
End-to-end accurate and high-throughput modeling of antibody-antigen complexes

Tomer Cohen

School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem, Israel
tomer.cohen13@mail.huji.ac.il

Dina Schneidman-Duhovny

School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem, Israel
dina.schneidman@mail.huji.ac.il

Abstract

Antibodies are produced by the immune system in response to infection or vaccination. While sequencing of the individual antibody repertoire is becoming routine, identifying the antigens they recognize requires costly low-throughput experiments. Even when the antigen is known, epitope mapping is still challenging: experimental approaches are low-throughput and computational ones are not sufficiently accurate. Recently, AlphaFold2 has revolutionized structural biology by predicting highly accurate protein structures and complexes. However, it relies on an evolutionary information that is not available for antibody-antigen interactions. Traditional computational epitope mapping is based on structure modeling (folding) of the antibodies followed by docking the predicted structure to the corresponding antigen. The problem with this sequential approach is that the folding step does not consider the structural changes of the antibody upon antigen binding and the docking step is inaccurate because the antibody is considered rigid. Here, we develop a deep learning end-to-end model, that given an antibody sequence and its corresponding antigen structure can simultaneously perform folding and docking tasks. The model produces the 3D coordinates of the entire antibody-antigen (Ab-Ag) or nanobody-antigen (Nb-Ag) complex, including the side chains. An accurate model is detected among the Top-5 and Top-100 predictions for 28% and 70% of the test set, respectively. In addition to mining antibody repertoires, such a method can have the potential to be used in antibody-based drug design, as well as in the vaccine design.

1 Introduction

Antibodies are part of the large and diverse repertoire of the immune receptors which are behind the specific antigen recognition mechanism [1, 2]. B cell sequencing provides a glimpse into the blood circulating antibody repertoires. However, the antigens and their epitopes remain unidentified. Moreover, antibodies are the most rapidly growing class of human therapeutics for a range of diseases, including cancer or viral infections. Epitope characterization is an important component of therapeutic antibody discovery and analysis of antibody repertoires [3, 4]. However, high-throughput experimental characterization of epitopes for thousands of Ab-Ag complexes remains beyond the reach of current technologies.

Computational methods for modeling Ab-Ag structures from individual components are high-throughput, but frequently suffer from high false positive rate, rarely converging to a single structure [5]. Modeling of Ab-Ag complexes most often proceeds in two separate steps. First, the antibody is folded from its sequence, using tools like comparative modeling, *ab initio* modeling [6, 7] or deep learning [8–11]. Second, the predicted antibody structure from the previous step is docked to an

antigen using shape complementary [12], physico-chemical constraints [13] and more recently deep learning models [14, 15]. There are two main bottlenecks: low accuracy of antibody modeling and Ab-Ag scoring functions. Docking algorithms produce thousands of models, over 99% of which are incorrect. These models are then re-ranked with some scoring function [16] to detect the accurate models. However this approach ignores the fact that antibodies have flexible loops (Complementarity Determining Regions - CDRs) that can change their conformation when binding to an antigen. Moreover, inaccuracies in antibody models introduce further errors in the docked configurations and their scores. Consequently, for Ab-Ag complexes docking algorithms succeed to find and rank an accurate complex among the Top-10 best scoring in about ~10-30% of the cases [5] with further decrease for modeled antibodies.

Recently, deep learning models have been successful in protein structure prediction [17–22]. Most recent structure prediction methods, including AlphaFold2 and RosettaFold, use deep learning models for end-to-end modeling, where the input is a sequence and the output is the 3D structure [21–23, 10]. There are also dedicated methods for antibody modeling that use deep learning, but they do not take the antigen into account [24, 8, 25, 9]. Moreover, AlphaFold2 can be applied for complex structure prediction using AlphaFold-Multimer (AFM) model [26]. However, it achieved only ~10% success rate on the Ab-Ag benchmarks [27] due to the lack of evolutionary information.

Here we address the Ab-Ag structure prediction by integrating the antibody folding task with the docking task. We design an end-to-end model that given an antigen structure and an antibody sequence produces accurate complex models. In contrast to docking, only dozens of models are produced with a much smaller fraction of incorrect ones, resulting in the Top-10 success rate of 37%.

2 Methods

Summary. The input to the network is the antibody sequence and the antigen structure. The network outputs the 3D structure of their complex, including backbone and side chain coordinates (Fig. 1). If no antigen structure is given as an input, only the antibody structure is predicted. The network simultaneously folds the antibody (antibody folding) and determines the orientation of the antigen with respect to the antibody (antigen docking). We achieve transformational invariance for antibody folding by aligning the training set structures on a single representative structure (as in NanoNet[11]). For antigen docking the invariance is accomplished by constructing amino acid reference frame for the antigen (using the N-C α -C atoms) and transforming it to the global reference frame. Reference frames are constructed for all interface amino acids during the training and for all surface accessible amino acids during the inference.

During the training, the network learns to predict (i) coordinates of the heavy and light chains separately (similarly to NanoNet), (ii) the transformation of the light chain with respect to the heavy chain, and (iii) the transformation of the antigen interface amino acid from the global reference frame to the correctly docked position. The loss is calculated as a weighted MSE (equivalent to the RMSD) on the backbone and side chain coordinates for the heavy, light, and antigen chains, separately (A.1). In addition, for antibody we add a term that enforces 3.77Å between consecutive C α atoms and an additional auxiliary loss that enforces that the intermediate layers structure is consistent with distances (A.1).

During inference, the antigen interface amino acids (the epitope) are usually unknown. Therefore, we generate complexes using each solvent accessible amino acid as an interface one, by transforming it to the global coordinate frame and predicting the complex structure. This iteration over amino acids enables us to obtain multiple complex models from the network. This requires us to build a scoring network that can receive a complex model and predict whether it is correct or not.

Network architecture. Each amino acid in the input antibody sequence is represented by a 24-dimensional feature vector consisting of 21 channels for amino acid type one-hot representation (20 standard + unknown), two additional channels indicating the chain to which the amino acid belongs to (heavy or light chain). The last channel is for indicating a placeholder for the light chain transformation. Consequently, each antibody corresponds to a sequence of length 281 consisting of 24-dimensional feature vectors, with 150 and 130 positions reserved for the heavy and light chains, respectively. The additional position is a placeholder for the transformation of the light chain with respect to the heavy chain. The input antigen structure is represented by a 54-dimensional feature vector consisting of 21 channels for amino acid type, one channel with solvent accessible area

normalized to [0,1], one channel indicating the current amino-acid used to construct a global reference frame, and additional 30 channels with backbone (N, C α , C, O) and side chain x,y,z coordinates (C β , and 5 additional atoms that define the χ 1-5 angles). Similarly to antibody representation, we have an additional placeholder for the transformation of the antigen with respect to the antibody. The antigen vector is limited to antigens of length 600. Consequently, each antigen corresponds to a sequence of length 601 consisting of 54-dimensional feature vectors.

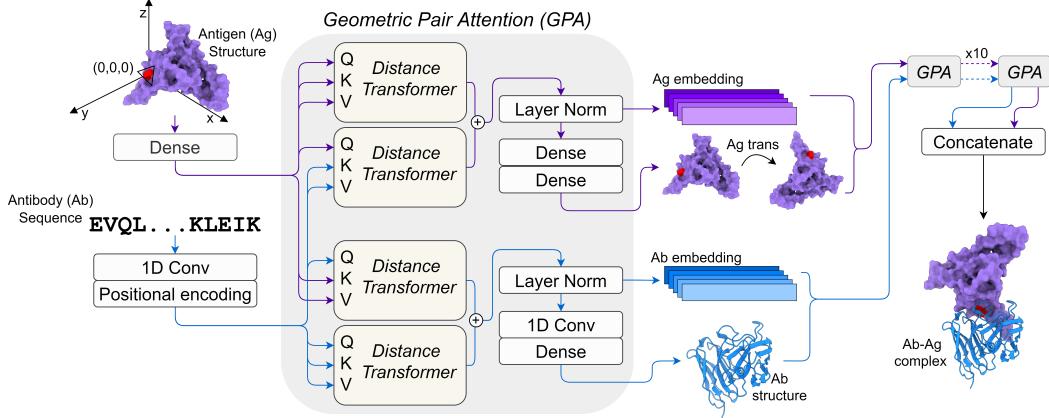


Figure 1: The Network architecture.

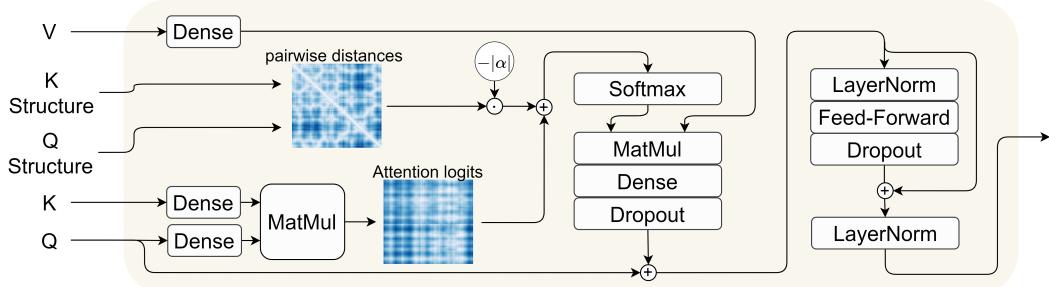


Figure 2: Distance transformer.

The network was designed to simulate the biological Ab-Ag recognition, consisting of four modules that enable inference based on the individual components (the antibody and the antigen) as well as their influence on each other. The network consists of several layers of Geometric Pair Attention (GPA) where each GPA layer integrates the information about antibody structure and antigen orientation from the previous layer using a dedicated Distance Transformer modules (Fig. 2)). Each of the four Distance Transformers is responsible for a different aspect of the Ab-Ag interaction. This design enables the network to pass information regarding the effect the antigen has on the antibody conformation and vice versa. The structural information is also incorporated into the attention mechanism by adding the learned pairwise distances to the attention logits and masking any attention logits of amino-acids with a distance greater than an adjusted threshold. Additionally, by splitting the Distance transformers to the 4 components we can train the network to predict antibody structures without an antigen.

The Distance Transformer module is designed similarly to the Evoformer in the AlphaFold2 [21]. The main principle is to view the problem as a pairwise-distance prediction problem or inference of the graph defined by amino acids. The Distance Transformer consists of 1D and 2D embedding that connect the sequence and pairwise distance representations. The pairwise distance are calculated in each layer based on the intermediate predicted structures and are added as a bias to the attention logits scaled by a learned scalar α . In the first GPA layer, all the pairwise distances are initialized to zeros.

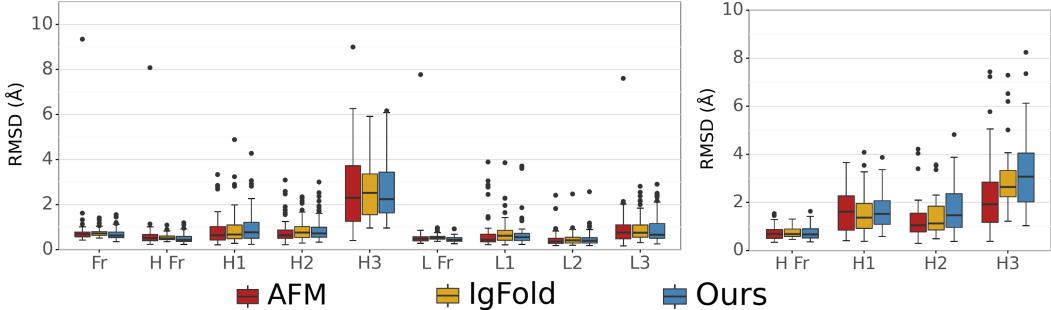


Figure 3: Antibody folding. **A.** RMSD (\AA) of different regions for the IgFold antibody test set (67 structures). **B.** RMSD (\AA) of different regions for the Igfold nanobody test set (21 structures).

Scoring. The input to the scoring network is the Ab-Ag complex generated by the folding and docking network with addition to the original vector representation used as input for the previous network. The transformational invariance is achieved by the antibody alignment as in the folding above. We use a similar architecture as before without changing the antibody structures or the antigen orientation during training. The network is trained to classify complexes into Acceptable quality or incorrect according to CAPRI criteria [28]).

Dataset. The antibody and nanobody structures were obtained from the SabDAB database [29, 30]. We used only structures with a resolution of 3.5\AA or better. We clustered this dataset using a sequence identity threshold of 99% for antibodies, keeping structures with identical sequence only if they appeared unbound and in complex with an antigen. This resulted in 2,823 structures (1,345 Ab-Ag, 380 Nb-Ag complexes, 920 antibody and 178 nanobody unbound structures). We reserved 8% of the data for validation. All the structures were aligned to a reference mAb frame using PyMOL. We used two test sets for evaluating the performance of the folding and docking tasks. The first test set for the evaluation of antibody folding is identical to IgFold [9] test set, containing 67 antibodies and 21 nanobodies (49 and 16 in complex with an antigen, respectively). The structures in this test set were published after the training of AlphaFold2 and IgFold (July 2021). The second test for the evaluation of the docking performance, contains the 65 complexes from the IgFold test set, as well as 66 complexes (54 antibodies and 12 nanobodies) from the extended unbound docking benchmark [31]. The structures from the docking benchmark were published prior to the training of AlphaFold2 and IgFold and therefore could not be used for the antibody folding evaluation. We removed all complexes with antigen length under 100 amino acids. This resulted in a test of 103 complexes (77 and 26 Ab-Ag and Nb-Ag, respectively) for the evaluation of the docking task. For both test sets we used the same antibody sequence identity cut-off as IgFold (99%) to exclude structures from the training set.

3 Results

Antibody folding accuracy. For the evaluation of predicted antibody structures we calculated the backbone RMSD of each region of the antibody (Frame H, H1, H2, H3, Frame L, L1, L2, L3) after alignment of the light or heavy chain frame regions (Table 1 ,2, Fig. 3). We compared our results to AFM and IgFold. For IgFold a single structure was produced, for AFM the best scoring one (pTM score) was evaluated, for our model the best scoring model was evaluated. We obtain high accuracy antibody models that are comparable to AFM and IgFold without relying on additional sequences and MSA. For nanobodies, AFM has a higher accuracy for CDR3 compared to IgFold [9] and our approach. We assume this is due to the small size and the composition of the nanobody test set, because the AlphaFold accuracy was lower for a larger test set [11].

Docking accuracy. Docking models were evaluated according to the DockQ quality measure classified into Incorrect, Acceptable, Medium, and High quality [32]. The success rate was calculated as the fraction of test set complexes with Acceptable, Medium, and High quality models among the top N best scoring models ($N = 1, 5, 10, 20, 50, 100$). The success rate of our method is $\sim 20\%$ when only the Top-1 prediction is considered (Fig. 4A). For Top-10 predictions the success rate is $\sim 37\%$. When Top-100 predictions are considered, our method reaches almost 70% success rate. The success

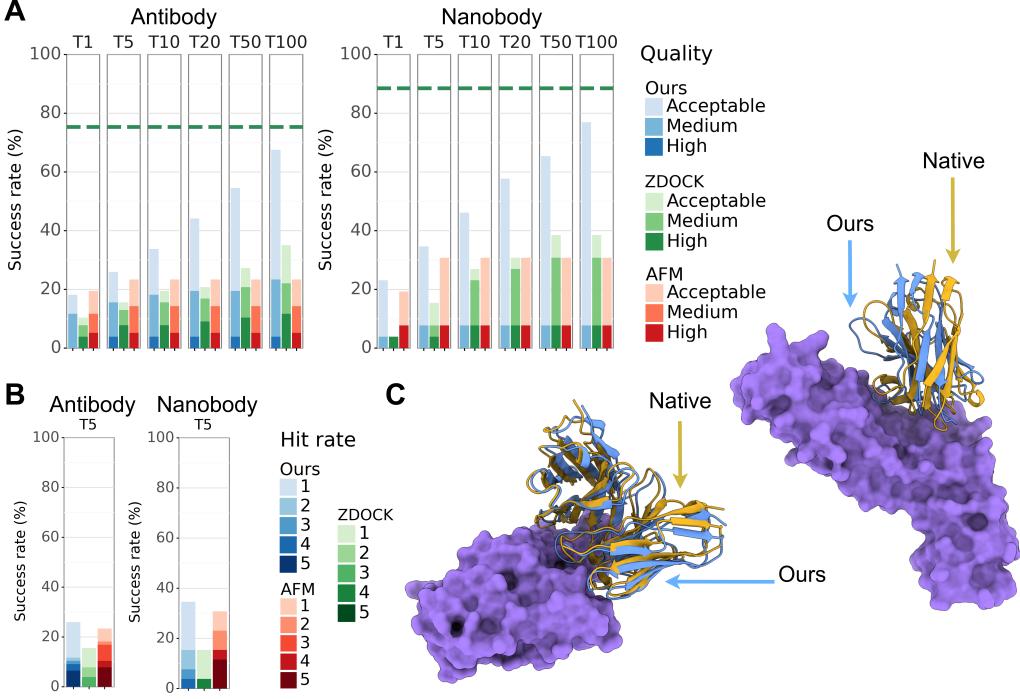


Figure 4: Docking performance for the 103 complexes in the docking test set: 77 Ab-Ag complexes, 26 Nb-Ag complexes. **A.** Success rate, the dotted line corresponds to the percentage of complexes with any acceptable models. **B.** Hit rate of the top-5 predictions. **C.** Top1 structures generated by our model (left - PDB 7s4s, DockQ of 0.78, right - PDB 6dbg, DockQ of 0.48).

Table 1: Antibody folding. Average RMSD (\AA) across frame regions and CDR loops for the antibody folding test set (67 structures).

Method	Fr	H Fr	H1	H2	H3	L Fr	L1	L2	L3
AFM	0.83	0.65	0.88	0.79	2.62	0.60	0.70	0.45	0.96
IgFold	0.75	0.54	0.94	0.87	2.66	0.55	0.77	0.47	0.95
Ours	0.69	0.50	1.01	0.91	2.62	0.46	0.78	0.49	0.95

rate is higher for nanobodies compared to antibodies. Most of the models are of Acceptable and Medium quality, with higher fraction of Medium quality for antibodies compared to nanobodies. In addition, we calculate the hit rate as the number of Acceptable or higher quality models among the Top-5 best scoring models. In \sim 8% of the antibody test set cases, all the Top-5 models are of Acceptable or higher quality (Fig. 4B). We compared our results in the docking task to the five models produced by AFM. The success rate of AFM is lower in the range of 20-31%. However, AFM is producing more models of High quality. This indicates that our models can be further refined when providing AlphaFold our predictions as templates [33]. We have also compared our results to ZDOCK 3.0.2 [34] with antibodies modeled by AFM and found that our performance is \sim 10-40% higher (Fig. 4A).

4 Conclusions

We developed an end-to-end model for accurate and high-throughput prediction of Ab-Ag complexes starting from antibody sequence and antigen structure. The approach is also applicable for Nb-Ag complexes with comparable accuracy and can also work with modeled antigens. The high accuracy was achieved by integrating antibody folding with antigen docking and designing transformationally invariant representation. We expect the method to be applicable in antibody drug design pipelines as well as in mining sequences antibody repertoires.

References

- [1] A. R. Rees. Understanding the human antibody repertoire. *MAbs*, 12(1):1729683, 2020. ISSN 1942-0870 (Electronic) 1942-0862 (Linking). doi: 10.1080/19420862.2020.1729683. URL <https://www.ncbi.nlm.nih.gov/pubmed/32097086>.
- [2] G. Georgiou, G. C. Ippolito, J. Beausang, C. E. Busse, H. Wardemann, and S. R. Quake. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol*, 32(2):158–68, 2014. ISSN 1546-1696 (Electronic) 1087-0156 (Linking). doi: 10.1038/nbt.2782. URL <https://www.ncbi.nlm.nih.gov/pubmed/24441474>.
- [3] Ulrich Reineke. Antibody epitope mapping using arrays of synthetic peptides. In *Antibody Engineering*, pages 443–463. Springer, 2004.
- [4] Johan Rockberg, John Löfblom, Barbara Hjelm, Mathias Uhlén, and Stefan Ståhl. Epitope mapping of antibodies using bacterial surface display. *Nature methods*, 5(12):1039–1045, 2008.
- [5] Francesco Ambrosetti, Brian Jiménez-García, Jorge Roel-Touris, and Alexandre MJJ Bonvin. Modeling antibody-antigen complexes by information-driven docking. *Structure*, 28(1):119–129, 2020.
- [6] B. D. Weitzner, D. Kuroda, N. Marze, J. Xu, and J. J. Gray. Blind prediction performance of rosettaantibody 3.0: grafting, relaxation, kinematic loop modeling, and full cdr optimization. *Proteins*, 82(8):1611–23, 2014. ISSN 1097-0134 (Electronic) 0887-3585 (Linking). doi: 10.1002/prot.24534. URL <https://www.ncbi.nlm.nih.gov/pubmed/24519881>.
- [7] G. Lapidoth, J. Parker, J. Prilusky, and S. J. Fleishman. Abpredict 2: a server for accurate and unstrained structure prediction of antibody variable domains. *Bioinformatics*, 35(9):1591–1593, 2019. ISSN 1367-4811 (Electronic) 1367-4803 (Linking). doi: 10.1093/bioinformatics/bty822. URL <https://www.ncbi.nlm.nih.gov/pubmed/30951584>.
- [8] Jeffrey A. Ruffolo, Jeremias Sulam, and Jeffrey J. Gray. Antibody structure prediction using interpretable deep learning. *Patterns*, 3(2):100406, 2022. ISSN 2666-3899. doi: <https://doi.org/10.1016/j.patter.2021.100406>. URL <https://www.sciencedirect.com/science/article/pii/S2666389921002804>.
- [9] Jeffrey A. Ruffolo, Lee-Shin Chu, Sai Pooja Mahajan, and Jeffrey J. Gray. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *bioRxiv*, 2022. doi: 10.1101/2022.04.20.488972. URL <https://www.biorxiv.org/content/early/2022/04/21/2022.04.20.488972>.
- [10] Brennan Abanades, Guy Georges, Alexander Bujotzek, and Charlotte M Deane. Ablooper: Fast accurate antibody cdr loop structure prediction with accuracy estimation. *bioRxiv*, 2021.
- [11] Tomer Cohen, Matan Halfon, and DIna Schneidman-Duhovny. Nanonet: Rapid and accurate end-to-end nanobody modeling by deep learning. *Frontiers in Immunology*, page 4319, 2022.
- [12] D. Duhovny, R. Nussinov, and H.J. Wolfson. Efficient unbound docking of rigid molecules. In R. Guigó and D. Gusfield, editors, *Second International Workshop, WABI 2002*, volume 2452/2002 of *Lecture Notes in Computer Science*, pages 185–200. Springer Berlin / Heidelberg, 2002. doi: 10.1007/3-540-45784-4.
- [13] Stephen R Comeau, David W Gatchell, Sandor Vajda, and Carlos J Camacho. Cluspro: a fully automated algorithm for protein–protein docking. *Nucleic acids research*, 32(suppl_2):W96–W99, 2004.
- [14] Xiao Wang, Genki Terashi, Charles W Christoffer, Mengmeng Zhu, and Daisuke Kihara. Protein docking model evaluation by 3d deep convolutional neural networks. *Bioinformatics*, 36(7):2113–2118, 2020.
- [15] Xiao Wang, Sean T Flannery, and Daisuke Kihara. Protein docking model evaluation by graph neural networks. *Frontiers in Molecular Biosciences*, page 402, 2021.
- [16] Guang Qiang Dong, Hao Fan, Dina Schneidman-Duhovny, Ben Webb, and Andrej Sali. Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics*, 29(24):3158–3166, 2013.
- [17] S. Wang, S. Sun, Z. Li, R. Zhang, and J. Xu. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol*, 13(1):e1005324, 2017. ISSN 1553-7358 (Electronic) 1553-734X (Linking). doi: 10.1371/journal.pcbi.1005324. URL <http://www.ncbi.nlm.nih.gov/pubmed/28056090>.
- [18] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Zidek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/s41586-019-1923-7. URL <https://www.ncbi.nlm.nih.gov/pubmed/31942072>.

- [19] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.
- [20] M. AlQuraishi. Machine learning in protein structure prediction. *Curr Opin Chem Biol*, 65:1–8, 2021. ISSN 1879-0402 (Electronic) 1367-5931 (Linking). doi: 10.1016/j.cbpa.2021.04.005. URL <https://www.ncbi.nlm.nih.gov/pubmed/34015749>.
- [21] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 2021. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/s41586-021-03819-2. URL <https://www.ncbi.nlm.nih.gov/pubmed/34265844>.
- [22] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, page eabj8754, 2021. doi: 10.1126/science.abj8754. URL <https://science.sciencemag.org/content/sci/early/2021/07/19/science.abj8754.full.pdf>.
- [23] Mohammed AlQuraishi. End-to-end differentiable learning of protein structure. *Cell systems*, 8(4):292–301, 2019.
- [24] J. A. Ruffolo, C. Guerra, S. P. Mahajan, J. Sulam, and J. J. Gray. Geometric potentials from deep learning improve prediction of cdr h3 loop structures. *Bioinformatics*, 36(Suppl):268–275, 2020. ISSN 1367-4811 (Electronic) 1367-4803 (Linking). doi: 10.1093/bioinformatics/btaa457. URL <https://www.ncbi.nlm.nih.gov/pubmed/32657412>.
- [25] Deniz Akpinaroglu, Jeffrey A. Ruffolo, Sai Pooja Mahajan, and Jeffrey J. Gray. Improved antibody structure prediction by deep learning of side chain conformations. *bioRxiv*, 2021. doi: 10.1101/2021.09.22.461349. URL <https://www.biorxiv.org/content/early/2021/09/22/2021.09.22.461349>.
- [26] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *BioRxiv*, pages 2021–10, 2022.
- [27] Rui Yin, Brandon Y Feng, Amitabh Varshney, and Brian G Pierce. Benchmarking alphafold for protein complex modeling reveals accuracy determinants. *Protein Science*, 31(8):e4379, 2022.
- [28] Raúl Méndez, Raphaël Leplae, Marc F Lensink, and Shoshana J Wodak. Assessment of capri predictions in rounds 3–5 shows progress in docking procedures. *Proteins: Structure, Function, and Bioinformatics*, 60(2):150–169, 2005.
- [29] James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi, and Charlotte M Deane. Sabdab: the structural antibody database. *Nucleic acids research*, 42(D1):D1140–D1146, 2014.
- [30] Constantin Schneider, Matthew IJ Raybould, and Charlotte M Deane. Sabdab in the age of biotherapeutics: updates including sabdab-nano, the nanobody structure tracker. *Nucleic acids research*, 50(D1):D1368–D1372, 2022.
- [31] Johnathan D. Guest, Thom Vreven, Jing Zhou, Iain Moal, Jeliazko R. Jeliazkov, Jeffrey J. Gray, Zhiping Weng, and Brian G. Pierce. An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. *Structure*, 29(6):606–621.e5, 2021. ISSN 0969-2126. doi: <https://doi.org/10.1016/j.str.2021.01.005>. URL <https://www.sciencedirect.com/science/article/pii/S0969212621000058>.
- [32] Sankar Basu and Björn Wallner. Dockq: a quality measure for protein-protein docking models. *PloS one*, 11(8):e0161879, 2016.
- [33] Usman Ghani, Israel Desta, Akhil Jindal, Omeir Khan, George Jones, Nasser Hashemi, Sergey Kotelnikov, Dzmitry Padhorny, Sandor Vajda, and Dima Kozakov. Improved docking of protein models by a combination of alphafold2 and cluspro. *BioRxiv*, pages 2021–09, 2022.

- [34] Brian G Pierce, Yuichiro Hourai, and Zhiping Weng. Accelerating protein docking in zdock using an advanced 3d convolution library. *PLoS one*, 6(9):e24657, 2011.

A Appendix

A.1 Loss function Definition

For the training of the network we defined the following loss function:

$$\mathcal{L} = \mathcal{L}_{AbStruct} + \mathcal{L}_{C\alpha} + 0.1 \cdot \mathcal{L}_{LightOrient} + 0.003 \cdot \mathcal{L}_{Dock} + 0.5 \cdot \mathcal{L}_{Aux} \quad (1)$$

where:

$$\mathcal{L}_{AbStruct}(y_{ab}, \hat{y}_{ab}) = MSE(y_{ab}, \hat{y}_{ab}) \quad (2)$$

Is the loss term regarding the antibody heavy and light chains structure. For each chain it is defined as the mean-squared-error (MSE) between the true 3D coordinates of the chain and the predicted coordinates by the model. Each of the chains is aligned to a reference frame region. This term is equivalent to the RMSD between the individual chains.

$$\mathcal{L}_{C\alpha}(\hat{y}_{ab}) = \frac{1}{N-1} \sum_{i=1}^{N-1} \text{Max}(|d(C\alpha_i, C\alpha_{i+1}) - 3.77| - 0.051, 0) \quad (3)$$

Is the loss term that optimizes the distance between consecutive $C\alpha$ atoms to 3.77 Å. Where $C\alpha_i$ are the 3D coordinates of the $C\alpha$ atom in the i amino acid of the predicted structure, and d is Euclidean distance. A standard deviation of 0.0051 Å was used.

$$\mathcal{L}_{LightOrient}(y_{abL}, \hat{y}_{abL}) = MSE(y_{abL}, \hat{y}_{abL}) \quad (4)$$

Is the loss term that optimizes the light chain orientation with respect to the heavy chain. It is defined as the MSE between the true 3D coordinates of the light chain and the predicted coordinates of the model after applying the light chain transformation predicted by the model.

$$\mathcal{L}_{Dock}(y_{ag}, \hat{y}_{ag}) = MSE(y_{ag}, \hat{y}_{ag}) \quad (5)$$

Is the loss term regarding the antigen orientation. It is defined as the MSE between the true 3D coordinates of the solved antigen interface and the predicted interface coordinates of the model. This term is equivalent to the interface RMSD.

$$\mathcal{L}_{Aux} = \mathcal{L}_{Aux}[\mathcal{L}_{AbStruct}] + \mathcal{L}_{Aux}[\mathcal{L}_{LightOrient}] + \mathcal{L}_{Aux}[\mathcal{L}_{Dock}] \quad (6)$$

where:

$$\mathcal{L}_{Aux}[f] = \sum_{i=1}^{L-1} 2^{i-L} \cdot f(y, \hat{y}_i) \quad (7)$$

L is the total number of GPA layers in the model, y is the true output relevant for this loss function and \hat{y}_i is the i 'th intermediate layer output. This term is enforces the model to have a structure representation of the complex in the intermediate layers as well so we can use the intermediate complex structure to calculate the pairwise distances inserted to the Distance Transformer.

A.2 Training setup

The docking and folding network was trained for 150 epochs with a batch size of 16 using a model checkpoint on the total validation loss. We used AdamW optimizer and an initial learning rate of 0.000025 with a cosine decay until 0.0001. In order to avoid bias of the network towards antibody structures that have more interface residues (and therefore represented more in the training set) we

used a sample weight of $\frac{1}{\#contacts}$ on each loss term regarding antibody structure (heavy and light chains, $C\alpha$ distance, light orientation).

The scoring network was trained for 20 epochs with a batch size of 64 using a model checkpoint on the validation F1 score. We used AdamW optimizer and an initial learning rate of 0.0005 with a cosine decay until 0.0001. For each class i we used a class weight equal to $\frac{\#total}{2 \cdot \#class_i}$

A.3 Scoring step

For the scoring task, we fed to our folding and docking network each of the residues of every Ab-At complex in the training and validation sets. To enrich the model with more positive samples and to avoid cases where there are no positive labels for some antibody, we generated additional positive samples for each Ab-Ag complex by generating a small random transformation for the antigen when bound in the right orientation.

Table 2: Nanobody folding. Average RMSD (\AA) across frame regions and CDR loops for the nanobody folding test set (21 structures).

Method	H Fr	H1	H2	H3
AFM	0.60	1.63	1.05	2.16
IgFold	0.65	1.63	1.48	3.04
Ours	0.60	1.75	1.81	3.34

Table 3: Docking success rate (%) for the 103 structures in the docking test set.

Method	Ab				Nb			
	T1	T5	T10	T100	T1	T5	T10	T100
AFM	19.48	23.37	23.37	23.37	19.23	30.77	30.77	30.77
ZDOCK	10.39	15.58	19.48	35.06	3.85	15.38	26.92	38.46
Ours	18.18	25.97	33.77	67.53	23.08	34.62	46.15	76.92