IPSJ Transactions on Bioinformatics Vol.17 40–47 (Apr. 2024)

[DOI: 10.2197/ipsjtbio.17.40]

Original Paper

Advancing Antibody-antigen Interface Analysis in Docking Scoring Functions for Precision Docking Analysis

Sangeetha Ratnayake[1,a)] Axel Martinelli[2] Toshinori Endo[3] Naoki Osada[3]

Received: November 6, 2023, Accepted: December 30, 2023

Abstract: Molecular docking simulations utilizing scoring functions are pivotal for assessing the stability of complex formations. The unique biochemical characteristics of antibody-antigen interfaces, however, present challenges in applying general parameter sets of scoring functions to these molecules, necessitating the customization of the scoring function to enhance prediction accuracy for structural configurations and binding affinities. In response to this, we have developed models within the Rosetta software framework, widely recognized for its utility in predicting antigen-antibody docking, to optimize the parameters of its scoring function. Through a quantitative evaluation of the shape of decoy distribution generated by Rosetta, we have been able to refine the parameters for each antibody-antigen complex, yielding a notable improvement in the prediction accuracy of the software for a given dataset. Furthermore, we have identified a distinct parameter set that is effective for the majority of complexes in our dataset, though not universally applicable. This study introduces a novel approach to customizing scoring functions, potentially contributing to advancements in drug discovery and deepening our understanding of the complexities inherent in antibody-antigen interactions at a molecular level.

Keywords: antibody-antigen complexes, scoring function, root mean square deviation, pearson's correlation coefficient, optimization, nelder-mead minimization

1. Introduction

Precise prediction of protein structures, particularly at the antibody-antigen (Ab-Ag) interfaces, holds significant importance for various reasons. Antibodies serve as essential proteins within the immune system, while antigens are foreign entities such as pathogens. The interaction between them is intricately defined by the unique characteristics of their respective structures. Gaining an in-depth understanding of the Ab-Ag interface is vital for developing effective vaccines, creating targeted therapies, and unraveling the complexities of autoimmune diseases. Accurate prediction of these structures promises substantial savings in both time and resources, as it could reduce the dependence on experimental validation. The considerable impact and relevance of this field have warranted its extensive exploration and discussion in academic and research areas.

Computational methods play a significant role in the field of protein structure prediction. Yasara [1], Lee-server [2], and Rosetta [3] are popular methods that largely rely on scoring functions for their effectiveness. These methods implement scoring functions for their efficiency, which evaluate the energy and stability across various protein conformations, facilitating the identification of the most accurate structures [3], [4], [5], [6]. A scoring function, which is meticulously designed to assess the congruency between two interacting molecular structures, is fundamental for gauging the energy and stability of molecular interactions. It encompasses a broad spectrum of factors, each encapsulating different aspects of molecular interactions such as van der Waals forces, hydrogen bonds, and electrostatic interactions, among others [3]. Scoring functions are formulated using a linear model with a set of weight parameters. Assigning appropriate weight parameters of scoring functions is of utmost significance, as it determines the relative contribution of each factor to the overall score [3], [6], [7], [8].

Although scoring functions prove to be beneficial tools, our latest research reveals that popular software like Rosetta may not deliver peak efficiency for Ab-Ag data analysis. The study concluded that discarding solvency parameters can moderately enhance performance [9], but this method does not fully resolve the underlying complexities. There is an immediate need for groundbreaking approaches tailored to optimize parameters for precise Ab-Ag docking.

In this study, our aim is to derive an optimal set of weight parameters to enhance the accuracy of the RosettaAb docking scoring function, specifically adapted for a dataset of Ab-Ag complexes. We aim to achieve this through four distinct analyses. Firstly, we identified crucial parameters within the Rosetta scoring function. This analysis led us to identify key parameters that have a substantial influence on the scoring function's ability to accurately detect high-quality docking positions. Secondly, we determined the optimal parametric weights for these key param-

[1] Division of Bioinformatics and Bioengineering, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido 060–0814, Japan
[2] BigOmics Analytics, Via Serafino Balestra 12, 6900 Lugano, Switzerland
[3] Faculty of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido 060–0814, Japan
[a)] sangeetha.ratnayake@gmail.com

eters. This was achieved by optimizing three different statistics that capture the shape of decoy distributions for each Ab-Ag complex in our dataset. Thirdly, we assessed the impact of the optimized weights by re-generating decoys using RosettaAb. This step involved both quantitative and qualitative evaluations of the decoy distributions to ascertain the effectiveness of the optimized weights. Finally, we investigated whether it is possible to establish a distinct set of weight parameters. These parameters would ideally enhance the prediction accuracy of Ab-Ag complexes across various scenarios, contributing to more reliable and precise computational predictions in the study of Ab-Ag interactions.

2. Methods

2.1 Structural Evaluation and Decoy Conformations

In this research, we employed 100 antibody-antigen (Ab-Ag) complex structures previously utilized in our prior study [9], obtained from the SabDab database [1]. Subsequently, we generated 500 docking conformations, referred to as decoys, through Rosetta docking using the default scoring function *ref*2015 [3]. This scoring function encompasses 20 predefined parameters along with their respective parametric weights. In each decoy generation, we documented parametric energies, total energy scores, and Root Mean Square Deviation (RMSD) values from the RosettaAb docking procedure. The docking involves explicit specification of the antibody and antigen light and heavy chains. The procedure aligns the entire decoy complex with the native complex by centering their masses, followed by rotation for optimal alignment with the native structure. Subsequently, RMSD is computed.

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 (**Fig. 1**).

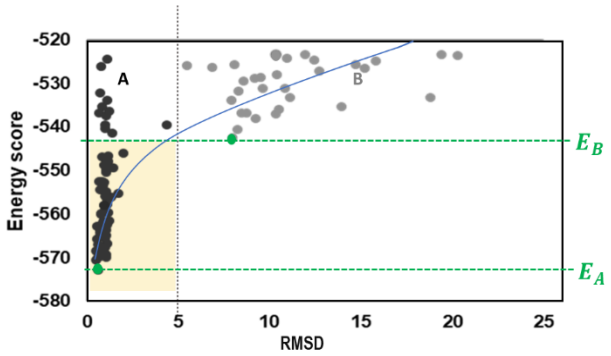


Fig. 1 Example of a decoy distribution and definition of parameters in Optimization Model 3. The decoy distribution was separated to two clusters A (decoys with RMSD < 5Å) and B (decoys with RMSD > 5Å). 5Å is considered as the threshold for near-nativeness of a decoy. Green line vertical line represents the 5Å cluster threshold line. $\boldsymbol{E_A}$: Decoy with the lowest energy score of the Cluster A. $\boldsymbol{E_B}$: Decoy with the lowest energy score of the cluster B. Green horizontal line is to separate the decoys with low energy score than $\boldsymbol{E_B}$, and the number of decoys in the yellow area is counted and accounted for in the M3 as denoted by D.

2.2 Identification of High Predictive Power Parameters in Rosetta Scoring Function

The Rosetta scoring function includes 20 weighted parameters (S1). To discern the relative influence of each parameter on the final energy score, the following analysis is undertaken. Initially, all parameter weights are standardized to a value of one, and the total energy score is computed for each structure. By comparing the total energy score obtained with all parameters to that achieved when a specific parameter is omitted, the significance of each variable is ascertained. This procedure was applied to all 100 complexes housed in the database, culminating in the determination of the average proportional contribution of each parameter.

2.3 Classification of the Decoy Distribution

An example of good decoy distribution is shown in Fig. 1. We defined two qualitative criteria to define a good or bad decoy distribution; 1) a Pearson's correlation coefficient (P) [11] 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).

2.4 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 [12] 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 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.

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

Optimization Model 3 (M3): In this model we execute a combined model of M1 and M3 with additional features of a decoy distribution (Fig. 1).

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 Fig. 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 are considered optimal.

2.5 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 2.3.

Net Reclassification Improvement (NRI_{net}) [13] 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) [14] and the k-means clustering algorithm [15] were applied to the set of parameters estimated in the 100 Ab-Ag complexes.

3. Results

3.1 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 (**Fig. 2**). 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 [3].

3.2 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 3 (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 Fig. 3 (e)–(f), the fre-

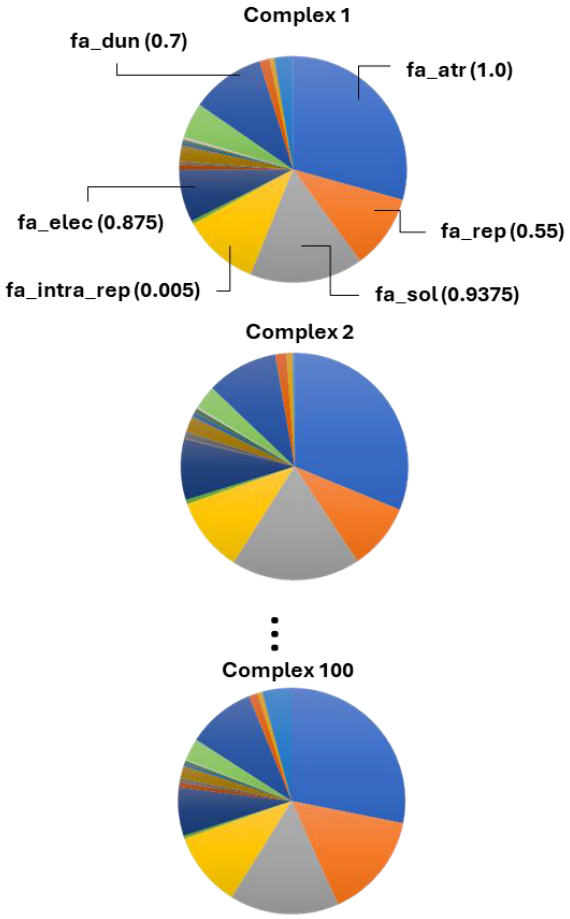


Fig. 2 Parametric Contributions to the Total Energy Score in Rosetta Scoring. We identified six crucial parameters, including Lennard-Jones Attractive Forces (fa_atr), Dihedral Angle Energy (fa_dun), Electrostatic Energy (fa_elec), Lennard-Jones Repulsion (fa_rep), Intra-residue Solvation Energy (fa_intra_rep), and Solvation Energy (fa_sol), with specific predefined weights of Rosetta $ref2015$ as specified within parentheses, that greatly impact the total energy score.

quencies were calculated for both post optimization and after validation.

We optimized the six key weight parameters discovered in Result Section 3.1, 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 (S3–S5). 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 (S2). All three models demonstrated a substantial improvement with a positive NRI value of 0.941.

3.3 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 3 (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 3 (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 [16], 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 ($t = 10.79$, $p = 2.06\times10-18$), followed by M3 ($t = 9.92$, $p = 1.60\times10-16$), and M2 ($t = 7.85$, $p = 5.00\times10-12$).

3.4 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 4** 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

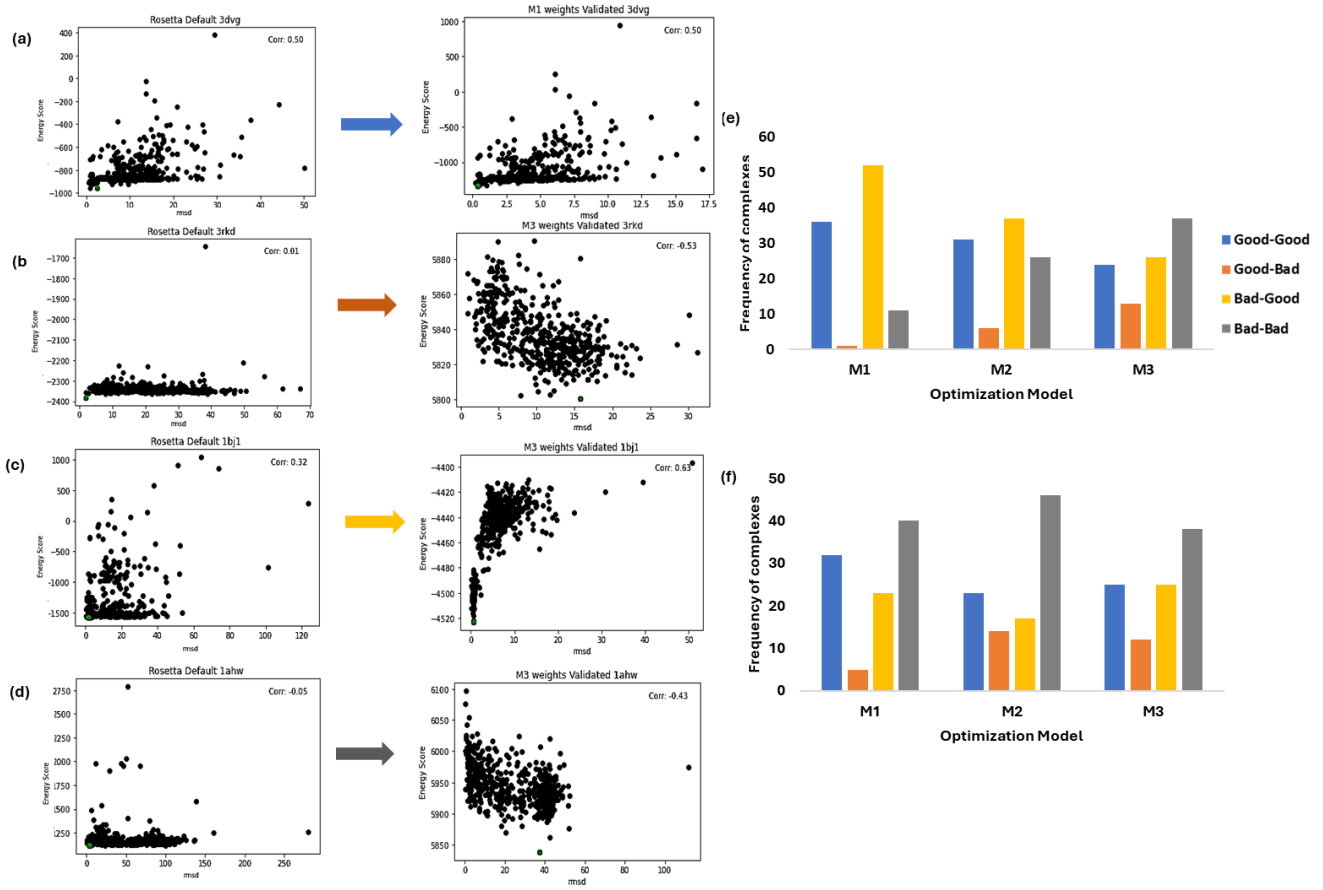


Fig. 3 Optimization Model Performance Comparison. On decoy distributions, x-axis represents decoy RMSD values and y-axis represents decoy energy scores. The decoy distributions were classified as, Good: Funnel-like and Bad: not funnel like distributions. (a)–(d) decoy distributions represent examples 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.

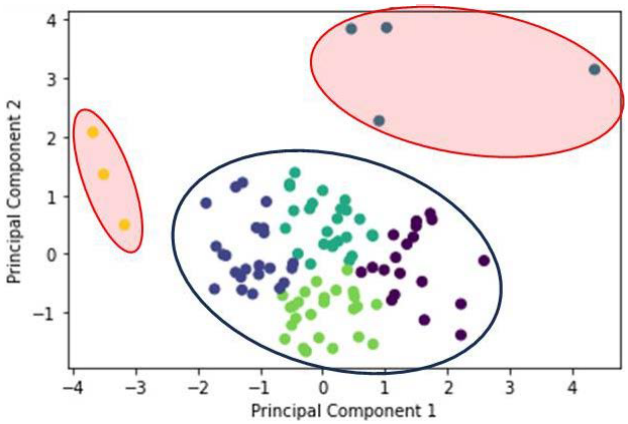


Fig. 4 Principal Component Analysis. 6 clusters were identified using K-means clustering. 2 clusters (marked in red) were considered as outliers. These datapoints (complexes) were excluded from the database during the calculation of mean values of parametric weights. Four remaining clusters seems closely related and together they create one big cluster (blue circle).

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, 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 $2.43\times10-14$. The NRI value between the Rosetta default and the distinctive weights was 0.924.

The identified through PCA (Fig. 4) were applied to a randomly selected independent dataset of 10 Ab-Ag complexes 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.

4. 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 article delves into an in-depth exploration of these critical parameters, shedding light on their importance in the field of Ab-Ag modeling and design.

4.1 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 [1], [17], [18], [19], [20], [21]. We have already shown that there was a significant improvement in scoring energy in the absence of the solvation parameter, fa_sol in our previous study 'Exploring the Interplay Between Scoring Functions and Physico-chemical Properties in Antibody-antigen Docking' [9]. The absence or presence of these parameters can affect the overall binding affinity during the docking process.

The study 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.

4.2 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.

4.3 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.

5. Conclusion

It is important to comprehend the critical variables that influence the computational docking scoring function. The optimization of the parametric weights of six key parameters in the Rosetta docking scoring function resulted in a significant performance improvement. This emphasizes the important influence these parameters have on the stability and energetics of Ab-Ag interactions. The data suggests that by optimizing parameters, such as binding affinity and stability, there is potential to improve therapeutic interventions for cancer and autoimmune diseases.

This research emphasizes the significance of continuous inves-

tigation into the factors that impact the precision of scoring functions. The research findings have significant implications for various fields such as structural biology, drug discovery, protein engineering, and computational biology. Accurate representations of protein-protein interactions are keys for advancing scientific knowledge and developing effective therapies in these areas.

6. Abbreviations

Ab: Antibody
Ag: Antigen
PCA: Principal Component Analysis
PCC: Pearson's Correlation Coefficient
PPI: Protein-Protein Interactions
RMSD: Root Mean Squared Deviation

Acknowledgments We extend our heartfelt gratitude to Prof. Luca Varani and Dr. Luca Simonelli of the Laboratory of Structural Biology at the Institute for Research in Biomedicine, Università Della Svizzera Italiana, Via Vela 6, 6500 Bellinzona, Switzerland. Their invaluable knowledge and unwavering guidance have been instrumental from the inception of this study. We also wish to express our thanks to Dr. Stefan Brennsteiner for his invaluable critical comments and insightful feedback on our work in the realm of scientific writing. Their contributions have been indispensable to the success of this endeavor.

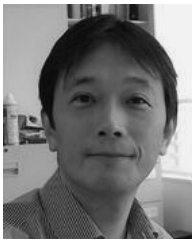
References

[1] Krieger, E., Koraimann, G. and Vriend, G.: 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, DOI: 10.1002/prot.10104 (2002).
[2] Arnittali, M., Rissanou, A.N. and Harmandaris, V.: Structure of Biomolecules Through Molecular Dynamics Simulations, *Procedia Computer Science*, Vol.156, pp.69–78, DOI: 10.1016/j.procs.2019.08.181 (2019).
[3] Alford, R.F. et al.: The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design, *J. Chem. Theory Comput.*, Vol.13, No.6, pp.3031–3048, DOI: 10.1021/acs.jctc.7b00125 (2017).
[4] Böhm, H.-J.: 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, DOI: 10.1007/BF00126743 (1994).
[5] Wang, R., Lai, L. and Wang, S.: Further development and validation of empirical scoring functions for structure-based binding affinity prediction, DOI: 10.1023/A:1016357811882 (2002).
[6] Leaver-Fay, A. et al.: Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules, *Methods Enzymol*, Vol.487, pp.545–574, DOI: 10.1016/B978-0-12-381270-4.00019-6 (2011).
[7] Wang, C., Schueler-Furman, O. and Baker, D.: Improved side-chain modeling for protein–protein docking, *Protein Science*, Vol.14, No.5, pp.1328–1339, DOI: 10.1110/ps.041222905 (2005).
[8] Lee, J., Scheraga, H.A. and Rackovsky, S.: New optimization method for conformational energy calculations on polypeptides: Conformational space annealing, *Journal of Computational Chemistry*, Vol.18, No.9, pp.1222–1232, DOI: 10.1002/(SICI)1096-987X(19970715)18:9<1222::AID-JCC10>3.0.CO;2-7 (1997).
[9] Ratnayake, S., Martinelli, A., Endo, T. and Osada, N.: Exploring The Interplay Between Scoring Functions and Physico-chemical Properties in Antibody-antigen Docking, *IPSJ Trans. Bioinformatics*, Vol.17, pp.10–17, DOI: 10.2197/ipsjtbio.17.10 (2024).
[10] Schneider, C., Raybould, M.I.J. and Deane, C.M.: SAbDab in the age of biotherapeutics: Updates including SAbDab-nano, the nanobody structure tracker, *Nucleic Acids Res.*, Vol.50, No.D1, pp.D1368–D1372, DOI: 10.1093/nar/gkab1050 (2022).
[11] Sedgwick, P.: Pearson's correlation coefficient, *BMJ*, Vol.345, p.e4483, DOI: 10.1136/bmj.e4483 (2012).
[12] Nelder, J.A. and Mead, R.: A Simplex Method for Function Minimization, *The Computer Journal*, Vol.7, No.4, pp.308–313, DOI: 10.1093/comjnl/7.4.308 (1965).
[13] Leening, M.J.G., Vedder, M.M., Witteman, J.C.M., Pencina, M.J. and Steyerberg, E.W.: Net reclassification improvement: Computation, interpretation, and controversies: A literature review and clinician's guide, *Ann. Intern. Med.*, Vol.160, No.2, pp.122–131, DOI: 10.7326/M13-1522 (2014).
[14] Maćkiewicz, A. and Ratajczak, W.: Principal components analysis (PCA), *Computers & Geosciences*, Vol.19, No.3, pp.303–342, DOI: 10.1016/0098-3004(93)90090-R (1993).
[15] MacQueen, J.: Some methods for classification and analysis of multivariate observations, *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley: University of California Press, pp.281–297 (1967), (Online), available from ⟨https://cir.nii.ac.jp/crid/1572261550390329472⟩ (accessed 2023-11-30).
[16] Schoeder, C.T. et al.: Modeling Immunity with Rosetta: Methods for Antibody and Antigen Design, *Biochemistry*, Vol.60, No.11, pp.825–846, DOI: 10.1021/acs.biochem.0c00912 (2021).
[17] Zhang, Z., Ehmann, U. and Zacharias, M.: Monte Carlo replica-exchange based ensemble docking of protein conformations, *Proteins: Structure, Function, and Bioinformatics*, Vol.85, No.5, pp.924–937, DOI: 10.1002/prot.25262 (2017).
[18] Aier, I., Varadwaj, P.K. and Raj, U.: Structural insights into conformational stability of both wild-type and mutant EZH2 receptor, *Sci. Rep.*, Vol.6, p. 34984, DOI: 10.1038/srep34984 (2016).
[19] Glielmo, A., Husic, B.E., Rodriguez, A., Clementi, C., Noé, F. and Laio, A.: Unsupervised Learning Methods for Molecular Simulation Data, *Chem. Rev.*, Vol.121, No.16, pp.9722–9758, DOI: 10.1021/acs.chemrev.0c01195 (2021).
[20] Gray, J.J. 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, DOI: 10.1016/S0022-2836(03)00670-3 (2003).
[21] Leman, J.K. et al.: Better together: Elements of successful scientific software development in a distributed collaborative community, *PLOS Computational Biology*, Vol.16, No.5, p.e1007507, DOI: 10.1371/journal.pcbi.1007507 (2020).

Sangeetha Ratnayake is a dedicated researcher in the field of computational biology and bioinformatics. She is currently pursuing her Ph.D. at Hokkaido University, Japan, focusing on the design of computational docking scoring functions tailored to antibody-antigen interfaces. Sangeetha has a strong foundation in data analysis, pattern recognition, and prediction modeling using Python and R. Her research areas are computational and structural bioinformatics, genomics, prediction modeling, bigdata analysis and machine learning. Her commitment to research-driven approaches and her expertise in computational biology make her a valuable contributor to the field.

Axel Martinelli is a highly accomplished scientist and Chief Scientific Officer (CSO) at BigOmics. With expertise in genomics and bioinformatics, they have made significant contributions to the study of parasite genomics, particularly malaria. Axel's research, including papers like "Rapid identification of genes controlling virulence and immunity in malaria parasites," showcases their deep understanding of the genetic factors underlying diseases. They hold advanced degrees in Bioinformatics, Biology, and Immunology from renowned universities. Axel's commitment to continuous learning is demonstrated through their completion of courses in Operations Management, Financial Accounting, and more. Their expertise and dedication position them as a respected figure in scientific research and leadership.

Toshinori Endo received his Ph.D. in Genetics from The Graduate University for Advanced Studies, SOKENDAI in 1996. He experienced post-docs at National Institute of Genetics from 1996, Biolabs, Harvard University from 1997, and Genome Sciences Center, RIKEN from 1999. He was a lecturer at the Medical Research Institute, Tokyo Medical and Dental University from 2000 to 2004. He became a full professor of bioinformatics at Graduate School of Information Science Technology, Hokkaido University in 2004. His major interests are molecular evolution, prediction, and design of gene functions.



Naoki Osada is an associate professor at Hokkaido University, Japan. He received his Ph.D. in 2002 at the University of Tokyo for his work on genomics and the molecular evolution of primates. He worked as a post-doctoral researcher at the University of Tokyo from 2002 to 2003, and the University of Chicago from 2003 to 2004. During that time, he studied bioinformatics and population genetics for speciation. He worked at the National Institute of Biomedical Innovation from 2005 to 2010 as a Scientific Researcher. From 2010 to 2015, he worked at the National Institute of Genetics as an assistant professor. He is currently working on a variety of research topics in molecular evolution and population genetics, including the study of human/primate evolution and gene expression evolution.

(Communicated by *Hiroyuki Kurata*)