TOWARDS MORE ACCURATE FULL-ATOM ANTIBODY CO-DESIGN

Jiayang Wu[1*] **Xingyi Zhang**[2*] **Xiangyu Dong**[3] **Kun Xie**[3] **Ziqi Liu**[4]
Wensheng Gan[1] **Sibo Wang**[3] **Le Song**[2,5†]
[1]Jinan University [2]MBZUAI [3]The Chinese University of Hong Kong
[4]University of Chinese Academy of Sciences [5]GenBio AI

ABSTRACT

Antibody co-design represents a critical frontier in drug development, where accurate prediction of both 1D sequence and 3D structure of complementarity-determining regions (CDRs) is essential for targeting specific epitopes. Despite recent advances in equivariant graph neural networks for antibody design, current approaches often fall short in capturing the intricate interactions that govern antibody-antigen recognition and binding specificity. In this work, we present Igformer, a novel end-to-end framework that addresses these limitations through innovative modeling of antibody-antigen binding interfaces. Our approach refines the inter-graph representation by integrating personalized propagation with global attention mechanisms, enabling comprehensive capture of the intricate interplay between local chemical interactions and global conformational dependencies that characterize effective antibody-antigen binding. Through extensive validation on epitope-binding CDR design and structure prediction tasks, Igformer demonstrates significant improvements over existing methods, suggesting that explicit modeling of multi-scale residue interactions can substantially advance computational antibody design for therapeutic applications.

1 MAIN

Antibodies are Y-shaped proteins that play a pivotal role in the immune system, capable of binding with high specificity to antigens, including pathogens and foreign molecules Murphy & Weaver (2016). The design of antibodies for specific epitopes in antigens is an essential yet challenging task with broad applications in therapeutic development and drug discovery Feldmann & Maini (2003); Adams & Weiner (2005); Mullard (2022); Wang et al. (2024b). The challenges arise primarily from the variability and structural complexity of complementarity-determining regions (CDRs), where binding primarily occurs He et al. (2024). Understanding and modeling the intricate interactions between antibodies and antigens is crucial for developing an effective computational antibody design approach Baran et al. (2017); Tennenhouse et al. (2024); Yan et al. (2024).

The past decade has witnessed a significant evolution in antibody design approaches. Traditional methods, including sequence-based models Alley et al. (2019); Shin et al. (2021) and energy-based optimization frameworks Pantazes & Maranas (2010); Lapidoth et al. (2015); Adolf-Bryfogle et al. (2018), provided initial solutions but fell short of fully capturing the structural and functional interactions between antibodies and antigens Ingraham et al. (2019). More recently, learning-based approaches, particularly equivariant graph neural networks, deep generative models, and diffusion models

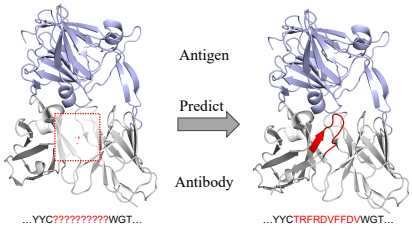


Figure 1: End-to-end antibody co-design.

[*]Equal contribution.
[†]Corresponding author: le.song@genbio.ai

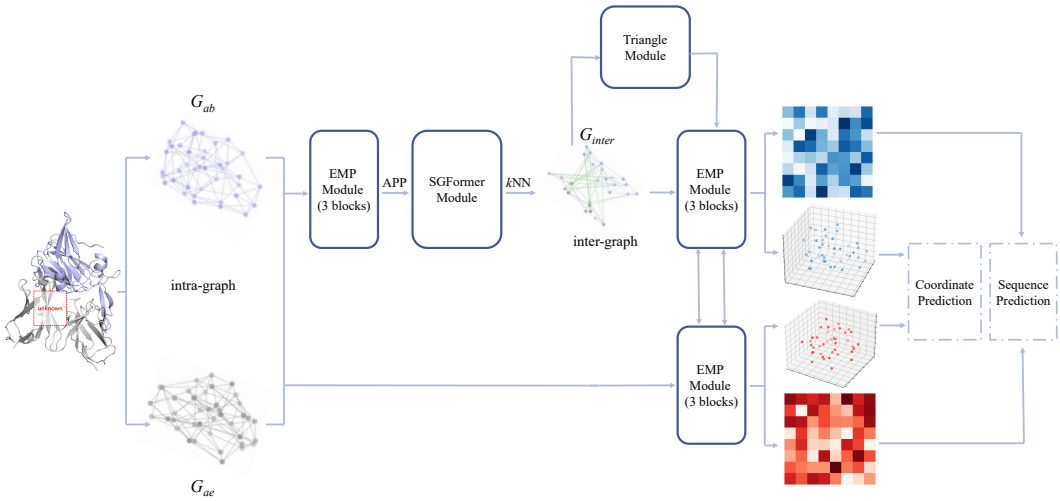


Figure 2: Framework of Igformer.

Saka et al. (2021); Jin et al. (2022b); Watson et al. (2022b); Luo et al. (2022); Kong et al. (2023a); Martinkus et al. (2023); Wang et al. (2024a), have emerged as powerful tools for simultaneously co-designing CDR 1D sequences and 3D structures. These models represent a significant advancement over traditional approaches by leveraging deep learning to model complex sequential and structural relationships. However, many of these methods simplify the problem by focusing exclusively on backbone modeling or specific CDR regions, overlooking side-chain interactions and the full-atom geometry that are crucial to accurately modeling antibody-antigen binding Ruffolo et al. (2023b); Martinkus et al. (2023).

Traditional computational antibody design relies on multi-stage pipelines that separately handle structure prediction, docking, and CDR generation. A recent work, dyMEAN Kong et al. (2023a), integrates these tasks into a unified end-to-end framework, providing a more streamlined approach to antibody co-design. However, despite these advances, dyMEAN does not fully take advantage of intricate antibody-antigen interactions, leaving opportunities for further improvement in capturing the detailed molecular interplay at binding interfaces. These limitations highlight the need for a more sophisticated approach that can simultaneously handle full-atom geometry and complex hierarchical interactions in antibody-antigen complexes. Further discussion of related works can be found in Appendix A.

To address these limitations, in this work, we introduce Igformer[1], a novel end-to-end framework for antibody co-design that advances antibody co-design through three key innovations. At its core, Igformer introduces a sophisticated antibody-antigen inter-graph refinement strategy that uniquely integrates personalized propagation with a global attention mechanism. The personalized propagation scheme models local chemical properties and geometric constraints, while the complementary global attention mechanism captures long-range structural dependencies. This novel dual-scale architecture enables Igformer to maintain atomic-level precision while effectively capturing intricate residue-level interactions within the antibody-antigen binding interface. Furthermore, Igformer incorporates E(3)-equivariance throughout the framework, ensuring that all structural predictions adhere to the physical and geometric constraints of biological systems.

Through comprehensive experimental evaluation, Igformer demonstrates superior effectiveness across multiple challenging antibody design tasks. For epitope-binding CDR design, Igformer shows a 2.02% improvements over state-of-the-art models in terms of amino acid recovery rate. For antigen-antibody complex structure prediction, Igformer achieves remarkable accuracy with an 11.84% reduction in RMSD compared to existing approaches. These substantial improvements validate the effectiveness of our technical innovations in addressing the challenges of full-atom geometry and sequence-structure co-design. Igformer thus represents a significant advancement in computational antibody design, offering promising new directions for accelerating drug discovery and therapeutic development.

[1] *Immunoglobulin Transformer (Igformer)*

Table 1: CDR-H3 design. Results of models with * are collected from the dyMEAN paper.

Model	Generation			Docking		
	AAR↑	TMscore↑	IDDT↑	CAAR↑	RMSD↓	DockQ↑
RosettaAb*	32.31%	0.9717	0.8272	14.58%	17.70	0.137
DiffAb*	35.31%	0.9695	0.8281	22.17%	23.24	0.158
MEAN*	37.38%	0.9688	0.8252	24.11%	17.30	0.162
HERN*	32.65%	-	-	19.27%	9.15	0.294
dyMEAN	42.64%	0.9728	0.8438	27.35%	8.42	0.408
Igformer	**43.50%**	**0.9757**	**0.8650**	**28.11%**	**7.15**	**0.450**

Table 2: AAR (%) on multiple CDRs design.

Model	CDR-L1	CDR-L2	CDR-L3	CDR-H1	CDR-H2	CDR-H3
dyMEAN	**75.55**	83.10	52.12	75.51	68.48	37.53
Igformer	75.20	**85.32**	**63.42**	**77.20**	**69.25**	**41.10**

2 RESULTS

2.1 IGFORMER

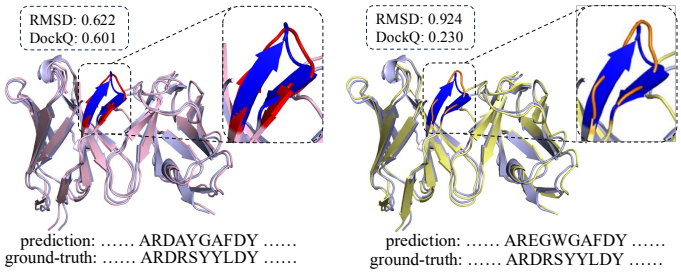
Figure 2 illustrates the overall framework of Igformer. Specifically, the *Equivariant Message Passing (EMP)* modules serve as foundational blocks of Igformer, integrating both spatial and biochemical properties during the message-passing process. Our theoretical analysis confirms the E(3)-equivariant property of EMP, which lays the theoretical foundation of Igformer.

The key innovation of Igformer lies in its sophisticated inter-graph refinement strategy. The process begins with the initialization of residue representations and coordinates, followed by intra-graph construction. Subsequently, based on the constructed intra-graph, Igformer captures the complex interactions within the binding range between antibody paratope and antigen epitope through two key modules: *Approximate Personalized Propagation (APP)*, which preserves local interaction information, and *Simplified Graph Transformer (SGFormer)*, which captures global binding dependencies. Next, the learned intricate binding patterns are utilized to refine the inter-graph, followed by *Triangle Multiplicative (TM)* module and *Axial Attention (AA)* module, which generate pairwise residue interactions through different mechanisms. Finally, the refined inter-graph, together with the intra-graph are processed through a *dual-scale EMP modules*, responsible for the final coordinate and sequence predictions. Systematical elaboration on our method and supplementary details can be found in Appendix C and appendix D.

Dataset. Following previous works on antibody design Jin et al. (2022a); Kong et al. (2023a;b), we train all models on the *Structural Antibody Database* (SAbDab) (Dunbar et al., 2013; Schneider et al., 2021) using the snapshot from November 2022. We split the dataset into training and validation sets with a ratio of 9:1, yielding $3,246$ antibodies for training and 365 antibodies for validation. The antibodies are clustered based on their CDR-H3 regions using a 40% sequence identity threshold, calculated using the BLOSUM62 substitution matrix Foote & Winter (1992). This clustering process, performed using MMseqs2, results in $1,644$ clusters in the training set and 182 clusters in the validation set. After that, we evaluate Igformer and other competitors on the RAbD benchmark Adolf-Bryfogle et al. (2018) for Tasks 1-3, which consists of 60 diverse antibody-antigen complexes. For Task 4, Igfold benchmark Ruffolo et al. (2023a) consisting of 51 antibody-antigen complexes is utilized. This test set selection prevents data leakage during the evaluation phase. Detailed introduction to datasets and training settings of IgFormer are listed in Appendix E.1 and Appendix E.2, respectively.

2.2 TASK 1: CDR-H3 DESIGN

We first evaluate each model for both sequence and structure prediction on CDR-H3, which plays a critical role in antigen binding. During training, the model simultaneously predicts CDR-H3 residue sequences and generates coordinates for the entire antibody structure. We compare Igformer against five SOTA methods: RosettaAb Adolf-Bryfogle et al. (2018), DiffAb Luo et al. (2022), HERN Jin et al. (2022a), MEAN Kong et al. (2023b), and dyMEAN Kong et al. (2023a). Evaluation metrics



(a) Igformer (b) dyMEAN

Figure 3: Antibody structures by Igformer & dyMEAN.

Table 3: Comparison on multiple CDR and full antibody design.

Multiple CDR Design				
Model	AAR↑	TMscore↑	lDDT↑	DockQ↑
dyMEAN	60.05%	0.9654	0.8029	0.3973
Igformer	**63.55%**	**0.9750**	**0.8311**	**0.4817**
Full Antibody Design				
Model	AAR↑	TMscore↑	lDDT↑	DockQ↑
dyMEAN	71.37%	0.9662	0.7471	0.4237
Igformer	**73.69%**	**0.9681**	**0.7580**	**0.4600**

include unaligned RMSD for CDR-H3 using CA atoms and sequence-based metrics for CDR-H3. Details of baselines and metrics can be found in Appendix E.3 and Appendix E.4, respectively.

Table 1 reports the results of the epitope-binding CDR-H3 design task on the RAbD dataset. As we can observe, Igformer demonstrates superior performance across all evaluation metrics. Specifically, it achieves an AAR of 43.50%, representing a 2.2% relative improvement over previous SOTA method dyMEAN, and substantially outperforms earlier approaches like RosettaAb and HERN. Most notably, Igformer excels in docking performance, achieving an RMSD of 7.15 and a DockQ score of 0.450, marking a relative improvement of 15.08% and 10.29% respectively over the second-best method. Figure 3 provides a comparative visualization of predictions generated by Igformer and dyMEAN. These results demonstrate the enhanced capability of Igformer to capture antibody-antigen interactions and generate more accurate binding interface predictions.

2.3 Task 2: Multiple CDR Design

In this set of experiments, we compare Igformer against dyMEAN, the current leading method, on sequence and structure prediction across all six CDRs using the RAbD dataset. The evaluation examines structural accuracy through AAR for individual CDR regions and assesses overall antibody structure quality using additional metrics. As shown in Tables 2-3, Igformer achieves better sequence recovery results in 5 out of 6 CDRs except for CDR-L1, where Igformer maintains comparable performance with dyMEAN. Moreover, Igformer consistently surpasses dyMEAN in overall performance by substantial margins. Notably, Igformer achieves relative improvements of 5.83% in overall AAR and 21.24% in DockQ score. These comprehensive results validate the effectiveness of Igformer in both sequence prediction and structural generation across multiple CDR regions.

2.4 Task 3: Full Antibody Design

In this set of experiments, we evaluate predictions across all regions, including both frameworks and CDRs on RAbD dataset, where the sequences and coordinates for the entire antibody are masked in the testing phase. Table 3 presents the performance comparison for full antibody design. These results align with the fundamental characteristic of antibody structure: framework regions exhibit higher conservation across different antibodies, making them more predictable than hyper-variable CDR regions. Our experimental findings confirm this biological characteristic, as both models achieve higher sequence recovery rates in full antibody design compared to CDR-specific prediction tasks. Notably, Igformer demonstrates consistent superiority over dyMEAN across all evaluation metrics. Igformer achieves a 3.25% relative improvement in AAR and an 8.57% relative enhancement in DockQ score. These significant improvements indicate that Igformer not only generates more accurate

Table 4: Comparison on complex structure prediction. Results with † are predicted using the ground-truth (GT) structure and serve as the upper bound of this task.

Model	Structure		Docking	
	TMscore↑	lDDT↑	RMSD↓	DockQ↑
IgFold⇒HDock*	0.9701	0.8439	16.32	0.202
IgFold⇒HERN*	0.9702	0.8441	9.63	0.429
GT⇒HERN	1.0000†	1.0000†	9.65	0.432
dyMEAN	**0.973**	0.8676	9.00	0.452
Igformer	**0.973**	**0.8677**	**7.88**	**0.522**

Table 5: Ablation study on CDR-H3 design.

Model	Generation			Docking		
	AAR↑	TMscore↑	lDDT↑	CAAR↑	RMSD↓	DockQ↑
Igformer	**43.50%**	**0.9757**	**0.8650**	**28.11%**	**7.15**	**0.450**
- APP	42.80%	0.9725	0.8412	27.60%	8.30	0.416
- SGFormer	43.45%	0.9743	0.8578	28.00%	7.82	0.431
- TM	43.32%	0.9731	0.8499	27.91%	8.40	0.417
- AA	43.06%	0.9735	0.8460	27.66%	8.12	0.425
- Dual EMP	42.75%	0.9728	0.8430	27.50%	8.22	0.419

sequence predictions but also maintains higher structural fidelity, leading to improved antigen binding affinity prediction.

2.5 Task 4: Complex Structure Prediction

In this set of experiments, we evaluate the model performance on the Igfold benchmark Ruffolo et al. (2023a), focusing on complex structure prediction when given complete antibody sequences, including CDR-H3. Evaluation metrics include CDR-H3-specific RMSD and comprehensive structural assessments of the entire antibody complex.

As demonstrated in Table 4, Igformer establishes new state-of-the-art performance in the complex structure prediction task, particularly excelling in docking metrics. Specifically, Igformer achieves relative improvements of 12.44% in RMSD and 15.49% in DockQ scores, respectively. Notably, even when compared to GT⇒HERN, which utilizes ground-truth structures for docking, Igformer demonstrates superior docking performance. These results validate the effectiveness of the proposed inter-graph refinement strategy in capturing the intricate residue interactions within antibody-antigen binding interfaces.

2.6 Ablation Study

We conduct comprehensive ablation studies to evaluate the individual contributions of key components in Igformer. Specifically, we systematically analyze APP and SGFormer introduced in Section C.3, triangle multiplicative module (TM), axial attention (AA), and dual EMP module presented in Section C.4. Table 5 illustrates the impact of removing each component from Igformer on epitope-binding CDR-H3 prediction tasks. Results of experiments on Task 4 are provided in Appendix F.1.

Our observations are as follows. Removing the APP reduces performance across all metrics, particularly affecting the DockQ score (0.416) and RMSD (8.30). The removal of SGFormer shows similar trends but with less impact, achieving a DockQ score of 0.431 and RMSD of 7.82. These results demonstrate their crucial role in capturing complicated interface interactions. The triangle multiplicative module and axial attention mechanism both contribute significantly to model performance. Without TM, the AAR drops to 43.32% and DockQ to 0.417, while removing AA results in an AAR of 43.06% and DockQ of 0.425. Most notably, replacing our dual EMP architecture with a single message passing framework results in significant performance degradation across all metrics, with AAR dropping to 42.79% and DockQ to 0.417, validating our design choice of separate intra- and inter-graph message passing processes.

3 CONCLUSION

In this paper, we propose Igformer, an end-to-end antibody co-design framework that simultaneously generates both 1D sequences and 3D structures for epitope-binding antibody CDRs. Igformer advances existing models through a novel intra-graph refinement process, which is capable of capturing intricate residue interactions within the epitope-paratope binding region. Through extensive experimental evaluation, we demonstrate the effectiveness of Igformer, providing a new perspective for end-to-end antibody co-design tasks.

ACKNOWLEDGMENTS

Sibo Wang is supported by the RGC GRF grant (No. 14217322), Hong Kong ITC ITF grant (No. MRP/071/20X), and Tencent Rhino-Bird Focused Research Grant.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

Gregory P Adams and Louis M Weiner. Monoclonal antibody therapy of cancer. *Nature biotechnology*, 23(9):1147–1157, 2005.

Jared Adolf-Bryfogle, Oleks Kalyuzhniy, Michael Kubitz, Brian D Weitzner, Xiaozhen Hu, Yumiko Adachi, William R Schief, and Roland L Dunbrack Jr. Rosettaantibodydesign (rabd): A general framework for computational antibody design. *PLoS computational biology*, 14(4):e1006112, 2018.

Sarah Alamdari, Nitya Thakkar, Rianne van den Berg, Neil Tenenholtz, Bob Strome, Alan Moses, Alex Xijie Lu, Nicolo Fusi, Ava Pardis Amini, and Kevin K Yang. Protein generation with evolutionary diffusion: sequence is all you need. *BioRxiv*, pp. 2023–09, 2023.

Ethan C Alley, Grigory Khimulya, Surojit Biswas, Mohammed AlQuraishi, and George M Church. Unified rational protein engineering with sequence-based deep representation learning. *Nature methods*, 16(12):1315–1322, 2019.

Dror Baran, M Gabriele Pszolla, Gideon D Lapidoth, Christoffer Norn, Orly Dym, Tamar Unger, Shira Albeck, Michael D Tyka, and Sarel J Fleishman. Principles for computational design of binding antibodies. *PNAS*, 114(41):10900–10905, 2017.

Sankar Basu and Björn Wallner. Dockq: a quality measure for protein-protein docking models. *PloS one*, 11(8):e0161879, 2016.

Cyrus Chothia, Arthur M. Lesk, Anna Tramontano, Michael Levitt, Sandra J. Smith-Gill, Gillian Air, Steven Sheriff, Eduardo A. Padlan, David Davies, William R. Tulip, Peter M. Colman, Silvia Spinelli, Pedro M. Alzari, and Roberto J. Poljak. Conformations of immunoglobulin hypervariable regions. *Nature*, 342(6252):877–883, 1989.

Frédéric A. Dreyer, Daniel Cutting, Constantin Schneider, Henry Kenlay, and Charlotte M. Deane. Inverse folding for antibody sequence design using deep learning. *CoRR*, abs/2310.19513, 2023.

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 Res.*, 42 (D1):D1140–D1146, 2013.

Marc Feldmann and Ravinder N Maini. Tnf defined as a therapeutic target for rheumatoid arthritis and other autoimmune diseases. *Nature medicine*, 9(10):1245–1250, 2003.

Jefferson Foote and Greg Winter. Antibody framework residues affecting the conformation of the hypervariable loops. *Journal of molecular biology*, 224(2):487–499, 1992.

Haohuai He, Bing He, Lei Guan, Yu Zhao, Feng Jiang, Guanxing Chen, Qingge Zhu, Calvin Yu-Chian Chen, Ting Li, and Jianhua Yao. De novo generation of sars-cov-2 antibody cdrh3 with a pre-trained generative large language model. *Nature Communications*, 15(1):6867, 2024.

John Ingraham, Vikas Garg, Regina Barzilay, and Tommi Jaakkola. Generative models for graph-based protein design. In *NeurIPS*, 2019.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Antibody-antigen docking and design via hierarchical structure refinement. In *ICML*, pp. 10217–10227, 2022a.

Wengong Jin, Jeremy Wohlwend, Regina Barzilay, and Tommi Jaakkola. Iterative refinement graph neural network for antibody sequence-structure co-design. In *ICLR*, 2022b.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon AA Kohl, Andrew J Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *nature*, 596 (7873):583–589, 2021.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923, 1976.

Diederik P Kingma. Adam: A method for stochastic optimization. In *ICLR*, 2015.

Johannes Klicpera, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. In *ICLR*, 2019.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. In *ICML*, pp. 17409–17429, 2023a.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Conditional antibody design as 3d equivariant graph translation. In *ICLR*, 2023b.

Gideon D Lapidoth, Dror Baran, Gabriele M Pszolla, Christoffer Norn, Assaf Alon, Michael D Tyka, and Sarel J Fleishman. Abdesign: A n algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins: Structure, Function, and Bioinformatics*, 83(8): 1385–1406, 2015.

Marie-Paule Lefranc, Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet, and Gérard Lefranc. Imgt unique numbering for immunoglobulin and t cell receptor variable domains and ig superfamily v-like domains. *Developmental & Comparative Immunology*, 27(1):55–77, 2003.

Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *NeurIPS*, 35:9754–9767, 2022.

Ali Madani, Bryan McCann, Nikhil Naik, Nitish Shirish Keskar, Namrata Anand, Raphael R Eguchi, Po-Ssu Huang, and Richard Socher. Progen: Language modeling for protein generation. *arXiv preprint arXiv:2004.03497*, 2020.

Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lddt: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.

Karolis Martinkus, Jan Ludwiczak, Wei-Ching Liang, Julien Lafrance-Vanasse, Isidro Hotzel, Arvind Rajpal, Yan Wu, Kyunghyun Cho, Richard Bonneau, Vladimir Gligorijevic, et al. Abdiffuser: full-atom generation of in-vitro functioning antibodies. In *NeurIPS*, 2023.

Asher Mullard. 2021 fda approvals. *Nat. Rev. Drug Discov.*, 21(2):83–88, 2022.

Kenneth Murphy and Casey Weaver. *Janeway's immunobiology*. Garland science, 2016.

Jo Erika T Narciso, Iris Diana C Uy, April B Cabang, Jenina Faye C Chavez, Juan Lorenzo B Pablo, Gisela P Padilla-Concepcion, and Eduardo A Padlan. Analysis of the antibody structure based on high-resolution crystallographic studies. *New biotechnology*, 28(5):435–447, 2011.

RJ Pantazes and Costas D Maranas. Optcdr: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Engineering, Design & Selection*, 23(11):849–858, 2010.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, pp. 8024–8035, 2019.

Thiruvarangan Ramaraj, Thomas Angel, Edward A Dratz, Algirdas J Jesaitis, and Brendan Mumey. Antigen–antibody interface properties: Composition, residue interactions, and features of 53 non-redundant structures. *BBA-Proteins and Proteomics*, 1824(3):520–532, 2012.

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. *Nature communications*, 14(1):2389, 2023a.

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. *Nature communications*, 14(1):2389, 2023b.

Koichiro Saka, Taro Kakuzaki, Shoichi Metsugi, Daiki Kashiwagi, Kenji Yoshida, Manabu Wada, Hiroyuki Tsunoda, and Reiji Teramoto. Antibody design using lstm based deep generative model from phage display library for affinity maturation. *Scientific reports*, 11(1):5852, 2021.

Constantin Schneider, Matthew I J Raybould, and Charlotte M Deane. Sabdab in the age of biotherapeutics: updates including sabdab-nano, the nanobody structure tracker. *Nucleic Acids Res.*, 50 (D1):D1368–D1372, 2021.

Jung-Eun Shin, Adam J Riesselman, Aaron W Kollasch, Conor McMahon, Elana Simon, Chris Sander, Aashish Manglik, Andrew C Kruse, and Debora S Marks. Protein design and variant prediction using autoregressive generative models. *Nature communications*, 12(1):2403, 2021.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *JMLR*, 15(1):1929–1958, 2014.

Ariel Tennenhouse, Lev Khmelnitsky, Razi Khalaila, Noa Yeshaya, Ashish Noronha, Moshit Lindzen, Emily K Makowski, Ira Zaretsky, Yael Fridmann Sirkis, Yael Galon-Wolfenson, et al. Computational optimization of antibody humanness and stability by systematic energy-based ranking. *Nature Biomedical Engineering*, 8(1):30–44, 2024.

Yuko Tsuchiya and Kenji Mizuguchi. The diversity of h 3 loops determines the antigen-binding tendencies of antibody cdr loops. *Protein Science*, 25(4):815–825, 2016.

Jue Wang, Sidney Lisanza, David Juergens, Doug Tischer, Ivan Anishchenko, Minkyung Baek, Joseph L Watson, Jung Ho Chun, Lukas F Milles, Justas Dauparas, et al. Deep learning methods for designing proteins scaffolding functional sites. *BioRxiv*, pp. 2021–11, 2021.

Rubo Wang, Fandi Wu, Xingyu Gao, Jiaxiang Wu, Peilin Zhao, and Jianhua Yao. Iggm: A generative model for functional antibody and nanobody design. *bioRxiv*, 2024a.

Shixia Wang, Kun-Wei Chan, Danlan Wei, Xiuwen Ma, Shuying Liu, Guangnan Hu, Saeyoung Park, Ruimin Pan, Ying Gu, Alexandra F Nazzari, et al. Human cd4-binding site antibody elicited by polyvalent dna prime-protein boost vaccine neutralizes cross-clade tier-2-hiv strains. *Nature Communications*, 15(1):4301, 2024b.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *BioRxiv*, pp. 2022–12, 2022a.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *BioRxiv*, 2022b.

Qihong Yan, Xijie Gao, Banghui Liu, Ruitian Hou, Ping He, Yong Ma, Yudi Zhang, Yanjun Zhang, Zimu Li, Qiuluan Chen, et al. Antibodies utilizing vl6-57 light chains target a convergent cryptic epitope on sars-cov-2 spike protein and potentially drive the genesis of omicron variants. *Nature Communications*, 15(1):7585, 2024.

Yumeng Yan, Huanyu Tao, Jiahua He, and Sheng-You Huang. The hdock server for integrated protein–protein docking. *Nature protocols*, 15(5):1829–1852, 2020.

Yang Zhang and Jeffrey Skolnick. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic acids research*, 33(7):2302–2309, 2005.

Table of Contents

A Related Work **11**

B Problem Formulation **11**

C Method **12**
C.1 Initialization 12
C.2 Equivariant Message Passing 13
C.3 Inter-Graph Refinement 14
C.4 Iterative Update 14
C.5 Igformer Pipeline 16
C.6 Prediction and Loss Function 16

D Supplementary **16**
D.1 Graph Construction 16
D.2 Residue Embedding 17
D.3 Template-based Coordinates 17
D.4 Epitope Selection 19
D.5 Proofs 20
D.6 Algorithms 22
D.7 Loss Function 23

E Additional Experimental Settings **24**
E.1 Datasets 24
E.2 Implementation Details 25
E.3 Baselines 27
E.4 Metrics 27

F Additional Experimental Results **28**
F.1 Additional Ablation Study 28
F.2 Hyperparameter analysis 28
F.3 Additional Case Study of Task 1 29

A Related Work

Antibody Co-design. Antibody co-design methodologies facilitate the simultaneous generation of sequence and structure in an end-to-end manner, thereby enabling the modeling of intricate dependencies between backbone conformation and amino acid composition. Specifically, RefineGNN Jin et al. (2022b) utilized an iterative refinement approach, alternately predicting atom coordinates and residue types in CDRs through auto-regression. DiffAb Luo et al. (2022) advanced the field by introducing a diffusion-based framework that explicitly targets specific antigens while considering atomic-level structures, enabling both sequence and structure generation for CDRs based on framework regions and target antigens. MEAN Kong et al. (2023b) and its successor dyMEAN Kong et al. (2023a) leveraged graph neural networks with E(3)-equivariant architectures for CDR prediction, with dyMEAN extending to full-atom geometry modeling including both backbone and side chains. AbDiffuser Martinkus et al. (2023) introduced an efficient SE(3)-equivariant architecture called APMixer for complete antibody structure generation, demonstrating success in wet-lab validation. Most recently, IgGM Wang et al. (2024a) expanded capabilities to end-to-end antibody design through a hybrid diffusion model that can generate complete antibody sequences and structures simultaneously without requiring template structures.Despite these advances, existing methods have largely simplified or overlooked the complex interactions at antibody-antigen binding interfaces. Our work addresses this limitation through a novel inter-graph refinement strategy that integrates personalized propagation into global attention mechanisms, enabling more accurate modeling of binding interface dynamics.

Sequence Design. The field of sequence design has evolved from LSTM-based approaches to more advanced language models. LSTM-based antibody design Saka et al. (2021) pioneered sequence generation for affinity maturation by employing a long short-term memory network trained on enriched phage display sequences to generate and prioritize antibody variants, using likelihood scores from the trained model to identify promising candidates and outperforming traditional frequency-based screening approaches. ProGen Madani et al. (2020) introduced a controllable protein generation language model trained on 280M sequences, enabling both high-quality sequence generation and fine-grained control through conditioning. The model employed a transformer architecture and demonstrated the ability to generate proteins with native-like structural and functional properties. More recently, EvoDiff Alamdari et al. (2023) introduced an evolutionary-scale diffusion framework operating directly in sequence space. The model combines discrete diffusion with evolutionary-scale sequence data to enable both unconditional generation of diverse, structurally-plausible proteins and flexible conditional generation via sequence inpainting.

Structure Design. Structure design has progressed from single-step prediction to sophisticated diffusion-based approaches. RFjoint Wang et al. (2021) initially pioneered functional site scaffolding with structure prediction networks, though limited by its single-step prediction approach. RFdiffusion Watson et al. (2022a) advanced this approach by fine-tuning the RoseTTAFold structure prediction network as a denoising network within a diffusion model framework, enabling iteratively protein structure building from random noise. IgFold Ruffolo et al. (2023a) introduced a specialized antibody structure prediction framework combining a pre-trained language model (AntiBERTy) with graph neural networks. This innovation enabled direct prediction of backbone atom coordinates, achieving accuracy comparable to AlphaFold 2 Jumper et al. (2021) while being significantly faster and eliminating the need for multiple sequence alignments or template structures. Most recently, AlphaFold 3 Abramson et al. (2024) has taken a broader approach by introducing a unified deep-learning framework capable of predicting structures for diverse biomolecular complexes. It achieves this by replacing evoformer in Alphafold2 with a simpler Pairformer and employing a diffusion-based structure module for direct atom coordinates prediction, representing a significant step toward a more generalized solution for biomolecular structure prediction.

B Problem Formulation

Antibodies are Y-shaped proteins with a distinctive molecular architecture: they comprise heavy and light chains, each containing constant domains that are conserved across antibodies and variable domains that determine binding specificity. Within these variable domains, the structure alternates between framework regions (FRs) that maintain structural stability and *complementarity-determining regions (CDRs)* that are responsible for antigen recognition, with CDR positions defined by the IMGT numbering scheme Lefranc et al. (2003). Among these regions, the CDR-H3 loop in the heavy chain

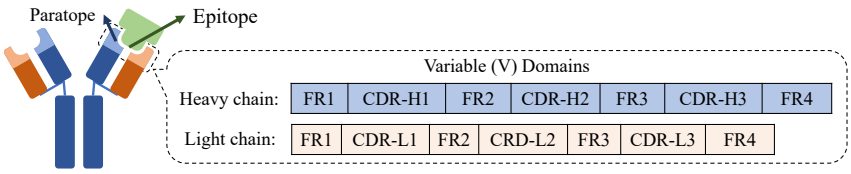


Figure 4: Illustration of antibody-antigen complex.

variable domain exhibits the highest variability and plays a pivotal role in antigen recognition Narciso et al. (2011); Tsuchiya & Mizuguchi (2016); Dreyer et al. (2023). In the context of antibody-antigen interactions, the binding surface of the antibody is termed the *paratope*, while its corresponding target region on the antigen is called the *epitope*, as illustrated in Figure 4.

The objective of antibody co-design is to simultaneously predict the amino acid sequence and 3D structure of antibody CDRs that optimally bind a given epitope on a target antigen, as illustrated in Figure 1. We formulate this problem using a dual-graph representation of the binding region. The first component is an antibody graph $G_{ab} = \{V_{ab}, E_{ab}\}$, while the second is an antigen epitope graph $G_{ae} = \{V_{ae}, E_{ae}\}$. The vertices V_{ab} and V_{ae} represent amino acid residues, where each residue v_i is characterized by its amino acid type s_i and a coordinate matrix $\boldsymbol{X}_i \in \mathbb{R}^{3\times c_i}$ encoding the 3D positions of its c_i atoms, including both backbone and side chain atoms. Spatial proximities between residues are captured through edges E_{ab} and E_{ae} constructed using a *k-nearest neighbors (kNN)* approach over all atomic positions. The complete graph construction process is detailed in Appendix D.1. Within this framework, we denote the paratope residues as a subset $V_p \subseteq V_{ab}$ of the antibody graph, where we specifically focus on *CDR-H3* as the paratope region.

Definition B.1. Given partial antibody sequence $\{s_i \mid v_i \in V_{ab}, v_i \notin V_p\}$ and the target antigen epitope sequence $\{s_j \mid v_j \in V_{ae}\}$ with corresponding structural information $\boldsymbol{X}_{ab}$ and $\boldsymbol{X}_{ae}$, the goal of antibody co-design is to predict both the 1D sequence $\{s_k \mid v_k \in V_p\}$ of the paratope and 3D structure $\{\boldsymbol{X}_k \mid v_k \in V_p\}$ of the entire antibody in the context of the antibody-antigen complex.

This formulation defines a joint optimization problem where the model simultaneously predicts both the amino acid sequence of the antibody paratope and its complete 3D coordinates within the antibody-antigen binding region. Through this formulation, we aim to harness the expressive power of equivariant graph neural architectures to capture the intricate spatial and chemical relationships that govern antibody-antigen interactions, enabling end-to-end structure-based antibody co-design.

C METHOD

In this section, we present Igformer, an end-to-end framework for antibody co-design that introduces a novel approach to modeling antibody-antigen interactions. At its core, Igformer advances previous attempts through a sophisticated inter-graph refinement mechanism that combines personalized propagation with global attention, enabling more accurate modeling of complex interactions within antibody-antigen binding interfaces. As illustrated in Figure 2, Igformer consists of three key components: (i) Equivariant Message Passing module, (ii) Inter-graph Refinement module, and (iii) Iterative Update module. In the following, we will detail the initialization process in Section C.1, elaborate on the Equivariant Message Passing module in Section C.2, followed by the Inter-Graph Refinement module in Section C.3, and the Iterative Update module in Section C.4. We then present the Igformer pipeline in Section C.5 and learning objectives in Section C.6.

C.1 INITIALIZATION

The foundation of Igformer lies in its comprehensive representation scheme that integrates both biochemical and structural information through embedding and coordinate initialization steps.

Feature initialization. For each residue v_i, we construct a dual-component embedding: a residue embedding $\boldsymbol{H}_i^{res} \in \mathbb{R}^d$ that encodes amino acid properties, and a position embedding $\boldsymbol{H}_i^{pos} \in \mathbb{R}^d$ that captures sequential context through distinct indices for antigen, heavy chain, and light chain regions. These components are combined additively to form the final residue representation: $\boldsymbol{H}_i = \boldsymbol{H}_i^{res} + \boldsymbol{H}_i^{pos}$, This initialization approach enables the simultaneous processing of amino acid

properties and their structural context while differentiating between functionally distinct antibody-antigen complex regions. More details of embedding initialization are provided in Appendix D.2.

Coordinate initialization. Igformer employs a structured coordinate initialization strategy that combines template-based modeling with precise alignment and normalization procedures. For the epitope region V_{ae}, we maintain the actual experimental coordinates, while other regions are initialized using templates (refer to Appendix D.3.1) that provide positions of backbone atoms (N, CA, C, O). Missing template coordinates are estimated through linear interpolation between the nearest known template positions along the sequence. We then optimize these initial structures through rigid-body alignment using the Kabsch algorithm Kabsch (1976), which minimizes structural discrepancies between the template and known coordinates. Following initialization, we apply a chain-specific processing pipeline to establish a consistent reference frame. Coordinates of each chain are centered relative to their respective virtual center of mass, and subsequently normalized to ensure uniform scale across the complex. This carefully calibrated initialization process provides a robust foundation for subsequent equivariant message passing and inter-graph refinement stages. For a comprehensive description of the coordinate initialization and processing procedures, we refer readers to Appendix D.3.

C.2 Equivariant Message Passing

Given input coordinates $\boldsymbol{X}$ and embeddings $\boldsymbol{H}$ of residues in a graph G, at the l-th layer, the *equivariant message passing (EMP)* module updates residue representations and coordinates $\boldsymbol{H}_i^{(l+1)}, \boldsymbol{X}_i^{(l+1)} = \textit{EMP}\left(\boldsymbol{H}_i^{(l)}, \boldsymbol{X}_i^{(l)}\right)$ through four sequential steps.

First, we compute multi-level pairwise residue similarities by combining residue and atomic information:

$$\Delta \boldsymbol{X}_{ij}^{(l)} = \boldsymbol{X}_i^{(l)} - \boldsymbol{X}_j^{(l)},\ sim_{ij}^{res} = \Delta \boldsymbol{X}_{ij}^{(l)} (\Delta \boldsymbol{X}_{ij}^{(l)})^\top,$$

$$sim_{ij(m,n)}^{atom} = \sum_{c=1}^{3} (\boldsymbol{X}_{i,m,c} - \boldsymbol{X}_{j,n,c})(\boldsymbol{X}_{i,m,c} - \boldsymbol{X}_{j,n,c})^\top,$$

where $m, n \in \{1, \cdots, 14\}^2$ denote different atoms in the residue. The multi-level positional information is processed through MLPs and combined into a similarity matrix:

$$\boldsymbol{S}_{i,j}^{(l)} = w \cdot \textit{MLP}_1(sim_{ij}^{res}) + (1 - w) \cdot \textit{MLP}_2(sim_{ij}^{atom}).$$

Next, we update edge features by incorporating node information and similarity scores:

$$\boldsymbol{H}_{e_{ij}}^{(l+1)} = \textit{EdgeMLP}\left(\boldsymbol{H}_i^{(l)} \oplus \boldsymbol{H}_j^{(l)} \oplus \boldsymbol{S}_{i,j}^{(l)} \oplus \boldsymbol{H}_{e_{ij}}^{(l)}\right),$$

where $\oplus$ denotes concatenation operation. The coordinates are subsequently updated through weighted aggregation of neighbor differences:

$$\boldsymbol{X}_i^{(l+1)} = \boldsymbol{X}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \Delta \boldsymbol{X}_{ij}^{(l)} \cdot \textit{CoordMLP}(\boldsymbol{H}_{e_{ij}}^{(l+1)}),$$

where $\mathcal{N}_i$ denotes the neighboring nodes of v_i in graph G. Finally, node features are updated by aggregating neighboring edge information with residual connections:

$$\boldsymbol{H}_i^{(l+1)} = \boldsymbol{H}_i^{(l)} + \textit{NodeMLP}\left(\boldsymbol{H}_i^{(l)} \oplus \sum_{j \in \mathcal{N}(i)} \boldsymbol{H}_{e_{ij}}^{(l+1)}\right)$$

This E(3)-equivariant (refer to Definition D.1 in Appendix D.5.1) message passing scheme ensures that both spatial and biochemical properties are properly captured and updated during the message passing process.

Theorem C.1. *For any transformation $T \in E(3)$, we have $\boldsymbol{H}_i^{(l+1)}, T(\boldsymbol{X}_i^{(l+1)}) = \textit{EMP}\left(\boldsymbol{H}_i^{(l)}, T(\boldsymbol{X}_i^{(l)})\right)$, where $T(X) := \boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}$ denotes the E(3) transformation of $\boldsymbol{X}$.*

All proofs of theorems are provided in Appendix D.5.

[2]Each residue is represented using 14 atoms in 3D space.

C.3 INTER-GRAPH REFINEMENT

The inter-graph $G_{inter} = (V_p, V_{ae}, E_{inter})$ represents interactions in the antibody-antigen binding interface between antibody paratope residues V_p and antigen epitope residues V_{ae}. The key innovation of Igformer lies in its sophisticated inter-graph refinement strategy, which combines personalized propagation with global attention to capture complex antibody-antigen interactions. Our approach integrates two complementary components: *approximate personalized propagation (APP)* for preserving local structural information and *simplified graph transformer (SGFormer)* for capturing global dependencies. Next, we elaborate on the graph generation and refinement process.

Approximate Personalized Propagation. We construct intra-graphs G_{ab} and G_{ae} by establishing connectivity between k-nearest residues based on atomic-level coordinate comparisons, ensuring that graph connectivity reflects spatial relationships within both antibody and antigen components. A detailed description of the intra-graph construction process is provided in Appendix D.1.

Given the residue embeddings $\boldsymbol{H}^{(0)}$ updated by EMP in the intra-graph, APP performs iterative message passing while maintaining a balance between local and global information:

$$\boldsymbol{H}^{(j+1)} = (1-\alpha) \cdot \hat{\boldsymbol{P}} \boldsymbol{H}^{(j)} + \alpha \cdot \boldsymbol{H}^{(0)}, \tag{1}$$

where $\hat{\boldsymbol{P}} = \hat{\mathbf{D}}^{-1/2} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-1/2}$ denotes the normalized transition matrix in the intra-graph with self-loops, and α controls the retention of initial node features. This propagation scheme enables the updated embeddings to capture neighborhood context while preserving crucial local structural information Klicpera et al. (2019).

Simplified Graph Transformer. Subsequently, we introduce a Simplified Graph Transformer (SGFormer) component that implements a global attention mechanism to dynamically model complex interactions between residues across the entire binding interface $V_{inter} = V_p \cup V_{ae}$. Specifically, for a pair of residues (v_i, v_j) in the given antibody-antigen binding region, we compute the attention weight:

$$a_{ij} = \frac{\exp\left(\boldsymbol{H}_i^\top \boldsymbol{W} \boldsymbol{H}_j\right)}{\sum_{k \in V} \exp\left(\boldsymbol{H}_i^\top \boldsymbol{W} \boldsymbol{H}_k\right)},$$

where $\boldsymbol{W}$ is a learnable weight matrix. Residue embeddings are updated through attention-weighted aggregation:

$$\boldsymbol{H}_i = \sum_{j \in V} a_{ij} \cdot \boldsymbol{H}_j.$$

This self-attention mechanism allows SGFormer to globally capture interactions between all residues in the binding region, thereby capturing both local information and long-range dependencies.

Inter-Graph Refinement. To model the binding interface dynamics, we encode bi-directional interactions between antibody paratope and antigen epitope residues into edge embeddings. For each residue pair (v_i, v_j), where v_i belongs to the antibody paratope and v_j to the antigen epitope. The updated pair-wise distance $dist_{ij}$ is:

$$dist_{ij} = \textit{EdgeMLP}\left(\boldsymbol{H}_i \oplus \boldsymbol{H}_j, \boldsymbol{H}_j \oplus \boldsymbol{H}_i\right).$$

These learned distances capture the complex chemical and spatial relationships between residues, enabling dynamic refinement of the inter-graph structure through k-nearest neighbor selection. The complete inter-graph refinement process is detailed in Appendix D.1.

Remark C.2. The integration of APP and SGFormer enables comprehensive refinement of both residue and edge representations. APP preserving crucial local structural properties while facilitating efficient information propagation, and the attention mechanism in SGFormer captures dynamic residue interactions across the entire binding interface. This dual-approach enables precise modeling of both local chemical interactions and long-range structural dependencies, which is crucial for accurate antibody structure prediction, as we will show in our experiments.

C.4 ITERATIVE UPDATE

Our framework implements a sophisticated iterative learning process that incorporates triangle attention mechanism Jumper et al. (2021) with a dual-EMP module to capture complex interactions within antibody-antigen binding interfaces. The process begins with the initial embeddings $\boldsymbol{H}_{full} =$

$\boldsymbol{H}_{ae} \cup \boldsymbol{H}_{ab}$ and extracts the interface-specific representations $\boldsymbol{H}_{inter} = \boldsymbol{H}_{ae} \cup \boldsymbol{H}_p$ from $\boldsymbol{H}_{full}$. The corresponding coordinates $\boldsymbol{X}_{full} = \boldsymbol{X}_{ae} \cup \boldsymbol{X}_{ab}$ and $\boldsymbol{X}_{inter} = \boldsymbol{X}_{ae} \cup \boldsymbol{X}_p$ are initialized according to Appendix D.3. For clarity, we denote $\boldsymbol{H}_{full}$ as $\boldsymbol{H}_{intra}$ and $\boldsymbol{X}_{full}$ as $\boldsymbol{X}_{intra}$ in subsequent discussions.

Triangle Multiplicative Module. Within the paratope region V_p containing n residues, we construct a normalized interaction matrix $\boldsymbol{Z} \in \mathbb{R}^{n \times n}$ to encode pairwise residue relationships for each residue pair (v_i, v_j):

$$\boldsymbol{Z}_{ij} = \textit{LayerNorm}\left(\textit{MLP}\left([\boldsymbol{H}_i \oplus \boldsymbol{H}_j]\right)\right).$$

The concatenation operation $\oplus$ ensures directional sensitivity through its non-commutative nature. This interaction matrix undergoes iterative refinement through two complementary mechanisms: triangle multiplicative module and axial attention.

The triangle multiplicative module processes embeddings through learnable normalized projections to generate pair-wise residue interactions in the paratope:

$$\boldsymbol{H}_l = \boldsymbol{W}_l \boldsymbol{Z}, \quad \boldsymbol{H}_r = \boldsymbol{W}_r \boldsymbol{Z}, \quad \boldsymbol{I}_{ij} = (\boldsymbol{H}_l)_i^\top (\boldsymbol{H}_r)_j,$$

where $\boldsymbol{W}_l$ and $\boldsymbol{W}_r$ are learnable weight matrices. The resulting pair-wise interactions are then modulated by a gating mechanism and projected into the embedding space:

$$f_{triangle}^{out}(\boldsymbol{H}) = \boldsymbol{W}_o \cdot \sigma(\boldsymbol{W}_g \boldsymbol{Z}) \cdot \boldsymbol{I},$$

where $\sigma(\cdot)$ is the sigmoid function.

Axial Attention. The row-wise axial attention mechanism computes dynamic residue relationships through a scaled dot-product attention:

$$f_{att}^{out}(\boldsymbol{H}) = \text{softmax}\left(\frac{\boldsymbol{Q}^\top \boldsymbol{K}}{\sqrt{d_k}}\right) \cdot \boldsymbol{V},$$

where residue embeddings in $\boldsymbol{Q} = \boldsymbol{W}_q \boldsymbol{H}$, $\boldsymbol{K} = \boldsymbol{W}_k \boldsymbol{H}$, and $\boldsymbol{V} = \boldsymbol{W}_v \boldsymbol{H}$ are generated by $f_{triangle}^{out}(\boldsymbol{H})$, and $\boldsymbol{W}_q$, $\boldsymbol{W}_k$, and $\boldsymbol{W}_v$ are learnable parameters. The final paratope outgoing representation integrates both attention and triangle multiplicative mechanism:

$$f_{out}(\boldsymbol{H}) = f_{triangle}^{out}(\boldsymbol{H}) + f_{att}^{out}(\boldsymbol{H}).$$

The representations of the paratope region are updated iteratively integrating both outgoing and incoming representations:

$$\boldsymbol{H}^{(k+1)} = \boldsymbol{H}^{(k)} + f_{out}(\boldsymbol{H}^{(k)}) + f_{in}(\boldsymbol{H}^{(k)}), \tag{2}$$

where $f_{in}(\boldsymbol{H})$ represents the ingoing update analogous to the outgoing formulation but adopts ingoing (column-wise) multiplicative interactions and column-wise attention. This bidirectional updating process ensures that paratope embedding $\boldsymbol{H}$ captures both pairwise geometric interactions via $f_{triangle}$ and dynamic global dependencies via f_{att}.

After Ks iterations, the diagonal elements of paratope embeddings calculated by Equation 2 are extracted to generate the final embeddings of paratope:

$$\boldsymbol{H}_p = \textit{MLP}\left(\textit{diag}\left(\boldsymbol{H}^{(K)}\right)\right), \tag{3}$$

Dual EMP Module. Next, we employ a dual-scale message-passing framework using two separate EMP modules to model intra-graph G_{intra} and inter-graph G_{inter} interactions:

$$\begin{aligned} \left(\boldsymbol{H}_{intra}^{(t+1)}, \boldsymbol{X}_{intra}^{(t+1)}\right) &= \textit{EMP}_{intra}\left(\boldsymbol{H}_{intra}^{(t)}, \boldsymbol{X}_{intra}^{(t)}\right), \\ \left(\boldsymbol{H}_{inter}^{(t+1)}, \boldsymbol{X}_{inter}^{(t+1)}\right) &= \textit{EMP}_{inter}\left(\boldsymbol{H}_{inter}^{(t)}, \boldsymbol{X}_{inter}^{(t)}\right). \end{aligned} \tag{4}$$

At each iteration, we update our graph representations according to Equation 4, and substitute the paratope embeddings in the intra-graph $\boldsymbol{H}_{intra}^{(t+1)}$ using $\boldsymbol{H}_{inter}^{(t+1)}$. On the other hand, coordinates for intra- and inter-graph are maintained separately. This dual-scale message-passing framework enables comprehensive modeling of both local residue relationships and global antigen-antibody interactions, with each scale optimized for its specific context. Finally, residue connections are incorporated in the updated embeddings for stable learning across different training iterations.

C.5 IGFORMER PIPELINE

We present the key stages of the Igformer learning pipeline, with detailed training algorithms provided in Appendix D.6. The process begins with the initialization of residue representations and coordinates. Following this, the intra-graph is constructed for personalized information propagation, which is then processed by the SGFormer for inter-graph refinement. The embeddings generated in this process serve exclusively for inter-graph refinement and are not involved in the representation learning process. Concurrently, a triangle attention module updates the representation of each residue in the paratope. Finally, a dual EMP module processes residue representations to generate the final coordinates and representations, which are then used for downstream structure and sequence prediction tasks.

The following theorem indicates that the coordinates generated by Igformer are E(3)-equivariant and the residue embeddings are invariant.

Theorem C.3. *Let $\hat{\boldsymbol{H}}_i, \hat{\boldsymbol{X}}_i = \textit{Igformer}\left(\boldsymbol{H}_i^{(0)}, \boldsymbol{X_i}\right)$ denote the embedding and coordinate of v_i generated by Igformer. For any transformation $T \in E(3)$, we have $\hat{\boldsymbol{H}}_i, T(\hat{\boldsymbol{X}}_i) = \textit{Igformer}\left(\boldsymbol{H}_i^{(0)}, T(\boldsymbol{X_i})\right)$, where $T(X) := \boldsymbol{QX} + \boldsymbol{b}$ denotes the E(3) transformation of $\boldsymbol{X}$.*

C.6 PREDICTION AND LOSS FUNCTION

Sequence Prediction. The i-th position in the amino acid sequence of the paratope is predicted using its embedding:

$$s_i = \textit{softmax}\left(\textit{MLP}\left(\boldsymbol{H}_i\right)\right)$$

Structure Prediction. The 3D structures of paratope are generated by the coordinates $\boldsymbol{X}$ computed in the last round of Dual EMP Module via Equation 4.

Loss Function. The loss function of Igformer consists of three components:

$$\mathcal{L} = \mathcal{L}_{seq} + \mathcal{L}_{struct} + \mathcal{L}_{interface} \tag{5}$$

Here, $\mathcal{L}_{seq}$ is the cross entropy loss that minimizes the dissimilarities between predicted and original 1D sequences. $\mathcal{L}_{struct}$ minimizes the difference between reconstructed and ground-truth 3D structures. $\mathcal{L}_{interface}$ optimizes the reconstructed inter-graph. Detailed descriptions of each component are provided in Appendix D.7.

D SUPPLEMENTARY

D.1 GRAPH CONSTRUCTION

Intra-Graph Construction. The intra-graph construction process integrates both sequence and structural information from the antibody-antigen complex. We begin by extracting information from the antigen epitope G_{ae} and antibody G_{ab}, where each residue i serves as a graph node v_i. Each node encapsulates both its amino acid type and 3D coordinate information, providing a comprehensive representation of its chemical and spatial properties. We formalize the intra-graph as $G_{Intra} = (V_{Intra}, E_{Intra})$, where $V_{Intra} = V_{ae} \cup V_{ab}$ represents the union of antigen epitope residues V_{ae} and antibody residues V_{ab}. Edge construction follows a distance-based approach, connecting residues within their respective chains (either epitope or antibody). For each residue v_i, we establish connections to its k-nearest neighbors v_j within the same chain based on Euclidean distances between their spatial coordinates. The shortest distance between two residues is defined as the minimum distance between any pair of atoms across two residues. This construction ensures that the intra-graph captures meaningful spatial relationships within both the antigen epitope and antibody domains while maintaining their distinct molecular contexts.

Inter-Graph Construction The inter-graph construction process integrates both sequence and structural information from the antibody-antigen binding region. We construct an inter-graph $G_{Inter} = (V_p, V_{ae}, E_{Inter})$ to model critical paratope-epitope interactions between antibody paratope residues V_p and antigen epitope residues V_{ae}. Edge construction in the inter-graph is based on embedding similarity between residue pairs. For paratope residues $v_i \in V_p$ and $v_j \in V_{ae}$ with embeddings $\boldsymbol{H}_i$ and $\boldsymbol{H}_j$, we compute bidirectional interaction features:

$$\boldsymbol{z}_{in} = \boldsymbol{H}_i \oplus \boldsymbol{H}_j, \quad \boldsymbol{z}_{out} = \boldsymbol{H}_j \oplus \boldsymbol{H}_i.$$

Table 6: Well-conserved residue positions used in template generation.

Chain	Positions	Count
Heavy (H)	8, 15, 23, 41, 44, 50, 52, 98, 100, 102, 104, 118, 119, 121, 126, 127	16
Light (L)	5, 6, 16, 23, 41, 70, 75, 76, 79, 89, 91, 98, 102, 104, 118, 119, 121, 122	18

These concatenated embeddings are processed through a feed-forward neural network (FFN) to generate interaction scores:

$$FFN(\boldsymbol{z}) = \sigma(\boldsymbol{W}_2 \cdot \sigma(\boldsymbol{W}_1 \cdot \boldsymbol{z} + \boldsymbol{b}_1) + \boldsymbol{b}_2), \tag{6}$$

where $\boldsymbol{W}_1 \in \mathbb{R}^{d\times 2d}$, $\boldsymbol{W}_2 \in \mathbb{R}^{d\times d}$, and $\boldsymbol{b}_1, \boldsymbol{b}_2 \in \mathbb{R}^d$ are learnable parameters, σ is the activation function. The final pair-wise distance of (v_i, v_j) combines interactions from both directions:

$$dist_{ij} = FFN(\boldsymbol{z}_{in}) + FFN(\boldsymbol{z}_{out}).$$

The edge set E_{inter} is then constructed by selecting k-nearest residues of each $v_i \in V_p$ based on these interaction scores. This construction focuses exclusively on paratope-epitope interactions, providing a refined inter-graph structure of the binding interface that captures the essential dynamics of antibody-antigen docking.

D.2 Residue Embedding

Our embedding framework captures comprehensive residue characteristics through two complementary components: residue-type and positional embeddings. First, we encode residue-type information through $\boldsymbol{H}_i^{res} \in \mathbb{R}^d$, which represents the biochemical properties of amino acid residue v_i. This embedding resides in a matrix of dimensions $N_{res} \times d$, where $N_{res} = 20$ represents the number of different amino acid types.

To preserve structural context, we implement position-specific embeddings $\boldsymbol{H}_i^{pos} \in \mathbb{R}^d$ that encode the sequential context of residues. This positional encoding maintains distinct indices for antigen, heavy chain, and light chain segments, enabling the model to differentiate between different components. The position embedding matrix has dimensions $N_{seq} \times d_{pos}$, where $N_{seq} = 192$ is the maximum length of the sequence, ensuring consistent positional representation across all residue types while maintaining chain-specific contexts.

Importantly, different segments of the sequence, such as the antigen, heavy chain, and light chain, utilize distinct position indices to allow the model to differentiate between these regions.

The final residue representation integrates both biochemical and positional information through a simple addition:

$$\boldsymbol{H}_i = \boldsymbol{H}_i^{res} + \boldsymbol{H}_i^{pos}. \tag{7}$$

D.3 Template-based Coordinates

D.3.1 Template Generation

The template generation process exploits a fundamental characteristic of antibody structures: the high conservation within framework regions (FRs). We define a residue as well-conserved when its amino acid type is maintained across at least 95% of antibodies in the dataset - a threshold optimized through empirical analysis to balance the number of conserved residues against their structural variability Kong et al. (2023a). Using this criterion, we identify 16 and 18 well-conserved residues in the heavy and light chains, respectively. To construct template coordinates $\boldsymbol{X}^{temp}$, we compute mean backbone atom coordinates (N, CA, C, O) in the main chains across all antibodies in the dataset, establishing an initial backbone template $\{\boldsymbol{X}_r \in \mathbb{R}^{3\times 4} | r \in \mathcal{W}_{temp}\}$, where $\mathcal{W}_{temp}$ denotes the positions of well-conserved residues as detailed in Table 6.

D.3.2 Coordinate Initialization

Our coordinate initialization strategy employs a template-based approach that preserves actual coordinates for the epitope region V_e while utilizing template-derived coordinates for paratope regions requiring prediction.

Each residue is represented using 14 atoms in 3D space, with coordinates denoted as $\boldsymbol{X}_{ae} = \{\boldsymbol{X}_i \in \mathbb{R}^3 \mid v_i \in V_{ae}\}$ for known positions and $\boldsymbol{X}^{temp} = \{\boldsymbol{X}_r \in \mathbb{R}^3 \mid j \in \mathcal{W}_{temp}\}$ for template-derived positions (refer to Appendix D.3.1).

For residues with missing coordinates in the template, linear interpolation is employed. The missing coordinate $\boldsymbol{X}_k$ for position k is calculated as follows:

$$\boldsymbol{X}_k = \boldsymbol{X}_{left} + \frac{k - left}{right - left}(\boldsymbol{X}_{right} - \boldsymbol{X}_{left}),$$

where *left* and *right* denote indices of nearest known template coordinates. The intersection of these indices is denoted as $\mathcal{W}_{align} = \mathcal{W}_{true} \cap \mathcal{W}_{temp}$.

To ensure interpolation remains within valid sequence boundaries, *left* and *right* are constrained as follows: $0 \leq left < k < right \leq N_{seq}$, where N_{seq} is the total length of the original antibody sequence. This ensures that interpolation remains within the structural constraints of the antibody sequence while preserving the geometric continuity of the coordinates.

To ensure structural coherence, we align template coordinates with true coordinates by optimizing a rigid-body transformation for shared indices $\mathcal{W}_{align}$. The optimal transformation, comprising rotation matrix $\boldsymbol{Q} \in \mathbb{R}^{3\times3}$ and translation vector $\boldsymbol{t} \in \mathbb{R}^3$, is determined by solving the following optimization problem:

$$\arg\min_{\boldsymbol{Q},\boldsymbol{t}} \sum_{i\in\mathcal{W}_{Align}, j\in\mathcal{W}_{temp}} \left|\boldsymbol{X}_i - (\boldsymbol{Q}\boldsymbol{X}_j + \boldsymbol{t})\right|^2, \tag{8}$$

subject to $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}$. After interpolation, we obtain the complete antibody coordinates as

$$\boldsymbol{X}_{ab}^{'} = \boldsymbol{X}^{temp} \cup \boldsymbol{X}^{interp}, \tag{9}$$

where $\boldsymbol{X}^{interp}$ represents the set of interpolated coordinates for positions that are within the sequence range but absent from the template. Formally,

$$\boldsymbol{X}^{interp} = \{\boldsymbol{X}_k \mid k = \{0, \cdots, N_{seq} - 1\} \text{ and } k \notin \mathcal{W}_{temp}\}. \tag{10}$$

This ensures that interpolation is applied only to residues that are part of the original antibody sequence but do not have known coordinates in the template. The resulting transformation for alignment is then applied globally:

$$\boldsymbol{X}_{ab} = \boldsymbol{Q}\boldsymbol{X}_{ab}^{'} + \boldsymbol{t}, \tag{11}$$

This initialization procedure primarily focuses on backbone atoms (N, CA, C, O), with side-chain atoms initially positioned at their respective CA coordinates, establishing a foundation for subsequent refinement.

D.3.3 Coordinate Processing

Following initial coordinate assignment, we implement a comprehensive normalization procedure to ensure consistent scale and reference frame across the antibody-antigen complex $\boldsymbol{X}$. For a system with N total atoms, we first define an indicator function $\mathbb{I}(i)$ that distinguishes between antigen and antibody atoms:

$$\mathbb{I}(i) = \begin{cases} 1 & \textit{if } v_i \in V_{ae} \\ 0 & \textit{if } v_i \in V_{ab} \end{cases}. \tag{12}$$

Using this indicator function, we compute chain-specific centers of mass:

$$\boldsymbol{X}_c = \frac{\sum_{i=1}^{N} \boldsymbol{X}_i \cdot \mathbb{I}(i)}{\sum_{i=1}^{N} \mathbb{I}(i)}. \tag{13}$$

The coordinates are then centered relative to their respective chain centers $\boldsymbol{X}_i^{'} = \boldsymbol{X}_i - \boldsymbol{X}_c$. Finally, we apply dimension-specific normalization to ensure consistent scale:

$$\boldsymbol{X}_i = \frac{\boldsymbol{X}_i^{'} - \mu_d}{\sigma_d},$$

where μ_d and σ_d represent the mean and standard deviation along dimension d across all atoms, with a target standard deviation of $\sigma = 10$. This processing pipeline ensures that all atomic coordinates are properly scaled and centered within their respective molecular contexts, facilitating subsequent structural prediction tasks. Finally, we update the global coordinates again based on Equations 12 and 13.

D.3.4 PARATOPE COORDINATES GENERATION

Our paratope coordinate generation framework implements a structured approach to initialize and refine binding interface positions. The process begins by anchoring initial paratope coordinates to the epitope center $\boldsymbol{X}_0 = \boldsymbol{X}_c$ where $\boldsymbol{X}_0$ represents the initial paratope coordinates and $\boldsymbol{X}_c$ represents the coordinates of the epitope center calculated by Equation 13, which are assigned as initial positions for all atoms in the paratope region. We then introduce controlled structural variations:

$$\boldsymbol{\epsilon}_i = \begin{cases} \boldsymbol{\epsilon}_{ca} & \textit{for CA atoms} \\ 0.1\boldsymbol{\epsilon}_o + \boldsymbol{\epsilon}_{ca} & \textit{for other atoms} \end{cases},$$

where $\boldsymbol{\epsilon}_o, \boldsymbol{\epsilon}_{ca} \sim \mathcal{N}(0, 1)$. This differential noise incorporation ensures that CA atoms exhibit greater conformational flexibility while maintaining structural coherence for other atoms through correlated, smaller perturbations. The final paratope coordinates are computed as:

$$\boldsymbol{X}_p = \boldsymbol{X}_0 + \boldsymbol{\epsilon}_i.$$

Our model maintains two distinct coordinate representations:

$$\boldsymbol{X}_{\textit{full}} = \boldsymbol{X}_{\textit{ae}} \cup \boldsymbol{X}_{\textit{ab}}, \tag{14}$$

where $\boldsymbol{X}_{\textit{full}}$ represents the complete antibody-antigen complex, with $\boldsymbol{X}_{\textit{ae}}$ denoting the epitope coordinates and $\boldsymbol{X}_{\textit{ab}}$ representing the full antibody structure, including framework and paratope regions.

Similarly, we define the focused binding interface representation as:

$$\boldsymbol{X}_{\textit{inter}} = \boldsymbol{X}_{\textit{ae}} \cup \boldsymbol{X}_p, \tag{15}$$

where $\boldsymbol{X}_{\textit{inter}}$ includes the epitope coordinates $\boldsymbol{X}_{\textit{ae}}$ and the generated paratope coordinates $\boldsymbol{X}_p$, without the full antibody framework.

This distinction ensures that the model can separately handle global structural representations $\boldsymbol{X}_{\textit{full}}$ and the more localized binding interactions $\boldsymbol{X}_{\textit{inter}}$ for improved prediction and analysis.

We employ the standardized IMGT/Chothia Chothia et al. (1989); Lefranc et al. (2003) to maintain structural consistency across antibody chains.This well-established system assigns independent position indices to each chain while preserving the original residue numbering from PDB structures. For the antibody heavy chain, the framework regions and CDRs follow specific positional ranges, with CDR-H1 typically spanning positions 23-35 and CDR-H3 occupying positions 104-118. Similarly, the light chain maintains its distinct numbering convention, with CDR-L1 typically located at positions 27-38 and CDR-L3 at positions 105-117. This independent numbering approach ensures clear differentiation between chain segments while facilitating accurate structural analysis and prediction. To maintain a clear distinction between antibody and antigen components, all antigen residues are uniformly assigned position 0, allowing the model to effectively distinguish between interacting molecular components.

D.4 EPITOPE SELECTION

We implement a distance-based approach for epitope selection that identifies the most relevant antigen residues interacting with the antibody CDR-H3. For each antigen residue $v_i \in V_{ag}$, we compute its minimal distance to any CDR-H3 residue v_j:

$$\hat{dist}(v_i) = \min_{v_j} dist_{i,j},$$

From these distances, we select the residues with top-k (typically $k = 48$) smallest distances to form the antigen epitope:

$$V_{\textit{ae}} = \{\arg\min_k \hat{dist}(v_i) \mid \forall v_i \in V_{ag}\}.$$

This selection process ensures that we focus on the most critical residues involved in antibody-antigen binding while maintaining computational efficiency. The fixed-size antigen epitope selection provides a consistent representation of the binding interface, enabling robust structural prediction and analysis.

D.5 PROOFS

In this section, we present the proofs of theorems in this paper, beginning with formal definitions of SE(3) equivariance and SE(3) invariance.

D.5.1 PRELIMINARIES

Definition D.1 (E(3)-equivariance and SE(3)-invariance)**.** Let $\varphi : \mathcal{X} \to \mathcal{Y}$ be a mapping function, and let $T \in E(3)$ denote a rigid transformation in three-dimensional space comprising rotation and reflection, $\boldsymbol{Q} \in O(3)$ and a translation $\boldsymbol{b} \in \mathbb{R}^3$. The transformation acts on coordinates as:

$$T(\boldsymbol{X}) := \boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}, \quad \boldsymbol{Q} \in O(3), \boldsymbol{b} \in \mathbb{R}^3.$$

Then:

- φ is **E(3)-equivariant**, if for any transformation $\boldsymbol{X} \mapsto \boldsymbol{X}_{\text{out}} = \varphi(\boldsymbol{X})$, we have $T(\boldsymbol{X}_{\text{out}}) = \varphi\big(T(\boldsymbol{X})\big)$.
- φ is **E(3)-invariant** if $\varphi\big(T(\boldsymbol{X})\big) = \varphi(\boldsymbol{X})$ for all $T \in \mathrm{E}(3)$, i.e. T acts as identity transformation in the output space.

In our model architecture, 3D coordinates are E(3)-equivariant, meaning they transform consistently with input transformations, while residue embeddings demonstrate E(3)-invariance, remaining unchanged under rigid transformations.

Notations. Next, we summarize the notations used throughout our analysis.

- Let $\{(\boldsymbol{H}_i^{(l)}, \boldsymbol{X}_i^{(l)})\}_{i=1}^n$ denote the residue embeddings $\boldsymbol{H}_i^{(l)} \in \mathbb{R}^d$ and 3D coordinates $\boldsymbol{X}_i^{(l)} \in \mathbb{R}^3$ at layer l.
- Let $\boldsymbol{H}_{e_{ij}}^{(l)}$ be the edge feature between residues v_i and v_j at layer l.
- Let $T(\boldsymbol{X}) := \boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}$ denote a rigid transformation in $E(3)$.

Below, we will show that Igformer maintains coordinate equivariance and embedding invariance under these transformations at each layer, following Definition D.1.

D.5.2 E(3)-EQUIVARIANCE OF THE COORDINATES

We begin by establishing the E(3)-equivariance property of coordinate predictions in the EMP module.

Lemma D.2 (Linear Transformation of Coordinate Differences)**.** *For coordinates differences* $\Delta\boldsymbol{X}_{ij}^{(l)} := \boldsymbol{X}_i^{(l)} - \boldsymbol{X}_j^{(l)}$*, applying* $T \in E(3)$ *to coordinates:* $\widetilde{\boldsymbol{X}}_i^{(l)} = T(\boldsymbol{X}_i^{(l)}) = \boldsymbol{Q}\,\boldsymbol{X}_i^{(l)} + \boldsymbol{b}$*, results in* $\Delta\widetilde{\boldsymbol{X}}_{ij}^{(l)} = \boldsymbol{Q}\,\Delta\boldsymbol{X}_{ij}^{(l)}$*.*

Proof. By direct substitution:

$$\widetilde{\boldsymbol{X}}_i^{(l)} - \widetilde{\boldsymbol{X}}_j^{(l)} = (\boldsymbol{Q}\boldsymbol{X}_i^{(l)} + \boldsymbol{b}) - (\boldsymbol{Q}\boldsymbol{X}_j^{(l)} + \boldsymbol{b}) = \boldsymbol{Q}\big(\boldsymbol{X}_i^{(l)} - \boldsymbol{X}_j^{(l)}\big) = \boldsymbol{Q}\Delta\boldsymbol{X}_{ij}^{(l)}.$$

The translation $\boldsymbol{b}$ cancels, leaving only the rotation $\boldsymbol{Q}$. □

Lemma D.3 (Geometry-Based Operations)**.** *Geometric Operations in Igformer, including pairwise distances* $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|$*, dot products, and dihedral angles, remain invariant under any* $T \in E(3)$*.*

Proof. A translation $\boldsymbol{b}$ always cancels in $\boldsymbol{X}_i - \boldsymbol{X}_j$. A rotation $\boldsymbol{Q} \in O(3)$ preserves Euclidean norms and dot products:$\|\boldsymbol{Q}\boldsymbol{v}\|^2 = \|\boldsymbol{v}\|^2$. Therefore, all geometric operations remain invariant under T. □

Theorem D.4 (E(3)-Equivariance of Coordinates in EMP)**.** *At layer l, Igformer updates coordinates through the following sequential operations:*

$$\begin{aligned}
\boldsymbol{S}_{ij}^{(l)} &= \textit{MLP}\big(\|\Delta\boldsymbol{X}_{ij}^{(l)}\|^2, \dots\big), \quad \textit{(geometry-based score)} \\
\boldsymbol{H}_{e_{ij}}^{(l+1)} &= \textit{EdgeMLP}\left(\boldsymbol{H}_i^{(l)} \oplus \boldsymbol{H}_j^{(l)} \oplus \boldsymbol{S}_{ij}^{(l)} \oplus \boldsymbol{H}_{e_{ij}}^{(l)}\right), \\
\boldsymbol{X}_i^{(l+1)} &= \boldsymbol{X}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \Delta\boldsymbol{X}_{ij}^{(l)} \cdot \textit{CoordMLP}\big(\boldsymbol{H}_{e_{ij}}^{(l+1)}\big).
\end{aligned}$$

This update rule is E(3)-equivariant on coordinates. Concretely, applying T to input coordinates $\{\boldsymbol{X}_i^{(l)}\}$ results in the same transformation T being applied to output coordinates $\{\boldsymbol{X}_i^{(l+1)}\}$.

Proof. By Lemma D.3, $\boldsymbol{S}_{ij}^{(l)}$ and consequently $\boldsymbol{H}_{e_{ij}}^{(l+1)}$ remain unchanged when input coordinates are transformed by T. Let $\boldsymbol{w}_{ij}^{(l+1)} := \textit{CoordMLP}\big(\boldsymbol{H}_{e_{ij}}^{(l+1)}\big)$. Then:

$$\boldsymbol{X}_i^{(l+1)} = \boldsymbol{X}_i^{(l)} + \sum_{j\in\mathcal{N}(i)} \Delta\boldsymbol{X}_{ij}^{(l)} \cdot \boldsymbol{w}_{ij}^{(l+1)}.$$

Under transformation $T : \boldsymbol{X}_i^{(l)} \mapsto \boldsymbol{Q}\boldsymbol{X}_i^{(l)} + \boldsymbol{b}$, we have:

$$\begin{aligned}
\widetilde{\boldsymbol{X}}_i^{(l+1)} &= \widetilde{\boldsymbol{X}}_i^{(l)} + \sum_{j\in\mathcal{N}(i)} \big[\widetilde{\boldsymbol{X}}_i^{(l)} - \widetilde{\boldsymbol{X}}_j^{(l)}\big] \cdot \boldsymbol{w}_{ij}^{(l+1)} && \big(\text{same } \boldsymbol{w}_{ij}^{(l+1)}\big) \\
&= (\boldsymbol{Q}\boldsymbol{X}_i^{(l)} + \boldsymbol{b}) + \sum_{j\in\mathcal{N}(i)} \Big[\boldsymbol{Q}(\boldsymbol{X}_i^{(l)} - \boldsymbol{X}_j^{(l)})\Big] \cdot \boldsymbol{w}_{ij}^{(l+1)} && \big(\text{Lemma D.2}\big) \\
&= \boldsymbol{Q}\,\boldsymbol{X}_i^{(l)} + \boldsymbol{b} + \boldsymbol{Q} \sum_{j\in\mathcal{N}(i)} (\boldsymbol{X}_i^{(l)} - \boldsymbol{X}_j^{(l)}) \cdot \boldsymbol{w}_{ij}^{(l+1)} \\
&= \boldsymbol{Q}\Big[\boldsymbol{X}_i^{(l)} + \sum_{j\in\mathcal{N}(i)} \Delta\boldsymbol{X}_{ij}^{(l)} \cdot \boldsymbol{w}_{ij}^{(l+1)}\Big] + \boldsymbol{b} \\
&= \boldsymbol{Q}\boldsymbol{X}_i^{(l+1)} + \boldsymbol{b} = T\big(\boldsymbol{X}_i^{(l+1)}\big).
\end{aligned}$$

Therefore, applying g to input coordinates leads to $T\big(\boldsymbol{X}_i^{(l+1)}\big)$ in output coordinates. By induction over all layers, Igformer maintains E(3)-equivariance through its coordinate computation in the EMP module. □

D.5.3 E(3)-Invariance of the Residue Embeddings

We now prove that residue embeddings $\boldsymbol{H}_i$ remain numerically invariant under global rigid transformation of the input 3D structure.

Lemma D.5 (Base Case for Embedding Invariance)**.** *Residue embeddings $\boldsymbol{H}_i^{(0)}$ are initialized using rigid independent features like residue identity, positional index, etc., making them inherently invariant under any transformation $T \in \mathrm{E}(3)$.*

Lemma D.6 (Edge Feature Update Invariance)**.** *For edge updates of the form:*

$$\boldsymbol{H}_{e_{ij}}^{(l+1)} = \textit{EdgeMLP}\Big(\boldsymbol{H}_i^{(l)} \oplus \boldsymbol{H}_j^{(l)} \oplus S_{ij}^{(l)} \oplus \boldsymbol{H}_{e_{ij}}^{(l)}\Big),$$

where $\boldsymbol{S}_{ij}^{(l)}$ represents geometry operations (distance, dot product, etc.). If $\boldsymbol{H}_i^{(l)}, \boldsymbol{H}_j^{(l)}$ are invariant and $\boldsymbol{S}_{ij}^{(l)}$ is unaffected by T (Lemma D.3), then $\boldsymbol{H}_{e_{ij}}^{(l+1)}$ maintains invariance.

Proof. The *EdgeMLP* receives numerically identical inputs regardless of coordinate transformation $\boldsymbol{Q}\boldsymbol{X}_i^{(l)} + \boldsymbol{b}$, ensuring $\boldsymbol{H}_{e_{ij}}^{(l+1)}$ remains unchanged. □

Theorem D.7 (E(3)-Invariance of Residue Embeddings)**.** *Given Igformer's residue embedding update rule:*

$$\boldsymbol{H}_i^{(l+1)} = \boldsymbol{H}_i^{(l)} + \textit{NodeMLP}\Big(\boldsymbol{H}_i^{(l)} \oplus \sum_{j\in\mathcal{N}(i)} \boldsymbol{H}_{e_{ij}}^{(l+1)}\Big),$$

or its self-attention variant, the embeddings $\boldsymbol{H}_i^{(l+1)}$ remain invariant under any global transformation $T \in \mathrm{E}(3)$. By extension, Igformer's final embeddings $\boldsymbol{H}_i$ maintain E(3)-invariance.

Proof. We prove this theorem through induction: **Base Case.** By Lemma D.5, initial embeddings $\boldsymbol{H}_i^{(0)}$ are independent of 3D coordinates.

Inductive Step. Assume $\{\boldsymbol{H}_i^{(l)}\}$ is already invariant under T.

- According to Lemma D.3, any geometry-based scalar $\boldsymbol{S}_{ij}^{(l)}$ remain unchanged under T.
- According to Lemma D.6, edge features $\boldsymbol{H}_{e_{ij}}^{(l+1)}$ maintain invariance under T.
- Consequently, all inputs to the residue-level update MLP/attention remain unchanged under T.

Therefore, $\boldsymbol{H}_i^{(l+1)}$ maintains invariance under T. By induction across layers, Igformer's residue embeddings are E(3)-invariant. □

Proof of Theorem C.1. Combining Theorems D.4 and D.7, we show that coordinates in the EMP module are E(3)-equivariant and embeddings are E(3)-invariant, which finishes the proof of Theorem C.1.

Lemma D.8. *The embeddings updated by inter-graph refinement and triangle multiplicative module (TMM) modules are invariant.*

Proof. The APP and SGFormer components update embeddings through weighted aggregation: $\boldsymbol{H}_i = \sum_{j \in V} \boldsymbol{H}_j$ with residue connections. These operations are independent of coordinates $\boldsymbol{X}_i$. Given that $\boldsymbol{H}_i$ is invariant under transformation T, the MLP/attention operations preserve E(3)-invariance during embedding updates.

The Triangle Multiplicative module processes pairwise embeddings $\boldsymbol{Z}_{ij} = \mathit{MLP}(\boldsymbol{H}_i, \boldsymbol{H}_j)$ without direct dependence on coordinates $\boldsymbol{X}_i$. Since $\boldsymbol{H}_i$ maintains invariance under T, the pairwise features and subsequent internal operations (row/column gating/attention) on $\boldsymbol{Z}$ preserves E(3)-invariance. The merging of $\boldsymbol{Z}$ into $\boldsymbol{H}_i$ through MLP/attention operations maintains this invariance property. □

D.5.4 Proof of Theorem C.3

Combining Theorem D.4, Theorem D.7, and Lemma D.8, we conclude:

Theorem D.9 (Igformer is E(3)-Equivariant (Coordinates) and E(3)-Invariant (Embeddings))**.** *Let* $\{\boldsymbol{H}_i^{(0)}, \boldsymbol{X}_i^{(0)}\}_{i \in V}$ *be the initial inputs (node features, coordinates). Suppose Igformer generetes final outputs as follows:*

$$\{\boldsymbol{H}_i^{(\mathit{final})}, \boldsymbol{X}_i^{(\mathit{final})}\} = \mathit{Igformer}\Big(\{\boldsymbol{H}_i^{(0)}, \boldsymbol{X}_i^{(0)}\}\Big).$$

Then, for any $T \in \mathrm{E}(3)$*, we have:*

$$\{\boldsymbol{H}_i^{(\mathit{final})}, T\big(\boldsymbol{X}_i^{(\mathit{final})}\big)\} = \mathit{Igformer}\Big(\{\boldsymbol{H}_i^{(0)}, T\big(\boldsymbol{X}_i^{(0)}\big)\}\Big).$$

Proof. The proof follows by chaining layer-wise properties:

- By Theorem D.4, the coordinate update at each layer maintains E(3)-equivariance. Through induction across layers, this property extends to the entire coordinate transformations.
- Similarly, Theorem D.7 and Lemma D.8 establish that residue feature update at each layer preserves E(3)-invariance. By induction, this invariance property carries through to the final residue embeddings.

Therefore, the complete Igformer model satisfies both equivariance and invariance properties as stated in the theorem. □

D.6 Algorithms

The training process iterates three times per batch, with each iteration updating the epitope coordinates and embeddings, followed by the reconstruction of intra-graph and inter-graph connections. The final iteration is utilized for loss function calculation. Algorithm 1 shows the pseudo-code of the iterative updating process in Igformer. Detailed descriptions from Lines 3-8 are provided in Sections C.3-C.4. In the first iteration ($T = 1$), we use the initialized embeddings for calculation. For subsequent iterations ($T \in \{2, 3\}$), both embeddings and coordinates are updated accordingly. (i) The full embedding update is given by $\boldsymbol{H}_{full} = \mathit{MLP}(\boldsymbol{H}_{intra}) + \boldsymbol{H}'_{full}$, where $\boldsymbol{H}'_{full} = \{\boldsymbol{H}_{ae}, \boldsymbol{H}'_{ab}\}$. The

Algorithm 1 Iterative update for antibody-antigen binding interface

Input: Initial embeddings, coordinates, number of iterations T, number of layers t
Output: Updated embeddings and coordinates

```
for T = 1 to 3 do
    Construct intra-graph and inter-graph according to Appendix D.1 using embeddings and coordinates calculated
    for t = 1 to 3 do
        if t = 1 then
            Triangle multiplicative update for paratope
        end if
        Dual EMP update
    end for
    Update the epitope coordinates and embeddings
    if T = 3 then
        Calculate training loss
    end if
end for
```

paratope region is specifically updated using attention weights: $\boldsymbol{H}_{ab}^{'} = \{\boldsymbol{P} \odot \boldsymbol{H}_p, \boldsymbol{H}_{ab\backslash p}\}$, where $\boldsymbol{P} = \mathit{MLP}(\boldsymbol{H}_p)$ assigns different weights to different residues through element-wise multiplication $\odot$, while maintaining embeddings of non-paratope regions $\boldsymbol{H}_{ab\backslash p}$. (ii) The coordinate update follows a similar strategy, replacing the original epitope coordinates: $\boldsymbol{X}_{full} = \{\boldsymbol{X}_{ae}^{'}, \boldsymbol{X}_{ab}\}$, ensuring consistent transformation patterns between embedding and coordinate spaces.

D.7 Loss Function

Our training objective combines multiple loss terms to ensure accurate prediction of both sequence and structure.

Sequence Loss. The Sequence Negative Log Likelihood (SNLL) loss evaluates sequence prediction accuracy for masked positions in CDR regions:

$$\mathcal{L}_{seq} = -\frac{1}{|V_p| \cdot |\mathcal{R}|} \sum_{v_i \in V_p} \sum_{r \in \mathcal{R}} y_{i,r} \log(\mathit{softmax}(\mathit{MLP}(\boldsymbol{Z}_i))_r), \tag{16}$$

where V_p is the set of masked residues in epitope-binding CDRs, $\boldsymbol{Z}_i$ is the output embedding of residue v_i, $\mathcal{R}$ is the set of all possible residue types, $y_{i,r}$ indicates the ground truth residue type through one-hot encoding. $\mathit{MLP}(\boldsymbol{Z}_i)$ outputs a vector of logits for each residue type and $\mathit{softmax}(\mathit{MLP}(\boldsymbol{Z}_i))_r$ gives the predicted probability for each residue type $r \in \mathcal{R}$.

Structural Loss. For structural accuracy, we employ a coordinate loss computed over the full antibody structure $\boldsymbol{X}_{ab}$:

$$\mathcal{L}_{coord} = \frac{1}{|V_{ab}|} \sum_{v_i \in V_{ab}} \ell_{huber}(\boldsymbol{X}_i^{pred} - \boldsymbol{X}_i^{true} T)$$

using the Huber loss function:

$$\ell_{huber}(x) = \begin{cases} 0.5x^2, & \mathit{if}\ |x| < \delta \\ \delta(|x| - 0.5\delta), & \mathit{otherwise} \end{cases}.$$

where T represents the optimal rigid transformation obtained through Kabsch alignment algorithm Kabsch (1976) using backbone atoms (N, Cα, C, O) of each residue, $\boldsymbol{X}_i^{pred}$ and $\boldsymbol{X}_i^{true}$ are the predicted and ground-truth 3D coordinates for residue v_i, $\delta = 1$ is the threshold parameter.

To maintain chemical validity, we include a backbone bond loss:

$$\mathcal{L}_{bond} = \frac{1}{|\mathcal{A}|} \sum_{b \in \mathcal{A}} \ell_{huber}(b^{pred} - b^{true})$$

where $\mathcal{A}$ contains all atoms in the antibody, b^{pred} and b^{true} are the predicted and ground-truth bond length derived from $\boldsymbol{X}^{pred}$ and $\boldsymbol{X}^{true}$, respectively. The final structure loss is the combination of the

coordinate loss and the backbone bond loss:

$$\mathcal{L}_{struct} = \mathcal{L}_{coord} + \mathcal{L}_{bond} \tag{17}$$

Interface Loss. The interface loss evaluates the quality of predicted interactions between antibody paratope and antigen epitope regions using $\boldsymbol{X}_{inter}$. This loss comprises two components: a structural accuracy term and an edge distance prediction term. The structural accuracy loss measures the deviation of predicted atomic coordinates from their true positions in the paratope region:

$$\mathcal{L}_{struct} = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \ell_{huber}(\boldsymbol{X}_k^{pred} - \boldsymbol{X}_k^{true}),$$

where $\mathcal{P}$ denotes the set of atoms in the antibody paratope region, and $\boldsymbol{X}_k^{pred}$ and $\boldsymbol{X}_k^{true}$ denote the predicted and true 3D coordinates for atom k, respectively. The edge distance prediction loss assesses the accuracy of predicted residue interactions within the binding interface:

$$\mathcal{L}_{edge} = \frac{1}{|E_{inter}|} \sum_{(v_i, v_j) \in E_{inter}} \ell_{huber}(f_\theta(\boldsymbol{h}_i, \boldsymbol{h}j) - dist_{ij}^*),$$

where E_{inter} is the set of epitope-paratope residue pairs at the interface with $v_i \in V_{ae}$ from the epitope and $v_j \in V_p$ from the paratope. The network f_θ predicts inter-residue distances based on residue embeddings $\boldsymbol{H}_i$ and $\boldsymbol{H}_j$ as input, $dist_{ij}^*$ is computed directly from coordinates as the minimum distance between atoms, $\boldsymbol{h}_i$ and $\boldsymbol{h}_j$, which are compared against true distances dij^* computed from atomic coordinates.

The total interface loss combines these components:

$$\mathcal{L}_{interface} = \mathcal{L}_{paratope} + \mathcal{L}_{edge} \tag{18}$$

Final Loss Function. By combining Equations 16-18, the final loss function is calculating following:

$$\mathcal{L}_{total} = \mathcal{L}_{seq} + \mathcal{L}_{struct} + \mathcal{L}_{interface}. \tag{19}$$

This comprehensive loss function ensures accurate prediction of both sequence and structure while maintaining chemical validity and interface quality.

E Additional Experimental Settings

E.1 Datasets

Our antibody dataset preprocessing pipeline leverages the Structural Antibody Database (SAbDab) snapshot from November 2022, with structures pre-numbered using the IMGT numbering scheme Lefranc et al. (2003). The IMGT system provides precise residue numbering for antibody structures, defining specific ranges for framework regions (FR) and complementarity determining regions (CDRs) in both heavy and light chains. For the heavy chain, these ranges encompass FR-H1 (1-26), CDR-H1 (27-38), FR-H2 (39-55), CDR-H2 (56-65), FR-H3 (66-104), CDR-H3 (105-117), and FR-H4 (118-129), with parallel definitions for the light chain regions. This standardized numbering scheme is crucial for consistent processing and analysis across all antibody structures.

We implement stringent filtering criteria, focusing exclusively on protein and peptide antigens while ensuring no chain overlap between antibody and antigen components. Other potential antigen types like small molecules, nucleic acids, or carbohydrates are excluded from our dataset. The validation process enforces three critical requirements: structural completeness, with verification of conserved residues (CYS23, CYS104, TRP41) in both chains; correct IMGT numbering across all regions; and comprehensive atomic coordinate validation, including backbone and sidechain atoms, with a 6.6 Å contact distance threshold for residue interactions.

The final dataset preprocessing stage involves sequence-based clustering using MMseqs2 with a 40% identity threshold, particularly focusing on CDR-H3 sequence similarity. This process yields 3,246 training and 365 validation antibodies. The curated SAbDab dataset serves as the foundation for both the RAbD Dunbar et al. (2013) (60 PDB structures) and the IgFold Ruffolo et al. (2023a) benchmark dataset (51 structures), ensuring consistent quality and standardization across all experimental evaluations.

Table 7: Structure of the EMP module.

Component	Layer Structure
Initialization Network	Linear(64 → 128), Dropout(0.1)
Radial Feature Network	Linear(196 → 128), Post-normalization
EdgeMLP	Input: source_features[128] ⊕ target_features[128] ⊕ radial_features[128] Linear(384 → 128), SiLU(), Dropout(0.1) Linear(128 → 128), SiLU(), Dropout(0.1)
CoordMLP	Linear(128 → 128), SiLU(), Dropout(0.1) Linear(128 → 3)
NodeMLP	Linear(64 → 128), SiLU(), Dropout(0.1) Linear(128 → 128), SiLU(), Dropout(0.1)

E.2 Implementation Details

We implement Igformer using PyTorch Paszke et al. (2019) and train the model on a single GeForce RTX 3090 Ti GPU using the Adam optimizer Kingma (2015). The model training process utilizes a batch size of 16, running for 150 epochs in task 1 and 300 epochs for tasks 2-4, while preserving the 10 best-performing model checkpoints throughout the training phase of each individual task. The model architecture employs a 64-dimensional embedding space for antibody feature representation, 14 channels (i.e., 14 atoms) for each residue, and 3 iterations. The graph structure connects each residue to its 9 nearest neighbors (k = 9). The model incorporates a dropout rate of 0.1 (dropout = 0.1) to enhance generalization. Training proceeds through 3 iterations per batch, with each iteration updating epitope coordinates and embeddings, followed by reconstruction of intra-graph and inter-graph connections. The final iteration is used for loss calculation. For structure analysis, the model defines interacting residues in the antibody-antigen interface using a binding distance threshold of 6.6 Å.

The optimization process employs gradient clipping at 1.0 and implements an adaptive learning rate schedule. The learning rate decays exponentially from 0.0013 to 0.0001 across the training epochs, with the decay factor determined as $ln(0.0001/0.0013)$, divided by the total number of training steps, ensuring a smooth convergence to the target learning rate.

Feature Initiation. Igformer employs two parallel embedding components to construct the initial residue representation. First, the residue type embedding transforms each amino acid into a 64-dimensional vector using a learned embedding matrix of size 25×64, where 25 accounts for the 20 types of amino acids and special tokens. Each amino acid is mapped to a unique embedding vector $\boldsymbol{H}_i^{res}$ that captures its chemical and physical properties. The positional information is encoded through a separate learned embedding matrix of size 192×64, which maps each position in the sequence to a 64-dimensional vector $\boldsymbol{H}_i^{pos}$. This enables the model to maintain sequential context for sequences up to 192 residues in length, allowing differentiation between identical amino acids at different positions. The final feature representation $\boldsymbol{H}_i$ is computed through element-wise addition of these two embeddings. This additive combination maintains the 64-dimensional feature space while integrating both residue identity and positional context, enabling the model to learn position-aware representations with comprehensive chemical information.

EMP Module. The EMP module iterates three times (3 blocks) and consists of several key components: feature initialization, feature similarity network, EdgeMLP, CoordMLP and NodeMLP with residual connections, as detailed in Table 7. Each layer incorporates dropout Srivastava et al. (2014) mechanisms after activation functions. The similarity features are normalized after being mapped from 196 to 128 dimensions, including sim_{ij}^{res} and sim_{ij}^{atom}. The iterative application of these components ensures efficient updates of both node features and coordinate transformations while maintaining stable training dynamics through the dropout mechanism. The EdgeMLP processes concatenated input features of dimension 384, comprising source_node features (128), target_node features (128), and similarity features (128). This design enables the model to effectively capture interactions between residues while preserving both structural and relational information crucial for antibody representation.

APP and SGFormer. As shown in Table 8, the inter-grpah refinment network consists of two main components. First, the APP module conducts 16 steps of propagation with $\alpha = 0.1$, , allocating 90% weight to neighbor-propagated information while retaining 10% of original residue features. This enables effective information spread across the graph while preserving essential original residue

Table 8: Structure of the inter-grpah refinement module.

Component	Layer Structure
APP	Propagation with residue connection (128 → 128)
SGFormer Layer (5 blocks)	Input: Linear(128 → 64) Query: Linear(64 → 64), Key: Linear(64 → 64), Value: Linear(64 → 64) LayerNorm, Residual ($\alpha = 0.9$), Dropout(0.5)
Output	Linear(64 → 128)

Table 9: Structure of the triangle module.

Component	Layer Structure
Feature Pair MLP	Input: features_i[128] $\oplus$ features_j[128] Linear(256 → 256), SiLU(), Dropout(0.1)
Triangle Multiply (outgoing/incoming)	Input: Matrix($n \times n \times 256$) Linear(256 → 256), SiLU(), Linear(256 → 256)
Triangle Attention (outgoing/incoming)	Input: Matrix($n \times n \times 256$) Query: Linear(256 → 256), Key: Linear(256 → 256), Value: Linear(256 → 256) LayerNorm()
Dimension Reduction	Input: Diagonal($n \times 256$) Linear(256 → 128), LayerNorm()

characteristics. The SGFormer component employs 5 single-head attention layers, computing attention once per layer rather than utilizing multiple attention heads. In its residual connections after each transformer layer, it uses $\alpha = 0.9$, emphasizing transformed features (90%) while maintaining minimal previous layer information (10%). Layer normalization is applied to the 64-dimensional feature space after each attention computation, working in conjunction with dropout (0.5) to regulate feature magnitudes and ensure training stability.

Triangle Module. The triangle network consists of two main components, as detailed in Table 9. First, the triangle multiplication module performs feature pair-wise propagation with directed connections. For each sequence of n residues, it constructs an $n \times n$ matrix by concatenating paired residue features (128-dim) and projecting them to 256-dimensional space via an MLP, establishing directed connections where $\boldsymbol{Z}_{ij} \neq \boldsymbol{Z}_{ji}$. The triangle attention component uses four sequential operations: multiply outgoing, multiply incoming, attention outgoing, and attention incoming. Each operation maintains the $n \times n \times 256$ dimensionality while enabling distinct information flows. The multiply operations use MLPs for feature transformation while preserving the triangle structure, whereas the attention operations employ single-head self-attention with query/key/value projections to capture long-range dependencies. The final step extracts diagonal elements and reduces them to 128 dimensions for network compatibility. The implementation incorporates a dropout rate of 0.1 and SiLU activation functions across all linear projections. Layer normalization is applied after attention operations, and an additional normalization step in the final dimension reduction. The entire module operates on a per-sequence basis, processing each unique sequence independently before the features are reassembled.

Dual EMP Module. The Dual EMP architecture extends the base EMP module by operating on two distinct graph structures simultaneously. It processes the intra-graph and inter-graph representations, along with their respective coordinates ($\boldsymbol{X}_{full}$ and $\boldsymbol{X}_{inter}$), while maintaining shared embeddings between them. A detailed description of this architecture is provided in Section C.4.

Embedding Distance. As detailed in Table 10, The edge distance MLP computes pairwise distances between residue embeddings. First, it concatenates pairs of 128-dimensional residue embeddings into 256-dimensional vectors. These concatenated features are processed through a two-layer MLP with SiLU activations. The final linear projection produces a scalar value representing the predicted pairwise distance or interaction strength.

Residue Type Prediction. As shown in Table 10, the residue type prediction MLP converts the learned 128-dimensional residue representations into probability distributions over 20 amino acid classes through a two-layer MLP with skip connections. The architecture processes features sequentially: an initial SiLU activation and linear projection preserve the 128-dimensional space, followed by a second SiLU activation and final linear transformation that maps to 20-dimensional amino acid class probabilities.

Table 10: Structure of the edge distance MLP and prediction MLP.

Component	**Layer Structure**
Edge Distance MLP	Input: residue_i[128] $\oplus$ residue_j[128]) SiLU(), Linear(256 $\rightarrow$ 128), SiLU(), Linear(128 $\rightarrow$ 1)
Prediction MLP	Input: hidden features[128] SiLU(), Linear(128 $\rightarrow$ 128), SiLU(), Linear(128 $\rightarrow$ 20)

E.3 BASELINES

For the baselines, we adopt the hyperparameters and training procedures from their official releases since all methods utilize SAbDab to construct training sets of similar scale and distribution. We save model parameters from the top 10 validation rounds, compute metrics for each model independently, and report the mean values as final results. State-of-the-art models on each task are included as competitors.

In Task 1, we evaluate against RosettaAb, an energy-based method using statistical functions Adolf-Bryfogle et al. (2018); DiffAb, a diffusion-based generative model Luo et al. (2022); MEAN, an equivariant attention network Kong et al. (2023b); HERN, an end-to-end framework limited to CDR-H3 Jin et al. (2022a); and dyMEAN, an adaptive multi-channel equivariant network Kong et al. (2023a).

For Task 4, we compare Igformer against IgFold⇒HDock, IgFold⇒HERN, GT⇒HERN, and dyMEAN, each employing different docking strategies. IgFold⇒HDock follows a two-stage pipeline where IgFold first predicts the antibody structure Ruffolo et al. (2023a), and HDock Yan et al. (2020), using knowledge-based scoring functions, performs docking. IgFold is a specialized version of AlphaFold tailored for antibody structure prediction, while HDock applies traditional docking approaches. In contrast, IgFold⇒HERN adopts a similar two-stage approach but replaces HDock with HERN Jin et al. (2022a), which takes the IgFold-predicted backbone structure as input and generates docked backbones, followed by Rosetta for side-chain packing Adolf-Bryfogle et al. (2018). GT⇒HERN further extends this pipeline by using ground truth antibody structures instead of predicted ones, allowing HERN to dock CDR-H3 along with other regions toward the epitope, thus providing an upper bound on performance by leveraging perfect structural information. Unlike these multi-stage pipelines, dyMEAN distinguishes itself as a fully end-to-end approach that directly models antigen-antibody interactions without requiring separate structure prediction and docking stages Kong et al. (2023a).

We adopt hyperparameters and training procedures from their official releases, as all methods utilize SAbDab for training. Specifically, HERN by Jin et al. Jin et al. (2022a) uses a larger hidden size of 256 with four layers and 16 RBF kernels for distance embedding. DiffAb by Luo et al. Luo et al. (2022) employs a hidden size of 128, a pair feature size of 64, six layers, and 100 diffusion steps. MEAN and dyMEAN share similar parameters, including an embedding size of 64, a hidden size of 128, three layers, three iterations, and nine nearest neighbors. dyMEAN introduces an additional parameter, d = 16, for atom type and position embedding. All baseline models use nine neighbors for KNN graph construction.

E.4 METRICS

We employ a comprehensive set of metrics to evaluate model performance in both antibody 1D sequence and 3D structure prediction tasks, encompassing both structural accuracy and sequence recovery:

- DockQ Basu & Wallner (2016): This metric evaluates antibody-antigen interaction quality, especially within the paratope-epitope binding region. While the model optimizes paratope-epitope interactions within a specific cutoff distance during training, its evaluation considers interactions with the complete antigen structure, ensuring a comprehensive assessment of binding interface prediction.
- Root Mean Square Deviation (RMSD): This metric measures structural accuracy by calculating unaligned distances between predicted and actual CA atoms in CDR regions, providing direct assessment of local structural precision.

Table 11: *Comparison of different paratopes.*

Paratope	TMscore↑	lDDT↑	RMSD↓	DockQ↑
Task 1				
H3	**0.9757**	**0.8650**	**7.15**	**0.45**
H3 + L3	0.9734	0.8512	8.05	0.41
Task 2				
H3	0.9706	0.8195	-	0.4255
H3 + L3	**0.9750**	**0.8311**	-	**0.4817**
Task 3				
H3	**0.9681**	**0.7580**	-	**0.4600**
H3 + L3	0.9667	0.7490	-	0.4367
Task 4				
H3	**0.9730**	**0.8677**	**7.88**	**0.522**
H3 + L3	0.9710	0.8631	8.25	0.4931

- TM-score Zhang & Skolnick (2005): This metric assesses global structural similarity between predicted and reference antibody structures, evaluating the overall quality of structure prediction across the entire antibody.
- Local Distance Difference Test (lDDT) Mariani et al. (2013): This metric provides an atomic-level assessment of local structural accuracy by comparing predicted and actual atomic positions across the entire antibody, offering detailed insights into structural fidelity.
- Amino Acid Recovery (AAR): This metric calculates the fraction of correctly predicted amino acids across the entire sequence, measuring overall sequence prediction accuracy.
- Contact AAR Ramaraj et al. (2012): This metric specifically evaluates prediction accuracy for residues in direct contact with the epitope, defined by a 6.6 Å distance threshold, focusing on binding interface accuracy.

F Additional Experimental Results

F.1 Additional Ablation Study

Choice of Paratope. We investigate the impact of paratope selection by comparing models using CDR-H3 alone versus both CDR-H3 and CDR-L3 regions. As shown in Table 11, the choice of paratope influences model performance across different tasks. For Task 1, using only CDR-H3 achieves superior performance with a TMscore of 0.9757, lDDT of 0.8650, RMSD of 7.15, and DockQ score of 0.45, compared to 0.9734, 0.8512, 8.05, and 0.41 respectively when including both CDR-H3 and L3. This trend continues in Task 3, where CDR-H3 alone yields better TMscore (0.9681 vs 0.9667), lDDT (0.7580 vs 0.7490), and DockQ (0.4600 vs 0.4367). Task 4 follows a similar pattern with improvements across all metrics when using only CDR-H3. Interestingly, Task 2 exhibits different characteristics, where including both CDR-H3 and CDR-L3 leads to improved performance, with TMscore increasing from 0.9706 to 0.9750, lDDT from 0.8195 to 0.8311, and DockQ from 0.4255 to 0.4817. These results underscore the importance of task-specific paratope selection for optimal performance.

Ablation Study on Task 4. We conduct additional experiments to evaluate the individual contributions of key components in Igformer for complex structure prediction. Table 12 presents the performance impact of removing each architectural component. Our observations are as follows. Both the triangle multiplicative module and axial attention prove critical to model performance. The removal of TM leads to the most significant degradation, with RMSD increasing to 8.12 and DockQ dropping to 0.447, while removing AA results in an RMSD of 8.06 and DockQ of 0.453. Additionally, replacing the dual EMP architecture with a single message passing framework diminishes performance, with RMSD rising to 7.99 and DockQ decreasing to 0.452. These systematic evaluations further validate the essential contribution of each component in Igformer.

F.2 Hyperparameter analysis

As illustrated in Figure 5, we analyze the impact of learnable weight factor w (refer to Section C.2) on similarity computation across Tasks 1-4. The DockQ, as a primary performance metric, consistently

Table 12: Ablation study on complex structure prediction.

Model	Generation		Docking	
	TMscore↑	lDDT↑	RMSD↓	DockQ↑
Igformer	**0.9730**	**0.8677**	**7.88**	**0.522**
- APP	0.9718	0.8603	7.95	0.449
- SGFormer	0.9725	0.8614	7.92	0.453
- TM	0.9709	0.8543	8.12	0.447
- AA	0.9708	0.8565	8.06	0.453
- Dual EMP	0.9725	0.8611	7.99	0.452

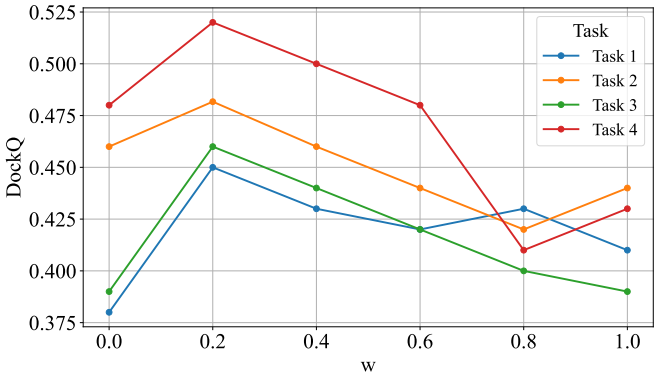


Figure 5: DockQ scores for varying w values across four tasks, with the peak observed at $w = 0.2$, emphasizing the importance of geometric distance in similairty matrix computation.

achieves its peak at $w = 0.2$ across all tasks. This optimal value indicates that atom-level similarity plays a predominant role in model performance. While both residue-level similarity (parameterized by w) and atom-level similarity contribute to modeling structural interactions, the optimal performance at $w = 0.2$ suggests that emphasizing atomic information (0.8) over residue-level features (0.2) most effectively captures antibody-antigen interactions. This finding provides crucial guidance for optimizing similarity matrix parameters in antibody sequence design and structure prediction tasks.

F.3 Additional Case Study of Task 1

Figure 6 illustrates more antibody structures generated by Igformer in Task 1.

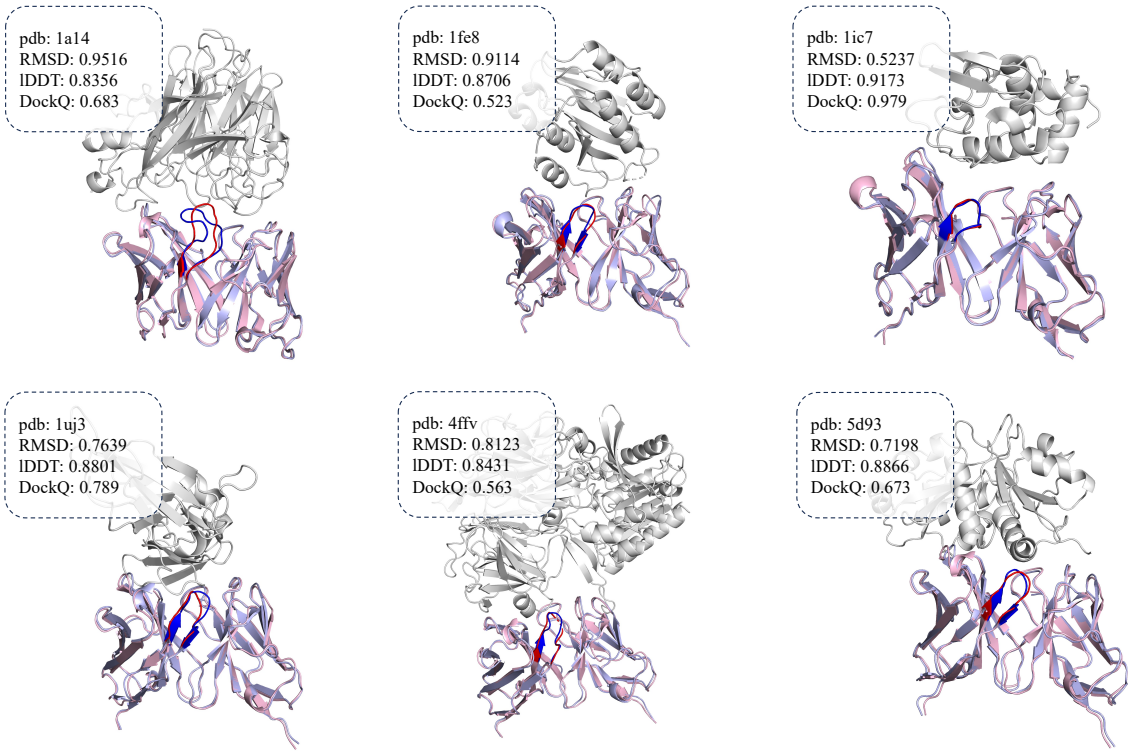


Figure 6: Additional antibody structures generated by Igformer.