https://doi.org/10.26565/2220-637X-2025-45-02 ISSN 2220-637X
УДК: 544.165+577.322

COMPARATIVE EVALUATION OF COMPUTATIONAL METHODS FOR MODELING SARS-COV-2 SPIKE PROTEIN RBD - ANTIBODY COMPLEXES

V. O. Morozova[a,*†], M. V. Prud[b,*†], A. Kyrychenko[c,*‡]

[*]V.N. Karazin Kharkiv National University, Education and Research Institute of Chemistry, 4 Svobody sq., Kharkiv 61022, Ukraine
[†]Enamine Ltd., Winston Churchill Street 78, Kyiv 02094, Ukraine
[‡]Institute of Functional Materials Chemistry, State Scientific Institution "Institute for Single Crystals" of National Academy of Sciences of Ukraine, 60 Nauky ave., 61072 Kharkiv, Ukraine

a) v.moroz.ukraine@gmail.com https://orcid.org/0009-0006-9615-5980
b) mykyta.prud@gmail.com https://orcid.org/0000-0002-0051-8104
c) a.v.kyrychenko@karazin.ua https://orcid.org/0000-0002-6223-0990

Computational modeling of protein-protein interactions (PPIs) is a high priority because it is essential for understanding nearly all cellular functions and for addressing the limitations of time-consuming experimental methods. This approach is crucial for drug discovery, as it enables rapid identification of disease targets, the design of therapeutics that modulate PPIs, and the simulation of interactions under realistic, crowded cellular conditions. The SARS-CoV-2 Spike protein's Receptor-Binding Domain (RBD) is the crucial region that directly engages the human ACE2 receptor to initiate infection. Therefore, studying the RBD and its interactions with neutralizing antibodies is paramount for understanding immune protection, assessing the threat posed by viral variants, and guiding the design of effective vaccines and antibody-based therapeutics. Here, we evaluated the performance of several computational tools for modeling complexes between the SARS-CoV-2 Spike protein receptor-binding domain (RBD) and various neutralizing antibodies. Recent breakthroughs in computational chemistry have moved beyond traditional protein-protein docking towards various generative methods and co-folding tools that predict a protein's 3D structure and binding poses from its sequence. Therefore, this motivates us to compare the performance of traditional docking methods, such as pyDockWEB and ClusPro, with that of novel, promising AI-driven tools, such as AlphaFold 3, Boltz-2, Protenix, and Chai-1. All these methods were systematically evaluated for their ability to reproduce the 3D structures of known Spike RBD-antibody complexes. We found that traditional docking tools, such as ClusPro and pyDockWeb, perform well at capturing correct protein-protein interactions for relatively small antibodies with well-defined interaction interfaces, but fall short at reproducing more complex protein-protein assemblies. AlphaFold 3 performs best at reproducing the 3D structures of the five studied RBD-antibody complexes among the four AI-driven prediction tools considered. Our study sheds light on understanding protein-protein interactions and provides a practical guide for accurate modeling of viral Spike protein RBD-antibody interactions.

Keywords: *protein folding, protein-protein interaction, molecular docking, AlphaFold, artificial intelligence.*

Introduction

Understanding protein-protein interactions is essential in designing selective and highly specific antibody, capable of binding to a unique antigen. Computational modeling these interactions helps identifying the structure of specific complexes between antibody and antigen, like between a lock and key, which further "tags" the invader for destruction by other immune cells or neutralizes it directly, providing long-term immunity after infections or vaccinations [1-3]. Therefore, prediction of the 3D structure of protein-protein complex opens up opportunities for developing novel therapies at low computational costs [4-6].

The Spike (S) protein of SARS-CoV-2 is a large, club-shaped trimeric glycoprotein on the virus's surface. It enables infection by binding to host ACE2 receptors, allowing viral entry and fusion [7]. The S protein has two domains: S1, which mediates virus entry via its receptor-binding domain (S-RBD) binding to the host receptor angiotensin-converting enzyme 2 (ACE2), and S2, which facilitates viral membrane fusion. It is the main target for vaccines and antibodies and forms the characteristic "corona" of the virus (see Figure 1a). Therefore, understanding how potent neutralizing monoclonal antibodies recognize the Spike protein is pivotal for developing new vaccines and antibody drugs. Although numerous efforts have been made to develop antibody-based vaccines and therapeutics, the emergence of new SARS-CoV-2 lineages has marked a new phase in this COVID-19 pandemic. Sev-

eral reports have suggested that these virus variants contain multiple genetic mutations in the S-RBD, which may change the shape of the S-RBD or interfere with how neutralizing antibodies recognize and bind to it, thereby potentially reducing the effectiveness of these antibodies [8].

Advancements in computational tools for biomolecular interactions are shifting from traditional, physics-based simulations towards novel "co-folding" architectures and geometric deep learning approaches. As of 2025, models that treat protein-ligand (PLI) and protein-protein interactions (PPI) as a unified structural prediction problem rather than two separate steps of folding and docking dominate the field.

In this study, we performed comparative evaluation of the performance of several computational tools for modeling complexes between the SARS-CoV-2 Spike protein RBD and various neutralizing antibodies utilizing web-based tools for molecular docking (pyDockWEB, ClusPro) as well as novel generative neural network models (AlphaFold 3, Boltz-2, Chai-1, Protenix).

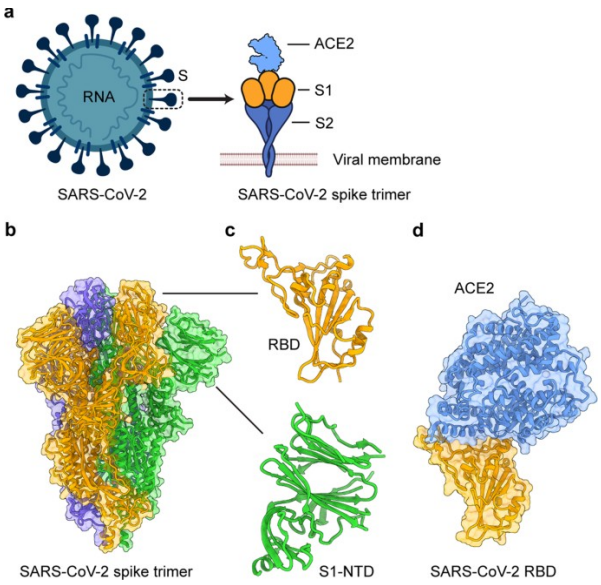

Figure 1. Structure of the SARS-CoV-2 spike (S) protein. (*a*) Schematic diagram of the SARS-CoV-2 virion and its S protein. (*b*) Cryo-EM structure of the S protein trimer (PDB 6VXX). The three subunits are color-codded by *orange*, *green*, and *blue*. (c) Detailed views of the receptor-binding domain (RBD) and N-terminal domain (NTD) in the S1 subunit. (d) Crystal structure of the SARS-CoV-2 RBD in complex with human ACE2 (PDB 6M0J). Human ACE2 is colored in *light blue*. Adopted with permission from [9].

Computational Methods

To model antibody-antigen complex structures, we used a docking protocol implemented in the PIPER programme and integrated into the ClusPro server [10]. The main methodological approach was the use of a specialised asymmetric potential ADARS (asymmetric Decoys as the Reference State), which accounts for the specific asymmetry of interfaces, where tyrosine, tryptophan and phenylalanine residues are much more common in the antibody paratope than in the antigen epitope. The process involved exhaustive global sampling of billions of conformations using a fast Fourier transform (FFT) approach, after which the 1,000 lowest-energy structures were subjected to RMSD clustering to identify the most populated clusters representing energy minima. The final models, defined as the centres of these clusters, were energy-minimised using the CHARMM force field to eliminate steric overlaps. This combined approach allows the identification of structures close to native ones, even without a priori information on the locations of CDR sites that determine complementarity [10-11].

Another classic tool for predicting the structure of protein-protein complexes used in this work was the pyDockWEB web server, based on a rigid docking algorithm [12]. The process of generating conformations (sampling) was carried out using a parallel implementation of the FTDock programme, which performs a global search in translational and rotational space by discretising molecules on orthogonal grids. To optimise fast Fourier transform (FFT) calculations, the grid size n was selected ac-

cording to the requirements of the FFTW library, which provides significant acceleration without loss of accuracy. The obtained models were ranked using the pyDock scoring function (version 3), primarily evaluates interactions based on electrostatics and desolvation energy with limited contribution from Van der Waals energy [12].

AlphaFold 3 was employed as a diffusion-based generative framework to predict the three-dimensional structures of biomolecular complexes, including protein–protein interactions [13]. The model processes interacting entities through a unified representation, synthesizing primary sequence data with learned structural priors [14]. This prediction task is performed as a denoising diffusion process, in which initially perturbed atomic coordinates are iteratively refined to reach physically plausible conformations. Its architecture integrates attention-based neural networks with geometric constraints, facilitating the simultaneous modeling of intra- and inter-molecular dynamics [15-16]. Evolutionary context, derived from multiple sequence alignments (MSAs), was incorporated during refinement when available. To rank the resulting complexes, model confidence was quantified using the predicted TM-score (pTM) and interface TM-score (ipTM), combined into a single aggregate metric.

Boltz-2 (version 2.2.0) served as an alternative diffusion-based generative model, utilizing a formulation conceptually similar to AlphaFold 3. This method maps interacting protein chains within a shared geometric coordinate system, applying successive denoising steps to transform stochastic initial states into energetically favorable assemblies. By explicitly accounting for inter-chain spatial orientations, Boltz-2 enables high-fidelity interface reconstruction without requiring rigid-body docking as a precursor. The diffusion-based sampling enabled exploration of diverse interaction geometries, guided by evolutionary data when applicable. Predicted complexes were ranked based on confidence indices derived from structural consistency and interface precision. In this study, predictions were generated via the TamarindBio web service (https://app.tamarind.bio/boltz) with both diffusion samples and recycling steps set to 10 [17-18].

Chai-1 (version 0.6.2) was utilized as a diffusion-based model optimized for the structural characterization of biomolecular assemblies and protein–protein interfaces [19]. The model facilitates the transition of perturbed conformations into coherent complex structures through an iterative denoising mechanism within a common geometric framework. This explicit modeling of inter-chain dependencies allows for accurate interface prediction independent of predefined docking poses. Evolutionary information was integrated to guide structural refinement, while the generative nature of the diffusion process enabled sampling of multiple plausible binding modes. Ranking was performed using internal metrics of global coherence and interface quality. Inference was performed using the TamarindBio service (https://app.tamarind.bio/chai), with 10 diffusion samples and 10 recycling steps [19].

Protenix is a trainable, open-source PyTorch implementation inspired by AlphaFold 3, designed for high-accuracy biomolecular structure prediction. The framework aims to provide an accessible, extensible research platform for developing and evaluating generative models in computational structural biology. Complexes were prioritized using internal confidence scores reflecting both structural stability and interface fidelity. Predictions were obtained using the implementation hosted on the TamarindBio web platform (https://app.tamarind.bio/protenix) [20].

The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC was used for visualization and RMSD analysis.

Results and discussion

The long-standing paradigm of traditional protein-protein docking, based on physics-based simulations that require a priori knowledge of the experimental 3D structure of each protein, is rapidly shifting towards various generative methods and co-folding tools that predict the protein's 3D structure and binding poses from protein sequences alone. Therefore, we compared two traditional web-based docking tools, such as PyDockWEB and ClusPro, with novel AI-driven approaches, such as AlphaFold 3, Boltz-2, Protenix, and Chai-1. All these conventional and generative methods were systematically evaluated against experimentally resolved Spike RBD-antibody structures summarized in Figure 2.

Benchmarking set of SARS-CoV-2 Spike protein RBD complexes with various neutralizing antibodies

As of January 2025, experimental structures of the SARS-CoV-2 S-protein/RBD complexes with neutralizing antibodies (NAbs) revealed up to a total of 540 PDB entries [21]. Therefore, our benchmark set comprised four representative SARS-CoV-2 S-RBD complexes with NAbs deposited by December 2024. These PDB entries were most likely used in the training sets of the evaluated AI-driven prediction tools. The PDB entry 9LAE was deposited after January 2025 and represents a new structure beyond the ML template datasets.

The crystal structure of CR3022, a neutralizing human antibody previously isolated from a convalescent SARS patient in complex with the SARS-CoV-2 S-protein RBD (3.1-Å resolution) was considered (Figure 2a). CR3022 recognizes a highly conserved epitopes of SARS-CoV-2 and SARS-CoV [22].

Other two examples are given by the co-crystal structures of neutralizing antibodies, CC12.1 and CV30 (see Figures 2b and 2c), which were previously isolated from a SARS-CoV-2–infected patient and were shown to be specific for the RBD [23-24].

The X-ray crystal structure of the human neutralizing antibody nCoV617 fragment in complex with the SARS-CoV-2 S-RBD (2.51 Å resolution) was identified from an isolated antibody derived from convalescent plasma (PDB: 7E3O) [25]. The crystal structure revealed that nCoV617 mainly binds to the back of the "ridge" of the RBD receptor and possesses common binding residues with ACE2, as shown in Figure 2d.

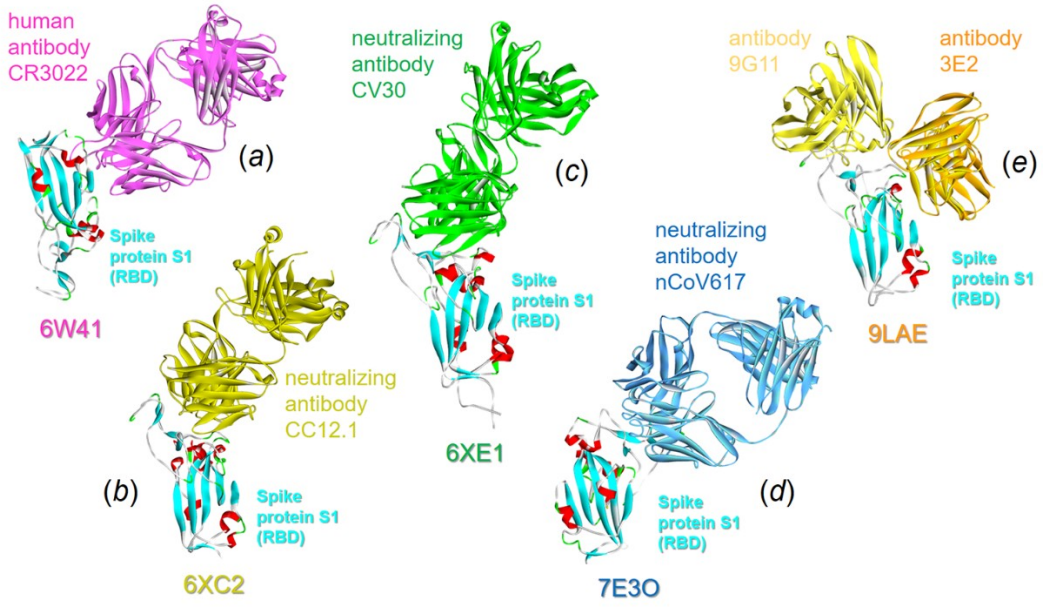

Figure 2. Crystal structure of SARS-CoV-2 receptor binding domain in complex with various neutralizing antibodies: (a) human antibody CR3022 (6W41) [22], (b) neutralizing antibody CC12.1 (6XC2) [23], (c) neutralizing antibody CV30 (PDB: 6XE1) [24], (d) neutralizing antibody nCoV617 (PDB: 7E3O) [25], (e) two antibodies 9G11 and 3E2 (PDB: 9LAE) [26].

Finally, we also included the recently deposited cryo-EM structure of the S-RBD complex (3.46 Å resolution) (Figure 2e), which combines "antibody cocktail" composed of new 9G11 antibody with the previously reported 3E2 one that broaden the neutralization spectrum against new variants of SARS-CoV-2 [26].

Traditional protein-protein docking

Traditional protein-protein docking tools, such as pyDockWEB, HADDOCK, and ClusPro, are controllable, interpretable, computationally efficient for screening, and can integrate experimental data directly. These tools have long been used by structural biologists and medicinal chemists as precise, tunable instruments for computational studies of complex protein-protein architectures. In our comparative evaluation, we benchmarked the performance of two of these tools – ClusPro and pyDockWEB in reproducing the available 3D-structure of the selected RBD-antibody complexes (see Figure 2).

As shown in Figure 3, the ClusPro reproduces complex structures with fairly high accuracy — the average backbone RMSD does not exceed 2.0 Å. The only exception is the crystal structure of 9LAE, a non-classical biomolecular complex that demonstrates simultaneous binding of the spike protein to two antibodies at once; the error relative to the crystal structure exceeds 10 Å. Such good performance of conventional methods in this metric can be explained by the fact that these methods use already known crystal structures as input, and in the process of protein-protein docking, only amino acid residues at the site of protein contact undergo conformational changes.

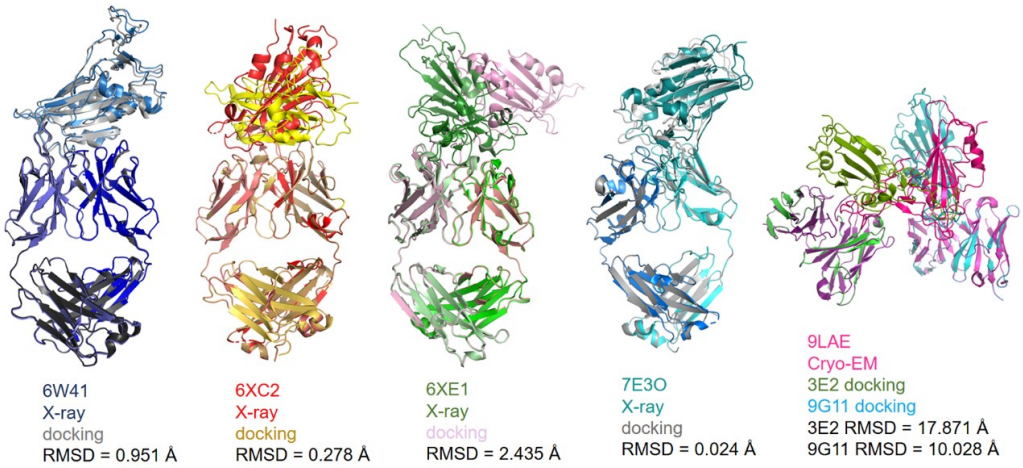

Figure 3. Performance of ClusPro in reproducing the 3D-structure across the studied RBD-antibody complexes. The overlap shows color-coded comparison of the experimental X-ray or Cryo-EM data and computed 3D-structure.

As with the previous method, pyDockWeb, for similar reasons, demonstrates quite good performance on this metric. However, minor differences are also present — some biomolecular complexes are reproduced better compared to the previous method (7E3O), while others are significantly worse (6XE1). Such differences between classical methods can be explained by their slightly different approaches. At the same time, the non-classical crystal structure 9LAE is reproduced poorly.

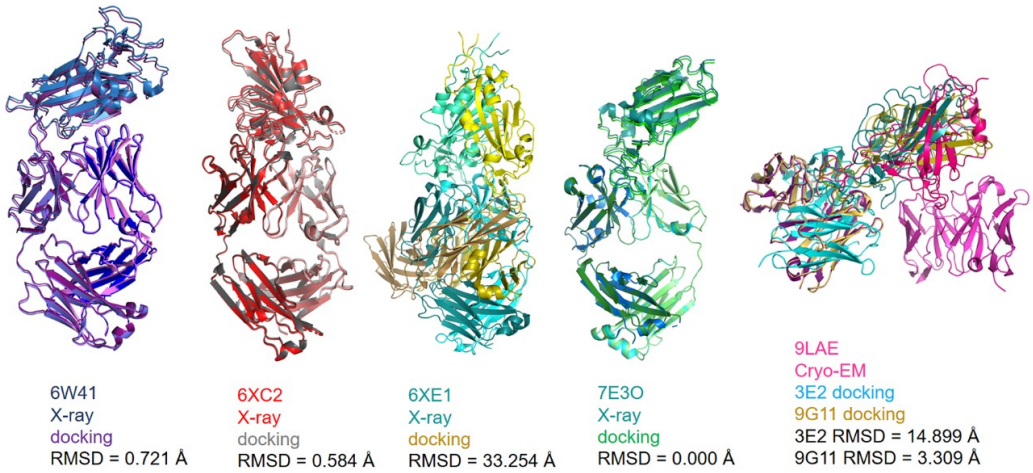

Figure 4. Performance comparison of pyDockWeb in reproducing the 3D-structure across the five studied RBD-antibody complexes.

A quantitative comparison of the two classical protein-protein docking methods is summarized in Table 1. The first obvious advantage of pyDockWeb is the availability of a docking score, which can also be used to assess the accuracy of the reconstructed structure (the simulated structure of 6XE1, which has the highest RMSD value, also has the highest docking score (-21.435 kcal/mol).

Table 1. Comparison of molecular docking of the 3D-structure of the Spike protein – antibody complexes using different web-tools.

Metrics	PDB: 6W41	PDB: 6XC2	PDB: 6XE1	PDB: 7E3O	PDB: 9LAE	
					antibody 3E2	antibody 9G11
ClusPro						
RMSD (Å)[a]	0.951	0.278	2.435	0.024	17.871	10.028
pyDockWeb						
RMSD (Å)[a]	0.721	0.584	33.254	0.000	14.899	3.309
Total Energy (kcal/ mol)	-45.884	-32.247	-21.435	-40.578	-41.700	-38.012

[a] – RMSD was calculated with respect to the corresponding X-Ray structure

Generative neural network methods for protein-protein folding and interactions

Recent breakthroughs have moved past simple "lock-and-key" docking to dynamic, data-driven pose and affinity prediction. Unlike traditional docking, which requires a pre-determined protein structure, various generative methods and co-folding tools predict the protein's 3D structure and the binding poses simultaneously. The appearance of these tools is the most significant paradigm shift in 2024–2025, so that in addition to conventional protein-protein docking methods, we also compared the performance of new generative methods based on neural network and artificial intelligence approaches [27]. A total of four methods, such as AlphaFold3, Boltz-2, Protenix, and Chai-1, were also compared.

The first considered method was AlphaFold 3 from DeepMind, whose authors were awarded the Nobel Prize in Chemistry in 2024. Unlike AlphaFold 2, this model is capable of uniformly predicting the structures of complex biomolecular complexes, including protein-protein, protein-DNA/RNA, and protein-ligand interactions, as well as taking into account ions and chemical modifications [16]. The use of a diffusion approach and updated architecture has significantly improved the accuracy of intermolecular contact modeling, making AlphaFold 3 a powerful tool for fundamental research, structural interpretation of biological processes, and applied problems in modern medicinal chemistry and CADD [14].

As shown in Figure 5, AlphaFold3 reproduced the structures of the first two complexes almost perfectly, with errors relative to the crystal structures of less than 0.5 Å in both cases. The third and fourth complexes were modeled with significantly lower accuracy. However, it should be noted that, unlike classical methods, these structures were generated from scratch, leading to error accumulation, among other things, due to structural differences outside the area of interest.

Similar to previous methods, the results of modeling the non-classical complex represented by PDB: 9LAE system also yielded the highest RSMD error rates.

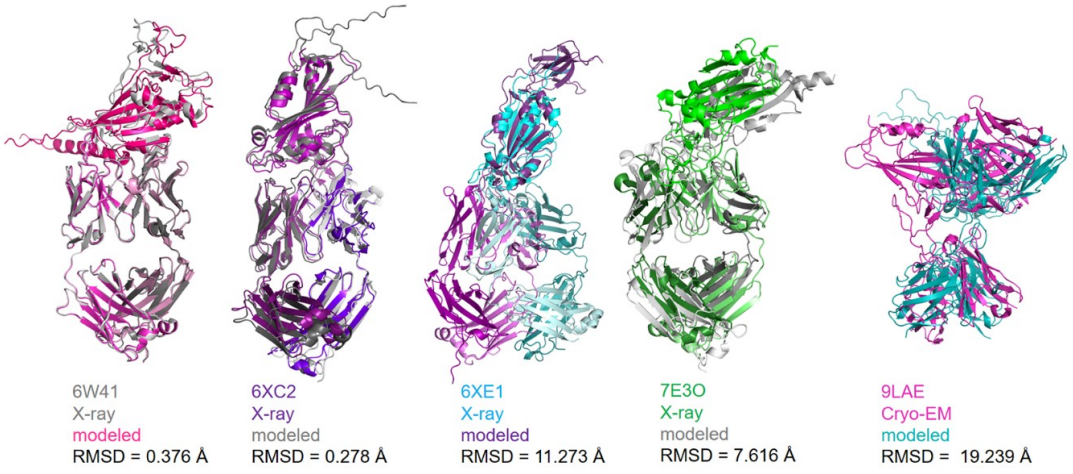

Figure 5. Performance comparison of AlphaFold 3 in reproducing the 3D-structure across the five benchmarked RBD-antibody complexes.

Chai-1 is a multimodal foundation model for molecular structure prediction that achieves state-of-the-art performance across diverse benchmarks, enabling unified modeling of proteins, small molecules, DNA, RNA, glycosylations, and other biomolecular entities [19]. Figure 6 shows the structure-prediction results of the studied systems using this tool. It can be seen that, similarly to AlphaFold 3, the first two structures were reproduced with significantly higher accuracy compared to the others. However, the RMSD errors in the computed results are significant: 11.5 and 4.0 Å. During visual analysis, it becomes clear that the high error values in the modeling of these two systems accumulated due to the effect described above, which is caused by the method's fundamental principles of operation — the generative model. For the next three systems, high RMSD values are a logical consequence of the incorrect reproduction of experimental X-ray structures.

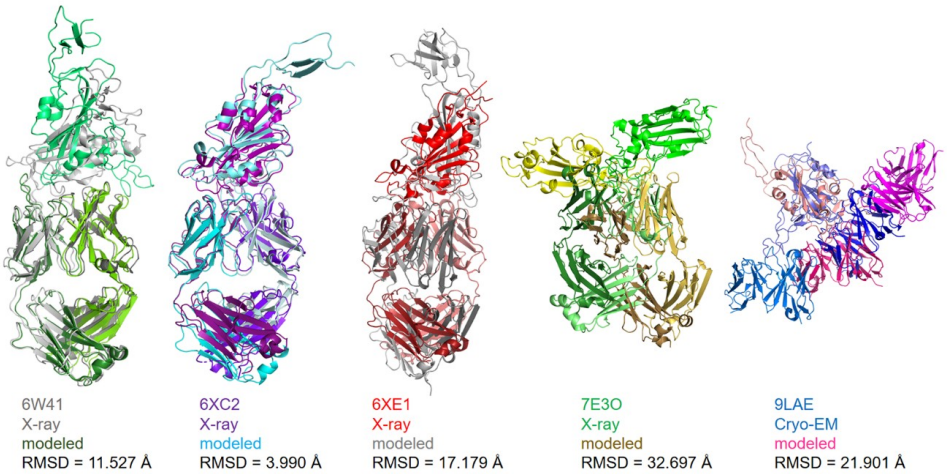

Figure 6. Performance comparison of Chai-1 in reproducing the 3D-structure of the benchmarked RBD-antibody complexes.

Boltz-2 is an advanced AI-based fundamental modeling tool that simultaneously predicts 3D structures of protein-ligand complexes and evaluates their binding affinity in a single GPU-accelerated workflow. It combines structure and affinity in a single process, returning both 3D structural coordinates with confidence metrics and estimates of the probability and strength of molecule binding, making it highly effective for virtual screening, ligand optimization, and drug discovery tasks compared to traditional physical simulations [18]. In addition, Boltz-2 is also capable of modeling protein-protein interactions, protein-RNA interactions, and other interactions. However, for such complex systems, it is not capable of predicting binding probability and efficiency.

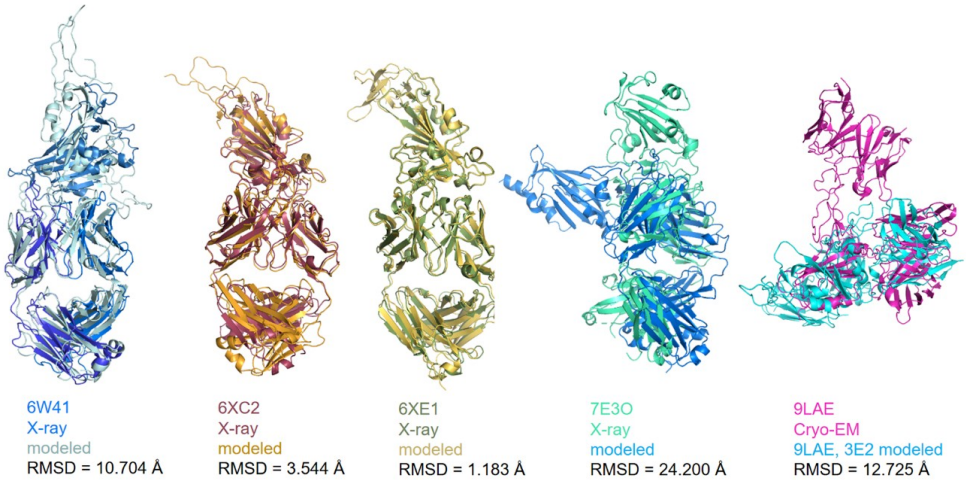

Figure 7. Performance comparison of Boltz-2 in reproducing the 3D-structure of the benchmarked RBD-antibody complexes.

All these features are outlined in Figure 7 for the comparison of the Boltz-2 evaluation of our benchmarked set of the protein-protein complexes. Except of the two structures 6XC2 and 6XE1, the prediction for other complexes suffered from large RMSD exceeding 10 Å.

Protenix is an open-source PyTorch reproduction and extension of AlphaFold 3; it predicts full 3D atomic coordinates for proteins, nucleic acids, small molecules, and their complexes, with support for constraint-guided modeling and lightweight variants, offering reproducible code, model weights, and flexible deployment while achieving strong performance across diverse structure prediction benchmarks.

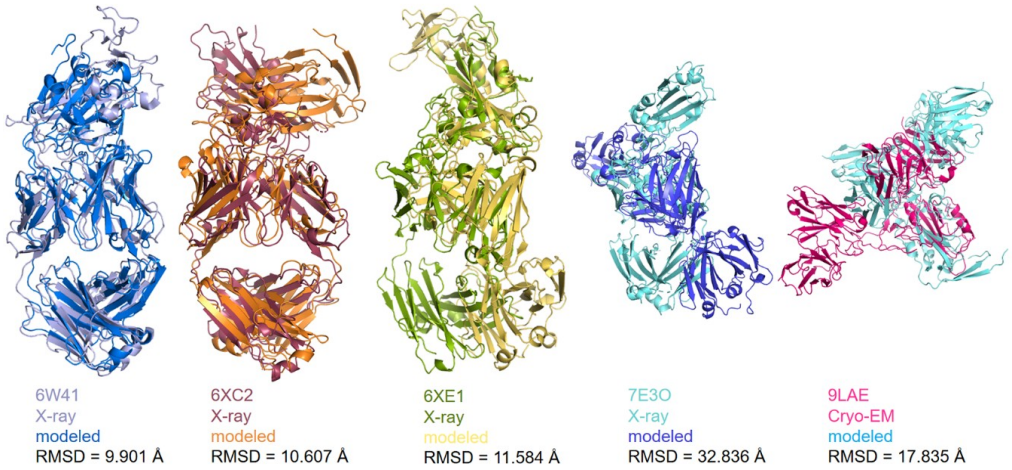

Figure 8. Performance comparison of Protenix in reproducing the 3D-structure across the five studied RBD-antibody complexes.

Backbone RMSD is often known to be dominated by large errors in flexible regions (like loops) and can misrepresent the quality of a well-predicted core or interface. These features are clearly observed in Figures 3-8.

Different generative tools employ distinct internal strategies to estimate prediction confidence, relying on tool-specific confidence, aggregate, or ranking scores. While these scores are designed to reflect model reliability, their definitions and underlying assumptions differ substantially between methods, complicating direct comparison. While internal confidence metrics are generally expected to correlate with structural accuracy, the strength and consistency of this relationship vary across tools and remain insufficiently quantified. To systematically assess how reliably these scores reflect actual model quality (as measured by RMSD), we introduce an independent metric, *Factor V*, that captures the agreement between the ranking-based confidence proposed by the tool and the structure-based real accuracy, as measured by RMSD after aligning the model with an X-ray structure. This unified framework enables an objective evaluation of tools` confidence score reliability, and facilitates cross-tool comparison. The main question is how well internal scoring schemes align with true structural performance.

Therefore, to compare the reliability of tools we defined a unified metric, *Factor V* by Eq. 1, which quantifies the agreement between the confidence derived from the model's predicted ranking score *(C)* and the confidence derived from the normalized RMSD *(R)*. *Factor V* closer to 1 indicates higher reliability, meaning the agreement between promised by the tool quality of the model (ranking score) and real RMSD that we obtained by alignment of the model to X-ray structure.

$$Factor\ V = 1 - |C - R| \tag{1}$$

where:

- C - is the confidence derived from the model's predicted by the tool ranking score.
- R - is the confidence derived from the normalized RMSD, reflecting real structural accuracy.

Both parameter C and R are mapped to a comparable [0, 1] scale using a sigmoid transformation, allowing direct comparison and computation among *Factor V*. The sigmoid function (logistic function), denoted σ, and is defined as (Eq. 2):

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

The sigmoid function (x) maps both the normalized deviation of the ranking score and normalized RMSD into the interval [0, 1], providing comparable confidence measures for evaluating model reliability.

The confidence from Ranking score C is computed as:

$$C = \sigma\left(\frac{S_{\text{rank}} - T_{\text{conf}}}{\alpha}\right) \tag{3}$$

- T_{conf} is the confidence threshold defines the minimal acceptable ranking score. T_{conf} = 0.6 defines the minimum acceptable ranking score for a prediction to be considered high-confidence.
- $\alpha = 0.05$ controls the steepness of the sigmoid function, ensuring a smooth transition around the threshold. α determines the sensitivity of the sigmoid transformation.
- $S_{\text{rank}} = 0.8\text{ipTM} + 0.2\text{pTM}$, ranking score S_{rank} is a weighted combination of ipTM and pTM, reflecting the overall predicted model quality. This weighting emphasizes the importance of inter-chain contacts (ipTM) in evaluating model reliability while still accounting for overall structural accuracy (pTM). pTM (predicted Template Modeling score) is a measure of the predicted overall structural accuracy of a protein model compared to a reference structure. The pTM score ranges from 0 to 1, where higher values indicate better agreement with the expected 3D fold. ipTM (inter-chain predicted Template Modeling score) is a variant of pTM that specifically evaluates inter-chain interactions in multimeric protein complexes. ipTM also ranges from 0 to 1, with higher scores indicating more accurate predicted inter-chain contacts.

C represents the soft confidence obtained from the model's predicted ranking score, normalized relative to the confidence threshold T_{conf} and scaled by α to control the steepness of the sigmoid function.

The RMSD-based confidence (R) is calculated as Eq. 4:

$$R = \sigma\left(\frac{T_{\text{RMSD}} - \text{RMSD}_{\text{norm}}}{\beta}\right) \tag{4}$$

where the normalized RMSD is defined by Eq. 5:

$$\text{RMSD}_{\text{norm}} = \frac{\log(1+\text{RMSD})}{1+\text{RMSD}_{\text{max}}} \tag{5}$$

- RMSD = 35 sets the normalization range.
- T_{RMSD} = 0.5 defines the threshold for structural accuracy.
- β = 0.05 controls the steepness of the sigmoid function for RMSD-based confidence.

$\text{RMSD}_{\text{norm}}$ scales the observed RMSD values logarithmically and normalizes them relative to the maximum observed RMSD, ensuring comparability across proteins. This normalization ensures comparability of RMSD values across proteins and scales them for the sigmoid transformation.

R represents the confidence associated with structural accuracy, obtained by applying a sigmoid transformation to the deviation of normalized RMSD from its threshold T_{RMSD}, scaled by β.

To visualize and evaluate method reliability, a 2×2 confusion matrix was constructed using the thresholds T_{conf} and T_{RMSD}:

	RMSD bad	RMSD good
Ranking score bad	TN	FP
Ranking score good	FN	TP

- Rows correspond to predicted confidence: good (≥0.6) or bad (<0.6).
- Columns correspond to structural accuracy: RMSD ≤ 0.5 (good) or >0.5 (bad).

- Each cell contains predictions in that category, and the mean *Factor V* per cell summarizes how well predicted confidence aligns with actual structural accuracy.

High *Factor V* in the TP cell indicates that high-confidence predictions are structurally accurate, whereas low *Factor V* in FP or FN cells highlights over- or underestimation of reliability.

This framework provides a quantitative and visual assessment of prediction reliability across proteins and allows comparison between different structure prediction methods.

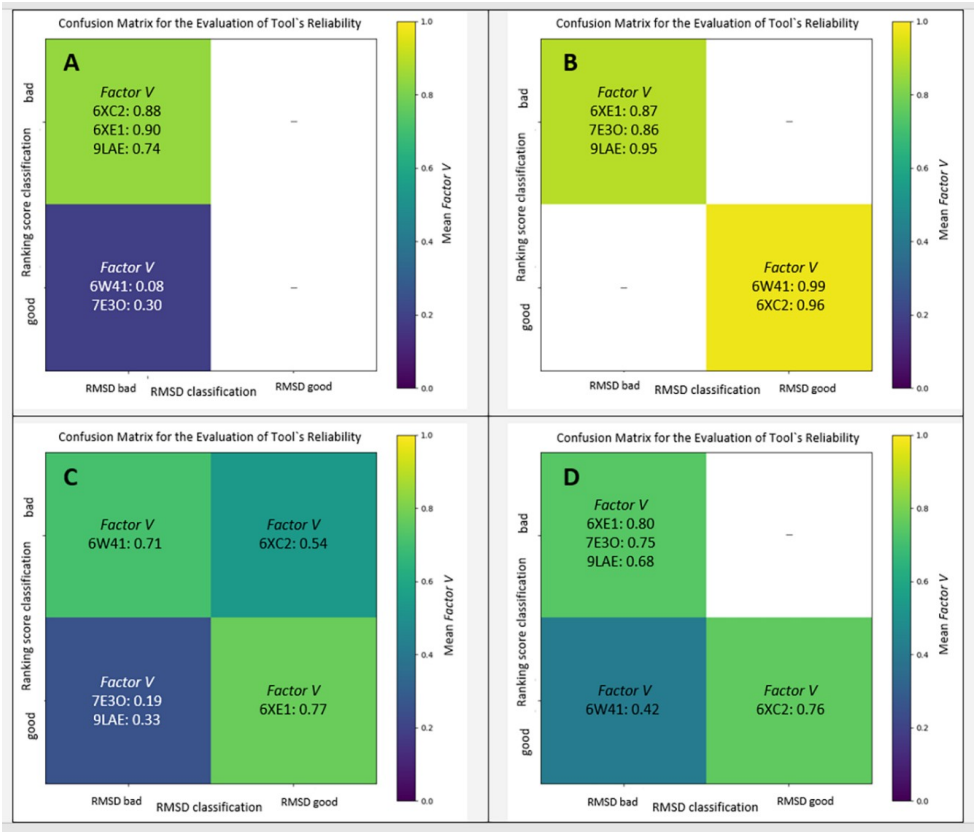

Figure 9. Confusion matrixes for the performance evaluation of four different generative tools. (A) Protenix, (B) AlphaFold 3, (C) Boltz-2, (D) Chai-1.

Figure 9 summarizes the reliability of the four studied AI-driven tools for predicting the 3D-structures of the benchmarked set of RBD-antibody complexes and the agreement between their scores and the quality represented by RMSD. Protenix exhibits a pronounced mismatch between Ranking score and structural accuracy. High *Factor V* values are predominantly observed in the Ranking score bad / RMSD bad quadrant, while structures classified as Ranking score good / RMSD bad show low *Factor V*. This indicates that the internal Ranking score does not reliably reflect true structural quality, limiting its usefulness as a confidence metric. AlphaFold 3 demonstrates strong agreement between Ranking score and RMSD. Nearly all predictions fall into the Ranking score good / RMSD good quadrant with consistently high *Factor V* values, indicating that the Ranking score serves as a robust and reliable proxy for structural accuracy. Boltz-2 shows a broadly distributed pattern across all four quadrants, reflecting partial decoupling between Ranking score and RMSD. Moderate *Factor V* values in mismatched classification cells suggest limited discriminative power of the Ranking score, although some predictive value is retained. Chai-1 displays moderate concordance between Ranking score and RMSD. Elevated *Factor V* values are observed in both correctly and incorrectly classified regions, indicating a tendency toward confident predictions that are not consistently associated with high structural accuracy. Consequently, the Ranking score provides only a partial indicator of model reliability.

All these findings suggest that in terms of *Factor V* metric, AlphaFold 3 over performed all other three models. AlphaFold 3 exhibits the highest consistency between internal confidence metrics and structural accuracy. In contrast, Protenix and Boltz-2 show weaker alignment, while Chai-1 occupies an intermediate position in terms of Ranking score reliability.

Table 2. Computational metrics for predicting the 3D structure of the Spike protein RBD – antibody complexes across five different systems.

Metrics	PDB: 6W41	PDB: 6XC2	PDB: 6XE1	PDB: 7E3O	PDB: 9LAE
Boltz-2					
RMSD (Å)[a]	10.625	11.704	1.183	1.789	12.725
pTM	0.6408	0.6450	0.7272	0.7081	0.7106
ipTM	0.5410	0.5533	0.6435	0.6653	0.6205
Ranking score	0,56	0,57	0,66	0,67	0,64
Factor V	0.71	0.54	0.77	0.19	0.33
Chai-1					
RMSD (Å)[a]	11.481	10.894	17.179	5.317	21.901
pTM	0.7268	0.8366	0.64677	0.6547	0.6565
ipTM	0.59248	0.7817	0.50268	0.51918	0.5392
Ranking score	0,62	0,79	0,53	0,55	0,56
Factor V	0.42	0.76	0.80	0.75	0.68
Protenix					
RMSD (Å)[a]	9.834	32.144	11.584	4.699	17.835
pTM	0.79	0.61	0.6035	0.677	0.6119
ipTM	0.737	0.4866	0.4725	0.6347	0.5312
Ranking score	0,75	0,51	0,50	0,64	0,55
Factor V	0.08	0.88	0.90	0.30	0.74
AlphaFold 3					
RMSD (Å)[a]	0.322	1.869	11.866	7.569	19.331
pTM	0.85	0.79	0.61	0.62	0.53
ipTM	0.82	0.75	0.49	0.53	0.43
Ranking score	0,83	0,76	0,51	0,55	0,45
Factor V	0.99	0.96	0.87	0.86	0.95

[a] – Backbone RMSD was calculated with respect to the corresponding X-Ray structure

Conclusions and Future Perspectives

Understanding protein-protein interactions (PPIs) is essential to advancing the discovery of lab-made antibodies for vaccines and disease treatment. Experimental methods for determining PPIs are often slow and expensive, which has driven significant efforts to develop novel computational approaches in this field. This is why modeling PPIs is becoming an important tool for developing therapies for cancer, autoimmune diseases, and infections [28-30]. In recent years, the theoretical prediction of PPIs has reached a critical turning point. The field has moved from simply predicting "if" two proteins interact to structure-aware co-modeling, where the 3D geometry of the interface is predicted alongside the interaction itself [31-32].

Systematic comparison of reproducibility of five experimentally resolved Spike RBD-antibody structures using different computational approaches revealed that traditional docking tools, such as ClusPro and pyDockWeb, performed well in capturing correct protein-protein interaction mapping for relatively small antibodies with well-defined interaction interfaces, as seen for the cases of 6W41, 6XC2, and 7E3O in Figures 3 and 4. However, rigid-body docking, which uses fixed protein shapes and lacks internal protein flexibility, hampers the performance of these tools for larger three-protein assemblies, such as the case of 9LAE.

Among four AI-driven prediction tools, AlphaFold 3 performs best at reproducing the 3D structures of the studied RBD-antibody complexes. AlphaFold 3 perfectly reproduced the 3D structures of the five RBD-antibody complexes, except for 9LAE. In terms of common RMSD criteria and the integrated *Factor V* metrics (Table 2), AlphaFold 3 outperformed other tools, such as Chai-1 and Protenix, for most of the studied complexes. Surprisingly, Botlz-2, which belongs to a family of models that emerged quickly after AF3's release, performed worse than its parent model, as evidenced by Figure 9 and Table 2. However, it should also be noted that Botlz-2 is not a mere clone of AF3. It was designed primarily for protein-ligand mapping, introducing some architectural innovations [18] that might alter its performance for PPIs.

Among five considered RBD-antibody complexes (see Figure 2), the structure 9LAE imposed some computational challenges for all benchmarked computational tools. While most tools captured the tertiary structure, predicting the correct PPIs remains beyond their current capability.

Finally, our findings suggest that while AI-guided methods are revolutionizing the field by providing astonishingly fast and often accurate predictions, traditional docking tools remain indispensable for hypothesis testing, detailed interaction analysis, and workflow integration where human insight and control are paramount. In last years, the development of SARS-CoV-2 Spike RBD complexes has moved toward variant-proof and computational-first design [8, 21]. Therefore, future perspective of modern computational pipelines of PPIs will combine both, using AI for rapid initial sampling and traditional methods for refinement and scoring.

Authors Contribution

V. O. Morozova and M. V. Prud have contributed equally to this work.

Acknowledgements

A.K. acknowledges the base financial support by the Ministry of Education and Science of Ukraine.

References

1. McDonnell R. T., Henderson A. N., Elcock A. H. Structure prediction of large rnas with AlphaFold 3 highlights its capabilities and limitations. *J. Mol. Biol.* **2024**, *436* (22), 168816. https://doi.org/10.1016/j.jmb.2024.168816
2. Kondo H. X., Takano Y. Structure comparison of heme-binding sites in heme protein predicted by AlphaFold 3 and AlphaFold 2. *Chem. Lett.* **2024**, *53* (8), upae148. https://doi.org/10.1093/chemle/upae148
3. He X.-h., Li J.-r., Shen S.-y., Xu H. E. AlphaFold 3 versus experimental structures: Assessment of the accuracy in ligand-bound G protein-coupled receptors. *Acta Pharm. Sinica* **2025**, *46* 1111–1122. https://doi.org/10.1038/s41401-024-01429-y
4. Binder J. L., Berendzen J., Stevens A. O., He Y., Wang J., Dokholyan N. V., Oprea T. I. AlphaFold illuminates half of the dark human proteins. *Curr. Opin. Struct. Biol.* **2022**, *74* 102372. https://doi.org/10.1016/j.sbi.2022.102372
5. Zhai S., Liu T., Lin S., Li D., Liu H., Yao X., Hou T. Artificial intelligence in peptide-based drug design. *Drug Discov. Today* **2025**, *30* (2), 104300. https://doi.org/10.1016/j.drudis.2025.104300
6. Duma Y., Kyrychenko A. Benchmarking Google DeepMind's AlphaFold 3 performance for protein 3D-structure prediction. *Kharkiv University Bulletin. Chemical Series* **2024**, *43*, 6-25. https://doi.org/10.26565/2220-637X-2024-43-01
7. Yevsieieva L. V., Lohachova K. O., Kyrychenko A., Kovalenko S. M., Ivanov V. V., Kalugin O. N. Main and papain-like proteases as prospective targets for pharmacological treatment of coronavirus SARS-CoV-2. *RSC Adv.* **2023**, *13* (50), 35500–35524. https://doi.org/10.1039/d3ra06479d
8. Chen Y., Zhao X., Zhou H., Zhu H., Jiang S., Wang P. Broadly neutralizing antibodies to SARS-CoV-2 and other human coronaviruses. *Nat. Rev. Immunol.* **2023**, *23* (3), 189-199. https://doi.org/10.1038/s41577-022-00784-3
9. Du L., Yang Y., Zhang X. Neutralizing antibodies for the prevention and treatment of COVID-19. *Cell Mol Immunol* **2021**, *18* (10), 2293-2306. https://doi.org/10.1038/s41423-021-00752-2
10. Kozakov D., Hall D. R., Xia B., Porter K. A., Padhorny D., Yueh C., Beglov D., Vajda S. The ClusPro web server for protein–protein docking. *Nat. Protocols* **2017**, *12* (2), 255-278. https://doi.org/10.1038/nprot.2016.169
11. Brenke R., Hall D. R., Chuang G.-Y., Comeau S. R., Bohnuud T., Beglov D., Schueler-Furman O., Vajda S., Kozakov D. Application of asymmetric statistical potentials to antibody–protein docking. *Bioinformatics* **2012**, *28* (20), 2608-2614. https://doi.org/10.1093/bioinformatics/bts493
12. Jiménez-García B., Pons C., Fernández-Recio J. Pydockweb: A web server for rigid-body protein–protein docking using electrostatics and desolvation scoring. *Bioinformatics* **2013**, *29* (13), 1698-1699. https://doi.org/10.1093/bioinformatics/btt262
13. Abramson J., Adler J., Dunger J., Evans R., Green T., Pritzel A., Ronneberger O., Willmore L., Ballard A. J., Bambrick J., Bodenstein S. W., Evans D. A., Hung C.-C., O'Neill M., Reiman D.,

Tunyasuvunakool K., Wu Z., Žemgulytė A., Arvaniti E., Beattie C., Bertolli O., Bridgland A., Cherepanov A., Congreve M., Cowen-Rivers A. I., Cowie A., Figurnov M., Fuchs F. B., Gladman H., Jain R., Khan Y. A., Low C. M. R., Perlin K., Potapenko A., Savy P., Singh S., Stecula A., Thillaisundaram A., Tong C., Yakneen S., Zhong E. D., Zielinski M., Žídek A., Bapst V., Kohli P., Jaderberg M., Hassabis D., Jumper J. M. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493-500. https://doi.org/10.1038/s41586-024-07487-w

14. Pei J., Andreeva A., Chuguransky S., Lázaro Pinto B., Paysan-Lafosse T., Dustin Schaeffer R., Bateman A., Cong Q., Grishin N. V. Bridging the gap between sequence and structure classifications of proteins with alphafold models. *J. Mol. Biol.* **2024**, *436* (22), 168764. https://doi.org/10.1016/j.jmb.2024.168764
15. Zonta F., Pantano S. From sequence to mechanobiology? Promises and challenges for AlphaFold 3. *Mechanobiol. Med.* **2024**, *2* (3), 100083. https://doi.org/10.1016/j.mbm.2024.100083
16. Weller J. A., Rohs R. Structure-based drug design with a deep hierarchical generative model. *J. Chem. Inf. Model.* **2024**, *64* (16), 6450-6463. https://doi.org/10.1021/acs.jcim.4c01193
17. Wohlwend J., Corso G., Passaro S., Reveiz M., Leidal K., Swiderski W., Portnoi T., Chinn I., Silterra J., Jaakkola T., Barzilay R. Boltz-1 democratizing biomolecular interaction modeling. *bioRxiv* **2024**, 2024.11.19.624167. https://doi.org/10.1101/2024.11.19.624167
18. Passaro S., Corso G., Wohlwend J., Reveiz M., Thaler S., Somnath V. R., Getz N., Portnoi T., Roy J., Stark H., Kwabi-Addo D., Beaini D., Jaakkola T., Barzilay R. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv* **2025**, 2025.06.14.659707. https://doi.org/10.1101/2025.06.14.659707
19. Chai D., Boitreaud J., Dent J., McPartlon M., Meier J., Reis V., Rogozhnikov A., Wu K. Chai-1: Decoding the molecular interactions of life. *bioRxiv* **2024**, 2024.10.10.615955. https://doi.org/10.1101/2024.10.10.615955
20. ByteDance A. M. L. A. I. S. T., Chen X., Zhang Y., Lu C., Ma W., Guan J., Gong C., Yang J., Zhang H., Zhang K., Wu S., Zhou K., Yang Y., Liu Z., Wang L., Shi B., Shi S., Xiao W. Protenix - advancing structure prediction through a comprehensive AlphaFold 3 reproduction. *bioRxiv* **2025**, 2025.01.08.631967. https://doi.org/10.1101/2025.01.08.631967
21. Cofas-Vargas L. F., Mendoza-Espinosa P., Montalvo-Sandoval F. D., Pérez-Rodríguez S., Rauda-Ceja J. A., Hernández-Peralta P., Durán-Vargas A., Trujillo-Roldán M. A., Valdez-Cruz N. A., García-Hernández E. A unified topology-based classification of SARS-CoV-2 RBD neutralizing antibodies systematizes affinity trends across variants. *mAbs* **2025**, *17* (1), 2575083. https://doi.org/10.1080/19420862.2025.2575083
22. Yuan M., Wu N. C., Zhu X., Lee C.-C. D., So R. T. Y., Lv H., Mok C. K. P., Wilson I. A. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, *368* (6491), 630-633. https://doi.org/10.1126/science.abb7269
23. Yuan M., Liu H., Wu N. C., Lee C.-C. D., Zhu X., Zhao F., Huang D., Yu W., Hua Y., Tien H., Rogers T. F., Landais E., Sok D., Jardine J. G., Burton D. R., Wilson I. A. Structural basis of a shared antibody response to SARS-CoV-2. *Science* **2020**, *369* (6507), 1119-1123. https://doi.org/10.1126/science.abd2321
24. Hurlburt N. K., Seydoux E., Wan Y.-H., Edara V. V., Stuart A. B., Feng J., Suthar M. S., McGuire A. T., Stamatatos L., Pancera M. Structural basis for potent neutralization of SARS-CoV-2 and role of antibody affinity maturation. *Nat. Commun.* **2020**, *11* (1), 5413. https://doi.org/10.1038/s41467-020-19231-9
25. Yang M., Li J., Huang Z., Li H., Wang Y., Wang X., Kang S., Huang X., Wu C., Liu T., Jia Z., Liang J., Yuan X., He S., Chen X., Zhou Z., Chen Q., Liu S., Li J., Zheng H., Liu X., Li K., Yao X., Lang B., Liu L., Liao H.-X., Chen S. Structural basis of a human neutralizing antibody specific to the SARS-CoV-2 spike protein receptor-binding domain. *Microbiology Spectrum* **2021**, *9* (2), e01352-21. https://doi.org/10.1128/Spectrum.01352-21
26. Jiang W., Jiang Y., Sun H., Deng T., Yu K., Fang Q., Ge H., Lan M., Lin Y., Fang Z., Zhang Y., Zhou L., Li T., Yu H., Zheng Q., Li S., Xia N., Gu Y. Structural insight into broadening SARS-CoV-2 neutralization by an antibody cocktail harbouring both NTD and RBD potent antibodies. *Emerging Microbes & Infections* **2024**, *13* (1), 2406300. https://doi.org/10.1080/22221751.2024.2406300

27. Prud M., Kyrychenko A. Studying ligand-protein interactions in the era of artificial intelligence: Benchmarking Boltz-1 for 3D-structure prediction of biomolecular complexes. *Kharkiv University Bulletin. Chemical Series* **2025**, (44(67)), 57-71. https://doi.org/10.26565/2220-637X-2025-44-06
28. Ivanov V., Lohachova K., Kolesnik Y., Zakharov A., Yevsieieva L., Kyrychenko A., Langer T., Kovalenko S. M., Kalugin O. N. Recent advances in computational drug discovery for therapy against coronavirus SARS-CoV-2. *ScienceRise: Pharm. Sci.* **2023**, 6(46), 4-24. https://doi.org/10.15587/2519-4852.2023.290318
29. Xia Y., Wang Y., Wang Z., Zhang W. A comprehensive review of molecular optimization in artificial intelligence-based drug discovery. *Quantitative Biology* **2024**, *12*, 15-29. https://doi.org/10.1002/qub2.30
30. Chen W., Liu X., Zhang S., Chen S. Artificial intelligence for drug discovery: Resources, methods, and applications. *Mol. Therapy. Nucl. Acids.* **2023**, *31,* 691-702. https://doi.org/10.1016/j.omtn.2023.02.019
31. Cui J., Yang S., Yi L., Xi Q., Yang D., Zuo Y. Recent advances in deep learning for protein-protein interaction: A review. *BioData Mining.* **2025**, *18* (1), 43. https://doi.org/10.1186/s13040-025-00457-6
32. Hosseini S., Golding G. B., Ilie L. Component puzzle protein–protein interaction prediction. *Briefings in Bioinformatics* **2025**, *26* (6), bbaf685. https://doi.org/10.1093/bib/bbaf685

Received 03.10.2025 *Accepted 17.11.2025* *Published 30.12.2025*

В. О. Морозова*†, М. В. Прудь*†, О. В. Кириченко*‡. Порівняльна оцінка обчислювальних методів для моделювання комплексів рецептор-зв'язуючий домен спайк-білка вірусу SARS-CoV-2 — антитіло.

* Харківський національний університет імені В.Н. Каразіна, Навчально-науковий інституту Хімії, майдан Свободи, 4, Харків, 61022, Україна

† Enamine Ltd., вулиця Вінстона Черчилля, 78, Київ 02094, Україна

‡ Інститут хімії функціональних матеріалів, НТК «Інститут монокристалів» НАН України, пр. Науки 60, Харків, 61072, Україна

Комп'ютерне моделювання взаємодій білок-білок (ВББ) є пріоритетним напрямком досліджень і є ключовим для розуміння більшості клітинних процесів, що дозволяє подолати обмеження складних експериментів. Комп'ютерне прогнозування має велике значення для розробки лікарських засобів, оскільки дозволяє швидко визначати мішені при захворюваннях, створювати терапевтичні агенти, що модулюють ВББ, та моделювати взаємодії у реалістичних умовах клітини. Рецептор-зв'язуючий домен (RBD) Спайк білка вірусу SARS-CoV-2 безпосередньо взаємодіє з людським рецептором ACE2 та ініціює проникнення вірусу в клітину. Вивчення цього домену та його взаємодій з нейтралізуючими антитілами є критично важливим для розуміння імунного захисту, оцінки потенційних загроз від нових варіантів вірусу та спрямування розробки ефективних вакцин і антитіло-орієнтованих терапевтичних засобів. У цій роботі ми порівняли ефективність кількох комп'ютерних методів для моделювання комплексів між RBD Спайк білка та різними нейтралізуючими антитілами. Нещодавні досягнення в комп'ютерній хімії виходять за рамки традиційного білок-білкового докінгу в напрямку різних генеративних методів та інструментів ко-фолдингу, які прогнозують 3D-структуру білка та його пози зв'язування на основі його послідовності. Тому метою роботи було порівняння ефективності традиційних методів докінгу, таких як pyDockWEB та ClusPro, з ефективністю нових, перспективних інструментів на основі штучного інтелекту, таких як AlphaFold 3, Boltz-2, Protenix та Chai-1. Всі ці методи були систематично оцінені на предмет їхньої здатності відтворювати 3D-структури відомих комплексів Spike RBD-антитіло. Показано, що традиційні методи докінгу, такі як ClusPro та pyDockWeb, добре відтворюють білок-білкову взаємодію для відносно малих антитіл з чітко визначеними інтерфейсами взаємодії, але не здатні відтворювати складніші білок-білкові комплекси. AlphaFold 3 показав найкращу відтворюваність 3D-структур п'яти досліджених комплексів RBD-антитіло серед чотирьох розглянутих інструментів прогнозування на основі штучного інтелекту. Наше дослідження проливає світло на розуміння білок-білкових взаємодій та надає практичний посібник для точного моделювання взаємодій вірусного шипоподібного білка RBD з антитілами.

Ключові слова: *фолдінг білка, білок-білкова взаємодія, молекулярний докінг, AlphaFold, штучний інтелект.*

Надіслано до редакції 03.10.2025 *Прийнято до друку 17.11.2025* *Опубліковано 30.12.2025*