学位論文

Enhancing Precision in Antibody-Antigen Complex Structure Prediction Through Parametric Optimization of RosettaAb Docking Scoring Function

RosettaAb Docking スコアリング関数のパラメータ最適化を通じた抗体-抗原複合体構造予測の精度向上

Sangeetha Ratnayake

Doctor of Philosophy

Laboratory of Information Biology
Department of Bioinformatics & Bioengineering
Graduate School of Information Science & Technology
Hokkaido University

January 2024

Acknowledgment

In my academic journey, profound thanks to Prof. Toshinori Endo for opening the door to the Laboratory of Information Biology, laying the foundation for today's milestone. Special appreciation to Assoc. Prof. Naoki Osada, a guiding compass in the research labyrinth. Dr. Axel Martinelli's unwavering belief and encouragement throughout my Ph.D. journey have been a guiding light, for which I'm truly grateful.

Deepest gratitude to Prof. Luca Varani and Dr. Luca Simonelli of the Laboratory of Structural Biology for invaluable knowledge and mentorship. Heartfelt thanks to my boyfriend, Dr. Stefan Brennsteiner, for constant support and insightful critiques, crucial to this journey's success.

In memory of my father, Gunapala Ratnayake, whose belief resonates even in his absence. To my mother and sisters, your unwavering encouragement nurtured my aspirations.

Sincere thanks to Narisa Rojanamonthien for transcending distances, propelling me to complete this Ph.D. To the entire Hult Prize Family, your support orchestrated this success. Grateful for the symphony of love and support.

Abstract

Understanding the structure of antibody-antigen (Ab-Ag) interactions is crucial for various scientific applications. Computational methods have emerged as efficient tools for studying these interactions, offering advantages over traditional methods. However, existing computational docking methods face challenges in accurately predicting Ab-Ag structures. One key limitation is the scoring function, designed for rigid and well-characterized protein structures, which often falls short in practice despite optimizations for Ab-Ag in programs. Rosetta is a widely used program for Ab-Ag docking.

To address this issue, a proposed solution involves using decoy distribution as a valuable indicator for assessing the goodness of fitting in docking simulations. A decoy, representing an alternative binding pose or conformation, is plotted on a distribution graph that showcases energy scores versus Root Mean Square Deviation (RMSD) values. This decoy distribution becomes crucial for evaluating existing scoring functions and seeking more optimized parameters, essential for accurately predicting Ab-Ag structures.

The thesis comprises four chapters. The first chapter provides background information, while the second chapter evaluates specific parameters within Rosetta-derived scoring functions, focusing on the energy landscape of generated structures. The third chapter develops models within the Rosetta framework to optimize scoring function parameters, using quantitative evaluations of decoy distributions to refine parameters for each Ab-Ag complex. This chapter introduces a novel approach to customizing scoring functions, potentially advancing drug discovery and deepening our understanding of molecular-level Ab-Ag interactions. The fourth chapter summarizes key findings, discusses further applications, and suggests areas for future investigation.

The application of the decoy distribution revealed that the default Rosetta approach proved ineffectual in 88 out of 100 cases, showcasing the influence of particular amino acids within antibody binding sites on its performance. The removal of solvation parameters slightly improved Rosetta's performance, but not to a sufficient extent. A new method was developed to optimize scoring function parameters for each Ab-Ag complex, resulting in significantly reduced RMSD values and the identification of parameters effective for most complexes.

The research outcomes hold implications for drug development, protein engineering, and computational biology. This work serves as a catalyst for innovation in medical research and therapeutic development, shedding light on the complexities of Ab-Ag interactions at the molecular level.

Table of Contents

Acknowledgment ..... **i**

Abstract ..... **ii**

Table of Contents ..... **iv**

List of Tables ..... **v**

List of Figures ..... **v**

List of Abbreviations ..... **vi**

Introduction & Motivation **1**

1.1. Studying Antibody-Antigen interactions *1*

1.2. Computational Docking and their Scoring Functions *1*

1.3. Challenges in Ab-Ag Modeling *5*

1.4. Goal of the Thesis *7*

1.5. Thesis outline and main contribution *8*

The Interplay Between Scoring Functions and Physico-chemical Properties in Antibody-Antigen Docking **9**

2.1. Methods *9*

2.1.1. Input Data Preparation 9

2.1.2. Ab-Ag docking and generating decoys 10

2.1.3. Energy scoring with customized weights 10

2.1.4. Extracting Antibody CDR Loop and Antigen Features 11

2.1.5. Analyzing Ab-Ag Interface Properties 11

2.1.6. Statistical Analysis 12

2.2. Results *12*

2.2.1. Docking scoring functions and docking funnels 12

2.2.2. Amino acid composition in the Ab-Ag interface 14

2.2.3. Significance of physico-chemical properties 15

2.3. Discussion *17*

2.3.1. Evaluation of docking scoring functions 17

2.3.2. Influence of physico-chemical properties in CDR loops and epitopes on docking 18

2.3.3. Correlation between physico-chemical properties and decoy distribution 19

2.4. Conclusion *20*

Analysis in Docking Scoring Functions for Precision Docking Analysis **21**

3.1. Methods *21*

3.1.1. Identification of high predictive power parameters in Rosetta scoring function. 21

3.1.2. Classification of the decoy distribution ... 22
3.1.3. Scoring function optimization models ... 23
3.1.4. Qualitative and quantitative analysis of the optimization model performance ... 24
3.2. Results ... *25*
3.2.1. Total energy score and the parametric contribution ... 25
3.2.2. Optimizing weight parameters by adjusting decoy distributions ... 26
3.2.3. Performance of optimization models ... 29
3.2.4. Efficacy of distinctive Parameters in Ab-Ag docking predictions ... 30
3.3. Discussion ... *31*
3.3.1. Influence of identified parameters on Rosetta docking scoring function ... 32
3.3.2. Optimization impact on lowest energy decoys ... 32
3.3.3. Evaluation of optimization models ... 33
3.4. Conclusion ... *34*
Conclusions ... **35**
4.1. Summary ... *35*
4.2. Topics for Future Research ... *36*
Reference ... **38**
Appendix ... **42**

List of Tables

Table 1: Net Reclassification Improvement. ... 29

List of Figures

Figure 1: Types of Decoy Distributions. ... 13
Figure 2: Quantitative Representation of Decoy Distribution Goodness. ... 13
Figure 3: Amino Acids Frequency comparison. ... 16
Figure 4: Significance of independent parameters in Generalized Linear Regression (GLM). ... 17
Figure 5: Example of a decoy distribution and definition of parameters in Optimization Model 3. ... 22
Figure 6: Parametric Contributions to the Total Energy Score in Rosetta Scoring. ... 26
Figure 7: Optimization Model Performance Comparison. ... 28
Figure 8: Principal Component Analysis. ... 30

List of Abbreviations

Ab	Antibody
Ag	Antigen
CDR	Complementary-determining Region
ELISA	Enzyme-Linked Immunosorbent Assay
Fab	Fragment Antigen Binding
ITC	Isothermal Titration Calorimetry
KNN	K-Nearest Neighbor
LLK	Lazaridis-Karplus
PCA	Principal Component Analysis
PCC	Pearson’s Correlation Coefficient
PPI	Protein-protein Interactions
RMSD	Root Mean Squared Deviation
SPR	Surface Plasmon Resonance

Chapter 1

Introduction & Motivation

1.1. Studying Antibody-Antigen interactions

In recent years, the field of biological medicines, especially antibody-based therapy, has revolutionized the treatment of various disorders. Antibodies, crucial components of the immune system, possess the unique ability to identify and neutralize foreign entities, including bacteria and viruses. Understanding the intricate interactions between antibodies and antigens has become a central focus in the development of antibody-based treatments [1], [2], [3].

Traditional experimental techniques, like X-ray crystallography, have been fundamental in revealing the structures of Ab-Ag complexes. However, these methods have limitations in terms of their cost, throughput, and applicability to all proteins. This is where computational docking steps in—a technique in computational biology and drug discovery that predicts the binding orientation and affinity of small molecules (ligands) with target biomolecules.

1.2. Computational Docking and their Scoring Functions

Computational docking serves as a cost-effective and efficient means to predict the binding structures of Ab-Ag interactions. The primary goal of this approach is to simulate the interaction between antigens and antibodies, allowing for the prediction of their three-dimensional structures upon binding [4]. This simulation facilitates the determination of binding affinity and interaction modes between the two proteins. Through the use of computational algorithms, researchers can analyze potential binding orientations and strengths among biomolecules, thereby enhancing the understanding of their interactions and enabling predictions about the binding outcomes. A prominent method for predicting protein-protein binding is the optimization of a "Scoring Function," specifically designed for rigid and well-characterized protein structures. These scoring functions, present in various systems [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], approximate the system's typical chemical potentials and are integrated into several docking software. The

following are some of commonly used computational docking software, along with the implemented scoring functions and included parameters.

- The FlexX algorithm, introduced by Kramer, Rarey, and Lengauer in 1999, represents a significant advancement in the field of protein-ligand docking [5]. A key feature of FlexX is its utilization of an incremental construction approach, enhancing the efficiency of protein-ligand docking simulations. Notably, the algorithm incorporates a modified version of the scoring function initially developed by Böhm. This modified scoring function includes parameters such as the binding energy of rotatable bonds and pairwise interactions of hydrophobic atoms. By considering these factors, FlexX contributes to a more comprehensive and accurate assessment of protein-ligand binding interactions, thereby advancing the understanding of molecular interactions crucial for drug discovery and design.
- The GOLD (Genetic Optimization for Ligand Docking) software stands out as a valuable tool in the realm of molecular docking simulations [6]. The key innovation lies in its Chemscore function, a scoring mechanism designed to enhance the accuracy of protein-ligand docking predictions. Chemscore takes into consideration crucial molecular interactions, including hydrophobic-hydrophobic contact area, hydrogen bonding, ligand flexibility, and metal interactions. By integrating these factors into the scoring function, GOLD provides a more comprehensive evaluation of the binding affinity between proteins and ligands. This approach contributes to improved precision in predicting ligand binding poses, offering valuable insights for drug discovery and design processes.
- Glide represents a groundbreaking approach to molecular docking and scoring [7]. Central to Glide's success is its utilization of an empirical scoring function that approximates ligand binding free energy. This scoring function incorporates various terms, including contributions from force fields such as electrostatic and van der Waals forces. Additionally, it considers terms that either reward or penalize interactions known to influence ligand binding, especially in hydrophobic regions of the interaction surface. By encompassing these diverse factors, Glide provides a rapid and accurate means of predicting ligand binding

poses, making it a valuable tool in virtual screening for drug discovery applications.

- AutoDock, is a notable software application in the field of molecular docking [8]. Central to AutoDock's capabilities is its utilization of an empirical free energy scoring function. Unlike some other scoring functions, AutoDock considers the intramolecular energetics of both unbound and bound conformations. This comprehensive approach allows AutoDock to provide a more nuanced assessment of ligand binding, incorporating the dynamic aspects of molecular interactions. By considering the energetic contributions from both the ligand and the receptor in various states, AutoDock contributes to a more accurate prediction of binding affinities and ligand binding poses, making it a valuable tool for computational studies in drug discovery and molecular biology.
- AutoDock Vina [9] represents a significant advancement in molecular docking techniques. At its core, AutoDock Vina employs an empirical free energy scoring function that enhances the speed and accuracy of docking simulations. This scoring function is designed to extract empirical information from both the conformational preferences of receptor-ligand complexes and experimental affinity measurements. By combining these sources of information, AutoDock Vina provides a more comprehensive and refined assessment of ligand binding, contributing to improved accuracy in predicting binding poses and affinities. The incorporation of efficient optimization algorithms and multithreading further enhances the software's computational efficiency, making it a valuable tool for researchers in the field of drug discovery and molecular modeling.
- Hex 8.0.0 [10], employs a Spherical Polar Fourier (SPF) correlation approach to calculate shape complementarity, coupled with an optimal in vacuo electrostatic contribution. This innovative technique enables Hex to assess the spatial arrangement of molecules, emphasizing excluded volume considerations and accounting for electrostatic interactions. The reliance on FFT-based algorithms and the utilization of graphics processors contribute to Hex's computational efficiency, making it a powerful tool for researchers engaged in protein-protein docking studies. The integration of these features underscores Hex's capability to

provide valuable insights into molecular interactions and contributes to the broader field of structural biology and drug discovery.

- FRODOCK 2.0 stands out as a valuable tool in the realm of protein-protein docking simulations [11]. The software employs a coarse-grained knowledge-based approach, combining both coarse-grained and atom-based potentials. FRODOCK 2.0 utilizes three key binding energies—van der Waals interactions, desolvation effects, and electrostatic interactions—to comprehensively assess the feasibility of protein-protein docking. By considering these crucial energy components, the software provides a nuanced perspective on the molecular interactions governing protein-protein binding. This coarse-grained knowledge-based strategy enhances the accuracy and efficiency of protein-protein docking simulations, making FRODOCK 2.0 a valuable resource for researchers in the fields of structural biology and drug discovery.
- Rosetta [12], [13], [15] employs an all-atom energy function to calculate the energy of all atomic interactions within a macromolecular structure. This energy calculation encompasses a combination of physical forces such as electrostatics and van der Waals interactions, as well as statistical terms like the probability of finding torsion angles in Ramachandran space. Rosetta's approach provides a comprehensive and detailed assessment of the energy landscape of macromolecules, facilitating molecular modeling and design studies. The software's accuracy and versatility make it a widely used resource in the scientific community for exploring the structural and energetic aspects of macromolecules, contributing significantly to the fields of structural biology and drug discovery.

The scoring functions in these systems are designed based on known protein structures and aim to approximate the typical chemical potentials within a system [5], [6], [7], [8], [9], [10], [11], [14]. They are integrated into the above docking software, enabling researchers to assess potential binding affinities and interaction modes (Appendix A). However, when applied to Ab-Ag interfaces, these scoring functions can fall short due to the unique characteristics of such interactions.

1.3. Decoy Distribution and Numerical Optimization

In this study, our evaluation and optimization methodology introduce a distinctive approach that centers around the use of decoy distributions as a critical indicator. This innovative method provides valuable insights into the accuracy and reliability of computational docking scoring functions, particularly in the context of Ab-Ag interactions.

Decoy distributions, representing the ensemble of potential docking poses generated by the computational method, serve as a pivotal metric in our assessment. Unlike traditional evaluation methods, which may rely on individual docking poses, the analysis of decoy distributions offers a more comprehensive view of the conformational landscape explored during the docking process. This approach aligns with the dynamic nature of PPI, providing a nuanced understanding of the potential binding orientations.

Numerical optimization plays a central role in fine-tuning the parameters of the Rosetta docking scoring function. The objective is to optimize the scoring function for improved accuracy in predicting Ab-Ag binding structures. By employing numerical optimization techniques, we navigate the complex parameter space with efficiency, addressing the challenges associated with manual tuning. This approach enhances the objectivity, robustness, and consistency of our analyses.

The integration of decoy distribution analysis with numerical optimization represents a unique contribution to the field of computational biology and drug discovery. This methodology goes beyond traditional scoring function assessments, offering a more holistic perspective on the challenges posed by Ab-Ag interactions. Emphasizing the significance of decoy distributions as an indicator and the role of numerical optimization in refining the scoring function becomes paramount for advancing the precision of Ab-Ag docking predictions.

In the subsequent sections, we delve into the specific results and implications of our approach, shedding light on how this unique evaluation and optimization methodology contributes to the broader understanding of Ab-Ag interactions.

1.4. Challenges in Ab-Ag Modeling

Protein-Protein Interactions (PPI) and Ab-Ag Interactions are fundamental processes that underpin various aspects of cellular function and immune responses. The realm of PPI, proteins engage in intricate associations, contributing to the formation of complexes, cellular signaling cascades, and the regulation of diverse biological pathways. These interactions can be transient or stable, playing essential roles in maintaining structural integrity and facilitating dynamic cellular processes. PPIs occur between different regions of proteins, involving domains, motifs, or active sites, and are subject to tight regulation influenced by cellular conditions, ligand binding, and post-translational modifications.

On the other hand, Ab-Ag Interactions are pivotal components of the immune system's ability to recognize and neutralize foreign entities. Antibodies, produced by B cells, exhibit specificity in binding to antigens, marking them for destruction or neutralization. The interaction between antibodies and antigens is characterized by the complementarity between the antibody's paratope and the antigen's epitope. These interactions can be transient, especially in the initial stages of immune response, and play a crucial role in initiating downstream immune effector mechanisms.

Studies in the field have uncovered intriguing details about the Ab-Ag interfaces. Notably, the Complementary-determined region (CDR) loops of antibodies, responsible for direct antigen interaction, exhibit peculiarities. Analyses have revealed an enrichment of aromatic and hydrophilic residues within these CDR loops [1], [2]. These peculiarities, especially the prevalence of aromatic residues like tyrosine (Tyr), have not been adequately incorporated into existing scoring functions designed for Ab-Ag interactions [3], [16], [17], [18]. This enrichment of specific amino acid residues enhances the binding affinity and specificity of antibodies to antigens, contributing to the effectiveness of the immune response. These structural features in the Ab-Ag interface underscore the intricate nature of immune recognition and highlight the selective pressure that has shaped the evolution of antibody molecules for efficient antigen binding. The distinct characteristics of Ab-Ag interactions further emphasize the specialized role of antibodies in recognizing and combating a diverse array of pathogens and foreign substances.

Another notable challenge in comprehending the performance of a docking scoring function arises from the subjective nature of selecting the decoy distribution's funnel-like shape. Historically, researchers have grappled with the interpretation of decoy distributions, often relying on visual assessments to identify funnel-like patterns. This subjective selection introduces a degree of ambiguity and potential bias in the evaluation process. The inherent complexity of protein-protein interactions, especially in the context of Ab-Ag complexes, necessitates a more objective and quantitative approach to assess the accuracy of docking predictions. Addressing this challenge requires methodologies that not only capture the nuanced features of decoy distributions but also provide a standardized and reproducible means of analyzing the intricate dynamics of molecular interactions.

1.5. Goal of the Study

In this thesis, we employed the versatile Rosetta software, recognized for its capabilities in protein modeling and docking [19]. Our study aimed to evaluate diverse scoring functions' impact on Ab-Ag binding predictions and enhance the accuracy of the RosettaAb docking scoring function. Traditional assessments often involve a subjective selection of the decoy distribution's funnel-like shape, introducing potential bias. To address this, our unique approach utilizes rigorous numerical optimization techniques for a more objective and quantitative evaluation of molecular interactions. The primary focus is on optimizing the RosettaAb docking scoring function, ensuring a refined model for predicting Ab-Ag binding structures. This not only overcomes challenges related to subjective interpretation but also enhances precision and reliability in computational docking scoring functions for Ab-Ag complexes.

Furthermore, our research demonstrates the applicability of numerical optimization to other Ab-Ag databases. Future investigations could delve into understanding the non-critical parameters of the scoring function, thereby refining precision. By addressing these limitations and broadening the scope of applications, we aspire to unlock the full potential of this refined scoring function. Such advancements are crucial for advancing our comprehension of Ab-Ag interactions and their pivotal roles in medical science.

1.6. Thesis outline and main contribution

Chapter 2 delves into the evaluation of decoy shape distribution within the RosettaAb docking scoring function. We assess the function's effectiveness in distinguishing favorable from unfavorable docking poses, aiming to identify critical parameters influencing its ability to locate high-quality docking positions. Our exploration not only unravels the intricacies of the docking scoring function but also sheds light on the concept of docking funnels, contributing to a nuanced understanding of molecular interactions in Ab-Ag complexes.

In the subsequent phase, we thoroughly examined the physical and chemical aspects of the Ab-Ag interface, specifically focusing on the CDR loops and epitopes. This analysis provided key insights into the composition of these regions, enhancing our understanding of Ab-Ag recognition. Within this exploration, we investigated the amino acid composition in the Ab-Ag interface, identifying crucial residues. Additionally, we utilized a Generalized Linear Model to discern significant associations between scoring function performance and physico-chemical properties, contributing to a more nuanced understanding of the intricate dynamics governing Ab-Ag interactions.

Chapter 3 focuses on the third analysis aimed at uncovering the critical factors influencing the accuracy of the RosettaAb docking scoring function for Ab-Ag complexes. The exploration involves assessing the correlation between specific properties and precise decoy distribution for predicting docking orientations. Additionally, the chapter includes supplementary investigations such as evaluating decoy shapes, optimizing parametric weights, and practically assessing assigned weights in the RosettaAb docking process. The study extends to scrutinizing the total energy score, parametric contributions, and optimizing weight parameters to identify universal parameters for the dataset. This comprehensive approach contributes to the ongoing refinement of the RosettaAb docking scoring function for improved predictions in Ab-Ag interactions.

Chapter 4 concludes this work, summarizes the work, and gives suggestions for further research.

Chapter 2

The Interplay Between Scoring Functions and Physico-chemical Properties in Antibody-Antigen Docking

This chapter explores the complex molecular interactions between antibodies and antigens, focusing on the dynamic field of biological medicines, specifically antibody-based therapies. This study investigates the performance of scoring functions and the intricate impact of physico-chemical properties on Ab-Ag interactions. By examining decoy distributions and conducting thorough analyses, our goal is to shed light on the intricate nature of Ab-Ag complexes. In this analysis, we will evaluate three scoring functions and highlight the significant influence of solvation parameters on docking performance. Furthermore, a detailed analysis of the amino acid composition in Ab-Ag interfaces highlights the presence of certain residues, emphasizing their association with high-affinity complexes. The chapter examines the relationship between physico-chemical properties and decoy distribution through a Generalized Linear Model (GLM) analysis. This opening chapter offers valuable insights and acknowledges its limitations. It also emphasizes the need for further research to enhance our understanding of Ab-Ag interactions and improve predictive models in the ever-changing field of drug discovery and design.

2.1. Methods

2.1.1. Input Data Preparation

We obtained the structures of Ab-Ag complexes from the SabDab Database and assembled a dataset comprising 100 such complexes. The structural resolution of these complexes is 5Å or lower, and the associated Ag sequences have a length of 25 amino acids or more (see Appendix B). Information regarding the CDR loops (L1, L2, L3, H1, H2, H3 fragments) was also extracted from the SabDab Database. The dataset exclusively includes antigens that are either peptides or proteins, while the antibody complexes are composed of human or mammalian protein sequences. Notably, our dataset is

characterized by high-quality structures, exhibiting fewer than 10% missing residues and an R-factor of less than 0.4.

2.1.2. Ab-Ag docking and generating decoys

To produce alternative binding poses, we employed RosettaAb Docking, a specialized program for the Ab-Ag interface in computational docking [15]. In the case of 100 selected Ab-Ag bound complexes, we generated 500 decoys for each complex, creating decoy distributions for every docking run. Subsequently, the overall energy score was computed using the Rosetta scoring function.

2.1.3. Energy scoring with customized weights

The Rosetta scoring function *ref2015* is expressed as a linear equation with predetermined weights, based on positive interactions between docking partners. It operates under the assumption that the energetically favored configuration represents the correct biological structure and thus possesses the lowest energy [16]. The Rosetta energy function estimates the conformational energy of biomolecules, denoted as ΔE_{total}, derived through a linear combination of energy terms E_i for the i-th parameter. These terms are functions of geometric degrees of freedom (θ), chemical identities (aa), and weights assigned to each term (w), as illustrated below.

$$\Delta E_{total} = \sum_i w_i E_i(\theta_i, aa_i) \quad (1)$$

The *ref2015* comprises 20 default parameters, representing various free energy components (see Appendix C). Among these, Rosetta incorporates solvation-based parameters to enhance protein modeling. *fa_sol* evaluates solvation energy, *fa_intra_rep* assesses internal clashes, *fa_intra_sol_xover4* combines solvation and repulsion terms, and *lk_ball_wtd* gauges ligand desolvation energy. We formulated three sets of weights based on the Rosetta scoring function: 1) The Rosetta default weights (W_{def}) served as the reference; 2) all parametric weights (w_i) were uniformly set to 1 (W_1); and 3) Rosetta energy scores were calculated by excluding four solvation-based parameters from the default scoring function (W_{nonsol}). Subsequently, using these three functions, the scoring energy for the generated decoys of each Ab-Ag complex was computed.

The assessment of the docking distribution's funnel-like configuration encompassed two criteria: examining the decoy with the smallest Root Mean Square Deviation (RMSD) and Pearson's correlation coefficient (PCC). The smallest RMSD among decoys offers insights into their resemblance to the original complex. A well-structured docking funnel is indicated when the lowest RMSD decoys also exhibit the lowest energy among all alternatives. Conversely, if this criterion is unmet, the funnel is considered inadequately shaped. Concerning the PCC criterion, we categorized the funnel shape as favorable when the value equaled or exceeded 0.4, and unfavorable if the value was less than 0.4. A more detailed assessment of the decoy distribution will be conducted in the following chapter.

2.1.4. Extracting Antibody CDR Loop and Antigen Features

Information on antigens' epitopes was obtained from The Immune Epitope Database (IEDB), a publicly accessible resource funded by NIAID (www.iedb.org) [20]. The database compiles experimental data on antibodies and T cell epitopes researched in various species of infectious diseases, allergy, autoimmunity, and transplantation. To predict antibody epitopes, we utilized BepiPred-2.0 [21], a sequential B-cell epitope predictor employing a Random Forest algorithm trained on crystal structures. Sequential prediction smoothing followed, considering residues with scores above the threshold (default 0.5) as epitope components.

2.1.5. Analyzing Ab-Ag Interface Properties

The distribution of hydrophobic and hydrophilic residues in each CDR loop and epitope was assessed using Kyte and Doolittle's (1982) hydrophobicity scale [22]. To understand polarity in the Ab-Ag interface, Zimmerman, Eliezer, and Simha's (1968) polarity index [23] determined high- and low-polar amino acid distribution in CDR loops and Ag epitopes.

Understanding the surface area of the Ab-Ag binding pocket is crucial. Therefore, we mathematically determined the area of the interacting interface. To understanding the Ab-Ag binding pocket is determining its surface area, we derived the Ab-Ag interface area (A_{Ab-Ag}) by multiplying the surface area accessible to water in the unfolded state of each CDR loop, following Miller et al. (1987) [24], and the fraction of accessible area lost (A_f) [30] during peptide folding.

In this calculation, we assumed the area lost upon folding contributes 100% to Ab-Ag

$$A_{Ab-Ag} = A_s A_f \quad (2)$$

binding. Antibody CDR loops typically feature enriched aromatic residues [25], and we computed their distribution in each Ab-Ag complex (Appendix D).

2.1.6. Statistical Analysis

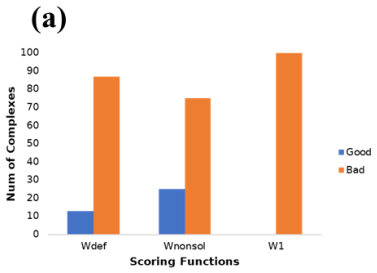
The Generalized Linear Model (GLM) is a statistical tool known for its adaptability in analyzing non-normally distributed data, allowing the modeling of response variables with varied distributions and relationships. We employed the GLM function from the statsmodels.api Python library to identify explanatory variables strongly correlated with the response variable—the funnelness of decoy distributions, represented by PCC [26] of the W_{def} and W_{nonsol} scoring functions (W_1 was excluded as all funnels were deemed inadequate).

The input dataset encompasses the percentage occurrence of diverse explanatory variables, including hydrophilicity, polarity, the presence of aromatic residues, and the surface area of each CDR loop and antigen epitope.

2.2. Results

2.2.1. Docking scoring functions and docking funnels

The performance of a scoring function on an Ab-Ag complex can be explained in terms of the distribution of the docking decoys. We evaluated three different scoring functions: a function with the Rosetta default weights (W_{def}), a function with all changeable weights set to 1 (W_1), and a function with Rosetta default weights except for the solvation-based weights set to 0 (W_{nonsol}). In molecular docking simulations of Ab-Ag binding, a decoy distribution illustrates the range of potential binding configurations. For every Ab-Ag complex in our dataset, we generated decoy distributions by mapping the RMSD value on the x-axis and the total energy score on the y-axis and for each scoring function, the respective docking distributions were created by plotting the generated decoys. As depicted in Figure 1, certain decoy distributions clearly display a funnel-like structure with low decoy energies, whereas others lack any unambiguous structure. When studying



(b)

W_{def}		W_1	
		G	B
	G	0	13
	B	0	87

(c)

W_{def}		W_{nonsol}	
		G	B
	G	8	5
	B	16	71

Figure 2: Quantitative Representation of Decoy Distribution Goodness. a) Distribution of good (blue) and bad (orange) funnels in all three scoring functions. b) and c) Counts of funnel-like shape improvement of two example structures respectively.

the accuracy of the decoy distribution we identified four characteristics that distinguish a good decoy distribution from a poor one.

Theoretically, a good distribution consists of decoys displaying both the lowest RMSD and the lowest energy scores. Figures 1a and 1b represent two distinctive representations of relatively bad decoy distribution which lack a funnel-like decoy distribution shape. Figure 1c is an example of a good distribution. The decoy distribution in Figure 1c is an excellent example of a "funnel" with a distinct form and low energy and RMSD decoys at its tip.

Namely, W_{def}, W_1, and W_{nonsol}, we counted the number of good and bad decoy distributions docking funnels (Figure 2b & 2c). We revealed the impact of different scoring functions on the decoy distributions for Ab-Ag complexes. Specifically, we observed that the W_1 scoring function yielded 100 percent poor decoy distributions, whereas the W_{nonsol} scoring function achieved a success rate of 25 percent (Appendix E).

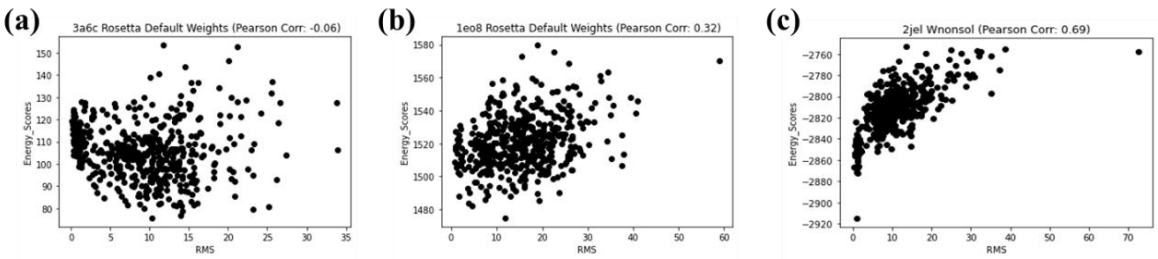


Figure 1: Types of Decoy Distributions. a) and b) examples of bad decoy distributions. c) An example of good funnel-like decoy distribution in scoring function.

To assess the relationship between accurate decoy distribution (a funnel-like distribution) and the scoring function, we conducted a chi-square test of independence. The obtained chi-square value of 11.544 exceeded the critical value of 3.841, rejecting the null hypothesis that the decoy distribution's accuracy is unrelated to the exclusion of solvation parameters from the default Rosetta scoring function (p<0.05). Excluding solvation parameters from the scoring function significantly influences the outcome, resulting in a more funnel-like decoy distribution. The analysis of three scoring functions highlights the significance of considering solvation effects to enhance protein-protein docking accuracy.

2.2.2. Amino acid composition in the Ab-Ag interface

The amino acid properties of the 100 Ab-Ag complexes in our dataset were analyzed, and the findings are depicted in Figure 3. A comparison between the amino acid frequencies observed in our study and those in the genome revealed three notable discoveries.

Firstly, serine (Ser) emerged as the most prevalent amino acid in antibodies, particularly in the light chains CDR loops, constituting 21.8% of the total composition. Interestingly, serine was completely absent in the heavy chains of antibodies. It was consistently present in the light chains of all 100 complexes but not observed in the heavy chains.

Secondly, despite being generally infrequent in eukaryotes, our study demonstrated a strong presence of tyrosine (Tyr) in antibody-heavy chain CDR loops, accounting for 15.6% of the total composition. Tyr residues were found in the light chain fragments of 97 complexes, the heavy chain fragments of 97 complexes, and 84 epitopes in our dataset. Notably, Tyr residues were more abundant in the heavy chain CDR loops compared to the epitope regions and the light chain.

Lastly, aromatic amino acids such as phenylalanine (Phe) and threonine (Thr) exhibited higher frequencies in the heavy chain CDR loops compared to their relatively low occurrence in eukaryotic genomes.

2.2.3. Significance of physico-chemical properties

To assess the impact of different explanatory variables on the accuracy of the decoy distribution, logistic regression with a generalized linear model (GLM) was performed on both the W_{def}, and W_{nonsol}scoring functions. Figure 3 illustrates the results, indicating that three variables demonstrated significant associations ($p<0.05$) with the predictive power of decoy distribution. These variables include the presence of non-aromatic residues in the L3 and H1 regions (0.016 and 0.044, respectively) in the W_{nonsol} scoring function, as well as the surface area of the Ab-Ag binding interface in the L2 region (0.047) in the W_{def} scoring function [Appendix F].

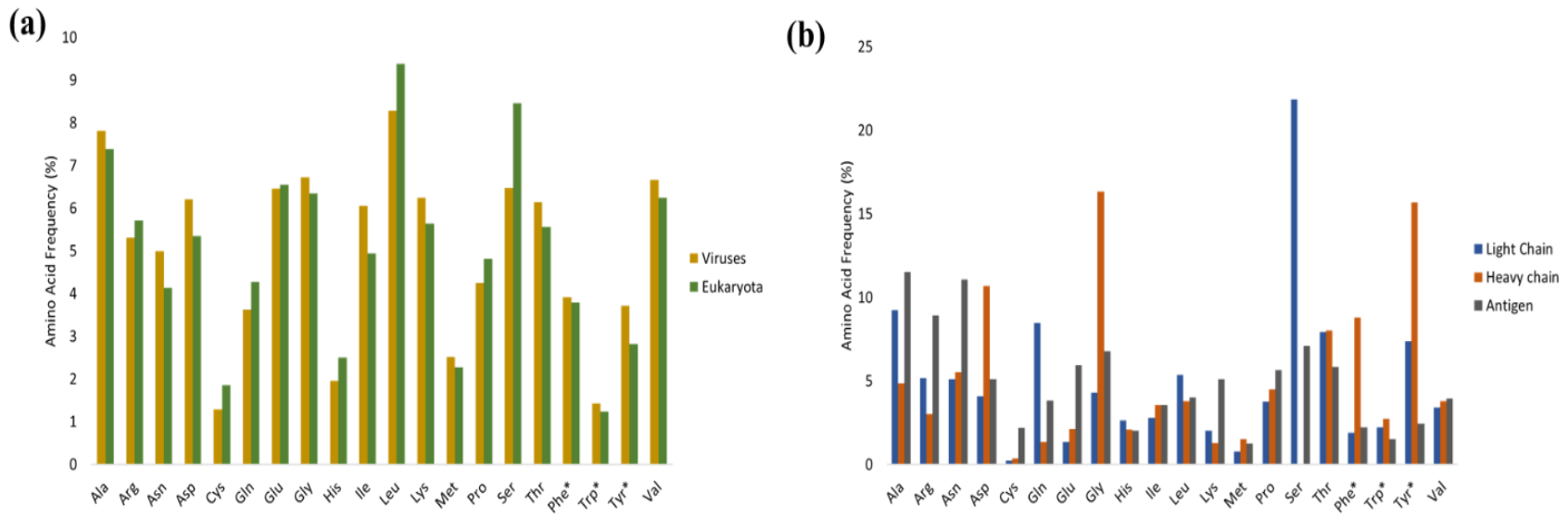


Figure 3: Amino Acids Frequency comparison. a) Amino acid frequency encoded in the genome of eukaryotes and viruses [41]. Yellow: Viruses and green: Eukaryotes. b) Amino acid frequencies of antibody CDR loops and Antigen epitope. Blue, orange, and gray bars reflect the frequency of residues on CDR loops (heavy chain, and light chain) and Ag, respectively. *Aromatic amino acids.

2.3. Discussion

Ab-Ag interactions play a crucial role in various biological processes, including immune response modulation, disease diagnosis, and therapeutic interventions. Understanding the physico-chemical properties underlying these interactions is essential for evaluating their specificity, affinity, and overall stability. This study aims to delve deeply into the exploration of these physico-chemical properties involved in Ab-Ag interactions, shedding light on their significance in the realm of drug discovery and design.

2.3.1. Evaluation of docking scoring functions

This research represents an initial effort to assess the efficacy of Rosetta scoring functions in handling Ab-Ag complexes, employing both quantitative and qualitative measures through the practical use of decoy distributions, a common practice among scientists. The performance of scoring functions in Ab-Ag complex docking was

Parameter		Wdef		Wnonsol	
		Coefficient	P	Coefficient	P
	constant	3.832	0.076	-0.983	0.745
Hydrophilicity	L1	-0.003	0.609	-0.010	0.266
	L2	-0.004	0.307	0.005	0.356
	L3	-0.005	0.210	-0.008	0.093
	H1	0.004	0.478	0.003	0.670
	H2	0.000	1.000	0.006	0.250
	H3	0.000	0.904	-0.001	0.793
	Ag	-0.014	0.086	0.002	0.852
Polarity	L1	-0.002	0.732	0.001	0.916
	L2	-0.006	0.114	0.008	0.104
	L3	0.007	0.076	-0.002	0.692
	H1	-0.001	0.732	-0.006	0.272
	H2	-0.004	0.228	0.001	0.804
	H3	0.002	0.620	0.001	0.766
	Ag	-0.003	0.666	-0.011	0.287
Non-aromatic residues	L1	-0.016	0.259	-0.006	0.756
	L2	-0.011	0.128	0.003	0.777
	L3	0.005	0.218	0.015	0.016*
	H1	-0.003	0.578	-0.013	0.044*
	H2	0.002	0.751	0.007	0.300
	H3	0.006	0.106	0.003	0.545
	Ag	0.012	0.273	0.010	0.514
Ab-Ag Surface Area	L1	0.000	0.719	0.000	0.635
	L2	-0.001	0.047*	0.000	0.954
	L3	0.000	0.456	0.000	0.752
	H1	0.000	0.361	0.000	0.816
	H2	0.001	0.116	0.001	0.098
	H3	0.000	0.896	0.000	0.756
	Ag	0.000	0.827	0.000	0.465

* Significant

Figure 4: Significance of independent parameters in Generalized Linear Regression (GLM). $1.0 < (P > |z|) < 0.05$ Green, and yellow colors reflect the positive and negative correlation of significant parameters respectively.

evaluated based on the distribution of docking decoys. The analysis of three scoring functions (W_{def}, W_1, and W_{nonsol}) revealed that the removal of solvation parameters had a significant impact on the quality of docking performance, highlighting the need to account for solvation effects to improve the accuracy of Ab-Ag docking predictions.

While the results indicate that excluding solvation parameters affects the quality of docking performance, they do not provide a comprehensive understanding of all factors that influence the precision of docking scoring functions. Other factors influencing docking accuracy, such as hidden scoring terms, energy functions, and additional factors, were not directly addressed in this study. There is a need for additional research and investigation into the broader landscape of docking scoring functions and their overall performance in various docking scenarios, but it is beyond the goal of this research.

2.3.2. Influence of physico-chemical properties in CDR loops and epitopes on docking

In analyzing 100 datasets comprising Ab-Ag bound complexes, we identified a conspicuous predominance of Tyr residues in the heavy chain CDR loops of antibodies, particularly enriched in the paratope regions. Conversely, Ser residues were completely absent from these heavy chain CDR loops, and cystine (Cys) residues were notably sparse in both heavy and light chain CDR loops.

Previous studies have underscored the integral roles of Tyr, Ser, and Cys in antibodies [27], [28], [29], [30], [31]; however, they have predominantly done so within the confines of controlled environments, involving mutations in the epitope interacting regions of nonhuman immunoglobulins [27], [28], [29], and have not provided CDR location-specific insights into amino acid composition. Furthermore, these investigations were performed on unbound immunoglobulins, leaving the stability of Ab-Ag interactions uncertain and suggesting that the molecular recognition mechanisms of naïve or unbound antibodies differ from those of co-evolved or affinity-matured complexes [32], [33], [34], [35], [36], [37], [38].

Our research conclusively demonstrates the significant correlation of Tyr, Ser, and Cys with high-affinity, specific in-situ Ab-Ag complexes. The distinct positioning of these amino acids within the CDR loops of engaged Ab-Ag pairs is pivotal, highlighting an

area that requires meticulous attention in the refinement of docking scoring algorithms for enhanced precision.

While the analysis provides valuable insights into the distribution and prevalence of specific amino acids within ab-ag complexes, there are limitations to consider. The results are confined to the specific dataset and conditions of the study and may not fully represent the entire population of ab-ag complexes or account for potential variations due to experimental setups, antibody types, antigen characteristics, or other factors. The observed variations in amino acid composition do not necessarily provide mechanistic insights into the functional implications or reasons for their presence or absence in CDR loops. Therefore, further research is necessary to explore the broader context, underlying mechanisms, functional implications, and generalizability of these findings.

2.3.3. Correlation between physico-chemical properties and decoy distribution

Through GLM analysis, we identified three critical variables impacting the predictive accuracy of decoy distribution within CDR segments: 1) the prevalence of non-aromatic residues in L3; 2) the scarcity of non-aromatic residues in H1; and 3) a smaller Ab-Ag binding interface in the L2 region. These findings highlight how the scoring function's accuracy fluctuates based on the abundance of non-aromatic residues and the interaction's surface area. Notably, the influence of aromatic residues persists even in scoring functions devoid of solvation parameters. While omitting solvation parameters can enhance Rosetta's performance with Ab-Ag structures, this alone doesn't fully negate the effects of irregular amino acids. Therefore, the development of scoring functions specialized for Ab-Ag interfaces is essential for more precise future Ab-Ag structural predictions.

The study offers important insights but is limited by its specific dataset and conditions, not fully encompassing the diversity of Ab-Ag complexes. A comprehensive understanding of the mechanisms dictating amino acid presence in CDR loops is missing, underscoring the need for broader research. This should include how amino acids influence the Rosetta scoring function, deeper exploration of Ab-Ag interactions, and enhancements in bio medicinal prediction models for wider applicability.

2.4. Conclusion

In this chapter we have found that omitting a specific parameter in Rosetta scoring function enhances performance, but we must carefully consider other factors for accurate predictions.

A deeper exploration of physicochemical characteristics, including hydrophilicity, hydrophobicity, polar versus nonpolar interactions, and specific amino acid residues, provides valuable insights into the stability, specificity, and efficacy of Ab-Ag complexes. The synthesis of these methodologies enhances the sophistication of computational techniques, advancing our ability to forecast and improve Ab-Ag interactions—a crucial aspect in drug innovation and formulation. These insights hold substantial promise for driving advancements in drug development, potentially leading to more effective therapeutic strategies and improved patient prognoses.

Exploring the chemical attributes of potential substitutes sheds light on the stability and effectiveness of these interactions. This knowledge improves our ability to forecast and enhance antibody interactions, crucial for drug development. While promising, further research is needed to fully grasp these interactions and improve therapeutic antibody production for better, targeted treatments.

Chapter 3

Analysis in Docking Scoring Functions for Precision Docking

In this chapter, we explore the details of the Rosetta scoring function and its influence on the prediction of Ab-Ag binding structures. The scoring function involves 20 weighted parameters, each with a distinct role in determining the overall energy score. A thorough analysis is conducted to understand the importance of these parameters. This involves carefully examining the weights and comparing the energy scores with and without specific parameters. An important aspect of this investigation involves categorizing decoy distributions and determining their quality based on Pearson's correlation coefficient and RMSD values. Next, three optimization models are presented to improve the shape of decoy distributions, highlighting the significance of specific parametric weights. The performance of these models is thoroughly evaluated, using both qualitative and quantitative measures. This study aims to identify and optimize key parameters that influence the Rosetta scoring function, with the goal of improving Ab-Ag docking predictions. This chapter presents a detailed analysis of the relationship between parameter weights, decoy distributions, and the accuracy of computational docking models. By doing so, it establishes a foundation for future advancements in therapeutic interventions and structural biology.

3.1. Methods

3.1.1. Identification of high predictive power parameters in Rosetta scoring function

The Rosetta scoring function incorporates 20 weighted parameters, as detailed in Appendix C. To assess the individual impact of each parameter on the ultimate energy score, a comprehensive analysis is conducted. Initially, all parameter weights are normalized to a value of one, and the total energy score is calculated for each structural configuration. Through the comparison of the overall energy score achieved with all parameters to that attained when a particular parameter is excluded, the importance of each variable is determined. This methodology is systematically applied to all 100

complexes stored in the database, resulting in the calculation of the average proportional contribution of each parameter.

3.1.2. Classification of the decoy distribution

An example of good decoy distribution is shown in Figure 5. We defined two qualitative criteria to define a good or bad decoy distribution; 1) a Pearson's correlation coefficient (P) [26] between RMSD and energy scores must be positive, and 2) The RMSD values of the lowest energy decoy should be at least 5Å smaller than the mean RMSD value of the distribution. This is an indication of the left skewness in the distribution.

If both conditions were met, the decoy distribution was classified as "Good" (Funnel-like). If either condition was unsatisfied, the decoy distribution was classified as "Bad" (not Funnel-like).

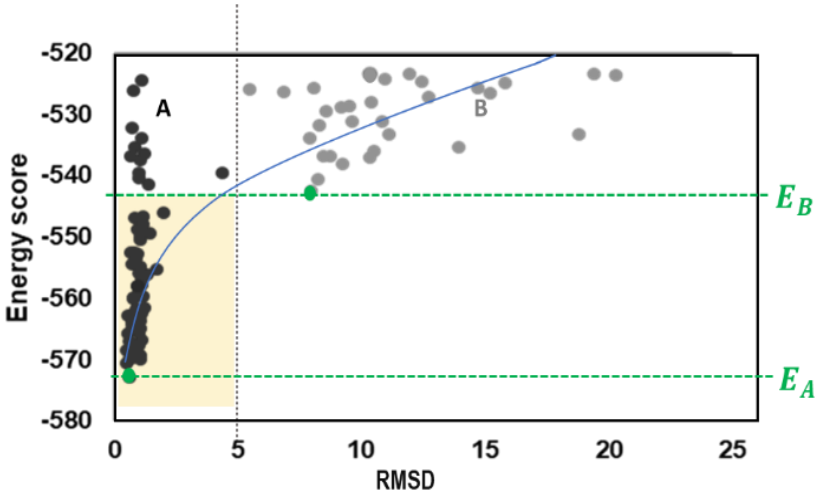


Figure 5: Example of a decoy distribution and definition of parameters in Optimization Model 3. The decoy distribution was separated to two clusters A (decoys with RMSD < 5Å) and B (decoys with RMSD > 5Å). 5Å is considered as the threshold for near-nativeness of a decoy. Green line vertical line represents the 5Å cluster threshold line. E_A: Decoy with the lowest energy score of the Cluster A. E_B: Decoy with the lowest energy score of the cluster B. Green horizontal line is to separate the decoys with low energy score than E_B, and the number of decoys in the yellow area is counted and accounted for in the M3 as denoted by D.

3.1.3. Scoring function optimization models

Three selection scores were developed to assess the quality of decoy distribution shape. These scores were then used to determine the optimal weights for each parameter in the dataset. Models were created to focus on the two key characteristics of a funnel-like decoy distribution.

The parameter space was explored using Nelder-Mead optimization [39] to find local optima of weights for all three models. The process will iterate through weight values from 0 to 1.05 with an increment of 0.01.

Optimization Model 1 (M1): This model is based on the first characteristic of a good decoy distribution, which is the positive Pearson's correlation coefficient between the total energy score and RMSD of decoys. Parameters were optimized to maximize the selection score (S_1), which equals to Pearson's correlation coefficient (P)

$$S_1 = P \tag{1}$$

Optimization Model 2 (M2): This optimization model is based on the fact that the decoy with the lowest RMSD should also have the lowest energy score among all the decoys. The selection score (S_2) for a given Ab-Ag complex is determined by subtracting the total energy score of the decoy with the lowest energy (E_L) from the total energy score of the decoy with the lowest RMSD value (E_R). S would be 0 if a decoy with the lowest energy also has the lowest RSMD.

$$S_2 = E_L - E_R \tag{2}$$

Weight sets with smaller S_2 values are considered optimal.

Optimization Model 3 (M3): In this model we execute a combined model of M1 and M3 with additional features of a decoy distribution (Figure 5).

The decoy distribution was divided into two clusters: A (decoys with RMSD < 5Å) and B (decoys with RMSD > 5Å), with the 5Å threshold indicating near-nativeness. The green vertical line represents the 5Å cluster threshold.

The selection score (S_3) is determined for a given Ab-Ag complex.

$$S_3 = E + D + P \tag{3}$$

where,

$$E = E_A - E_B \tag{4}$$

where, E_A denotes the decoy with the lowest energy score in cluster A, and E_B represents the lowest energy score decoy in cluster B. *D* is the count of decoys (yellow highlighted area in Figure 1) in cluster A with lower energy than the lowest energy decoy in cluster B.

The statistics *E*, *D*, and *P* were individually normalized to their respective *Z*-scores. Normalizing the contribution of each parameter ensured that they all had an equal impact on the overall score. In this model, weight sets with the highest S_3 value is considered optimal.

3.1.4. Qualitative and quantitative analysis of the optimization model performance

To re-estimate the Rosetta total energy score with the identified weights, we applied the determined optimal weight parameters for the RosettaAb docking scoring function, and re-regenerated 500 decoys for each Ab-Ag complex in the dataset. We constructed the respective distributions for these decoys. We re-categorized the decoy distributions as “Good” and “Bad” based on the same classification method explained in method 3.1.3.

Net Reclassification Improvement (NRI_{net}) [40] gauges the improvement in classifying decoy distributions as "Good" or "Bad" when using optimized weight parameters compared to default parameters in Rosetta Performance.

$$NRI_{net} = NRI_{good} - NRI_{bad} \tag{5}$$

$$NRI_{net} = \frac{(a-b)}{(a+b)} - \frac{(c-d)}{(c+d)} \tag{6}$$

where, NRI_{good} as represents in Table 1, the difference between the probability of "Good" decoy distribution predicted by the optimization model and the probability predicted by Rosetta, and NRI_{bad} represents the difference between the probability of "Bad" decoy distributions predicted by the Rosetta and the probability predicted by the optimization model for individuals without events. It quantifies the net enhancement in classification accuracy provided by the optimization models.

In order to obtain a the parameter set that can be applied to majority of the Ab-Ag complexes in our dataset, Principal Component Analysis (PCA) [41] and the k-means clustering algorithm [42] were applied to the set of parameters estimated in the 100 Ab-Ag complexes.

3.2. Results

3.2.1. Total energy score and the parametric contribution

We first identified weight parameters with high contribution to output in the Rossetta scoring function. Parameters contributing more than 5% to the total energy score were selected (Figure 6). The six parameters are: Lennard-Jones attractive between atoms in different residues (*fa_atr*), Lennard-Jones repulsive between at oms in different residues (*fa_rep*), Lazaridis-Karplus solvation energy (*fa_sol*), Intra-residue Lazaridis-Karplus solvation energy (*fa_intra_rep*), Coulombic electrostatic potential with a distance-dependent dielectric (*fa_elec*), and Probability of amino acid, given torsion values for phi and psi (*fa_dun*). They respectively have Rosetta *ref2015* model predefined weights of 1.0, 0.55, 0.9375, 0.005, 0.875, and 0.7 [12] (Appendix C).

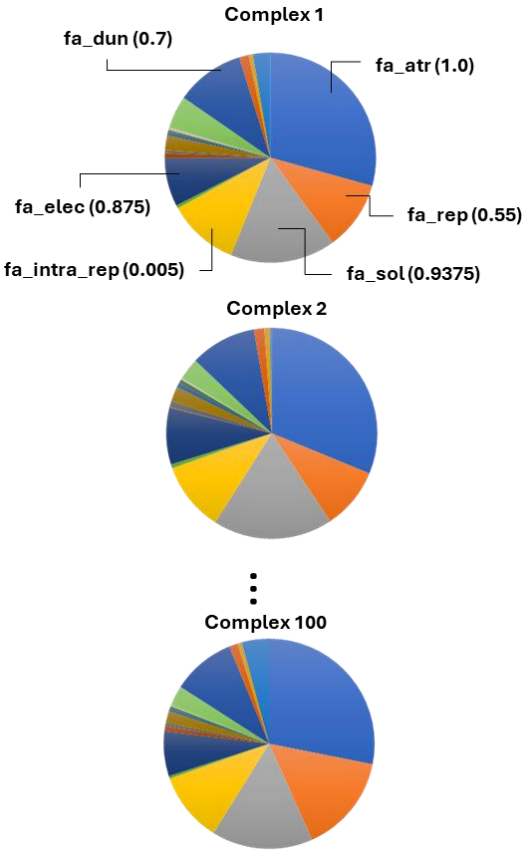


Figure 6: Parametric Contributions to the Total Energy Score in Rosetta Scoring. We identified six crucial parameters, including Lennard-Jones Attractive Forces (*fa_atr*), Dihedral Angle Energy (*fa_dun*), Electrostatic Energy (*fa_elec*), Lennard-Jones Repulsion (*fa_rep*), Intra-residue Solvation Energy (*fa_intra_rep*), and Solvation Energy (*fa_sol*), with specific predefined weights of Rosetta *ref2015* as specified within parentheses, that greatly impact the total energy score.

3.2.2. Optimizing weight parameters by adjusting decoy distributions

We assumed that an ideal scoring function with optimal weight parameters can produce a good pattern of decoy distributions. We classified a good or bad decoy distribution using conditional judgment of the following two criteria. 1) Pearson's Correlation Coefficient of the distribution must be positive, 2) RMSD values were considered relative by subtracting the lowest energy decoy's RMSD from the mean RMSD, with a ≤5Å difference criterion.

Figure 7 (a-d) shows graphs depicting different patterns of decoy distributions. During the initial decoy generation with Rosetta default weight parameters, we obtained 37 good

(funnel-like) decoy distributions and 63 bad (non-funnel like) distributions among 100 Ab-Ag complexes. We calculated the frequency of decoy distribution goodness change compared to the Rosetta initial decoy distributions. As presented in Figure 8 a-b, the frequencies were calculated for both post optimization and after validation.

We optimized the six key weight parameters discovered in Result section 3.2.1, using the three different optimization models (M1, M2, and M3), each capturing different aspects of the goodness of decoy distribution, using a numerical optimization method (see methods). Note that, in this analysis, we keep using the same predicted structures as the initial prediction, which means the value of RMSD for each decoy is invariant while the total energy score changes depending on parameters. A smaller energy score indicates a stronger binding affinity.

Using default parameters in Rosetta, we found that 37 complexes had good decoy distributions, while 63 complexes had bad decoy distributions. We examined a total of one hundred complexes using RosettaAb Docking after assigning specific weights to them. 55 complexes in Model M1 had favorable decoy distributions (Appendix G). Model M2 generated 40 complexes with good decoy distributions, while Model M3 produced 50 complexes with favorable decoy distributions.

While the optimized parameters did not uniformly enhance the shape of decoy distributions in certain structures, the three methods, namely M1, M2, and M3, generally improved the overall shape of distributions across the dataset. To evaluate the improvement, we calculated NRI (Table 1) between default parameters and each of three models. All three models demonstrated a substantial improvement with a positive NRI value of 0.941.

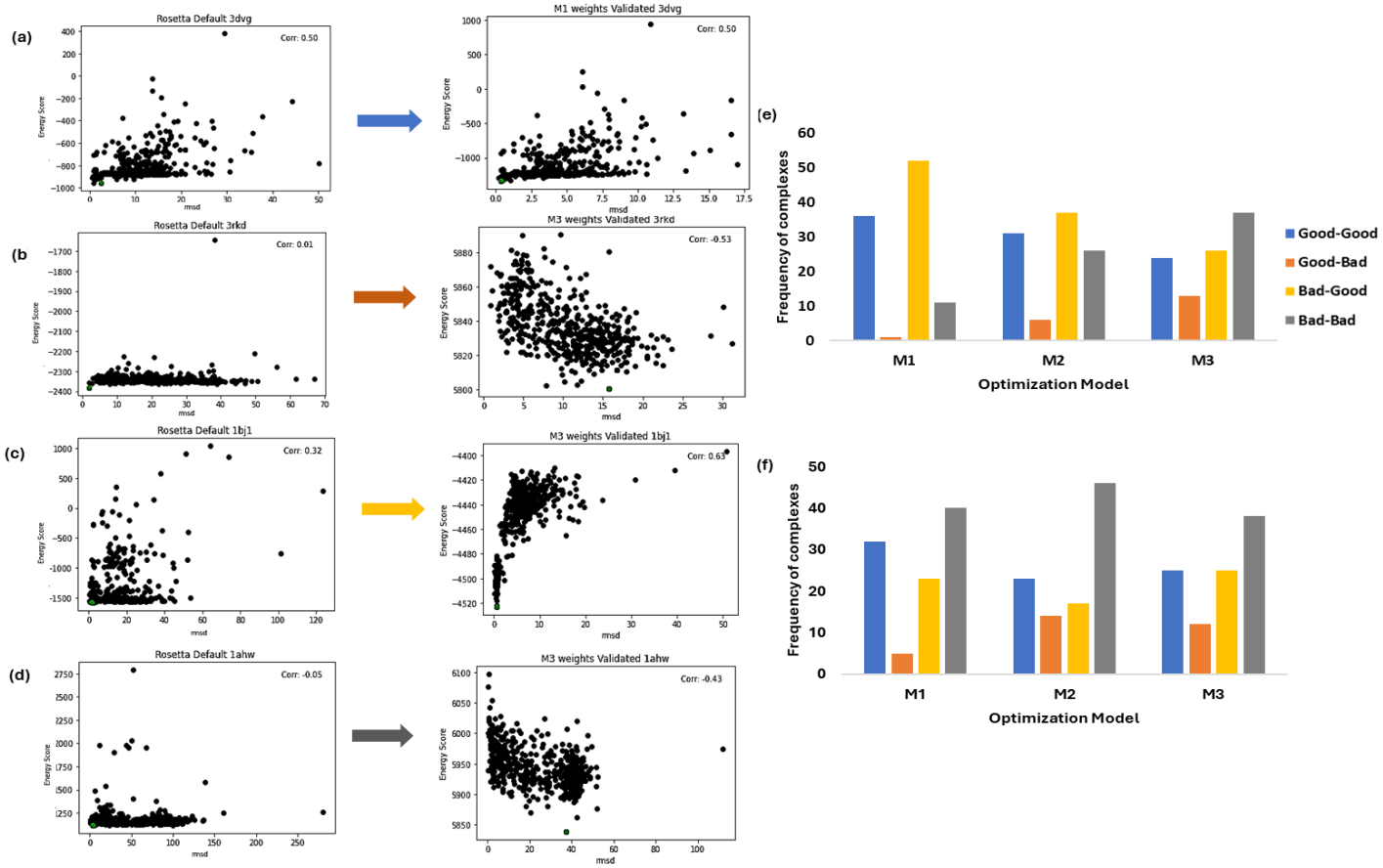


Figure 7: Optimization Model Performance Comparison. On decoy distributions, x-axis represents decoy RMSD values and y-axis represents decoy energy scores. The decoy distributions were classified as, Good: Funnel-like and Bad: not funnel like distributions. a-d decoy distributions represent examples of Rosetta Default decoy distributions and post validation decoy distributions. a) Good – Good; b) Good – Bad; c) Bad – Good; d) Bad – Bad. e) Frequency of Ab-Ag complexes in each category respective to each selection optimization model after optimization; f) Frequency of complexes in each category after validation.

Table 1: Net Reclassification Improvement. Post-optimization analyses for the three models (M1, M2, and M3) demonstrated considerable enhancements in predicting the goodness of decoy distributions. The positive Net Reclassification Improvement (NRI) values underscored the effectiveness of the optimization process.

Model	Good-Good (a)	Good-Bad (b)	Bad-Good (c)	Bad-Bad (d)	NRI
M1	32	23	5	38	0.941
M2	23	14	14	46	0.683
M3	25	25	12	40	0.52
Post Validation	33	26	4	37	0.924

3.2.3. Performance of optimization models

We subsequently re-generated decoys using the optimized parameters for each Ab-Ag complex. We assumed that, if the optimized parameters improve the performance of RossettaAb, the shape of decoy distributions would become better and generated decoys should be close to the real structure hence showing overall small RMSD. Although we used the former criteria for optimization and evaluation in the previous analyses, the latter results directly show the improvement of RossettaAb docking performance.

Figure 7 e and f display the performance of optimization for M1, M2, and M3 decoys. The bars indicate the change in the decoy distribution after optimization, showing whether it improved (Bad to Good), deteriorate (Good to Bad), or unchanged (Good to Good and Bad to Bad). The numerical optimization resulted in improvements in "Good-Good" distributions: 32 in M1, 23 in M2, and 25 in M3. There were 40 "Bad-Bad" distributions in M1, 46 in M2, and 38 in M3.

Figure 7.f summarizes the frequency distributions post-validation. M1 had the highest number of "Good-Good" distributions with 36, followed by M2 with 31, and M3 with 24. M1 had 1 distribution, M2 had 6 distributions, and M3 had 13 distributions for "Good-Bad." The "Bad-Good" had 52 in M1, 37 in M2, and 26 in M3. In the song "Bad-Bad," M1, M2, and M3 had 11, 26, and 37 distributions, respectively.

We compared the mean RMSD values of the lowest energy decoys between the default and optimized models, and the mean values decreased in the 87%, 79%, and 69% of cases

with models M1, M2, and M3, respectively, indicating significant improvements in all three optimization models. The level of improvement was also assessed by conducting a paired t-test [15], where the mean RMSD value for each structure of the default model and the optimized model were paired. The highest t statistics and the lowest p-values were observed for M1 (t = 10.79, p = 2.06×10−18), followed by M3 (t = 9.92, p = 1.60×10−16), and M2 (t = 7.85, p = 5.00×10−12).

3.2.4. Efficacy of distinctive Parameters in Ab-Ag docking predictions

We finally sought the possibility of finding a distinctive parameter set that potentially improves the prediction accuracy of Ab-Ag complex structures using RossettaAb. To this end, we first visualized the pattern of parameter sets that were individually optimized for 100 Ab-Ag complexes and performed clustering. Figure 8 illustrates the results of PCA with k-means clustering, yielding six distinct clusters. The six clusters consist of 2 outlier clusters (marked in red) and 4 closely related clusters that could be merged into one large cluster (marked in blue circle). We therefore expected the parameter sets shared with the large cluster would perform generally well for most of our dataset. The mean of the parametric weight parameters of the remaining complexes was calculated; the mean parameters for the *fa_atr*, *fa_dun*, *fa_elec*, *fa_rep*, *fa_intra_rep*, and *fa_sol* is 0.997, 0.331,

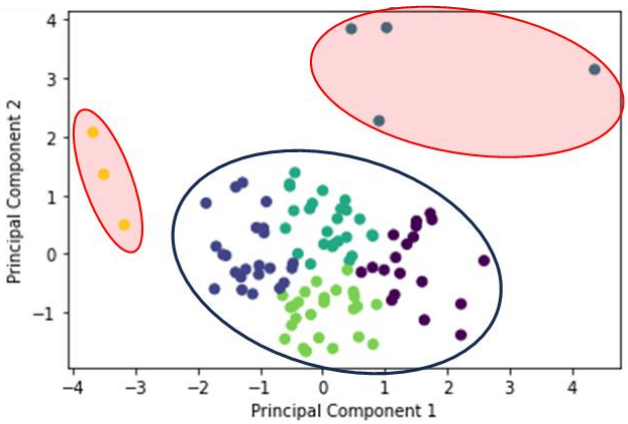


Figure 8: Principal Component Analysis. 6 clusters were identified using K-means clustering. 2 clusters (marked in red) were considered as outliers. These datapoints (complexes) were excluded from the database during the calculation of mean values of parametric weights. Four remaining clusters seems closely related and together they create one big cluster (blue circle).

0.428, 0.831, 0.55, and 0.098, respectively. We employed Rosetta to reevaluate the overall performance of our 100 Ab-Ag complex dataset by applying mean weights.

Applying the mean weight parameters among the large cluster, out of the 100 complexes analyzed, 91 complexes exhibited improvement as indicated by a smaller mean RMSD compared to the Rosetta default. On the other hand, 6 complexes showed an increase in mean RMSD, while 3 complexes did not show any change in RMSD when compared to the Rosetta default. The level of improvement was also assessed using a paired t-test. The results showed a *t*-statistic of 8.924 and a *p*-value of 2.43×10-14. The NRI value between the Rosetta default and the distinctive weights was 0.924.

The identified through PCA (Figure 8) were applied to a randomly selected independent dataset of 10 Ab-Ag complexes (Appendix I) from the SabDab database using Rosetta. Out of the 10 complexes analyzed, default decoy distributions from Rosetta were unfavorable in 8 cases, showing only 2 instances with funnel-like patterns. Upon further assessment, it was found that 2 complexes displayed favorable distributions, while 4 consistently exhibited unfavorable patterns using both Rosetta Default and different parametric weights. Four complexes that had unfavorable distributions when characterized with Rosetta default weights showed improvement when using the new weights, resulting in favorable distributions. Mean RMSD values consistently decreased for all 10 complexes. Parametric weights can have different effects on energy scoring for datasets. However, numerical optimization models can help identify more appropriate weights for Ab-Ag datasets.

3.3. Discussion

Computational docking scoring functions are essential in modeling the interaction Ab-Ag. The accuracy of these models relies on specific weight parameters within the scoring function. It's important to grasp the significance of these parameters in achieving accurate predictions for Ab-Ag binding, which, in turn, impacts binding affinity, specificity, and stability. This study delves into an in-depth exploration of these critical parameters, shedding light on their importance in the field of Ab-Ag modeling and design.

3.3.1. Influence of identified parameters on Rosetta docking scoring function

In analyzing the Rosetta energy scores for 100 datasets, we identified a dominance of six parameters in the Rosetta scoring function that particularly contributed to the energy score. The set includes parameters for *fa_atr*, *fa_dun*, *fa_elec*, *fa_rep*, *fa_intra_rep*, and *fa_sol*. Previous studies have underscored the importance of identifying the influential variables and adjusting their weights is crucial for accurately predicting binding affinity, stability, and other important attributes in this system [19], [43], [44], [45], [46], [47]. We have already shown that there was a significant improvement in scoring energy in the absence of the solvation parameter, *fa_sol* in the previous chapter. The absence or presence of these parameters can affect the overall binding affinity during the docking process.

This chapter highlights the importance of the six found criteria but also acknowledges the possibility of additional factors that may affect the accuracy of the scoring function, which have not yet been explored. Specifically, the complexity associated with solvation parameters necessitates a thorough investigation to fully comprehend the holistic nature of docking scoring functions and to determine their adaptability in various situations. The extraction of valuable knowledge from vast research is crucial for advancing drug development efforts and enhancing our understanding of biomolecular interactions, particularly within the field of molecular medicine.

3.3.2. Optimization impact on lowest energy decoys

Mathematical optimization is a powerful tool that enhances decision-making in a data-driven, objective manner while reducing bias and subjectivity. It ensures consistency across analyses by consistently applying the same algorithm to diverse datasets. This approach improves reliability by promoting objectivity, efficiency, and robustness.

In our analysis, we employed numerical optimization for several reasons. It offers significant advantages when fine-tuning six weights using three different statistics (M1, M2, and M3) to assess various aspects of decoy distribution quality. This process occurs in a complex parameter space, where manual tuning becomes impractical due to the multitude of possible combinations.

Applying the identified distinct parametric weights to any dataset may yield variable effects on energy scoring. Nonetheless, the numerical optimization models introduced in this study offer a means to identify more suitable parametric weights for any Ab-Ag dataset. Numerical optimization is indispensable for determining the optimal weight configuration, whether for minimizing or maximizing an objective function.

Despite the advantages of using robust numerical optimization methods shown in this study, it has limitations in scope and assumptions. The analysis is narrow, overlooking broader optimization challenges and assuming resistance to data noise and uncertainties. Future work should validate these results across various settings, explore broader applications, and consider the ethical implications of these methods.

3.3.3. Evaluation of optimization models

This study aims to find an optimization model for determining the best parametric weights of the Rosetta scoring function in analyzing Ab-Ag complexes. It uses qualitative and quantitative methods, specifically by analyzing decoy distributions. Optimization model performance was evaluated using docking decoy distribution. The analysis of models M1, M2, and M3 showed that all three models can enhance docking accuracy. This emphasizes the importance of identifying specific distinct parametric weights for the dataset being studied, to improve the accuracy of Ab-Ag docking prediction.

The Paired *t*-test revealed that the M1 model outperformed M2 and M3. The M1 model's unique parametric weights improved the docking simulation's performance by 4%. The results suggest that the simple PCC can be used to assess the accuracy of a scoring function and determine the optimal parametric weights for a specific Ab-Ag complex using only six selected parameters from the Rosetta scoring function. However, it does not provide a complete understanding of all the other parameters and their reliability.

The study provides valuable insights but is constrained by its narrow dataset and conditions, which do not fully represent the range of Ab-Ag complexes. Further research is needed to fully understand the relationship between the parameters of the scoring function and their reliability. This study examines the impact of various parameters on

the Rosetta scoring function, investigates the influence of Ab-Ag interaction on decoy distributions, and proposes improvements to docking scoring models for broader use.

3.4. Conclusion

In conclusion, a deep understanding of the pivotal variables shaping computational docking scoring functions is imperative. The enhancement achieved through the fine-tuning of six key parameters within the Rosetta docking scoring function underscores their critical role in influencing the stability and energetics of Antibody-Antigen (Ab-Ag) interactions. This study illuminates the potential for optimizing parameters, including binding affinity and stability, to revolutionize therapeutic interventions in conditions like cancer and autoimmune diseases.

The significance of ongoing exploration into factors affecting scoring function precision cannot be overstated. The research outcomes carry profound implications across diverse domains, including structural biology, drug discovery, protein engineering, and computational biology. The accurate portrayal of protein-protein interactions emerges as a cornerstone for advancing scientific knowledge and forging effective therapies in these multifaceted fields. As we continue to unravel the intricacies of scoring functions, this knowledge serves as a catalyst for innovation, pushing the boundaries of what is achievable in the realms of medical research and therapeutic development.

Chapter 4

Conclusions

4.1. Summary

Recent years have seen a revolution in treating disorders through antibody-based therapy. Antibodies, pivotal in the immune system, play a crucial role in identifying and neutralizing foreign entities. Understanding these interactions is vital for developing effective treatments. Traditional methods like X-ray crystallography have limitations, leading to the importance of computational docking in predicting how small molecules bind to biomolecules.

This work is motivated by the understudied intricacies of Ab-Ag interactions and the imperative need to bridge the gap between traditional experimental techniques and computational approaches, particularly in the context of drug discovery and design. The revolutionary advancements in antibody-based therapy underscore the significance of comprehending the physico-chemical properties governing these interactions. By exploring the efficacy of Rosetta scoring functions, uncovering the prevalence of specific amino acids in CDR loops, and evaluating critical variables influencing decoy distribution, this study aims to contribute valuable insights to enhance our understanding of Ab-Ag complexes. The motivation lies in addressing the limitations of existing methodologies, propelling the field towards more accurate predictions and broader applicability in the realm of molecular medicine.

This study proposes enhanced scoring functions, targeting specific parameters in the Rosetta scoring function, to improve accuracy in predicting Ab-Ag complexes. It introduces numerical optimization methods, including models (M1, M2, and M3), for fine-tuning these functions. The following are the main conclusions of this thesis:

1. Efficacy of Rosetta Scoring Functions:

- Solvation parameters significantly impact docking performance in Ab-Ag complexes.

- Factors beyond solvation, such as hidden scoring terms and energy functions, require further investigation for improved accuracy.

2. Amino Acid Correlation in CDR Loops:

- Specific amino acids (Tyr, Ser, and Cys) in heavy chain CDR loops correlate with high-affinity Ab-Ag complexes.

3. Critical Variables Influencing Decoy Distribution:

- GLM analysis identifies key factors (abundance of non-aromatic residues in L3, scarcity in H1, and smaller Ab-Ag binding interface in L2) influencing predictive accuracy.

4. Persistent Influence of Aromatic Residues:

- Aromatic residues continue to influence scoring functions, even in the absence of solvation parameters, highlighting the need for specialized Ab-Ag interfaces.

5. Optimization Impact on Docking Accuracy:

- Numerical optimization methods, specifically models M1, M2, and M3, enhance docking accuracy with the M1 model showing a 4% improvement.

- Identified parametric weights and the Pearson correlation coefficient (PCC) serve as tools for assessing scoring function accuracy, necessitating further parameter exploration for reliability.

4.2. Topics for Future Research

Although the findings of the research have significant consequences across a variety of fields, there are still many aspects that need further investigations.

- Comprehensive Exploration of Hidden Scoring Terms: Investigate and uncover additional hidden scoring terms and energy functions influencing the accuracy of Rosetta scoring functions in Ab-Ag complexes.
- In-Depth Analysis of Amino Acid Presence in CDR Loops: Conduct a broader study to understand the mechanistic implications of specific amino acids (Tyr, Ser, and Cys) in heavy chain CDR loops, exploring their functional roles and reasons for presence or absence.
- Enhancement of Numerical Optimization Methods: Further refine and expand numerical optimization methods for determining optimal parametric weights,

considering broader applications and potential challenges, such as resistance to data noise and uncertainties.

- Validation of Findings Across Diverse Settings: Validate research results across various experimental settings, including different Ab-Ag complexes, experimental setups, antibody types, and antigen characteristics, to assess the generalizability of the identified parameters and their impact on scoring functions.
- Ethical Implications of Computational Models in Drug Development: Investigate the ethical implications of using computational models, such as Rosetta scoring functions, in drug development, particularly in terms of their reliability, potential biases, and implications for decision-making in the development of antibody-based therapies.

Reference

[1] J. V. Kringelum, M. Nielsen, S. B. Padkjær, and O. Lund, "Structural analysis of B-cell epitopes in antibody:protein complexes," *Mol Immunol*, vol. 53, no. 1–2, pp. 24–34, Jan. 2013, doi: 10.1016/j.molimm.2012.06.001.

[2] T. Ramaraj, T. Angel, E. A. Dratz, A. J. Jesaitis, and B. Mumey, "Antigen–antibody interface properties: Composition, residue interactions, and features of 53 non-redundant structures," *Biochim Biophys Acta*, vol. 1824, no. 3, pp. 520–532, Mar. 2012, doi: 10.1016/j.bbapap.2011.12.007.

[3] I. S. Mian, A. R. Bradwell, and A. J. Olson, "Structure, function and properties of antibody binding sites," *Journal of Molecular Biology*, vol. 217, no. 1, pp. 133–151, Jan. 1991, doi: 10.1016/0022-2836(91)90617-F.

[4] G. M. Morris and M. Lim-Wilby, "Molecular Docking," in *Molecular Modeling of Proteins*, vol. 443, Humana Press, 2008. Accessed: Nov. 29, 2023. [Online]. Available: https://link.springer.com/protocol/10.1007/978-1-59745-177-2_19

[5] B. Kramer, M. Rarey, and T. Lengauer, "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking," *Proteins*, vol. 37, no. 2, pp. 228–241, Nov. 1999, doi: 10.1002/(sici)1097-0134(19991101)37:2<228::aid-prot8>3.0.co;2-8.

[6] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, "Improved protein-ligand docking using GOLD," *Proteins*, vol. 52, no. 4, pp. 609–623, Sep. 2003, doi: 10.1002/prot.10465.

[7] T. A. Halgren *et al.*, "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening," *J Med Chem*, vol. 47, no. 7, pp. 1750–1759, Mar. 2004, doi: 10.1021/jm030644s.

[8] A. D. Hill and P. J. Reilly, "Scoring functions for AutoDock," *Methods Mol Biol*, vol. 1273, pp. 467–474, 2015, doi: 10.1007/978-1-4939-2343-4_27.

[9] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading," *J Comput Chem*, vol. 31, no. 2, pp. 455–461, Jan. 2010, doi: 10.1002/jcc.21334.

[10] G. Macindoe, L. Mavridis, V. Venkatraman, M.-D. Devignes, and D. W. Ritchie, "HexServer: an FFT-based protein docking server powered by graphics processors," *Nucleic Acids Res*, vol. 38, no. Web Server issue, pp. W445–W449, Jul. 2010, doi: 10.1093/nar/gkq311.

[11] D. Tobi, "Designing coarse grained-and atom based-potentials for protein-protein docking," *BMC Structural Biology*, vol. 10, no. 1, p. 40, Nov. 2010, doi: 10.1186/1472-6807-10-40.

[12] R. F. Alford *et al.*, "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design," *J. Chem. Theory Comput.*, vol. 13, no. 6, pp. 3031–3048, Jun. 2017, doi: 10.1021/acs.jctc.7b00125.

[13] P. Agrawal, H. Singh, H. K. Srivastava, S. Singh, G. Kishore, and G. P. S. Raghava, "Benchmarking of different molecular docking methods for protein-peptide docking," *BMC Bioinformatics*, vol. 19, no. 13, p. 426, Feb. 2019, doi: 10.1186/s12859-018-2449-y.

[14] H.-J. Böhm, "The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional

structure," *J Computer-Aided Mol Des*, vol. 8, no. 3, pp. 243–256, Jun. 1994, doi: 10.1007/BF00126743.
[15] C. T. Schoeder *et al.*, "Modeling Immunity with Rosetta: Methods for Antibody and Antigen Design," *Biochemistry*, vol. 60, no. 11, pp. 825–846, Mar. 2021, doi: 10.1021/acs.biochem.0c00912.
[16] Y. Ofran, A. Schlessinger, and B. Rost, "Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes," *J Immunol*, vol. 181, no. 9, pp. 6230–6235, Nov. 2008, doi: 10.4049/jimmunol.181.9.6230.
[17] C.-M. Yu *et al.*, "Rationalization and Design of the Complementarity Determining Region Sequences in an Antibody-Antigen Recognition Interface," *PLOS ONE*, vol. 7, no. 3, p. e33340, Mar. 2012, doi: 10.1371/journal.pone.0033340.
[18] H.-P. Peng *et al.*, "Antibody CDR amino acids underlying the functionality of antibody repertoires in recognizing diverse protein antigens," *Sci Rep*, vol. 12, no. 1, Art. no. 1, Jul. 2022, doi: 10.1038/s41598-022-16841-9.
[19] J. K. Leman *et al.*, "Better together: Elements of successful scientific software development in a distributed collaborative community," *PLOS Computational Biology*, vol. 16, no. 5, p. e1007507, May 2020, doi: 10.1371/journal.pcbi.1007507.
[20] R. Vita *et al.*, "The Immune Epitope Database (IEDB): 2018 update," *Nucleic Acids Res*, vol. 47, no. D1, pp. D339–D343, Jan. 2019, doi: 10.1093/nar/gky1006.
[21] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes," *Nucleic Acids Res*, vol. 45, no. Web Server issue, pp. W24–W29, Jul. 2017, doi: 10.1093/nar/gkx346.
[22] J. KYTE and R. DOOLITTLE, "A simple method for displaying the hydropathic character of a protein - PubMed," *Journal of Molecular Biology*, vol. Volume 157, no. Issue 1, p. Pages 105-132, Jan. 1982, doi: https://doi.org/10.1016/0022-2836(82)90515-0.
[23] J. M. Zimmerman, Eliezer, and R. Simha, "The characterization of amino acid sequences in proteins by statistical methods - ScienceDirect," *Journal of Theoretical Biology*, vol. 21, no. 2, pp. 170–201, 1968, doi: https://doi.org/10.1016/0022-5193(68)90069-6.
[24] S. Miller, J. Janin, A. M. Lesk, and C. Chothia, "Interior and surface of monomeric proteins," *Journal of Molecular Biology*, vol. 196, no. 3, pp. 641–656, Aug. 1987, doi: 10.1016/0022-2836(87)90038-6.
[25] G. Rose, A. Geselowitz, Lesser, R. Lee, and M. Zehfus, "Hydrophobicity of amino acid residues in globular proteins.," *Science*, vol. 229(4716), no. 834, p. 8, Aug. 1985, doi: 10.1126/science.4023714.
[26] P. Sedgwick, "Pearson's correlation coefficient," *BMJ*, vol. 345, p. e4483, Jul. 2012, doi: 10.1136/bmj.e4483.
[27] E. A. Padlan, "On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands," *Proteins*, vol. 7, no. 2, pp. 112–124, 1990, doi: 10.1002/prot.340070203.
[28] K. Tsumoto, K. Ogasahara, Y. Ueda, K. Watanabe, K. Yutani, and I. Kumagai, "Role of Tyr Residues in the Contact Region of Anti-lysozyme Monoclonal

Antibody HyHEL10 for Antigen Binding (*)," *Journal of Biological Chemistry*, vol. 270, no. 31, pp. 18551–18557, Aug. 1995, doi: 10.1074/jbc.270.31.18551.
[29] M. Shiroishi *et al.*, "Structural Consequences of Mutations in Interfacial Tyr Residues of a Protein Antigen-Antibody Complex: THE CASE OF HyHEL-10-HEL*," *Journal of Biological Chemistry*, vol. 282, no. 9, pp. 6783–6791, Mar. 2007, doi: 10.1074/jbc.M605197200.
[30] A. Steinmaurer, I. Wimmer, T. Berger, P. S. Rommer, and J. Sellner, "Bruton's Tyrosine Kinase Inhibition in the Treatment of Preclinical Models and Multiple Sclerosis," *Current Pharmaceutical Design*, vol. 28, no. 6, pp. 437–444.
[31] S. Koide and S. S. Sidhu, "The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins," *ACS Chem Biol*, vol. 4, no. 5, pp. 325–334, May 2009, doi: 10.1021/cb800314v.
[32] Y. Hagihara and D. Saerens, "Engineering disulfide bonds within an antibody," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1844, no. 11, pp. 2016–2023, Nov. 2014, doi: 10.1016/j.bbapap.2014.07.005.
[33] S. Birtalan, Y. Zhang, F. A. Fellouse, L. Shao, G. Schaefer, and S. S. Sidhu, "The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies," *Journal of Molecular Biology*, vol. 377, no. 5, pp. 1518–1528, Apr. 2008, doi: 10.1016/j.jmb.2008.01.093.
[34] F. A. Fellouse, C. Wiesmann, and S. S. Sidhu, "Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition," *Proceedings of the National Academy of Sciences*, vol. 101, no. 34, pp. 12467–12472, Aug. 2004, doi: 10.1073/pnas.0401786101.
[35] J. N. Martins, J. C. Lima, and N. Basílio, "Selective Recognition of Amino Acids and Peptides by Small Supramolecular Receptors," *Molecules*, vol. 26, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/molecules26010106.
[36] J. Janin, "Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition," *Structure*, vol. 7, no. 12, pp. R277–R279, Jan. 1999, doi: 10.1016/S0969-2126(00)88333-1.
[37] H. Ma, C. Ó'Fágáin, and R. O'Kennedy, "Antibody stability: A key to performance - Analysis, influences and improvement," *Biochimie*, vol. 177, pp. 213–225, Oct. 2020, doi: 10.1016/j.biochi.2020.08.019.
[38] T. Lazaridis and M. Karplus, "Effective energy function for proteins in solution," *Proteins: Structure, Function, and Bioinformatics*, vol. 35, no. 2, pp. 133–152, 1999, doi: 10.1002/(SICI)1097-0134(19990501)35:2<133::AID-PROT1>3.0.CO;2-N.
[39] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, Jan. 1965, doi: 10.1093/comjnl/7.4.308.
[40] M. J. G. Leening, M. M. Vedder, J. C. M. Witteman, M. J. Pencina, and E. W. Steyerberg, "Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide," *Ann Intern Med*, vol. 160, no. 2, pp. 122–131, Jan. 2014, doi: 10.7326/M13-1522.
[41] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Computers & Geosciences*, vol. 19, no. 3, pp. 303–342, Mar. 1993, doi: 10.1016/0098-3004(93)90090-R.

[42] M. J, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1967*, University of California Press, Berkeley: University of California Press, 1967, pp. 281–297. Accessed: Nov. 30, 2023. [Online]. Available: https://cir.nii.ac.jp/crid/1572261550390329472

[43] E. Krieger, G. Koraimann, and G. Vriend, “Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field,” *Proteins: Structure, Function, and Bioinformatics*, vol. 47, no. 3, pp. 393–402, 2002, doi: 10.1002/prot.10104.

[44] Z. Zhang, U. Ehmann, and M. Zacharias, “Monte Carlo replica-exchange based ensemble docking of protein conformations,” *Proteins: Structure, Function, and Bioinformatics*, vol. 85, no. 5, pp. 924–937, 2017, doi: 10.1002/prot.25262.

[45] I. Aier, P. K. Varadwaj, and U. Raj, “Structural insights into conformational stability of both wild-type and mutant EZH2 receptor,” *Sci Rep*, vol. 6, p. 34984, Oct. 2016, doi: 10.1038/srep34984.

[46] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, “Unsupervised Learning Methods for Molecular Simulation Data,” *Chem. Rev.*, vol. 121, no. 16, pp. 9722–9758, Aug. 2021, doi: 10.1021/acs.chemrev.0c01195.

[47] J. J. Gray *et al.*, “Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations,” *Journal of Molecular Biology*, vol. 331, no. 1, pp. 281–299, Aug. 2003, doi: 10.1016/S0022-2836(03)00670-3.

Appendices

The QR code provided below contains additional materials that play a significant role in our academic work. These appendices are carefully put together to give a deep understanding of different aspects of our study:

Through the given QR code, you will be able to access the following appendices:

- **Appendix A: Existing docking software and the scoring functions** - https://sangeetha-ratnayake.notion.site/Appendix-A-ca1a4c55cec84763878a60987100aae4
- **Appendix B: The dataset of 100 antibody-antigen complexes used in the study** - https://sangeetha-ratnayake.notion.site/Appendix-B-bcc2f9c529094981835c8d88e3fa31d0
- **Appendix C: Rosetta ref2015 Default Weights** - https://sangeetha-ratnayake.notion.site/Appendix-C-5232f6012b1e483fb58f363e28916c42
- **Appendix D: Physico-chemical properties of 20 Amino Acids** - https://sangeetha-ratnayake.notion.site/Appendix-D-fb09d07d72414ae8833e5284f207aedc
- **Appendix E: Rosetta-derived scoring functions and their performance** - https://sangeetha-ratnayake.notion.site/Appendix-E-385342ac93744ecb947113be0687052c
- **Appendix F: General Linear Model Regression Results** - https://sangeetha-ratnayake.notion.site/Appendix-F-ccbf15568bd540819bdd57f64bc64936
- **Appendix G: Optimization and Validation Results** - https://sangeetha-ratnayake.notion.site/Appendix-G-657829c5cb50443d9bfdae166c01e087
- **Appendix H: Validation Dataset of 10 Antibody-Antigen Complexes** - https://sangeetha-ratnayake.notion.site/Appendix-H-f795e860abf14d519e40f4fcd6d0ff73