146/annurev-biochem-062917-012102.

[35] Justin R Klesmith, John-Paul Bacik, Emily E Wrenbeck, Ryszard Michalczyk, and Timothy A Whitehead. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proceedings of the National Academy of Sciences*, 114(9):2265–2270, 2017. DOI: 10.1073/pnas.161443711.

[36] Óscar Álvarez-Machancoses, Enrique J De Andrés-Galiana, Juan Luis Fernández-Martínez, and Andrzej Kloczkowski. Robust prediction of single and multiple point protein mutations stability changes. *Biomolecules*, 10(1):67, 2019. DOI: 10.3390/biom10010067.

[37] Yesol Sapozhnikov, Jagdish Suresh Patel, F Marty Ytreberg, and Craig R Miller. Statistical modeling to quantify the uncertainty of foldx-predicted protein folding and binding stability. *BMC Bioinformatics*, 24(1):426, 2023. DOI: 10.5281/zenodo.7897628.

[38] David T Jones. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292(2):195–202, 1999. DOI: 10.1006/jmbi.1999.3091.

[39] Ambrish Roy, Alper Kucukural, and Yang Zhang. I-tasser: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4):725–738, 2010. DOI: 10.1038/nprot.2010.5.

[40] Michele Vendruscolo and Christopher M Dobson. Towards complete descriptions of the free–energy landscapes of proteins. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1827):433–452, 2005. DOI: 10.1098/rsta.2004.1501.

[41] Hongyi Zhou and Yaoqi Zhou. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science*, 11(11):2714–2726, 2002. DOI: 10.1110/ps.0217002.

[42] Raphael Guerois, Jens Erik Nielsen, and Luis Serrano. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of Molecular Biology*, 320(2):369–387, 2002. DOI: 10.1016/S0022-2836(02)00442-4.

[43] Peter Kollman. Free energy calculations: applications to chemical and biochemical phenomena. *Chemical Reviews*, 93(7):2395–2417, 1993. DOI: 10.1021/cr00023a004.

[44] Benjamin BV Louis and Luciano A Abriata. Reviewing challenges of predicting protein melting temperature change upon mutation through the full analysis of a highly detailed dataset with high-resolution structures. *Molecular Biotechnology*, 63(10):863–884, 2021. DOI: 10.1007/s12033-021-00349-0.

[45] Till Siebenmorgen and Martin Zacharias. Computational prediction of protein–protein binding affinities. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(3):e1448, 2020. DOI: 10.1002/wcms.1448.

[46] Julien Michel and Jonathan W Essex. Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *Journal of Computer-Aided Molecular Design*, 24:639–658, 2010. DOI: 10.1007/s10822-010-9363-3.

[47] Christian B Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096): 223–230, 1973. DOI: 10.1126/science.181.4096.22.

[48] Björn Wallner and Arne Elofsson. Can correct protein models be identified? *Protein Science*, 12(5):1073–1086, 2003. DOI: 10.1110/ps.0236803.

[49] Pascal Benkert, Silvio CE Tosatto, and Dietmar Schomburg. Qmean: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, 71 (1):261–277, 2008. DOI: /10.1002/prot.21715.

[50] Pascal Benkert, Marco Biasini, and Torsten Schwede. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3):343–350, 2011. DOI: 10.1093/bioinformatics/btq662.

[51] Emidio Capriotti, Piero Fariselli, and Rita Casadio. I-mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, 33(suppl_2): W306–W310, 2005. DOI: 10.1093/nar/gki375.

[52] MD Shaji Kumar, K Abdulla Bava, M Michael Gromiha, Ponraj Prabakaran, Koji Kitajima, Hatsuho Uedaira, and Akinori Sarai. Protherm and pronit: thermodynamic databases for proteins and protein–nucleic acid interactions. *Nucleic Acids Research*, 34(suppl_1):D204–D206, 2006. DOI: 10.1093/nar/gkj103.

[53] Yves Dehouck, Jean Marc Kwasigroch, Dimitri Gilis, and Marianne Rooman. Popmusic 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, 12(1):1–12, 2011. DOI: 10.1186/1471-2105-12-151.

[54] Martina Audagnotto, Werngard Czechtizky, Leonardo De Maria, Helena Käck, Garegin Papoian, Lars Tornberg, Christian Tyrchan, and Johan Ulander. Machine learning/molecular dynamic protein structure prediction approach to investigate the protein conformational ensemble. *Scientific Reports*, 12(1):10018, 2022. DOI: 10.1038/s41598-022-13714-z.

[55] Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, 2002. DOI: 10.1038/nsb0902-646.

[56] Michael Levitt. The birth of computational structural biology. *Nature Structural Biology*, 8 (5):392–393, 2001. DOI: 10.1038/87545.

[57] Jianwen Fang. A critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. *Briefings in Bioinformatics*, 21(4):1285–1292, 2020. DOI: 10.1093/bib/bbz071.

[58] Castrense Savojardo, Pier Luigi Martelli, Rita Casadio, and Piero Fariselli. On the critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. *Briefings in Bioinformatics*, 22(1):601–603, 2021. DOI: 10.1093/bib/bbz102.

[59] Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020. DOI: 10.1038/s41586-019-1923-7.

[60] Mohammed AlQuraishi. End-to-end differentiable learning of protein structure. *Cell Systems*, 8(4):292–301, 2019. DOI: 10.1016/j.cels.2019.03.006.

[61] Sarah Webb et al. Deep learning for biology. *Nature*, 554(7693):555–557, 2018. DOI: 10.1038/d41586-018-02174-z.

[62] Ron O Dror, Robert M Dirks, JP Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics*, 41:429–452, 2012. DOI: 10.1146/annurev-biophys-042910-155245.

[63] Adam Hospital, Josep Ramon Goñi, Modesto Orozco, and Josep L Gelpí. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, pages 37–47, 2015. DOI: 10.2147/AABC.S70333.

[64] Aditi Garg and Debnath Pal. Exploring the use of molecular dynamics in assessing protein variants for phenotypic alterations. *Human Mutation*, 40(9):1424–1435, 2019. DOI: 10.1002/humu.23800.

[65] Davide Chicco. Ten quick tips for machine learning in computational biology. *BioData Mining*, 10(1):35, 2017. DOI: 10.1186/s13040-017-0155-3.

[66] William Gilpin, Yitong Huang, and Daniel B Forger. Learning dynamics from large biological data sets: machine learning meets systems biology. *Current Opinion in Systems Biology*, 22: 1–7, 2020. DOI: 10.1016/j.coisb.2020.07.009.

[67] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019. DOI: 10.1038/s42256-019-0048-x.

[68] Zachary C Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, 2018. DOI: 10.1145/3236386.3241340.

[69] John Moult, Jan T Pedersen, Richard Judson, and Krzysztof Fidelis. A large-scale experiment to assess protein structure prediction methods. *Proteins: Structure, Function, and Bioinformatics*, 23(3):ii–iv, 1995. DOI: 10.1002/prot.340230303.

[70] Rohit Shukla and Timir Tripathi. Molecular dynamics simulation in drug discovery: opportunities and challenges. *Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design*, pages 295–316, 2021. DOI: 10.1007/978-981-15-8936-2_12.

[71] Punam Sonar, Luca Bellucci, Alessandro Mossa, Pétur O Heidarsson, Birthe B Kragelund, and Ciro Cecconi. Effects of ligand binding on the energy landscape of acyl-coa-binding protein. *Biophysical Journal*, 119(9):1821–1832, 2020. DOI: 10.1016/j.bpj.2020.09.016.

[72] Egidijus Kazlauskas, Vytautas Petrauskas, Vaida Paketurytė, and Daumantas Matulis. Standard operating procedure for fluorescent thermal shift assay (ftsa) for determination of protein–ligand binding and protein stability. *European Biophysics Journal*, 50(3-4):373–379, 2021. DOI: 10.1007/s00249-021-01537-1.

[73] Lucia Fusani, David S Palmer, Don O Somers, and Ian D Wall. Exploring ligand stability in protein crystal structures using binding pose metadynamics. *Journal of Chemical Information and Modeling*, 60(3):1528–1539, 2020. DOI: 10.1021/acs.jcim.9b00843.

[74] Marina A Pak, Nikita V Dovidchenko, Satyarth Mishra Sharma, and Dmitry N Ivankov. The new mega dataset combined with a deep neural network makes progress in predicting the impact of single mutations on protein stability. *BioRxiv*, pages 2022–12, 2023. DOI: 10.1101/2022.12.31.522396.

[75] Fabrizio Pucci, Martin Schwersensky, and Marianne Rooman. Artificial intelligence challenges for predicting the impact of mutations on protein stability. *Current Opinion in Structural Biology*, 72:161–168, 2022. DOI: /10.1016/j.sbi.2021.11.001.

[76] Rakesh Trivedi and Hampapathalu Adimurthy Nagarajaram. Intrinsically disordered proteins: an overview. *International Journal of Molecular Sciences*, 23(22):14050, 2022. DOI: 10.3390/ijms232214050.

[77] Juan J Galano-Frutos, Francho Nerín-Fonz, and Javier Sancho. Calculation of protein folding thermodynamics using molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 63(24):7791–7806, 2023. DOI: 10.1021/acs.jcim.3c01107.

[78] Helmut Grubmüller, Helmut Heller, Andreas Windemuth, and Klaus Schulten. Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6(1-3):121–142, 1991. DOI: 10.1080/08927029108022142.

[79] Elena della Valle, Paolo Marracino, Stefania Setti, Ruggero Cadossi, Micaela Liberti, and Francesca Apollonio. Magnetic molecular dynamics simulations with velocity verlet algorithm. In *2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, pages 1–4. IEEE, 2017. DOI: 10.23919/URSIGASS.2017.8105168.

[80] Justin A Lemkul. Preparing and analyzing polarizable molecular dynamics simulations with the classical drude oscillator model. *Computational Design of Membrane Proteins*, pages 219–240, 2021. DOI: 10.1007/978-1-0716-1468-6_13.

[81] Nikolay Kondratyuk, Vsevolod Nikolskiy, Daniil Pavlov, and Vladimir Stegailov. Gpu-accelerated molecular dynamics: State-of-art software performance and porting from nvidia cuda to amd hip. *The International Journal of High Performance Computing Applications*, 35 (4):312–324, 2021. DOI: 10.1177/10943420211008288.

[82] Jumin Lee, Xi Cheng, Sunhwan Jo, Alexander D MacKerell, Jeffery B Klauda, and Wonpil Im. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Biophysical Journal*, 110(3):641a, 2016. DOI: 10.1021/acs.jctc.5b00935.

[83] James C Phillips, David J Hardy, Julio DC Maia, John E Stone, João V Ribeiro, Rafael C Bernardi, Ronak Buch, Giacomo Fiorin, Jérôme Hénin, Wei Jiang, et al. Scalable molecular dynamics on cpu and gpu architectures with namd. *The Journal of Chemical Physics*, 153(4), 2020. DOI: 10.1063/5.0014475.

[84] Francesco Rao and Martin Karplus. Protein dynamics investigated by inherent structure analysis. *Proceedings of the National Academy of Sciences*, 107(20):9152–9157, 2010. DOI: 10.1073/pnas.0915087107.

[85] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996. DOI: 10.1016/0263-7855(96)00018-5.

[86] Warren L DeLano et al. Pymol: An open-source molecular graphics tool. , 2002. URL `https://legacy.ccp4.ac.uk/newsletters/newsletter40/11_pymol.pdf`.

[87] Yinglong Miao, Ferran Feixas, Changsun Eun, and J Andrew McCammon. Accelerated molecular dynamics simulations of protein folding. *Journal of Computational Chemistry*, 36(20): 1536–1549, 2015. DOI: 10.1002/jcc.23964.

[88] Sara Calhoun and Valerie Daggett. Structural effects of the l145q, v157f, and r282w cancer-associated mutations in the p53 dna-binding core domain. *Biochemistry*, 50(23):5345–5353, 2011. DOI: 10.1021/bi200192j.

[89] Zhuoqiang Guo, Denghui Lu, Yujin Yan, Siyu Hu, Rongrong Liu, Guangming Tan, Ninghui Sun, Wanrun Jiang, Lijun Liu, Yixiao Chen, et al. Extending the limit of molecular dynamics with ab initio accuracy to 10 billion atoms. In *Proceedings of the 27th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, pages 205–218, 2022. DOI: 10.1145/3503221.3508425.

[90] Maria Pechlaner and Wilfred F van Gunsteren. On the use of intra-molecular distance and angle constraints to lengthen the time step in molecular and stochastic dynamics simulations of proteins. *Proteins: Structure, Function, and Bioinformatics*, 90(2):543–559, 2022. DOI: 10.1002/prot.26251.

[91] Shunzhou Wan, Agastya P Bhati, and Peter V Coveney. Comparison of equilibrium and nonequilibrium approaches for relative binding free energy predictions. *Journal of Chemical Theory and Computation*, 19(21):7846–7860, 2023. DOI: 10.1021/acs.jctc.3c00842.

[92] Matteo Aldeghi, Bert L de Groot, and Vytautas Gapsys. Accurate calculation of free energy changes upon amino acid mutation. *Computational Methods in Protein Evolution*, pages 19–47, 2019. DOI: 10.1007/978-1-4939-8736-8_2.

[93] John C Kendrew, G Bodo, Howard M Dintzis, RG Parrish, Harold Wyckoff, and David C Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, 1958. DOI: 10.1038/181662a0.

[94] Kurt Wüthrich. Nmr with proteins and nucleic acids. *Europhysics News*, 17(1):11–13, 1986. DOI: 10.1051/epn/19861701011.

[95] Christopher M Dobson. Protein misfolding, evolution and disease. *Trends in Biochemical Sciences*, 24(9):329–332, 1999. DOI: 10.1016/S0968-0004(99)01445-0.

[96] Michael Levitt and Arieh Warshel. Computer simulation of protein folding. *Nature*, 253(5494): 694–698, 1975. DOI: 10.1038/253694a0.

[97] J Andrew McCammon, Bruce R Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977. DOI: 10.1038/267585a0.

[98] Valerie Daggett and Michael Levitt. Protein unfolding pathways explored through molecular dynamics simulations. *Journal of Molecular Biology*, 232(2):600–619, 1993. DOI: 10.1006/jmbi.1993.1414.

[99] Aaron R Dinner, Andreaj Sali, and Martin Karplus. The folding mechanism of larger model proteins: role of native structure. *Proceedings of the National Academy of Sciences*, 93(16):8356–8361, 1996. DOI: 10.1073/pnas.93.16.835.

[100] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. DOI: 10.1093/nar/28.1.235.

[101] Leonardo G Ferreira, Ricardo N Dos Santos, Glaucius Oliva, and Adriano D Andricopulo. Molecular docking and structure-based drug design strategies. *Molecules*, 20(7):13384–13421, 2015. DOI: 10.3390/molecules200713384.

[102] Nikolaj Blom, Thomas Sicheritz-Pontén, Ramneek Gupta, Steen Gammeltoft, and Søren Brunak. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, 4(6):1633–1649, 2004. DOI: 10.1002/pmic.200300771.

[103] Frank Noé, Gianni De Fabritiis, and Cecilia Clementi. Machine learning for protein folding and dynamics. *Current Opinion in Structural Biology*, 60:77–84, 2020. DOI: 10.1016/j.sbi.2019.12.005.

[104] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. DOI: 10.1038/s41586-021-03819-2.

[105] Nadine Homeyer and Holger Gohlke. Free energy calculations by the molecular mechanics poisson- boltzmann surface area method. *Molecular Informatics*, 31(2):114–122, 2012. DOI: 10.1002/minf.201100135.

[106] Michael K Gilson, Tiqing Liu, Michael Baitaluk, George Nicola, Linda Hwang, and Jenny Chong. Bindingdb in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*, 44(D1):D1045–D1053, 2016. DOI: 10.1093/nar/gkv1072.

[107] Irina S Moreira, Pedro A Fernandes, and Maria J Ramos. Hot spots—a review of the protein–protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*, 68(4):803–812, 2007. DOI: 10.1002/prot.21396.

[108] S Kumar and R Nussinov. How do thermophilic proteins deal with heat? *Cellular and Molecular Life Sciences CMLS*, 58:1216–1233, 2001. DOI: 10.1007/PL00000935.

[109] Ruth Nussinov and Chung-Jung Tsai. Allostery in disease and in drug discovery. *Cell*, 153(2): 293–305, 2013. DOI: 10.1016/j.cell.2013.03.034.

[110] K Gunasekaran, Buyong Ma, and Ruth Nussinov. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function, and Bioinformatics*, 57(3):433–443, 2004. DOI: doi.org/10.1002/prot.20232.

[111] Nobuhiko Tokuriki and Dan S Tawfik. Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology*, 19(5):596–604, 2009. DOI: 10.1016/j.sbi.2009.08.003.

[112] Mark A DePristo, Daniel M Weinreich, and Daniel L Hartl. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Reviews Genetics*, 6(9):678–687, 2005. DOI: 10.1038/nrg1672.

[113] Peter E Wright and H Jane Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, 16(1):18–29, 2015. DOI: 10.1038/nrm3920.

[114] Irwin D Kuntz, Jeffrey M Blaney, Stuart J Oatley, Robert Langridge, and Thomas E Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161 (2):269–288, 1982. DOI: 10.1016/0022-2836(82)90153-X.

[115] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G Wolynes. Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, 1997. DOI: 10.1146/annurev.physchem.48.1.545.

[116] Jiarui Chen and Shirley WI Siu. Machine learning approaches for quality assessment of protein structures. *Biomolecules*, 10(4):626, 2020. DOI: 10.3390/biom10040626.

[117] David Baker and Andrej Sali. Protein structure prediction and structural genomics. *Science*, 294(5540):93–96, 2001. DOI: 10.1126/science.1065659.

[118] Matthew Jacobson and Andrej Sali. Comparative protein structure modeling and its applications to drug discovery. *Annual Reports in Medicinal Chemistry*, 39(85):259–274, 2004. DOI: 10.1016/S0065-7743(04)39020-2.

[119] Björn Wallner and Arne Elofsson. Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics*, 21(23):4248–4254, 2005. DOI: 10.1093/bioinformatics/bti702.

[120] Min-yi Shen and Andrej Sali. Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15(11):2507–2524, 2006. DOI: 10.1110/ps.062416606.

[121] Hongyi Zhou and Jeffrey Skolnick. Goap: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophysical Journal*, 101(8):2043–2052, 2011. DOI: 10.1016/j.bpj.2011.09.012.

[122] Jian Zhang and Yang Zhang. A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. *PloS One*, 5 (10):e15386, 2010. DOI: 10.1371/journal.pone.0015386.

[123] Sohee Kwon, Jonghun Won, Andriy Kryshtafovych, and Chaok Seok. Assessment of protein model structure accuracy estimation in casp14: Old and new challenges. *Proteins: Structure, Function, and Bioinformatics*, 89(12):1940–1948, 2021. DOI: 10.1002/prot.26192.

[124] Arjun Ray, Erik Lindahl, and Björn Wallner. Improved model quality assessment using proq2. *BMC Bioinformatics*, 13:1–12, 2012. DOI: 10.1186/1471-2105-13-224.

[125] Karolis Uziela, Nanjiang Shu, Björn Wallner, and Arne Elofsson. Proq3: Improved model quality assessments using rosetta energy terms. *Scientific Reports*, 6(1):33509, 2016. DOI: 10.1038/srep33509(2016).

[126] Karolis Uziela, David Menéndez Hurtado, Nanjiang Shu, Björn Wallner, and Arne Elofsson. Proq3d: improved model quality assessments using deep learning. *Bioinformatics*, 33(10): 1578–1580, 2017. DOI: 10.1093/bioinformatics/btw819.

[127] Renzhi Cao, Debswapna Bhattacharya, Jie Hou, and Jianlin Cheng. Deepqa: improving the estimation of single protein model quality with deep belief networks. *BMC Bioinformatics*, 17:1–9, 2016. DOI: 10.1186/s12859-016-1405-y.

[128] Georgy Derevyanko, Sergei Grudinin, Yoshua Bengio, and Guillaume Lamoureux. Deep convolutional networks for quality assessment of protein folds. *Bioinformatics*, 34(23):4046–4053, 2018. DOI: 10.1093/bioinformatics/bty494.

[129] Guillaume Pagès, Benoit Charmettant, and Sergei Grudinin. Protein model quality assessment using 3d oriented convolutional neural networks. *Bioinformatics*, 35(18):3313–3319, 2019. DOI: 10.1093/bioinformatics/btz122.

[130] Rin Sato and Takashi Ishida. Protein model accuracy estimation based on local structure quality assessment using 3d convolutional neural network. *PloS One*, 14(9):e0221347, 2019. DOI: 10.1371/journal.pone.0221347.

[131] Yuma Takei and Takashi Ishida. P3cmqa: single-model quality assessment using 3dcnn with profile-based features. *Bioengineering*, 8(3):40, 2021. DOI: 10.3390/bioengineering8030040.

[132] Federico Baldassarre, David Menéndez Hurtado, Arne Elofsson, and Hossein Azizpour. Graphqa: protein model quality assessment using graph convolutional networks. *Bioinformatics*, 37(3):360–366, 2021. DOI: 10.1093/bioinformatics/btaa714.

[133] Chen Chen, Xiao Chen, Alex Morehead, Tianqi Wu, and Jianlin Cheng. 3d-equivariant graph neural networks for protein model quality assessment. *Bioinformatics*, 39(1):btad030, 2023. DOI: 10.1093/bioinformatics/btad030.

[134] Naozumi Hiranuma, Hahnbeom Park, Minkyung Baek, Ivan Anishchenko, Justas Dauparas, and David Baker. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature Communications*, 12(1):1340, 2021. DOI: 10.1038/s41467-021-21511-x.

[135] Md Hossain Shuvo, Sutanu Bhattacharya, and Debswapna Bhattacharya. Qdeep: distance-based protein model quality estimation by residue-level ensemble error classifications using stacked deep residual neural networks. *Bioinformatics*, 36(Supplement_1):i285–i291, 2020. DOI: 10.1093/bioinformatics/btaa455.

[136] Modesto Orozco. A theoretical view of protein dynamics. *Chemical Society Reviews*, 43(14): 5051–5066, 2014. DOI: 10.1039/C3CS60474H.

[137] Janita Thusberg, Ayodeji Olatubosun, and Mauno Vihinen. Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*, 32(4):358–368, 2011. DOI: 10.1002/humu.21445.

[138] Peter D Stenson, Matthew Mort, Edward V Ball, Katy Evans, Matthew Hayden, Sally Heywood, Michelle Hussain, Andrew D Phillips, and David N Cooper. The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics*, 136:665–677, 2017. DOI: 10.1007/s00439-017-1779-6.

[139] Dennis J Selkoe. Folding proteins in fatal ways. *Nature*, 426(6968):900–904, 2003. DOI: 10.1038/nature02264.

[140] Cyriac Kandoth, Michael D McLellan, Fabio Vandin, Kai Ye, Beifang Niu, Charles Lu, Mingchao Xie, Qunyuan Zhang, Joshua F McMichael, Matthew A Wyczalkowski, et al. Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471):333–339, 2013. DOI: 10.1038/nature12634.

[141] Maria Petrosino, Leonore Novak, Alessandra Pasquo, Roberta Chiaraluce, Paola Turina, Emidio Capriotti, and Valerio Consalvi. Analysis and interpretation of the impact of missense variants in cancer. *International Journal of Molecular Sciences*, 22(11):5416, 2021. DOI: 10.3390/ijms22115416.

[142] Angel L Pey, François Stricher, Luis Serrano, and Aurora Martinez. Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases. *The American Journal of Human Genetics*, 81(5):1006–1024, 2007. DOI: 10.1086/521987.

[143] Shahid Iqbal, Fuyi Li, Tatsuya Akutsu, David B Ascher, Geoffrey I Webb, and Jiangning Song. Assessing the performance of computational predictors for estimating protein stability changes upon missense mutations. *Briefings in Bioinformatics*, 22(6):bbab184, 2021. DOI: 10.1093/bib/bbab184.

[144] Elizabeth H Kellogg, Andrew Leaver-Fay, and David Baker. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics*, 79(3):830–838, 2011. DOI: 10.1002/prot.22921.

[145] Joost Schymkowitz, Jesper Borg, Francois Stricher, Robby Nys, Frederic Rousseau, and Luis Serrano. The foldx web server: an online force field. *Nucleic Acids Research*, 33(suppl_2): W382–W388, 2005. DOI: 10.1093/nar/gki387.

[146] Javier Delgado, Leandro G Radusky, Damiano Cianferoni, and Luis Serrano. Foldx 5.0: working with rna, small molecules and a new graphical interface. *Bioinformatics*, 35(20):4168–4169, 2019. DOI: 10.1093/bioinformatics/btz184.

[147] Gen Li, Shailesh Kumar Panday, and Emil Alexov. Saafec-seq: a sequence-based method for predicting the effect of single point mutations on protein thermodynamic stability. *International Journal of Molecular Sciences*, 22(2):606, 2021. DOI: 10.3390/ijms22020606.

[148] Chi-Wei Chen, Jerome Lin, and Yen-Wei Chu. istable: off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics*, 14:1–14, 2013. DOI: 10.1186/1471-2105-14-S2-S5.

[149] Chi-Wei Chen, Meng-Han Lin, Chi-Chou Liao, Hsung-Pin Chang, and Yen-Wei Chu. istable 2.0: predicting protein thermal stability changes by integrating various characteristic modules. *Computational and Structural Biotechnology Journal*, 18:622–630, 2020. DOI: 10.1016/j.csbj.2020.02.021.

[150] Piero Fariselli, Pier Luigi Martelli, Castrense Savojardo, and Rita Casadio. Inps: predicting the impact of non-synonymous variations on protein stability from sequence. *Bioinformatics*, 31(17):2816–2821, 2015. DOI: 10.1093/bioinformatics/btv291.

[151] Lukas Folkman, Bela Stantic, Abdul Sattar, and Yaoqi Zhou. Ease-mm: sequence-based prediction of mutation-induced stability changes with feature-based multiple models. *Journal of Molecular Biology*, 428(6):1394–1405, 2016. DOI: 10.1016/j.jmb.2016.01.012.

[152] Corrado Pancotti, Silvia Benevenuta, Valeria Repetto, Giovanni Birolo, Emidio Capriotti, Tiziana Sanavia, and Piero Fariselli. A deep-learning sequence-based method to predict protein stability changes upon genetic variations. *Genes*, 12(6):911, 2021. DOI: 10.3390/genes12060911.

[153] Xuan Lv, Jianwen Chen, Yutong Lu, Zhiguang Chen, Nong Xiao, and Yuedong Yang. Accurately predicting mutation-caused stability changes from protein sequences using extreme gradient boosting. *Journal of Chemical Information and Modeling*, 60(4):2388–2395, 2020. DOI: 10.1021/acs.jcim.0c00064.

[154] Shahid Iqbal, Fang Ge, Fuyi Li, Tatsuya Akutsu, Yuanting Zheng, Robin B Gasser, Dong-Jun Yu, Geoffrey I Webb, and Jiangning Song. Prost: Alphafold2-aware sequence-based predictor to estimate protein stability changes upon missense mutations. *Journal of Chemical Information and Modeling*, 62(17):4270–4282, 2022. DOI: 10.1021/acs.jcim.2c00799.

[155] Arun Prasad Pandurangan, Bernardo Ochoa-Montano, David B Ascher, and Tom L Blundell. Sdm: a server for predicting effects of mutations on protein stability. *Nucleic Acids Research*, 45(W1):W229–W235, 2017. DOI: 10.1093/nar/gkx439.

[156] Catherine L Worth, Robert Preissner, and Tom L Blundell. Sdm—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Research*, 39(suppl_2):W215–W222, 2011. DOI: 10.1093/nar/gkr363.

[157] Douglas EV Pires, David B Ascher, and Tom L Blundell. mcsm: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, 30(3):335–342, 2014. DOI: 10.1093/bioinformatics/btt691.

[158] Douglas EV Pires, David B Ascher, and Tom L Blundell. Duet: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research*, 42(W1):W314–W319, 2014. DOI: 10.1093/nar/gku411.

[159] Ludovica Montanucci, Emidio Capriotti, Giovanni Birolo, Silvia Benevenuta, Corrado Pancotti, Dennis Lal, and Piero Fariselli. Ddgun: an untrained predictor of protein stability

changes upon amino acid variants. *Nucleic Acids Research*, 50(W1):W222–W227, 2022. DOI: 10.1093/nar/gkac325.

[160] Ludovica Montanucci, Emidio Capriotti, Yotam Frank, Nir Ben-Tal, and Piero Fariselli. Ddgun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. *BMC Bioinformatics*, 20:1–10, 2019. DOI: 10.1186/s12859-019-2923-1.

[161] Castrense Savojardo, Piero Fariselli, Pier Luigi Martelli, and Rita Casadio. Inps-md: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics*, 32 (16):2542–2544, 2016. DOI: 10.1093/bioinformatics/btw192.

[162] Carlos HM Rodrigues, Douglas EV Pires, and David B Ascher. Dynamut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Science*, 30(1):60–69, 2021. DOI: 10.1002/pro.3942.

[163] Josef Laimer, Heidi Hofer, Marko Fritz, Stefan Wegenkittl, and Peter Lackner. Maestro-multi agent stability prediction upon point mutations. *BMC Bioinformatics*, 16(1):1–13, 2015. DOI: 10.1186/s12859-015-0548-6.

[164] Shannon Stefl, Hafumi Nishi, Marharyta Petukh, Anna R Panchenko, and Emil Alexov. Molecular mechanisms of disease-causing missense mutations. *Journal of Molecular Biology*, 425(21): 3919–3936, 2013. DOI: 10.1016/j.jmb.2013.07.014.

[165] C George Priya Doss, Chiranjib Chakraborty, Luonan Chen, Hailong Zhu, et al. Integrating in silico prediction methods, molecular docking, and molecular dynamics simulation to predict the impact of alk missense mutations in structural perspective. *BioMed Research International*, 2014, 2014. DOI: 10.1155/2014/895831.

[166] Maria Antonietta La Serra, Pietro Vidossich, Isabella Acquistapace, Anand K Ganesan, and Marco De Vivo. Alchemical free energy calculations to investigate protein–protein interactions: the case of the cdc42/pak1 complex. *Journal of Chemical Information and Modeling*, 62(12): 3023–3033, 2022. DOI: 10.1021/acs.jcim.2c00348.

[167] Alexander D Wade, Agastya P Bhati, Shunzhou Wan, and Peter V Coveney. Alchemical free energy estimators and molecular dynamics engines: Accuracy, precision, and reproducibility. *Journal of Chemical Theory and Computation*, 18(6):3972–3987, 2022. DOI: 10.1021/acs.jctc.2c00114.

[168] Dharmeshkumar Patel, Jagdish Suresh Patel, and F Marty Ytreberg. Implementing and assessing an alchemical method for calculating protein–protein binding free energy. *Journal of Chemical Theory and Computation*, 17(4):2457–2464, 2021. DOI: 10.1021/acs.jctc.0c01045.

[169] Jianxin Duan, Dmitry Lupyan, and Lingle Wang. Improving the accuracy of protein thermostability predictions for single point mutations. *Biophysical Journal*, 119(1):115–127, 2020. DOI: 10.1016/j.bpj.2020.05.020.

[170] Vytautas Gapsys, Servaas Michielssens, Daniel Seeliger, and Bert L De Groot. pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of Computational Chemistry*, 36(5):348–354, 2015. DOI: 10.1002/jcc.23804.

[171] Vytautas Gapsys and Bert L de Groot. pmx webserver: a user friendly interface for alchemistry. *Journal of Chemical Information and Modeling*, 57(2):109–114, 2017. DOI: 10.1021/acs.jcim.6b00498.

[172] Vytautas Gapsys, Servaas Michielssens, Jan Henning Peters, Bert L de Groot, and Hadas Leonov. Calculation of binding free energies. *Molecular Modeling of Proteins*, pages 173–209, 2015. DOI: 10.1007/978-1-4939-1465-4_9.

[173] Gavin E Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90:1481–1487, 1998. DOI: 10.1023/A:1023208217925.

[174] Vytautas Gapsys, Servaas Michielssens, Daniel Seeliger, and Bert L de Groot. Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. *Angewandte Chemie International Edition*, 55(26):7364–7368, 2016. DOI: 10.1002/anie.201510054.

[175] Shuyu Wang, Hongzhou Tang, Yuliang Zhao, and Lei Zuo. Bayestab: Predicting effects of mutations on protein stability with uncertainty quantification. *Protein Science*, 31(11):e4467, 2022. DOI: 10.1002/pro.4467.

[176] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005. DOI: 10.1002/jcc.20291.

[177] Szilárd Páll, Artem Zhmurov, Paul Bauer, Mark Abraham, Magnus Lundborg, Alan Gray, Berk Hess, and Erik Lindahl. Heterogeneous parallelization and acceleration of molecular dynamics simulations in gromacs. *The Journal of Chemical Physics*, 153(13), 2020. DOI: 10.1063/5.0018516.

[178] Romelia Salomon-Ferrer, Andreas W Gotz, Duncan Poole, Scott Le Grand, and Ross C Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald. *Journal of Chemical Theory and Computation*, 9(9):3878–3888, 2013. DOI: 10.1021/ct400314y.

[179] Mehmet Akdel, Douglas EV Pires, Eduard Porta Pardo, Jürgen Jänes, Arthur O Zalevsky, Bálint Mészáros, Patrick Bryant, Lydia L Good, Roman A Laskowski, Gabriele Pozzati, et al. A structural biology community assessment of alphafold2 applications. *Nature Structural & Molecular Biology*, 29(11):1056–1067, 2022. DOI: 10.1038/s41594-022-00849-w.

[180] Qisheng Pan, Thanh Binh Nguyen, David B Ascher, and Douglas EV Pires. Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures. *Briefings in Bioinformatics*, 23(2):bbac025, 2022. DOI: 10.1093/bib/bbac025.

[181] Gwen R Buel and Kylie J Walters. Can alphafold2 predict the impact of missense mutations on structure? *Nature Structural & Molecular Biology*, 29(1):1–2, 2022. DOI: 10.1038/s41594-021-00714-2.

[182] Marina A Pak, Karina A Markhieva, Mariia S Novikova, Dmitry S Petrov, Ilya S Vorobyev, Ekaterina S Maksimova, Fyodor A Kondrashov, and Dmitry N Ivankov. Using alphafold to predict the impact of single mutations on protein stability and function. *PloS One*, 18(3): e0282689, 2023. DOI: 10.1371/journal.pone.0282689.

[183] Hao-Bo Guo, Alexander Perminov, Selemon Bekele, Gary Kedziora, Sanaz Farajollahi, Vanessa Varaljay, Kevin Hinkle, Valeria Molinero, Konrad Meister, Chia Hung, et al. Alphafold2 models indicate that protein sequence determines both structure and dynamics. *Scientific Reports*, 12 (1):10696, 2022. DOI: 10.1038/s41598-022-14382-9.

[184] John Moult, Jan T Pedersen, Richard Judson, and Krzysztof Fidelis. A large-scale experiment to assess protein structure prediction methods. *Proteins: Structure, Function, and Bioinformatics*, 23(3):ii–iv, 1995. DOI: 10.1002/prot.340230303.

[185] J. Lemkul. From proteins to perturbed hamiltonians: A suite of tutorials for the gromacs-2018 molecular simulation package [article v1.0]. *Living Journal of Computational Molecular Science*, 1:5068, 2019. DOI: 10.33011/livecoms.1.1.5068.

[186] Gregory R Bowman. Accurately modeling nanosecond protein dynamics requires at least microseconds of simulation. *Journal of Computational Chemistry*, 37(6):558–566, 2016. DOI: 10.1002/jcc.23973.

[187] Lili Duan, Xiaona Guo, Yalong Cong, Guoqiang Feng, Yuchen Li, and John ZH Zhang. Accelerated molecular dynamics simulation for helical proteins folding in explicit water. *Frontiers in Chemistry*, 7:540, 2019. DOI: 10.3389/fchem.2019.00540.

[188] Hui-Hsu Tsai, Chung-Jung Tsai, Buyong Ma, and Ruth Nussinov. In silico protein design by combinatorial assembly of protein building blocks. *Protein Science*, 13(10):2753–2765, 2004. DOI: 10.1110/ps.04774004.

[189] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28): 6474–6487, 2001. DOI: 10.1021/jp003919d.

[190] Herman JC Berendsen, James PM Postma, Wilfred F van Gunsteren, and Jan Hermans. Interaction models for water in relation to protein hydration. In *Intermolecular forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, april 13–16, 1981*, pages 331–342. Springer, 1981. DOI: 10.1007/978-94-015-7658-1_21.

[191] Dawei Zhang and Raudah Lazim. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports*, 7(1):44651, 2017. DOI: 10.1038/srep44651.

[192] Imlimaong Aier, Pritish Kumar Varadwaj, and Utkarsh Raj. Structural insights into conformational stability of both wild-type and mutant ezh2 receptor. *Scientific Reports*, 6(1):34984, 2016. DOI: 10.1038/srep34984.

[193] Robert T McGibbon, Kyle A Beauchamp, Matthew P Harrigan, Christoph Klein, Jason M Swails, Carlos X Hernández, Christian R Schwantes, Lee-Ping Wang, Thomas J Lane, and Vijay S Pande. Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528–1532, 2015. DOI: 10.1016/j.bpj.2015.08.015.

[194] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12):2577–2637, 1983. DOI: 10.1002/bip.360221211.

[195] Robert B Best, Gerhard Hummer, and William A Eaton. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*, 110(44):17874–17879, 2013. DOI: 10.1073/pnas.1311599110.

[196] Adam Zemla. Lga: a method for finding 3d similarities in protein structures. *Nucleic Acids Research*, 31(13):3370–3374, 2003. DOI: 10.1093/nar/gkg571.

[197] Xiaohong Zhuang, Judah R Makover, Wonpil Im, and Jeffery B Klauda. A systematic molecular dynamics simulation study of temperature dependent bilayer structural properties. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838(10):2520–2529, 2014. DOI: 10.1016/j.bbamem.2014.06.010.

[198] Castrense Savojardo, Maria Petrosino, Giulia Babbi, Samuele Bovo, Carles Corbi-Verge, Rita Casadio, Piero Fariselli, Lukas Folkman, Aditi Garg, Mostafa Karimi, et al. Evaluating the predictions of the protein stability change upon single amino acid substitutions for the fxn cagi5 challenge. *Human Mutation*, 40(9):1392–1399, 2019. DOI: 10.1002/humu.23843.

[199] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature Methods*, 19 (6):679–682, 2022. DOI: 10.1038/s41592-022-01488-1.

[200] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010. DOI: 10.1002/prot.22711.

[201] Robert B Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix- coil transition of polypeptides. *The Journal of Physical Chemistry B*, 113(26): 9004–9015, 2009. DOI: 10.1021/jp901540t.

[202] 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, 1983. DOI: 10.1063/1.445869.

[203] Vytautas Gapsys, Daniel Seeliger, and Bert L de Groot. New soft-core potential function for molecular dynamics based alchemical free energy calculations. *Journal of Chemical Theory and Computation*, 8(7):2373–2382, 2012. DOI: 10.1021/ct300220p.

[204] Jianting Gong, Juexin Wang, Xizeng Zong, Zhiqiang Ma, and Dong Xu. Prediction of protein stability changes upon single-point variant using 3d structure profile. *Computational and Structural Biotechnology Journal*, 21:354–364, 2023. DOI: 10.1016/j.csbj.2022.12.008.

[205] Aljoscha M Hahn and Holger Then. Measuring the convergence of monte carlo free-energy calculations. *Physical Review E*, 81(4):041117, 2010. DOI: 10.1103/PhysRevE.81.041117.

[206] Michael R Shirts, Eric Bair, Giles Hooker, and Vijay S Pande. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Physical Review Letters*, 91(14):140601, 2003. DOI: 10.1103/PhysRevLett.91.140601.

[207] Charles H Bennett. Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976. DOI: 10.1016/0021-9991(76)90078-4.

[208] Alex N Bullock, Julia Henckel, and Alan R Fersht. Quantitative analysis of residual folding and dna binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. *Oncogene*, 19(10):1245–1256, 2000. DOI: 10.1038/sj.onc.1203434.

[209] Fabrizio Pucci, Katrien V Bernaerts, Jean Marc Kwasigroch, and Marianne Rooman. Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics*, 34(21):3659–3665, 2018. DOI: 10.1093/bioinformatics/bty348.

[210] Carsten Kutzner, Christian Kniep, Austin Cherian, Ludvig Nordstrom, Helmut Grubmuúller, Bert L de Groot, and Vytautas Gapsys. Gromacs in the cloud: A global supercomputer to speed up alchemical drug design. *Journal of Chemical Information and Modeling*, 62(7): 1691–1711, 2022. DOI: 10.1021/acs.jcim.2c00044.

[211] Sumaiya Iqbal, Eduardo Pérez-Palma, Jakob B Jespersen, Patrick May, David Hoksza, Henrike O Heyne, Shehab S Ahmed, Zaara T Rifat, M Sohel Rahman, Kasper Lage, et al. Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants. *Proceedings of the National Academy of Sciences*, 117(45):28201–28211, 2020. DOI: 10.1073/pnas.2002660117.

[212] Jianguo Liu and Xianjiang Kang. Grading amino acid properties increased accuracies of single point mutation on protein stability prediction. *BMC Bioinformatics*, 13(1):1–11, 2012. DOI: 10.1186/1471-2105-13-44.