ABodyBuilder3: Improved and scalable antibody structure predictions

Henry Kenlay [1] **Frédéric A. Dreyer** [1] **Daniel Cutting** [1] **Daniel Nissley** [1] **Charlotte M. Deane** [1 2]

Abstract

Accurate prediction of antibody structure is a central task in the design and development of monoclonal antibodies, notably to understand both their developability and their binding properties. In this article, we introduce ABodyBuilder3, an improved and scalable antibody structure prediction model based on ABodyBuilder2. We achieve a new state-of-the-art accuracy in the modelling of CDR loops by leveraging language model embeddings, and show how predicted structures can be further improved through careful relaxation strategies. Finally, we incorporate a predicted Local Distance Difference Test into the model output to allow for a more accurate estimation of uncertainties.

1. Introduction

Immunoglobulin proteins play a key role in the active immune system, and have emerged as an important class of therapeutics (Lu et al., 2020). They are constructed from two heavy and two light chains, separated into distinct domains. The tip of each of the two antibody binding arms is defined as the variable region, and contains six complementarity-determining regions (CDRs) across the heavy and light chains which make up most of the antigen-binding site. As part of an immune response, B cells undergo clonal expansion, which, coupled with somatic hypermutations and recombinations, leads to an accumulation of mutations in the DNA encoding the CDR loops. The remaining domains compose the constant region and are primarily involved in effector function.

Understanding the three-dimensional structure of antibodies is critical to assessing their properties (Chungyoun & Gray, 2023) and developability (Raybould et al., 2019; 2024). The framework regions connecting the CDR loops are relatively conserved and thus easily predicted from sequence similarity. Similarly, five of the CDR loops tend to cluster along canonical forms (Adolf-Bryfogle et al., 2014; Wong et al., 2019) and are thus relatively straightforward to model. The third loop of the heavy chain (CDRH3), for which the coding sequence is created during the recombination of the V, D, and J gene segments (Roth, 2014), is however more challenging due to its much larger sequence and length diversity. As the CDRH3 loop often drives antigen recognition, e.g. (Narciso et al., 2011; Tsuchiya & Mizuguchi, 2016), improving the accuracy with which its structure can be predicted from sequence is a key component to advancing rational antibody design.

Experimental protein structure determination remains a costly and slow process (Slabinski et al., 2007), such that only a small fraction of known antibody sequences have experimentally resolved three-dimensional structural information (Dunbar et al., 2013b; Schneider et al., 2022). One approach to circumvent these experimental limitations is through structure prediction methods, which have had immense success in reaching experimental accuracy on general protein structures (Jumper et al., 2021a; Baek et al., 2021; Lin et al., 2023).

Structure models are also a necessary element to accurately predict biophysical properties of proteins and advance the field of rational therapeutic design. Several dedicated tools have emerged to model specifically the variable region of antibodies. Among them are IgFold (Ruffolo et al., 2023), which is based on a language model, DeepAb (Ruffolo et al., 2022), which uses an attention mechanism, ABlooper (Abanades et al., 2022a), which predicts backbone atom positions using a graph neural network, ABodyBuilder (Leem et al., 2016), a homology modelling pipeline, and ABodyBuilder2 (Abanades et al., 2022b), which uses a modified version of the AlphaFold-Multimer architecture (Evans et al., 2022). More recently, xTrimoPGLM-Ab (Chen et al., 2023) has shown promising results on antibody structures by combining a General Language Model framework (Zeng et al., 2023) with a modified AlphaFold2 architecture.

In this article, we introduce ABodyBuilder3, an antibody structure prediction model based on ABodyBuilder2 (Abanades et al., 2022b). As shown in Figure 1, ABodyBuilder2 consists of an embedding representation of

[1]Exscientia, Oxford Science Park, Oxford, OX4 4GE, UK [2]Department of Statistics, University of Oxford, Oxford OX1 3LB, UK. Correspondence to: Henry Kenlay <hkenlay@exscientia.co.uk>.

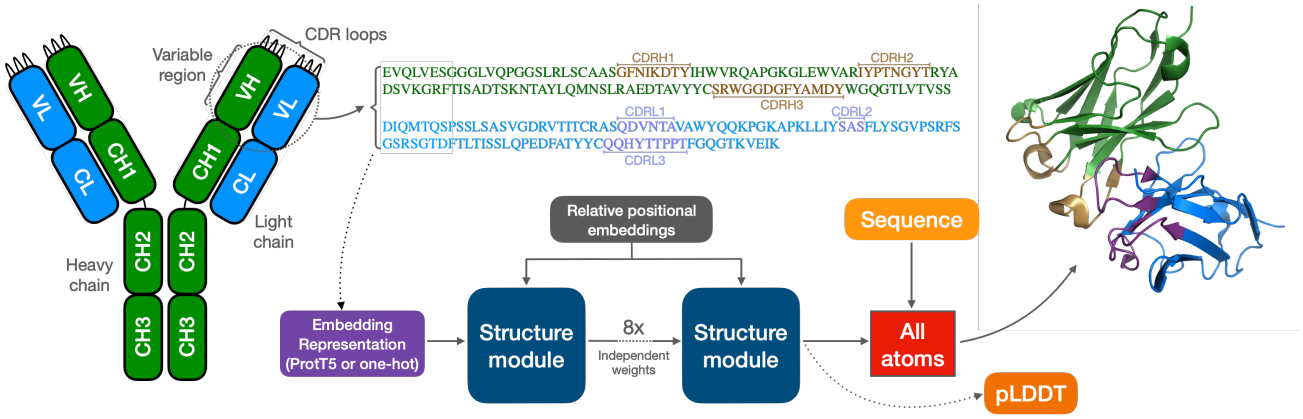

Figure 1. Left: Overview of an antibody structure, with the variable region and CDR loops shown. Right: Schematic representation of the ABodyBuilder3 architecture, with 8 sequential and independent update blocks providing the final atomic coordinates and uncertainty predictions from an embedding representation of the variable region sequence.

the variable region sequence, which is provided as input to a sequence of eight structure modules that update the node features and residue coordinates through an invariant point attention layer and a backbone update layer, respectively. We detail key changes to the implementation, data curation, sequence representation and structure refinement that improve the scalability and accuracy of the model. Additionally, we introduce an uncertainty estimation based on the predicted local-distance difference test (pLDDT), which outperforms the previous ensemble-based estimate. Together, these enhancements provide a substantial improvement in the quality of antibody structure predictions and open the possibility of a scalable and precise assessment of large numbers of therapeutic candidates.

2. Improved structure modelling and evaluation

Rapid prototyping is paramount to generating insights and improving the design of machine learning models. We develop an efficient and scalable implementation of the ABodyBuilder2 architecture which makes use of vectorisation to improve hardware utilisation, leveraging optimisations from the OpenFold project (Ahdritz et al., 2022). This is in contrast to the implementation of ABodyBuilder2, which generates minibatch gradients by computing a gradient for each minibatch sample sequentially before averaging (i.e. accumulated gradients). ABodyBuilder2 also uses double precision which is not well optimised on modern GPU hardware compared to lower precision data-types. We find the model can be trained robustly using bfloat16 precision for weights and use mixed precision for training, providing faster computational throughput and an efficient memory footprint. Our implementation is more than three times faster, and can be scaled easily across multiple GPUs using a distributed data parallel strategy.

We use the Structural Antibody Database (SAbDab) (Dunbar et al., 2014), a dataset of experimentally resolved antibody structures, to train our models on all available data up to January 2024. We perform an initial filtering to remove nanobodies, structures with resolution above 3.5Å, and outliers more than 3.5 standard deviations from the mean for any of the six summary statistics given by ABangle (Dunbar et al., 2013a). Furthermore, we filter out ultra-long CDRH3 loops, which predominantly come from bovine antibodies (de Los Rios et al., 2015) by removing any sequence with a CDRH3 of over 30 residues. We label residues using IMGT numbering generated via ANARCI (Dunbar & Deane, 2016). In an attempt to remove potential structural outliers, we also remove antibodies from species which occur less than 15 times in SAbDab.

Following ABodyBuilder2, we train ABodyBuilder3 in two stages. The first stage uses a dropout of 10%, whereas the second stage has zero dropout and adds a structural violation loss component based on AlphaFold2. For both the first and second stage of training we select weights based on the lowest validation loss. We use a validation set of 150 structures and a test set of 100 structures, which are significantly larger than those used in ABodyBuilder2 and lead to a more robust assessment of model capabilities. We retain the original ABodyBuilder2 test set of 34 structures as a subset of our test set to allow for direct comparisons. As a primary interest is the modelling of antibodies with high humanness in the context of therapeutic antibody development, we constrain the validation and additional test structures to be annotated as human. We require a resolution below 2.5Å and a CDRH3 length of less than 22 for the selection of our validation structures. Furthermore, we

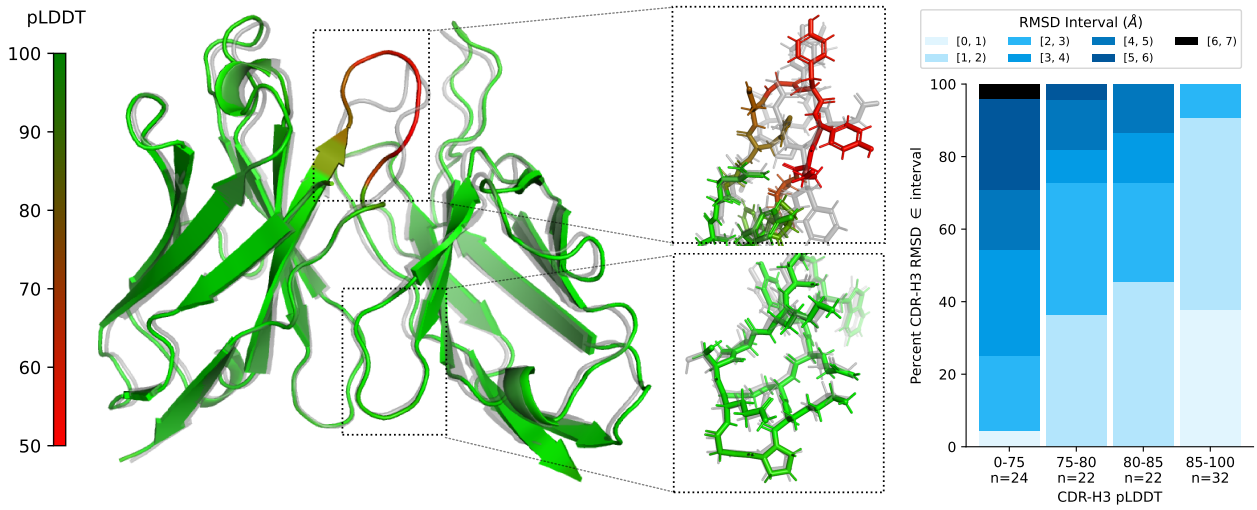

Figure 2. Left: Structure predicted by ABodyBuilder3, with colouring indicating the pLDDT uncertainty estimate. The ground truth (7T0J) is shown in grey. Right: Distribution of CDRH3 RMSD across different bins of the CDRH3 pLDDT score.

remove any structures from the training data that share an identical sequence in any of the CDR regions with any of the validation or test sets.

We consider two physics-based refinement strategies, OpenMM (Eastman et al., 2017) and YASARA (Krieger & Vriend, 2015), to fix stereochemical errors and provide realistic structures. We find that minimization in the YASARA2 forcefield (Krieger et al., 2009) in explicit water leads to improved accuracy of all regions, particularly in the framework. Further details and comparisons between the minimization methods are given in Appendix A.

In Table 1, the first three rows show a comparison of the original ABodyBuilder2 model with our baseline model obtained with our improved implementation and dataset curation. We give the root mean squared deviation (RMSD[1]) for each region of the variable domain, and provide results with both refinement strategies. Note here that ABodyBuilder2 predictions are obtained by taking the closest structure to the mean of an ensemble of four models. This ensemble of models is selected from ten distinct trainings of which six models are then discarded. By comparison, our baseline consists of a single model without any need for model selection and ensemble prediction.

3. Language model representation

Inspired by the success of language model embeddings being used to model protein structure, e.g. (Lin et al., 2023; Ruffolo et al., 2023), we experiment with replacing the one-hot-encoding used as the residue representation in ABodyBuilder2 with a language model embedding. We use the ProtT5 model (Elnaggar et al., 2021), an encoder-decoder text-to-text transformer model (Raffel et al., 2020) pretrained on billions of protein sequences, to generate a residue level embedding of each antibody. As this language model was trained on single chains, we embed the heavy and light chain separately by passing them through the ProtT5 model, and concatenate their residue representations to obtain a per-residue embedding of the full variable region. We also explored antibody-specific models such as the paired IgT5 and IgBert models (Kenlay et al., 2024b), but ultimately found that general protein language models achieved higher performance. This might be because antibody language models introduce potential dataset contamination and overfitting during the language model pre-training. To train our language model-based structure prediction model, all parameters are kept identical to ABodyBuilder2, except for a lower initial learning rate of $5 \cdot 10^{-4}$, and a reduction of the minimum learning rate to 0 in the scheduler, which we found to improve stability on learning rate resets.

In Table 1, we show the performance of our ABodyBuilder3 model, comparing the one-hot encoding with the ProtT5 embedding representation which we denote as ABodyBuilder3-LM. One can observe a small reduction in RMSD using the language model representation, notably in the modelling of the CDRH3 and CDRL3 loops.

4. Uncertainty estimation

The ABodyBuilder2 model uses an ensemble of four models to provide a confidence score from the diversity between predictions. This approach has an increased computational burden, as multiple models are required both at

[1]The RMSD is computed using backbone atoms after aligning each chain independently.

	CDRH1	CDRH2	CDRH3	Fw-H	CDRL1	CDRL2	CDRL3	Fw-L
ABodyBuilder2	0.84	0.73	2.54	0.56	0.55	0.36	0.88	0.53
Baseline (OpenMM)	0.92	0.75	2.53	0.60	0.67	0.35	0.96	0.58
Baseline (Yasara)	0.90	0.74	2.49	0.59	0.58	0.37	0.92	0.57
ABodyBuilder3	0.87	0.70	2.42	0.58	0.61	0.39	0.93	0.58
ABodyBuilder3-LM	0.87	0.75	2.40	0.57	0.59	0.37	0.89	0.58

Table 1. Modelling accuracy as measured by mean RMSD in Angstroms, given for each CDR loop and framework region. Baseline models refer to a single model with our optimised implementation of the ABodyBuilder2 architecture trained on our updated curated dataset whereas ABodyBuilder2 refers to the original version which uses an ensemble of models and the prediction closest to the mean.

	CDRH1	CDRH2	CDRH3	Fw-H	CDRL1	CDRL2	CDRL3	Fw-L
ABodyBuilder2	0.41	0.38	0.57	0.50	0.47	0.48	0.72	0.40
ABodyBuilder3	0.58	0.26	0.61	0.48	0.60	0.20	0.68	0.67
ABodyBuilder3-LM	0.69	0.36	0.73	0.39	0.72	0.52	0.68	0.58

Table 2. Pearson correlation between average uncertainty prediction for a region and the corresponding mean RMSD. Uncertainties for ABodyBuilder2 are derived from an ensemble of four models, while all ABodyBuilder3 uncertainties are directly predicted by a pLDDT head.

training and inference time. We instead estimate the intrinsic model accuracy by predicting the per-residue lDDT-Cα scores (Mariani et al., 2013), as implemented in the AlphaFold2 model (Jumper et al., 2021b). This introduces a very small increase in the number of parameters, but circumvents the need for an ensemble of models. The pLDDT is obtained from the final single representation of the structure module and predicts a projection into 50 bins by a multilayer perceptron with softmax activation. Training is achieved by discretising the predicted structure with per-residue lDDT-Cα against the ground truth structure and computing the cross-entropy loss, which is added to the original ABodyBuilder2 loss with a weight of 0.01. A pLDDT score for the full variable domain, or for specific regions, is obtained as an average of the corresponding per-residue pLDDT scores.

In Table 2, we give the Pearson correlation between the pLDDT score and the RMSD, averaged over each region of the variable domain. The ABodyBuilder2 model, with an uncertainty score obtained from the ensemble model, has lower correlation with RMSD than our pLDDT score. It is interesting to note here that the ABodyBuilder3-LM model, which uses ProtT5 embeddings as input, achieves a higher correlation than the one-hot encoding representation model, notably in the modelling of the CDRH3 uncertainty. We note however that when considering the Spearman correlation, shown in Appendix A, the difference between models is less marked. We provide a guideline for thresholding pLDDT for modelling the CDRH3 region in Figure 2 (right), applied here on the full test set. Incorporating a threshold of a pLDDT above 85, we retain approximately 32% of structures, with over 80% of those retained having a CDR-H3 RMSD below 2Å.

5. Conclusions

In this article, we present ABodyBuilder3, a state-of-the-art antibody structure prediction model based on ABodyBuilder2. We incorporated several improvements to the implementation, notably enhancing hardware acceleration through vectorisation, which significantly improve the scalability of our model. We also made changes to the data processing and structure refinement that lead to more accurate predictions.

In addition, we show how leveraging a language model representation of the antibody sequence can improve the modelling of CDRH3. Finally, we demonstrate how the use of pLDDT head, combined with protein language model embeddings, can be used as a substitute for an ensemble of models approach, which require substantially more training and inference compute.

It would be interesting to explore the use of self-distillation, which has shown to improve accuracy in general protein structure prediction models (Jumper et al., 2021a), by pre-training our model on a large dataset of synthetic structures predicted from the paired Observed Antibody Space (Kovaltsuk et al., 2018; Greenshields-Watson et al., 2023). To even further improve the accuracy of the predictions and of the uncertainty estimates, one could also consider combining the pLDDT approach introduced in this article with an ensemble of models, though this would be at the cost of increased training and inference compute.

We release the code[2] and model weights for ABodyBuilder3 (Kenlay et al., 2024a).

[2] github.com/Exscientia/ABodyBuilder3

Acknowledgements

We are grateful to Brennan Abanades for numerous helpful discussions and advice throughout this project, as well to Constantin Schneider, Claire Marks and Aleksandr Kovaltsuk for useful comments.

References

Abanades, B., Georges, G., Bujotzek, A., and Deane, C. M. ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation. *Bioinformatics*, 38(7):1877–1880, 01 2022a. ISSN 1367-4803. doi: 10.1093/bioinformatics/btac016. URL https://doi.org/10.1093/bioinformatics/btac016.

Abanades, B., Wong, W. K., Boyles, F., Georges, G., Bujotzek, A., and Deane, C. M. Immunebuilder: Deep-learning models for predicting the structures of immune proteins. *bioRxiv*, 2022b. doi: 10.1101/2022.11.04.514231. URL https://www.biorxiv.org/content/early/2022/11/04/2022.11.04.514231.

Adolf-Bryfogle, J., Xu, Q., North, B., Lehmann, A., and Dunbrack, Roland L., J. PyIgClassify: a database of antibody CDR structural classifications. *Nucleic Acids Research*, 43(D1):D432–D438, 11 2014. ISSN 0305-1048. doi: 10.1093/nar/gku1106. URL https://doi.org/10.1093/nar/gku1106.

Ahdritz, G., Bouatta, N., Kadyan, S., Xia, Q., Gerecke, W., O'Donnell, T. J., Berenberg, D., Fisk, I., Zanichelli, N., Zhang, B., et al. Openfold: Retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *bioRxiv*, pp. 2022–11, 2022.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021. doi: 10.1126/science.abj8754. URL https://www.science.org/doi/abs/10.1126/science.abj8754.

Chen, B., Cheng, X., ao Geng, Y., Li, S., Zeng, X., Wang, B., Gong, J., Liu, C., Zeng, A., Dong, Y., Tang, J., and Song, L. xtrimopglm: Unified 100b-scale pre-trained transformer for deciphering the language of protein. *bioRxiv*, 2023. doi: 10.1101/2023.07.05.547496. URL https://www.biorxiv.org/content/early/2023/07/14/2023.07.05.547496.

Chungyoun, M. and Gray, J. J. Ai models for protein design are driving antibody engineering. *Current Opinion in Biomedical Engineering*, pp. 100473, 2023.

de Los Rios, M., Criscitiello, M. F., and Smider, V. V. Structural and genetic diversity in antibody repertoires from diverse species. *Current opinion in structural biology*, 33:27–41, 2015.

Dunbar, J. and Deane, C. M. Anarci: antigen receptor numbering and receptor classification. *Bioinformatics*, 32(2):298–300, 2016.

Dunbar, J., Fuchs, A., Shi, J., and Deane, C. M. Abangle: characterising the vh–vl orientation in antibodies. *Protein Engineering, Design & Selection*, 26(10):611–620, 2013a.

Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., and Deane, C. M. SAbDab: the structural antibody database. *Nucleic Acids Research*, 42 (D1):D1140–D1146, 11 2013b. ISSN 0305-1048. doi: 10.1093/nar/gkt1043. URL https://doi.org/10.1093/nar/gkt1043.

Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., and Deane, C. M. Sabdab: the structural antibody database. *Nucleic acids research*, 42 (D1):D1140–D1146, 2014.

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., and Pande, V. S. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7):1–17, 07 2017. doi: 10.1371/journal.pcbi.1005659. URL https://doi.org/10.1371/journal.pcbi.1005659.

Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., et al. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7112–7127, 2021.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. Protein complex prediction with alphafold-multimer. *bioRxiv*, 2022. doi: 10.1101/2021.10.04.463034.

URL https://www.biorxiv.org/content/early/2022/03/10/2021.10.04.463034.

Greenshields-Watson, A., Abanades, B., and Deane, C. M. Investigating the ability of deep learning-based structure prediction to extrapolate and/or enrich the set of antibody cdr canonical forms. *bioRxiv*, 2023. doi: 10.1101/2023.12.08.570786. URL https://www.biorxiv.org/content/early/2023/12/09/2023.12.08.570786.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021a. doi: 10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021b.

Kenlay, H., Dreyer, F., Cutting, D., Nissley, D., and Deane, C. ABodyBuilder3: Improved and scalable antibody structure predictions, May 2024a. URL https://doi.org/10.5281/zenodo.11354577.

Kenlay, H., Dreyer, F. A., Kovaltsuk, A., Miketa, D., Pires, D., and Deane, C. M. Large scale paired antibody language models, 2024b.

Kovaltsuk, A., Leem, J., Kelm, S., Snowden, J., Deane, C. M., and Krawczyk, K. Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires. *The Journal of Immunology*, 201(8):2502–2509, 10 2018. ISSN 0022-1767. doi: 10.4049/jimmunol.1800708. URL https://doi.org/10.4049/jimmunol.1800708.

Krieger, E. and Vriend, G. New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry*, 36(13):996–1007, 2015. doi: https://doi.org/10.1002/jcc.23899. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23899.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., and Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in casp8. *Proteins: Structure, Function, and Bioinformatics*, 77(S9):114–122, 2009.

Leem, J., Dunbar, J., Georges, G., Shi, J., and Deane, C. Abodybuilder: Automated antibody structure prediction with data–driven accuracy estimation. *mAbs*, 8, 07 2016. doi: 10.1080/19420862.2016.1205773.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., and Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://www.science.org/doi/abs/10.1126/science.ade2574.

Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., Tsai, H.-Z., Li, H.-J., and Wu, H.-C. Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, 27(1):1, 2020. doi: 10.1186/s12929-019-0592-z. URL https://doi.org/10.1186/s12929-019-0592-z.

Mariani, V., Biasini, M., Barbato, A., and Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 08 2013. ISSN 1367-4803. doi: 10.1093/bioinformatics/btt473. URL https://doi.org/10.1093/bioinformatics/btt473.

Narciso, J. E., Uy, I., Cabang, A., Chavez, J., Pablo, J., Padilla-Concepcion, G., and Padlan, E. Analysis of the antibody structure based on high-resolution crystallographic studies. *New biotechnology*, 28:435–47, 04 2011. doi: 10.1016/j.nbt.2011.03.012.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1), jan 2020. ISSN 1532-4435.

Raybould, M. I. J., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A. P., Bujotzek, A., Shi, J., and Deane, C. M. Five computational developability guidelines for therapeutic antibody profiling. *Proceedings of the National Academy of Sciences*, 116(10):4025–4030, 2019. doi: 10.1073/pnas.1810576116. URL https://www.pnas.org/doi/abs/10.1073/pnas.1810576116.

Raybould, M. I. J., Turnbull, O. M., Suter, A., Guloglu, B., and Deane, C. M. Contextualising the developability risk of antibodies with lambda light chains using enhanced therapeutic antibody profiling. *Communications Biology*, 7(1):62, 2024. doi: 10.1038/

s42003-023-05744-8. URL https://doi.org/10.1038/s42003-023-05744-8.

Roth, D. B. V(d)j recombination: Mechanism, errors, and fidelity. *Microbiology Spectrum*, 2(6): 10.1128/microbiolspec.mdna3–0041–2014, 2014. doi: 10.1128/microbiolspec.mdna3-0041-2014. URL https://journals.asm.org/doi/abs/10.1128/microbiolspec.mdna3-0041-2014.

Ruffolo, J. A., Sulam, J., and Gray, J. J. 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.

Ruffolo, J. A., Chu, L.-S., Mahajan, S. P., and Gray, J. J. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications*, 14(1):2389, 2023. doi: 10.1038/s41467-023-38063-x. URL https://doi.org/10.1038/s41467-023-38063-x.

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 Research*, 50(D1):D1368–D1372, 02 2022. ISSN 0305-1048. doi: 10.1093/nar/gkab1050. URL https://doi.org/10.1093/nar/gkab1050.

Slabinski, L., Jaroszewski, L., Rodrigues, A. P., Rychlewski, L., Wilson, I. A., Lesley, S. A., and Godzik, A. The challenge of protein structure determination—lessons from structural genomics. *Protein Science*, 16(11):2472–2482, 2007. doi: https://doi.org/10.1110/ps.073037907. URL https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.073037907.

Tsuchiya, Y. and Mizuguchi, K. The diversity of h3 loops determines the antigen-binding tendencies of antibody cdr loops. *Protein science : a publication of the Protein Society*, 25, 01 2016. doi: 10.1002/pro.2874.

Wong, W. K., Leem, J., and Deane, C. M. Comparative analysis of the cdr loops of antigen receptors. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. doi: 10.3389/fimmu.2019.02454. URL https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.02454.

Zeng, A., Liu, X., Du, Z., Wang, Z., Lai, H., Ding, M., Yang, Z., Xu, Y., Zheng, W., Xia, X., Tam, W. L., Ma, Z., Xue, Y., Zhai, J., Chen, W., Zhang, P., Dong, Y., and Tang, J. Glm-130b: An open bilingual pre-trained model, 2023.

A. Structure refinement

Most approaches to protein or antibody structure prediction, including AlphaFold2 (Jumper et al., 2021a), AlphaFold2-Multimer (Evans et al., 2022), and ABodyBuilder2 (Abanades et al., 2022b), that predominantly rely on deep learning also use a final physics-based refinement step to fix stereochemical errors and provide realistic structures. In most cases, this refinement takes the form of an in-vacuo minimization that may neglect important aspects of the real system in favor of expediency. To test the influence of different types of minimization on the quality of ABodyBuilder3 output structures, we compare an OpenMM refinement with a refinement using YASARA2, shown in Figure 3. These results confirm that while minimization in vacuo is sufficient to improve many structures, a minimization in the YASARA2 forcefield (Krieger et al., 2009; Krieger & Vriend, 2015) in explicit water allows for further improvements across all the regions, particularly in the framework, whilst improving model quality according to the z-score produced by YASARA2.

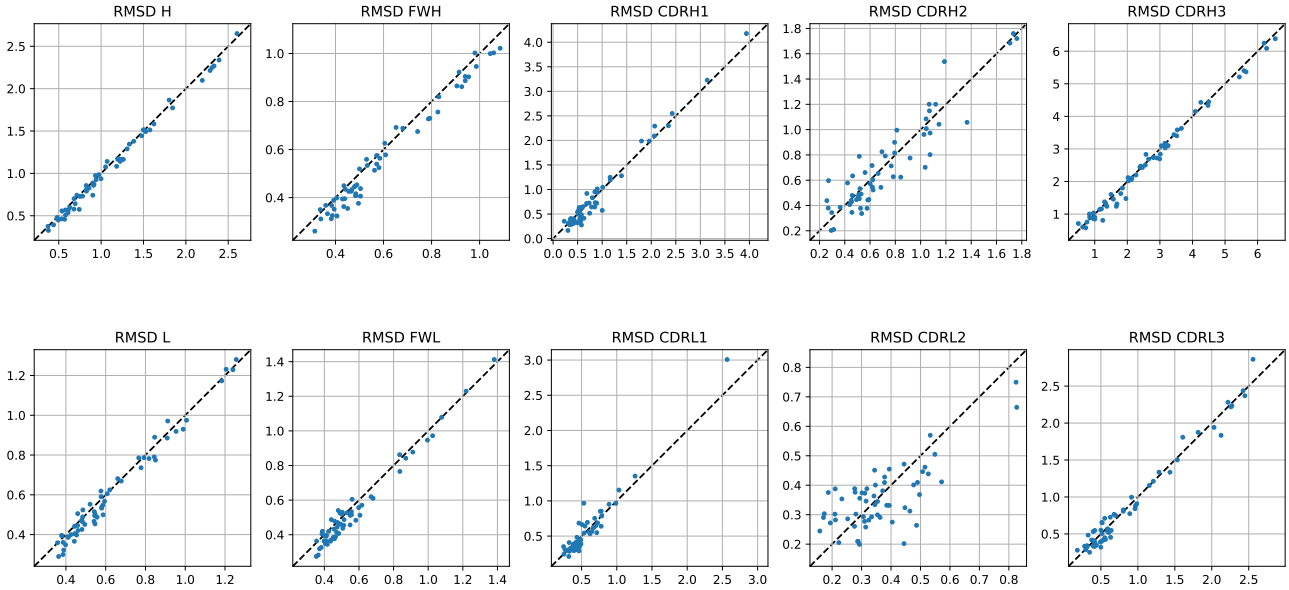

Figure 3. YASARA2 refinement (x-axis) compared to OpenMM refinement (y-axis).

B. Spearman correlation of uncertainties

We show the Spearman correlation between uncertainty predictions and RMSD in Table 3.

	CDRH1	CDRH2	CDRH3	Fw-H	CDRL1	CDRL2	CDRL3	Fw-L
ABodyBuilder2	0.45	0.30	0.75	0.50	0.56	0.21	0.73	0.42
ABodyBuilder3	0.48	0.23	0.63	0.50	0.37	0.01	0.59	0.59
ABodyBuilder3-LM	0.48	0.40	0.73	0.27	0.43	0.20	0.60	0.53

Table 3. Spearman correlation between average uncertainty prediction for a region and the corresponding mean RMSD.