Ibex: Pan-immunoglobulin structure prediction

Frédéric A. Dreyer [1] **Karolis Martinkus** [1] **Jan Ludwiczak** [1] **Brennan Abanades** [2] **Robert G. Alberstein** [1] **Pranav Rao** [1] **Jae Hyeon Lee** [1] **Richard Bonneau** [1] **Andrew M. Watkins** [1] **Franziska Seeger** [1]

Abstract

We introduce Ibex, an immunoglobulin protein structure prediction model that achieves state-of-the-art accuracy in modeling the binding domains of antibodies, nanobodies and T-cell receptors. Ibex can model both bound and unbound conformations of the protein, having been trained on labeled *apo* and *holo* pairs. Using a private dataset of hundreds of antibody structures, we evaluate the out-of-distribution performance of common structure prediction tools, showing improved robustness of Ibex compared to existing specialized structure prediction models.

1. Introduction

The accurate prediction of immunoglobulin protein structures, most notably among them antibodies and T-cell receptors, is of critical importance to the design of better biologics and the acceleration of drug discovery. A central task for antibody design is understanding the interactions between residues in the complementarity determining region (CDR) and the antigen, particularly those involving the third CDR loop on the heavy chain (CDR H3) (Narciso et al., 2011). The CDR H3 loop is uniquely diverse due to it being constructed by the recombination of three different genes, and can undergo substantial conformation change upon binding, making it challenging to model (Greenshields-Watson et al., 2025). We aim to improve the prediction accuracy on CDR loops of antibodies by training a pan-immunoglobulin structure prediction model, through the curation of a wider corpus of related structural data and the explicit incorporation of their binding state.

We introduce Ibex, a state-of-the-art antibody, TCR and nanobody structure prediction model that achieves comparable performance to Boltz-1 (Wohlwend et al., 2024) at a fraction of the computational cost. We benchmark Ibex against several recent models on an internal benchmark of several hundred unpublished antibody structures, showing improved performance compared to previous specialized and general protein models.

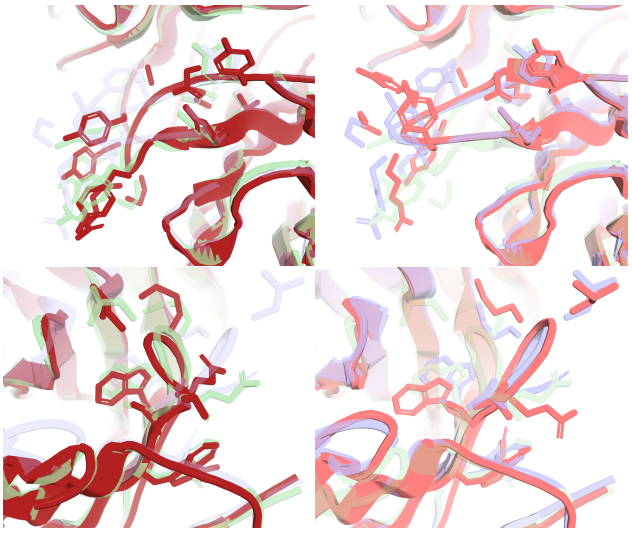

Figure 1. Holo ground truth and predictions shown on the left, with *apo* poses shown on the right. The top row shows an antibody (PDB codes: `2fr4` and `1xf3`), with *holo* Ibex predictions in dark red and *apo* predictions in salmon, superimposed on their respective ground truth structures shown in green and blue. The view is centered on the H3 loop, with side chains shown for the loop residues. The bottom row shows the $\beta 3$ loop of a TCR (`6eqb` and `4jfh`).

Ibex is trained on a combined dataset of antibodies, nanobodies, TCRs and other immunoglobulin-like domains. Each structure in the training dataset is labeled as *apo* or *holo*, and both the bound or unbound conformation can be predicted at inference. Examples of *apo* and *holo* predictions for the same sequences are shown in Figure 1, with the corresponding ground truth overlayed in grey.

2. Model

Ibex follows a similar architecture to ABodyBuilder2 (Abanades et al., 2023) and uses several structure module blocks from AlphaFold2 (Jumper et al., 2021) to refine structures from an input sequence embedding. An overview of the model is given in Figure 2.

We use as input a combination of ESM-C sequence embeddings (ESM Team, 2024), the one-hot encoding of the

[1]Prescient Design, Genentech, South San Francisco, CA, USA [2]Large Molecule Research, Roche, Penzberg, Germany. Correspondence to: Frédéric A. Dreyer <dreyer.frederic@gene.com>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

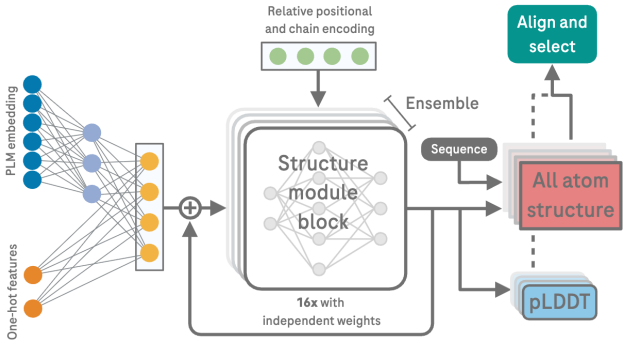

Figure 2. Overview of the Ibex model. The one-hot encoded residue features are concatenated with a projection of language model embeddings and used as input to 16 structure module blocks. The final residue representation is then used to predict all atomic coordinates and uncertainties.

sequence, and a residue-level encoding of the chain and conformation feature. The conformation token is provided as a label during training, and used as an additional input at inference, to disambiguate between *apo* and *holo* structures. The language model embeddings are passed to a two-layer MLP and concatenated with the remaining node features after applying a layer normalization. Pairwise edge features are set to the one-hot encoding of relative positions in the sequence, concatenated with an encoding of which chains both residues belong to. For each invariant point attention layer in the structure module, this pair representation is concatenated with a distance feature map and provided as input. The core of the model consists of 16 consecutive and independent structure module blocks, which update residue coordinates and node features through an invariant point attention layer. A key change from ABodyBuilder3 is the introduction of a residual connection between the initial residue representation embedding and each structure module block, which improves training convergence and ensures information from the conformation token is preserved throughout the network. The residue representation from the last structure module block is used to predict backbone atom coordinates. The original input sequence is then used to reconstruct side-chain atoms from idealized coordinates using predicted chi-angles. Uncertainties are modeled through a predicted local-distance difference test (pLDDT) head, which predicts a projection of local confidence into 50 bins.

3. Data

Ibex is intended to be a pan-immunoglobulin structure prediction model, and we therefore aim to collate a comprehensive dataset of all available immunoglobulin protein structures, with explicit labeling of their binding state. The training data for Ibex is constructed from three different sources, which are clustered and combined dynamically throughout the training process. Each structure is also annotated with a conformation token, which labels them as *apo* or *holo* based on the presence of an antigen to which the variable region is bound. The model implicitly learns the distinctive structural motifs and folding patterns that differentiate antibodies, TCRs, and nanobodies.

The first source is a dataset of structures of antibody variable regions, nanobodies and TCR variable regions. These are curated from the structural antibody database (SAbDab) (Dunbar et al., 2014) and the structural TCR database (STCRDab) (Leem et al., 2018). Sequences are numbered using Anarci (Dunbar & Deane, 2016), and we use the North definition (North et al., 2011) to delineate the CDR residues. Structures are characterized as *apo* if no antigen chain is indicated in the SAbDab or STCRDab metadata, otherwise they are labeled as *holo*. We remove structures with a resolution above 3.5Å, as well as structures with a CDRH3 or β3 loop of more than 35 residues. We also remove any antibody variable domains for which one of the six Abangle (Dunbar et al., 2013) VH-VL orientation angle or distance values are more than five standard deviations from the mean computed over SAbDab. We are left with 14k structures, 760 of which are matched apo/*holo* pairs for which both bound and unbound conformations are known. All structures are clustered based on the sequence of their concatenated CDR loops, using mmseqs2 (Steinegger & Söding, 2017) with 95% sequence identity, resulting in 4.2k unique clusters.

Our second dataset consists of immunoglobulin-like domains found in the Protein Data Bank. We identify any immunoglobulin-like structures according to the ECOD (Cheng et al., 2014) X-groups "Immunoglobulin-like beta-sandwich" and "jelly-roll" with PDB codes absent from both SAbDab and STCRDab. These are processed as individual chains, and loop residues are defined with the DSSP algorithm (Kabsch & Sander, 1983). We assign apo labels to chains for which no heavy atoms are found within 10Å of a loop residue. We obtain a dataset of 22k single domain structures with resolution below 3.5Å, which are clustered at 80% sequence identity into 3k unique clusters.

Finally, we curate a dataset of predicted structures from paired sequences from the Observed Antibody Space (OAS) (Kovaltsuk et al., 2018; Olsen et al., 2022) using ESMFold (Lin et al., 2023) and Boltz-1 (Wohlwend et al., 2024). Starting from 1.7 million unique paired sequences from OAS, we cluster them using concatenated CDR loops with a 60% sequence identity threshold, as well as on the H3 sequence with 50% minimum sequence identity. We then consider all cluster representatives from the concatenated CDR clustering which are also in distinct H3 clusters,

		CDR H1	CDR H2	CDR H3	Fw H	CDR L1	CDR L2	CDR L3	Fw L
Antibodies	Boltz-1	*0.63*	**0.52**	2.96	*0.47*	**0.54**	**0.38**	**0.96**	**0.52**
	ABodyBuilder3	0.71	0.65	*2.86*	0.51	0.67	0.49	1.13	0.59
	ESMFold	0.70	0.99	3.15	0.65	1.04	0.50	1.29	0.61
	Ibex	**0.61**	*0.57*	**2.72**	**0.45**	*0.57*	*0.43*	*0.98*	**0.52**
Nanobodies	Boltz-1	*1.59*	0.97	**2.83**	**0.61**	-	-	-	-
	NanoBuilder2	1.74	1.22	3.31	0.76	-	-	-	-
	ESMFold	*1.55*	**0.94**	3.60	0.63	-	-	-	-
	Ibex	1.62	*0.96*	*3.12*	*0.62*	-	-	-	-
		CDR β1	CDR β2	CDR β3	Fw β	CDR α1	CDR α2	CDR α3	Fw α
TCRs	Boltz-1	*0.59*	**0.53**	2.40	0.66	**0.95**	0.87	2.05	**0.58**
	TCRBuilder2+	0.72	0.79	1.85	0.68	1.18	1.04	*2.00*	0.94
	ESMFold	0.68	0.66	2.49	0.71	1.35	0.96	2.31	0.75
	Ibex	**0.57**	*0.57*	**1.84**	**0.60**	*1.02*	**0.85**	**1.93**	*0.71*

Table 1. Mean test RMSD in angstrom, evaluated separately for each region on a test set of antibodies, nanobodies and TCRs.

resulting in a highly diversified set of 91k paired sequences. Using ESMFold, we predict the corresponding structures, and filter out any for which one of the Abangle value is more than three standard deviations from the SAbDab mean, leaving us with 51k ESMFold structures. We randomly sample 10k of the 40k sequences for which ESMfold structures failed our Abangle filter, and predict their structure with Boltz-1, which are again filtered through the same procedure, resulting in 9k accepted Boltz-1 structures. We perform a short relaxation with OpenMM of all predicted structures. All predicted structures are labeled as unbound. This provides us with a dataset of 60k data points with very high CDR diversity which we use for distillation and to improve generalization.

4. Training

We train Ibex in three stages, using a curriculum learning strategy (Bengio et al., 2009), where training is slowly specialized towards experimental antibody and TCR structures after pre-training on a larger corpus of predicted structures and related experimental protein data.

The first two stages use a Frame Aligned Point Error (FAPE) loss along with a backbone torsion angle and plDDT losses. The FAPE loss is clamped at 10Å, and 30Å when it is computed between CDR and framework residues, or between loop and non-loop annotated residues in the case of immunoglobulin-like single domains. Following a similar approach to AlphaFold2 (Jumper et al., 2021), the total loss term is the sum of the average backbone FAPE loss across each layer, the full atom FAPE loss from the final structure, as well as side-chain and backbone torsion angle losses and a pLDDT loss. The pLDDT loss consists of a cross-entropy loss on the discretised per-residue lDDT-$C\alpha$.

The third stage introduces structural violation losses. These follow AlphaFold2-Multimer (Evans et al., 2021) and penalize bond length violations, bond angle violations and steric clashes of non bonded atoms.

Across each stage, we use an AdamW optimizer (Loshchilov & Hutter, 2017). The first stage is trained with a constant learning rate of $2 \cdot 10^{-4}$ for 3000 epochs. The second stage lasts 2000 epochs, and starts with a learning rate of $2 \cdot 10^{-4}$ and slowly anneals to $2 \cdot 10^{-5}$ from 200 epochs onwards using a lambda scheduler. The third stage is trained with a constant learning rate of $5 \cdot 10^{-5}$ for 1000 epochs.

An epoch consists of randomly sampling N structures according to weighted probabilities assigned to each data point, where N is set to the total size of the SAbDab and STCRDab structures used. We assign a probability inversely proportional to the cluster size to each data point to improve o.o.d. generalization (Loukas et al., 2024). Matched pairs of *apo/holo* structures are identified and sampled separately from the unpaired SAbDAb and STCRDab structures to ensure they appear together in a batch. We upsample nanobodies and TCR by a factor 1.5, while paired *apo/holo* are upsampled by a factor 2. In the first stage, the weighting of different data sources is split as 30% SAbdab and STCRDab, 40% predicted structures, and 30% from the broader immunoglobulin protein dataset. This is changed to 70% SAbDAb and STCRDab, 20% predicted structures and 10% immunoglobulin proteins in the second stage. Finally, in the third stage, we use 95% of SAbDab and STCRDab structures and 5% of predicted structures, this time restricting only to Boltz-1 samples and discarding the ESMFold and broader immunoglobulin data.

The final Ibex model consists of an ensemble of eight independently trained models. Predictions of all eight models

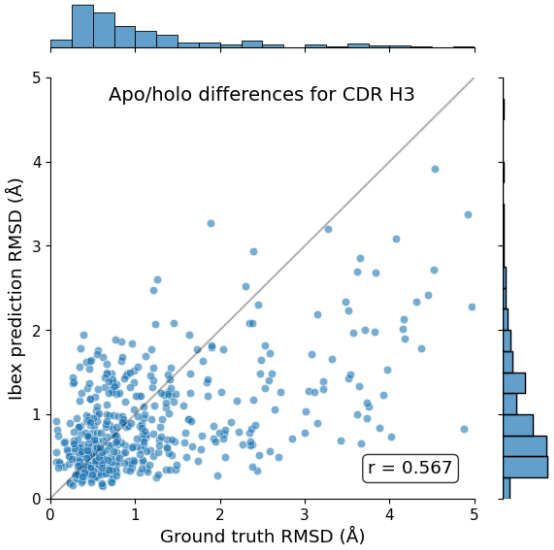

Figure 3. Distribution of RMSD between matching *apo* and *holo* pairs, comparing ground truth structures to Ibex predictions.

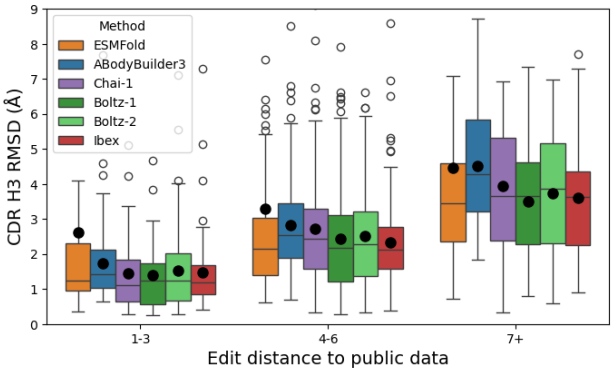

Figure 4. CDR H3 RMSD as a function of edit distance to the closest matching H3 loop in SAbDAb. The box shows the lower and upper quartile, with the median represented as a horizontal line and the mean as a filled circle. Outliers are shown as empty circles, with ESMFold having outliers outside of the plotted range.

are aligned, and the model returns the prediction closest to the mean. Each model in the ensemble is trained using a different validation set of 150 antibody structures randomly selected from the SAbDab clusters.

5. Results

Our immunoglobulin structure prediction model outperforms existing antibody-specific models in the prediction of the CDRH3 loop. A summary table of average RMSD values over a test set of antibodies, nanobodies and TCRs is shown in Table 1, showing Ibex achieving the best results across most modalities. For consistent benchmarking, we train Ibex using the combined test set of antibodies, nanobodies and TCRs from the ImmuneBuilder suite (Abanades et al., 2023), removing 6 nanobodies from the test set for which a matching sequence was observed in the NanoBody2 train or validation split. All structures from the dataset that are in a cluster that contains a test set structure are removed from training. Further comparisons are given in Appendix A

We are also able to recapitulate the conformation shift between bound and unbound structures of existing antibodies for which both *apo* and *holo* conformations are known. This is shown in Figure 3, where we present the CDRH3 RMSD between matched *apo* and *holo* pairs, both for predicted structures obtained from Ibex and for the ground truth PDB structures. Most of these pairs are part of the training data.

We further evaluate our model on a diverse private dataset of over a thousand antibody structures. This dataset consists of 1103 structures resolved for internal studies, often antibodies with no known public structure, spanning H3 loop lengths from 5 to 22 residues. The resolution of these structures ranges from 1.1Å to 3.6Å with an average resolution of 2.3Å. In Figure 4, we show the CDRH3 RMSD on these internal structures as a function of the edit distance to the closest matching H3 loop in SAbDab, for the 300 structures that have unique H3 loops not represented in SAbDab. Here we observe that while ABodyBuilder3 achieves high performance on the test set, it is not as robust out of distribution compared to state-of-the-art general protein structure prediction models. In contrast, Ibex shows comparable performance to Boltz-1, which we attribute to distillation from predicted structures and the use of a broader corpus of training data.

6. Discussion

We introduce Ibex, an immunoglobulin protein structure prediction tool that can model antibodies, nanobodies and T-cell receptors, achieving state-of-the-art accuracy in antibody structure prediction while introducing the capability to model both bound and unbound conformations of the binding domain. This explicit modeling of the binding state allows us to extend the data landscape and improve prediction accuracy in cases where several ground truth structures exist for the same sequence, previously leading to ambiguous minimization objectives.

We evaluate our model on an internal dataset of hundreds of antibody structures, showing improved performance compared to existing antibody and protein structure prediction models, at a fraction of the compute cost of the latter. Ibex provides an important stepping stone towards the design of better biologics, by allowing for fast and accurate predictions of the structure of their binding domains.

Acknowledgements

We are grateful to David Konerding, Joseph Kleinhenz, and Henri Dwyer for engineering support, and to the Structural Biology teams at Genentech and Roche for providing structures used in benchmarking. We thank Henry Kenlay, Allen Goodman, Josh Southern, Andrew Leaver-Fay, Nathan Frey, Lian Huang, as well as the entire Prescient Design team for useful discussions.

Impact Statement

This paper presents work whose goal is to advance the field of machine learning for biological sciences. Our work has potential impact on drug discovery, with notable applications to the design and evaluation of biopharmaceuticals.

References

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. *Communications Biology*, 6(1):575, 2023.

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pp. 41–48, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. doi: 10.1145/1553374.1553380. URL https://doi.org/10.1145/1553374.1553380.

Cheng, H., Schaeffer, R. D., Liao, Y., Kinch, L. N., Pei, J., Shi, S., Kim, B.-H., and Grishin, N. V. Ecod: an evolutionary classification of protein domains. *PLoS computational biology*, 10(12):e1003926, 2014.

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, 2013.

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.

ESM Team. Esm cambrian: Revealing the mysteries of proteins with unsupervised learning, 2024. URL https://evolutionaryscale.ai/blog/esm-cambrian.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pp. 2021–10, 2021.

Greenshields-Watson, A., Vavourakis, O., Spoendlin, F. C., Cagiada, M., and Deane, C. M. Challenges and compromises: Predicting unbound antibody structures with deep learning. *Current Opinion in Structural Biology*, 90:102983, 2025. ISSN 0959-440X. doi: https://doi.org/10.1016/j.sbi.2025.102983. URL https://www.sciencedirect.com/science/article/pii/S0959440X25000016.

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, 2021.

Kabsch, W. and Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983. doi: https://doi.org/10.1002/bip.360221211. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360221211.

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, 2018.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Leem, J., de Oliveira, S. H. P., Krawczyk, K., and Deane, C. M. Stcrdab: the structural t-cell receptor database. *Nucleic acids research*, 46(D1):D406–D412, 2018.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130, 2023.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Loukas, A., Martinkus, K., Wagstaff, E., and Cho, K. Generalizing to any diverse distribution: uniformity, gentle finetuning and rebalancing. *arXiv preprint arXiv:2410.05980*, 2024.

Narciso, J. E. T., Uy, I. D. C., Cabang, A. B., Chavez, J. F. C., Pablo, J. L. B., Padilla-Concepcion, G. P., and Padlan, E. A. Analysis of the antibody structure based on high-resolution crystallographic studies. *New*

Biotechnology, 28(5):435–447, 2011. ISSN 1871-6784. doi: https://doi.org/10.1016/j.nbt.2011.03.012. URL https://www.sciencedirect.com/science/article/pii/S1871678411000744. Antibodies: From Basics to Therapeutics.

North, B., Lehmann, A., and Dunbrack Jr, R. L. A new clustering of antibody cdr loop conformations. *Journal of molecular biology*, 406(2):228–256, 2011.

Olsen, T. H., Boyles, F., and Deane, C. M. Observed antibody space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Science*, 31(1):141–146, 2022.

Steinegger, M. and Söding, J. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, 35(11):1026–1028, 2017.

Wohlwend, J., Corso, G., Passaro, S., Reveiz, M., Leidal, K., Swiderski, W., Portnoi, T., Chinn, I., Silterra, J., Jaakkola, T., and Barzilay, R. Boltz-1 democratizing biomolecular interaction modeling. *bioRxiv*, 2024. doi: 10.1101/2024.11.19.624167. URL https://www.biorxiv.org/content/early/2024/11/20/2024.11.19.624167.

A. Detailed comparison

We provide here a detailed view of the summary table 1. In figure 5, we show the RMSD on the ImmuneBuilder antibody test set separated by region, and comparing Ibex predictions to ABodyBuilder3. Figure 6 gives a comparison of Ibex to NanoBuilder2 on the test set of nanobodies. Here the 6 structures for which an identical match was identified in the train or validation split of NanoBuilder2, which are PDB codes `7n4n`, `7omt`, `7q6c`, `7rg7`, `7zmv`, and `7zxu`, are shown in grey. These points were excluded from the average presented in table 1. In Figure 7, we provide a comparison of Ibex to TCRBuilder2+ on the test set of 21 TCR structures.

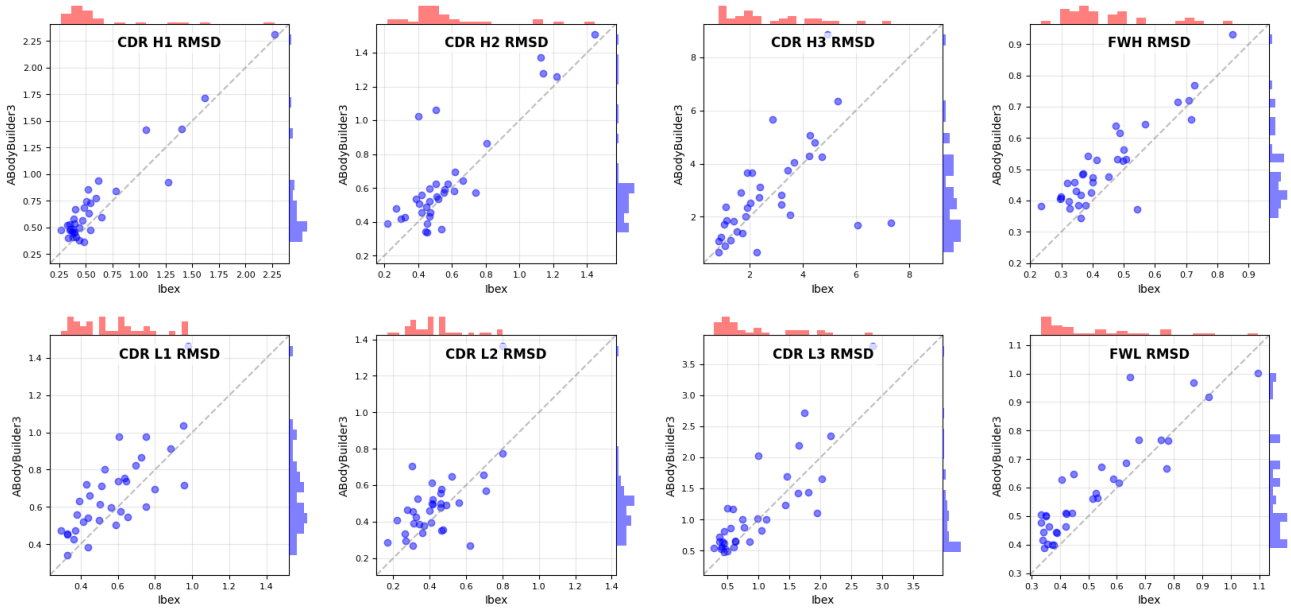

Figure 5. CDR and framework RMSD in angstrom, comparing Ibex against ABodyBuilder3 on the ImmuneBuilder test set of 34 antibodies.

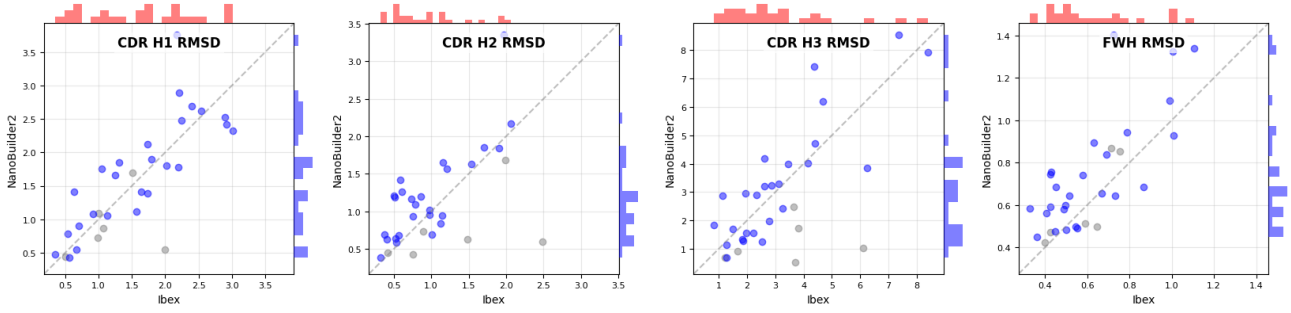

Figure 6. CDR and framework RMSD in angstrom, comparing Ibex against NanoBuilder2 on the ImmuneBuilder test set of 32 nanobodies. The 6 datapoints for which identical CDR H3 sequences were identified in the train or validation split are shown in grey and are excluded from the displayed histogram.

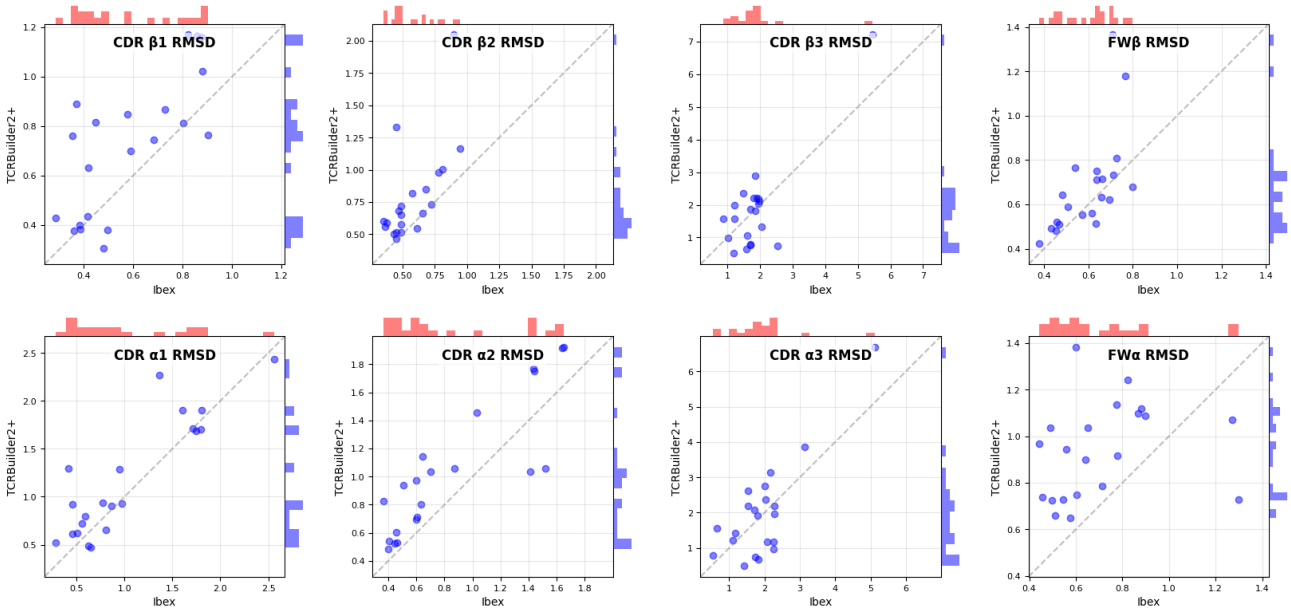

Figure 7. CDR and framework RMSD in angstrom, comparing Ibex against ABodyBuilder3 on the ImmuneBuilder test set of 21 TCRs.

B. Ablation studies

In this section we consider the impact on out-of-distribution robustness of several architecture and data choices. To this end, we retrain a single model checkpoint (i.e. without ensemble) for 400 epochs in each three stage. This is shown as ”base” in Figure 8. Ablation studies of similar checkpoints trained without the immunoglobulin-like data, without the predicted data, as well as trained only on SAbDab and STCRDab, are also shown. We can observe that both the predicted structures and the immunoglobulin-like data lead to improved robustness at large edit distance from known public CDR H3 loops, with the model trained on combined data performing best.

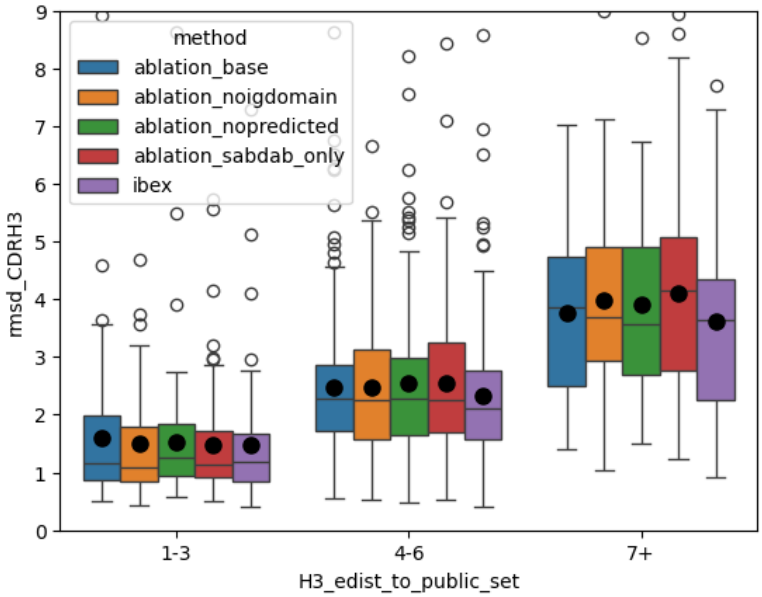

Figure 8. CDR H3 RMSD as a function of edit distance to the closest matching H3 loop in SAbDAb. We show ablation studies of the data used, removing the predicted data, the immunoglobulin-like data, and both, from the training of a single Ibex checkpoint.