IgFlow: Flow Matching for *De Novo* Antibody Design

Sanjay Nagaraj[*]
Absci
snagaraj@absci.com

Amir Shanehsazzadeh
Absci
ashane@absci.com

Hyun Park
Absci
hpark@absci.com

Jonathan King
Absci
jking@absci.com

Simon Levine
Absci
sgottreich@absci.com

Abstract

In this work, we present IgFlow, an SE(3)-flow matching model for *de novo* design of antibody structures. We focus on generating novel variable domain regions of the antibody and assess the performance of our model on 1) unconditional heavy and light chain generation and 2) framework-conditional loop design of the complementarity-determining regions (CDRs). Our results show that IgFlow generated antibodies are structurally similar to naturally observed antibodies. We compare our approach to IgDiff, an SE(3)-diffusion model for unconditional variable domain generation, on designability. Furthermore, we benchmark IgFlow and IgDiff on two conditional CDR inpainting tasks commonly encountered in antibody design. We find that IgDiff and IgFlow are both performant at unconditionally designing antibodies, and that IgFlow conditionally designs full CDR loops with higher self-consistency than IgDiff. Overall, our approach offers an alternative approach for antibody generation with additional computational benefits, including sample data efficiency and inference speed.

1 Introduction

Generative models have gained prominence in protein engineering given their ability to learn the distribution of natural protein space and generate novel, realistic proteins [1] [2]. Sequence and structure-based generative approaches have emerged as the two main frameworks, as well as frontier generative models with multimodal architectures that integrate the two [3] [4]. Recent state-of-the-art generative biomolecular design methods are based on either diffusion [5] [6] or flow-matching paradigms [7]. Within the protein design space, RFdiffusion [8] and FoldFlow-2 [9] have emerged as the best performing models, both of which leverage geometric constraints such as SE(3)-equivariance.

Within the space of protein engineering, a problem of major therapeutic interest is *de novo* antibody design [10] [11] [12]. Specifically, accurate, target-conditional immunoglobulin design is a critical step toward *de novo* therapeutic biologics. Immunoglobulins, or antibodies, are heterodimers composed of two heavy and two light chains, where chains are composed of constant and variable domains [13]. Each variable domain contains hypervariable loops called complementarity-determining regions (CDRs) which are critical to antigen recognition and binding. The most structurally variable and active in binding of these regions is Heavy Chain CDR3 (HCDR3) [14]. Therefore, antibody design tasks are generally framed around conditional design of CDRs, with a heavy emphasis on accurate determination of the HCDR3.

In this work, we build foundations for antibody design using flow-matching based approaches by extending the flow-matching protein model FrameFlow [15] as well as its discrete flow matching

[*]Work conducted during internship at Absci

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

extension MultiFlow [16]. We make architectural modifications to the base FrameFlow model and retrain the model on the Structural Antibody Database (SAbDab) [17], a subset of the Protein Data Bank containing structures of antibodies and antibody-antigen complexes. We train IgFlow under two settings, to first unconditionally generate paired heavy and light chain variable region backbones, and then to design specific CDR regions conditioned on the remainder of the framework. We compare designability of samples from our IgFlow variants with samples from IgDiff [18], a recent SE(3)-diffusion model for antibody variable domain generation based on FrameDiff [19]. We also investigate the structural realism, diversity, and novelty of IgFlow-designed structures. Finally, we compare the performance of IgFlow when conditionally designing all CDR regions and designing just the HCDR3 conditioned on the remainder of the variable domain.

2 Methods

Our approach builds upon recent advancements in SE(3)-diffusion and flow matching to develop a generative model for antibody structures. Specifically, we adapt flow matching techniques, as introduced by Lipman et al. [7], and the protein backbone frame parameterization used by Yim et al. [19], [20], for antibody design. This enables efficient, simulation-free training of continuous normalizing flows (CNFs) for antibody structure generation. We apply this method to antibodies with a focus on accurate generative modeling of CDRs and overall antibody geometry.

2.1 SE(3)-Flow Matching for Antibody Structure Generation

We model the generative process for the continuous backbone frames using SE(3)-flow matching in the same manner as FrameFlow [15]. A full description of representations is given in the Appendix. Given an initial distribution $p_0(T_0) = p_{\text{noise}}(T_0)$, we evolve the frames through a learned time-dependent vector field $v(T_t, t)$. This vector field governs the ordinary differential equation (ODE) for the structural flow with initial condition $T_0 \sim p_0$:

$$\frac{dT_t}{dt} = v(T_t, t)$$

The target distribution is the distribution $p_1(T_1)$, the distribution of antibody structures. We train a neural network to approximate the flow by regressing against the conditional field $u(T_t, t|T_1)$, which interpolates between initial (noisy) frames and final (target) frames through a series of probability distributions. Thus, the loss seeks to predict clean (target) frames from noised frames and is given by:

$$L_{\text{SE(3)}} = \mathbb{E}_{t, p_1(T_1), p_0(T_0)} \left[\|u(T_t, t|T_1) - v(T_t, t)\|^2 \right]$$

After training, antibody structures may be flexibly sampled by backward integration of the ODE by specifying a number of timesteps and variance schedule. We refer to models trained with the SE(3)-Flow Matching objective as IgFlow.

2.2 Discrete Flow Matching for Sequence Generation

For the antibody sequence, we employ masked-state interpolation discrete flow matching following MultiFlow [16], which constructs a probability flow that interpolates from a noise distribution $p_0(S_0) = p_{\text{noise}}(S_0)$ to target distribution $p_1(S_1) = p_{\text{data}}(S_1)$. This probability flow is built through a Continuous Time Markov Chain (CTMC) with a specified stochasticity level that can be adjusted at inference time to improve generation [16]. This process is given by:

$$p_t(S_t) = \mathbb{E}_{p_{\text{data}}(S_1)} \left[p_{t|1}(S_t|S_1) \right]$$

As introduced in MultiFlow, we parametrize this conditional flow towards x_1 from a specified mask state, M: $p_{t|1}^{\texttt{mask}} = \mathbf{Cat}(t\delta\{S_1, S_t\} + (1-t)\delta\{M, S_t\})$ with endpoint $S_1 = t_1$.

Similarly to FrameFlow, the final loss objective associated with this process seeks to predict the denoised amino acid distribution $p(S_1)$ from masked residues $p_{t|1}(S_t|S_1)$ through unmasking:

$$L_{\text{seq}} = \mathbb{E}_{t, p_1(S_1), p_0(S_0)} \left[\|u(S_t, t|S_1) - v(S_t, t)\|^2 \right]$$

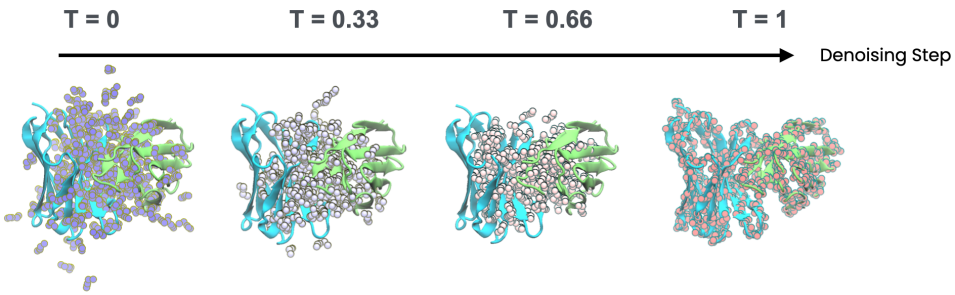

Figure 1: **IgFlow denoising step progression from noise to a highly realistic antibody structure**. $T = 0$ and $T = 1$ represent the beginning and ending denoising steps. The cyan and lime colored protein structure indicate the antibody's heavy and light chain, respectively. The backbone frame color changes from blue to red at each time step to indicate denoising progression.

2.3 Joint Structure and Sequence Flow Matching

To ensure that the generated antibody sequences and structures are compatible, we combine the SE(3)-flow matching for the backbone frames with discrete flow matching for learning residue identities. We condition the sequence flow on the generated backbone frames and vice versa to ensure this compatibility. We sample the same noisy timestep for both sequence and structure during training allowing the model to learn to denoise on the same variance schedule. We term models trained with the joint structure and sequence flow matching objective as IgFlow-Seq.

2.4 Training and Sampling Procedure

We train IgFlow on a dataset of 5142 antibody structures derived from SAbDab. We include corresponding sequences when training IgFlow with joint structure-sequence flow matching. Once trained, IgFlow can generate antibody structures and optionally sequences by backward integration of the learned ODEs for the appropriate flows. Starting from noise, we evolve the system over time to produce a final antibody structure, as shown in Figure 1, and optionally its corresponding sequence (not shown).

All other architecture, training, and inference details are described in the Appendix.

3 Experiments

3.1 Unconditional Generation of Antibody Variable Domains

We investigate IgFlow's ability to produce feasible antibodies when asked to design paired heavy and light chains. To do so, we train two different IgFlow models, one with structure flow matching (IgFlow) and one with joint structure-sequence flow matching (IgFlow-Seq). We use the same 100 heavy and light chain lengths as IgDiff, which were sampled uniformly between the 2.5 and 97.5 percentiles of lengths in the OAS [21]. We then sample 8 structures per combination of chain lengths for a total of 800 unconditional structures.

We benchmark antibody feasibility using the novelty and diversity metrics introduced in IgDiff. Novelty is calculated by computing the HCDR3 RMSD to the closest match by TM-score in our SAbDab training set with the same HCDR3 length [22]. Figure 2 (left) shows that IgFlow generated samples have similar median HCDR3 RMSD (1.94 Å) to the SAbDab validation set (2.05 Å). IgFlow-Seq samples have slightly lower median HCDR3 RMSD (1.74 Å) than the SAbDab baseline. However, both models seem capable of generating novel HCDR3 loops based on their distribution shapes. In the diversity benchmark, we use AbMPNN [23], an antibody-specific inverse folding model based on ProteinMPNN [24], to design sequences for each unconditionally generated heavy and light chain. We annotate these sequences with IMGT numbering using ANARCI [25] [26] and then compute pairwise RMSD between samples that contain CDR regions of the same chain length. Figure 2 (right) highlights the CDR distribution spread across regions which shows higher HCDR than LCDR diversity, with HCDR3 being the highest. Furthermore, we conduct an extensive

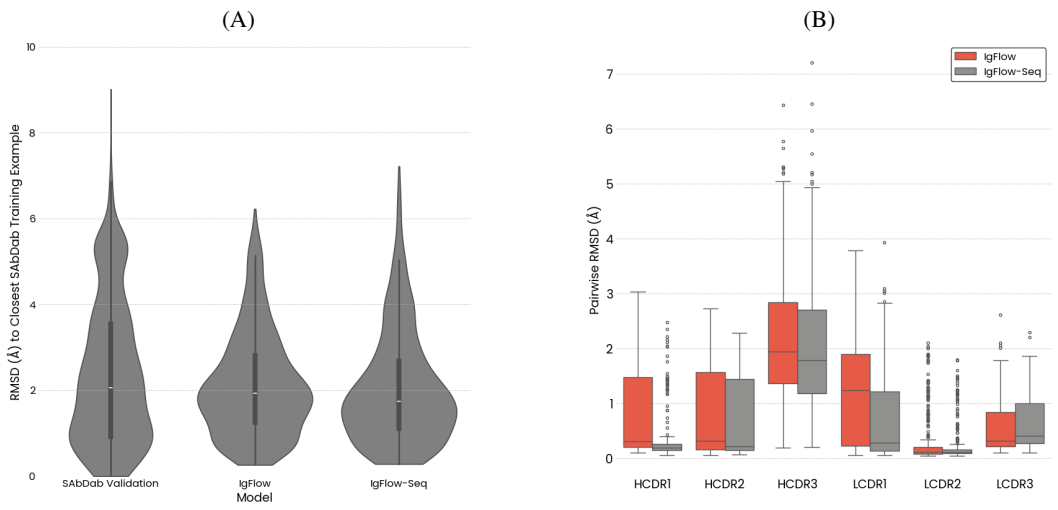

Figure 2: **800 unconditionally generated IgFlow samples are of good quality based on their novelty and diversity**. A) Both IgFlow variants show similar HCDR3 RMSD as the SAbDab baseline to their closest example in the training set as measured by TM-score, indicating that the generative model has learned an accurate distribution of the natural antibody space. For reference, we align on just the HCDR3 prior to RMSD computation. B) Pairwise RMSD between generated samples showcases high HCDR3 diversity and lower diversity for the remaining CDRs, mimicking the diversity of natural antibodies. IgFlow shows higher diversity than IgFlow-Seq for most regions, except the LCDR3 region.

length analysis in Figure 3 which highlights heavy chain length driving HCDR3 length, as expected, given that CDR3 contributes most to diversity and length of the heavy chain [27]. Together, this indicates that IgFlow is capable of designing diverse CDRs across regions while adhering to the general properties of the different CDR regions.

Next, we run the designability analysis referenced in IgDiff. For each sampled structure, we use AbMPNN to design 20 sequences using $T = 0.2$. We predict a structure for each sequence using a folding model, and compute the root mean squared deviation (RMSD) between each model predicted structure and the originally sampled structure, termed self-consistency RMSD (scRMSD). Finally, we compute the ratio of model designs with minimum scRMSD < 2 Å [†] between all predicted structures and the sampled structure, which we term the scRMSD success rate. We use two distinct folding models, ABodyBuilder2 (ABB2) [29] and ESMFold [30] to assess designability and reduce bias from any specific folding model. Over 800 generated structures, Table 3 highlights that IgDiff performs the best (all-CDR), but IgFlow-Seq and IgFlow still generate a high proportion of designable structures (all-CDR scRMSD success rates of 0.88 and 0.72 respectively).

We further compare the backbone dihedral angle distributions of unconditional IgFlow and IgFlow-Seq samples with the natural antibody distribution from SAbDab in Figure 4. This analysis shows that IgFlow adheres to the allowable dihedral distribution.

Lastly, we note that inference for IgFlow is approximately twice as fast as IgDiff, as shown in Table 2. This highlights the potential of flow matching to enhance antibody design through scaling.

3.2 Conditional Generation of CDR loops

Conditional CDR loop design is a key problem in antibody engineering that can enable generation of diverse, targeted antibody libraries in drug discovery [31]. We consider two realistic loop design tasks: full design of all 6 CDRs conditioned on the remainder of the variable domain, and HCDR3 design conditioned on the remainder of the CDRs and the rest of the variable domain. We note that the conditional designs explored here are different than those presented in IgDiff, as the CDR lengths remain fixed for the design. We build a set of 50 reference antibody structures from SAbDab that are in neither the training set of IgDiff nor IgFlow for the analysis of these two tasks. We train an IgFlow

[†]Based on [28] that cites this scRMSD cutoff as a stringent metric for self-consistency.

Folding Model		All	HCDR1	HCDR2	HCDR3	LCDR1	LCDR2	LCDR3
ABB2	*IgFlow*	0.34	0.98	0.87	0.96	0.99	1.0	0.69
	IgDiff	0.21	0.98	0.54	0.85	1.0	1.0	0.71
ESMFold	*IgFlow*	0.35	0.97	0.84	0.97	1.0	1.0	0.68
	IgDiff	0.08	0.96	0.41	0.78	1.0	1.0	0.57

Table 1: **scRMSD success rate comparison of IgDiff and IgFlow on the all-CDR conditional design task across the various CDR regions.** Higher scRMSD success rates indicate more favorable designability for the given region. "All" refers to the proportion of designs where all 6 CDRs are independently deemed designable. IgFlow performs well in design across all regions except LCDR3 as assessed by either folding model. IgDiff's performance is dependent on choice of folding model, and shows a 4-fold worse performance on all CDR designability than IgFlow when using ESMFold.

variant to design CDRs conditioned on the remainder of the framework and generate 8 structures per seed structure. We run conditional inference with IgDiff following their setup by imputing all frames at each timestep t in the undesigned region with the appropriately noised frame for t to similarly generate 8 structures per seed.

We evaluate the scRMSD success rate and RMSD with respect to ground truth structures of sampled designs using ABB2/ESMFold predicted structures from AbMPNN inverse-folded sequences. Table 1 shows IgFlow's superior performance in the all-CDR design task, covering 41 of 50 antibodies from our conditional benchmark (with 9 excluded due to AbMPNN or ANARCI annotation failures). Consistent with our unconditional design results, IgDiff's designs are more successful using ABB2 compared to ESMFold, while IgFlow performs similarly regardless of folding model. This suggests ABB2 may inflate IgDiff's performance, likely due to its training set using ABB2-predicted structures. Both models perform poorly on LCDR3, with IgDiff also struggling on HCDR2.

In their 10-structure conditional benchmark, the IgDiff authors cite HCDR1 as the most challenging, contributing to poor scRMSD (0.04) on the all-CDR task. Based on our results, we hypothesize that design challenges may be more related to specific reference structures than the model itself. Figure 6 shows IgFlow outperforms IgDiff on HCDR3 in the full CDR design setting, with $34/41 \approx 82.9\%$ of targets having lower median scRMSD for IgFlow designs. For HCDR3 design, Table 4 indicates similar performance between IgDiff and IgFlow, with Figure 7 displaying per-target scRMSDs. Between the all-CDR and HCDR3 tasks, IgDiff's scRMSD success increases from $5/41 \approx 12\%$ to $23/41 \approx 56.1\%$, while IgFlow's success rises from $16/41 \approx 39.0\%$ to $24/41 \approx 58.5\%$, demonstrating IgFlow's versatility in HCDR3 design across different conditions.

Finally, we analyze the RMSD between the designed loops and the ground truth target structure to assess how well IgFlow can reproduce known CDR loop structures. As shown in Figure 5, IgFlow achieves the lowest RMSD for LCDR2 design, while displaying higher RMSD values for HCDR1, LCDR1, and LCDR3 regions. Contrary to previous analyses of ML-based CDR prediction, the model struggles most with LCDR3 and LCDR1, which could be due to longer LCDRs or their involvement in antigen binding in our benchmark [32]. Further research is needed to examine how CDR lengths affect RMSD and scRMSD success rates.

While the success rates of IgFlow and IgDiff on the conditional tasks indicate that reliable CDR loop inpainting remains challenging, IgFlow provides a promising foundation for future models to improve CDR design capabilities.

4 Conclusion

In this work, we utilize an SE(3)-flow matching architecture for the design of antibody variable domains under conditional and unconditional settings. We find that IgFlow is a fast, sample-efficient model capable of generating realistic antibodies to its training distribution, outperforming IgDiff in certain conditional settings. Future directions will involve training IgFlow on larger amounts of experimental and synthetic antibody structures and investigating antigen-aware antibody structure-sequence co-design.

References

[1] John Ingraham et al. *Generative Models for Graph-Based Protein Design*. 2019. URL: https://openreview.net/forum?id=SJgxrLLKOE.

[2] Tim Kucera, Matteo Togninalli, and Laetitia Meng-Papaxanthos. "Conditional generative modeling for de novo protein design with hierarchical functions". In: *Bioinformatics* 38.13 (May 2022), pp. 3454–3461. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btac353. eprint: https://academic.oup.com/bioinformatics/article-pdf/38/13/3454/44268843/btac353.pdf. URL: https://doi.org/10.1093/bioinformatics/btac353.

[3] Josh Abramson et al. "Accurate structure prediction of biomolecular interactions with AlphaFold 3". In: *Nature* 630.8016 (May 2024), pp. 493–500. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07487-w. URL: http://dx.doi.org/10.1038/s41586-024-07487-w.

[4] Thomas Hayes et al. "Simulating 500 million years of evolution with a language model". In: (July 2024). DOI: 10.1101/2024.07.01.600583. URL: http://dx.doi.org/10.1101/2024.07.01.600583.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. 2020. DOI: 10.48550/ARXIV.2006.11239. URL: https://arxiv.org/abs/2006.11239.

[6] Yang Song et al. *Score-Based Generative Modeling through Stochastic Differential Equations*. 2020. DOI: 10.48550/ARXIV.2011.13456. URL: https://arxiv.org/abs/2011.13456.

[7] Yaron Lipman et al. *Flow Matching for Generative Modeling*. 2023. DOI: 10.48550/ARXIV.2210.02747. URL: https://arxiv.org/abs/2210.02747.

[8] Joseph L. Watson et al. "De novo design of protein structure and function with RFdiffusion". In: *Nature* 620.7976 (July 2023), pp. 1089–1100. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06415-8. URL: http://dx.doi.org/10.1038/s41586-023-06415-8.

[9] Guillaume Huguet et al. *Sequence-Augmented SE(3)-Flow Matching For Conditional Protein Backbone Generation*. 2024. DOI: 10.48550/ARXIV.2405.20313. URL: https://arxiv.org/abs/2405.20313.

[10] Amir Shanehsazzadeh et al. "In vitro validated antibody design against multiple therapeutic antigens using generative inverse folding". In: (Dec. 2023). DOI: 10.1101/2023.12.08.570889. URL: http://dx.doi.org/10.1101/2023.12.08.570889.

[11] Jeffrey A. Ruffolo et al. "Adapting protein language models for structure-conditioned design". In: (Aug. 2024). DOI: 10.1101/2024.08.03.606485. URL: http://dx.doi.org/10.1101/2024.08.03.606485.

[12] Nathaniel R. Bennett et al. "Atomically accurate de novo design of single-domain antibodies". In: (Mar. 2024). DOI: 10.1101/2024.03.14.585103. URL: http://dx.doi.org/10.1101/2024.03.14.585103.

[13] Harry W. Schroeder and Lisa Cavacini. "Structure and function of immunoglobulins". In: *Journal of Allergy and Clinical Immunology* 125.2 (Feb. 2010), S41–S52. ISSN: 0091-6749. DOI: 10.1016/j.jaci.2009.09.046. URL: http://dx.doi.org/10.1016/j.jaci.2009.09.046.

[14] Yuko Tsuchiya and Kenji Mizuguchi. "The diversity of <scp>H</scp>3 loops determines the antigen-binding tendencies of antibody <scp>CDR</scp> loops". In: *Protein Science* 25.4 (Jan. 2016), pp. 815–825. ISSN: 1469-896X. DOI: 10.1002/pro.2874. URL: http://dx.doi.org/10.1002/pro.2874.

[15] Jason Yim et al. "Fast protein backbone generation with SE (3) flow matching". In: *arXiv preprint arXiv:2310.05297* (2023).

[16] Andrew Campbell et al. "Generative Flows on Discrete State-Spaces: Enabling Multimodal Flows with Applications to Protein Co-Design". In: *arXiv preprint arXiv:2402.04997* (2024).

[17] James Dunbar et al. "SAbDab: the structural antibody database". In: *Nucleic Acids Research* 42.D1 (Nov. 2013), pp. D1140–D1146. ISSN: 1362-4962. DOI: 10.1093/nar/gkt1043. URL: http://dx.doi.org/10.1093/nar/gkt1043.

[18] Daniel Cutting et al. *De novo antibody design with SE(3) diffusion*. 2024. eprint: arXiv:2405.07622.

[19] Jason Yim et al. "SE(3) diffusion model with application to protein backbone generation". In: (2023). eprint: arXiv:2302.02277.

[20] Jason Yim et al. “Improved motif-scaffolding with SE(3) flow matching”. In: *Transactions on Machine Learning Research* (2024). ISSN: 2835-8856. URL: https://openreview.net/forum?id=fa1ne8xDGn.

[21] Tobias H. Olsen, Fergus Boyles, and Charlotte M. Deane. “Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences”. In: *Protein Science* 31.1 (Oct. 2021), pp. 141–146. ISSN: 1469-896X. DOI: 10.1002/pro.4205. URL: http://dx.doi.org/10.1002/pro.4205.

[22] Y. Zhang. “TM-align: a protein structure alignment algorithm based on the TM-score”. In: *Nucleic Acids Research* 33.7 (Apr. 2005), pp. 2302–2309. ISSN: 1362-4962. DOI: 10.1093/nar/gki524. URL: http://dx.doi.org/10.1093/nar/gki524.

[23] Frédéric A. Dreyer et al. *Inverse folding for antibody sequence design using deep learning*. 2023. DOI: 10.48550/ARXIV.2310.19513. URL: https://arxiv.org/abs/2310.19513.

[24] J. Dauparas et al. “Robust deep learning–based protein sequence design using ProteinMPNN”. In: *Science* 378.6615 (Oct. 2022), pp. 49–56. ISSN: 1095-9203. DOI: 10.1126/science.add2187. URL: http://dx.doi.org/10.1126/science.add2187.

[25] James Dunbar and Charlotte M. Deane. “ANARCI: antigen receptor numbering and receptor classification”. In: *Bioinformatics* 32.2 (Sept. 2015), pp. 298–300. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv552. URL: http://dx.doi.org/10.1093/bioinformatics/btv552.

[26] Marie-Paule Lefranc et al. “IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains”. In: *Developmental amp; Comparative Immunology* 29.3 (Jan. 2005), pp. 185–203. ISSN: 0145-305X. DOI: 10.1016/j.dci.2004.07.003. URL: http://dx.doi.org/10.1016/j.dci.2004.07.003.

[27] Sara D’Angelo et al. “Many Routes to an Antibody Heavy-Chain CDR3: Necessary, Yet Insufficient, for Specific Binding”. In: *Frontiers in Immunology* 9 (Mar. 2018). ISSN: 1664-3224. DOI: 10.3389/fimmu.2018.00395. URL: http://dx.doi.org/10.3389/fimmu.2018.00395.

[28] Joseph L. Watson et al. “Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models”. In: (Dec. 2022). DOI: 10.1101/2022.12.09.519842. URL: http://dx.doi.org/10.1101/2022.12.09.519842.

[29] Brennan Abanades et al. “ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins”. In: *Communications Biology* 6.1 (May 2023). ISSN: 2399-3642. DOI: 10.1038/s42003-023-04927-7. URL: http://dx.doi.org/10.1038/s42003-023-04927-7.

[30] Zeming Lin et al. “Evolutionary-scale prediction of atomic-level protein structure with a language model”. In: *Science* 379.6637 (Mar. 2023), pp. 1123–1130. ISSN: 1095-9203. DOI: 10.1126/science.ade2574. URL: http://dx.doi.org/10.1126/science.ade2574.

[31] Sai Pooja Mahajan et al. “Hallucinating structure-conditioned antibody libraries for target-specific binders”. In: *Frontiers in Immunology* 13 (Oct. 2022). ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.999034. URL: http://dx.doi.org/10.3389/fimmu.2022.999034.

[32] Henry Kenlay et al. *ABodyBuilder3: Improved and scalable antibody structure predictions*. 2024. eprint: arXiv:2405.20863.

[33] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (July 2021), pp. 583–589. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2. URL: http://dx.doi.org/10.1038/s41586-021-03819-2.

[34] G. N. Ramachandran et al. “Stereochemistry of polypeptide chain configurations.””. In: *J. mol. Biol* 7 (1963), pp. 95–99.

A Appendix

A.1 Additional Experimental Details

A.1.1 Representations

We model the structure of the antibody backbone by closely following the SE(3)-flow matching framework introduced by FrameFlow. Using the backbone frame parametrization in AlphaFold2, each residue in the antibody is represented by a rigid frame consisting of a rotation matrix $r_i \in \mathrm{SO}(3)$ and a translation vector $x_i \in \mathbb{R}^3$ forming a group element $T_i = (r_i, x_i) \in \mathrm{SE}(3)$ [33]. The full antibody backbone of N residues is then represented by a sequence of these frames $\{T_1, \ldots, T_N\} \in \mathrm{SE}(3)^N$. We refer to FrameDiff [19] for a full description of the frame representation for proteins. Importantly, we follow FrameFlow's convention of representing the protein structure within the zero center of mass (CoM) subspace of $\mathbb{R}^{N\times 3}$ by subtracting the CoM from all datapoints. This is critical to ensuring the distribution of generated frames is SE(3)-invariant. All other details follow FrameFlow.

Following MultiFlow for conditional flow-matching of sequences, each residue $s_i \in \{s_1, \ldots, s_N\}$ is associated with an amino acid type s_i belonging to the discrete set of twenty standard amino acids.

When training IgFlow for unconditional variable domain generation, we follow the minimal set of FrameFlow features introduced for general protein design. We find that in the conditional CDR generation setting, it is helpful to utilize additional equivariant features like backbone dihedrals and residue centrality and invariant features like relative sequence positional information and secondary structure information. We also include self-conditioning on distances and backbone dihedral angles.

A.1.2 Model Architecture

Based on the empirical findings of MultiFlow, we modify the FramePred architecture [19] developed in FrameDiff and used in FrameFlow and MultiFlow. We increase the number of Invariant Point Attention (IPA)-transformer blocks from 6 to 8 and increase the number of transformer layers per block from 2 to 4, which increases the depth of the network. We also add dropout to each transfomer layer.

A.1.3 Training and Sampling Procedure

Training and validation splits were made using SAbDab by clustering at 40% antigen sequence similarity with 10% of clusters being reserved for validation. Note that our training set of 5142 antibodies represents an almost 30-fold decrease in training set size compared to IgDiff, which is trained on $148,832$ ABB2 predicted structures based on sequences from the Observed Antibody Space (OAS) [18] [21]. We preprocess each antibody structure by trimming to just its variable domain.

For each training step, we sample a time $t \in [0, 1]$, noise the structure and sequence (for IgFlow-Seq) to t, and compute the loss over the structure flow or jointly over the structure and sequence flows. We additionally supervise IgFlow over auxiliary losses to discourage violations (i.e. chain breaks and clashes) and improve sequence recovery.

The total loss is formally defined as:

$$L_{\text{total}} = \begin{cases} \lambda_{\text{structure}} L_{\text{SE(3)}} + \lambda_{\text{sequence}} L_{\text{seq}} + \lambda_{\text{aux}} L_{\text{aux}} & \text{for IgFlow-Seq} \\ \lambda_{\text{structure}} L_{\text{SE(3)}} + \lambda_{\text{aux}} L_{\text{aux}} & \text{for IgFlow} \end{cases}$$

For IgFlow, we set $\lambda_{\text{structure}} = 1$. For IgFlow-Seq, we set $\lambda_{\text{sequence}} = 1$ and $\lambda_{\text{structure}} = 1$. For all IgFlow models, we use $\lambda_{\text{aux}} = 0.1$ for both a violation loss, that penalizes chain breaks and clashes, and a cross-entropy loss for sequence recovery.

All models are trained using an AdamW optimizer with a learning rate of 0.0001, batch size of 1, and for a total of 25 epochs on 8 A100 GPUs.

A.2 Figures

Model	Sampling speed (sec)
IgFlow	11.63
IgFlow-Seq	15.43
IgDiff	25.88

Table 2: **IgDiff versus IgFlow inference speeds calculated from a sample of 800 generated structures**. IgFlow can increase structure sampling capabilities which is better for large-scale biomolecular prediction campaigns.

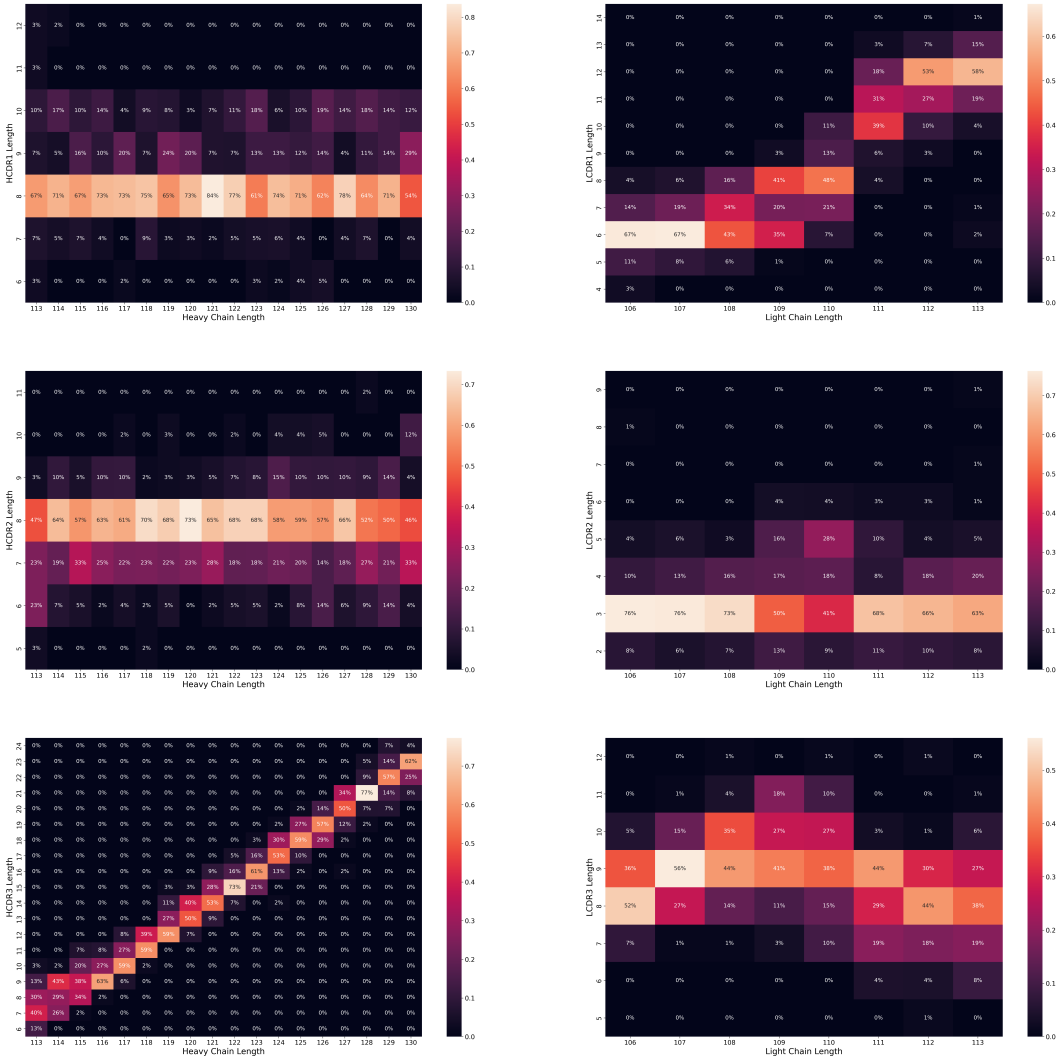

Figure 3: **Distribution of CDR loop lengths across designs of specified heavy and light chain lengths**. We note that HCDR length changes seem to directly correlate and possibly drive HCDR3 length, with smaller associations observed for the other chains. Empirically, these chain distributions are similar to those reported by IgDiff, indicating that both models learn similar CDR patterns even though they are trained on different antibody sets.

Folding Model		IgDiff	IgFlow	IgFlow-Seq
ABB2	*Joint CDR scRMSD*	0.96 (0.88)	0.49	0.65
	All CDR scRMSD	0.99	0.73	0.87
ESMFold	*Joint CDR scRMSD*	0.92	0.48	0.69
	All CDR scRMSD	0.99	0.72	0.88

Table 3: **Quality of 800 unconditionally generated samples from IgFlow variants and IgDiff as evaluated by success rate on two scRMSD tests across two different folding models, ABody-Builder2 (ABB2) and ESMFold**. The Joint CDR scRMSD test requires all 6 CDRs to independently have RMSD < 2Å between the refolded structure and the sampled structure. The All CDR scRMSD test requires that the amortized scRMSD across CDRs is less than 2Å, which is a slightly less stringent metric. We directly used unconditionally sampled structures released in the IgDiff paper for this computation and thus report the ABB2 scRMSD from the IgDiff paper in parentheses for reference.

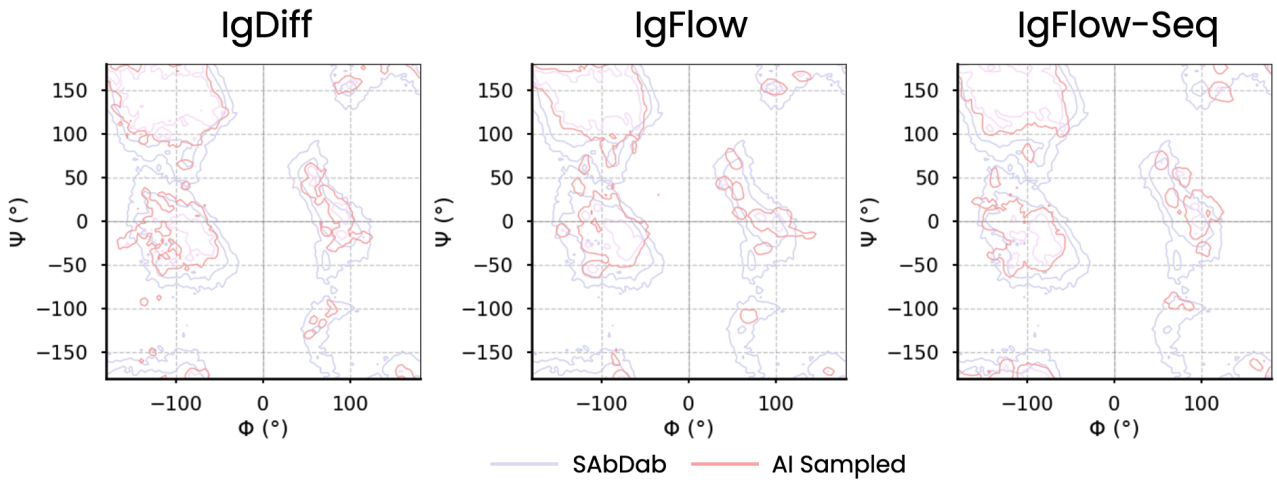

Figure 4: **Ramachandran plot [34] of dihedral angle distributions of IgDiff, IgFlow, and IgFlow-Seq (red contour) as compared to our SAbDab training set (blue contour)**. All 800 unconditionally generated samples for each model type are displayed. For SAbDab, we compute dihedral angles of only the antibody Fv region. We see similar model-generated dihedral hotspots appearing across model types. The KL-divergence (lower is better) of IgDiff is 0.84, IgFlow is 0.70, and IgFlow-Seq is 0.80, indicating that the samples generated by IgFlow are the best, followed by IgFlow-Seq and IgDiff in terms of matching the distribution of SAbDab. Although no model has full coverage over the SAbDab distribution, it is noteworthy that IgFlow and IgFlow-Seq have minimal dihedral angle placement in non-allowed regions (i.e., outside contour), in contrast to IgDiff, where the violation is more conspicuous.

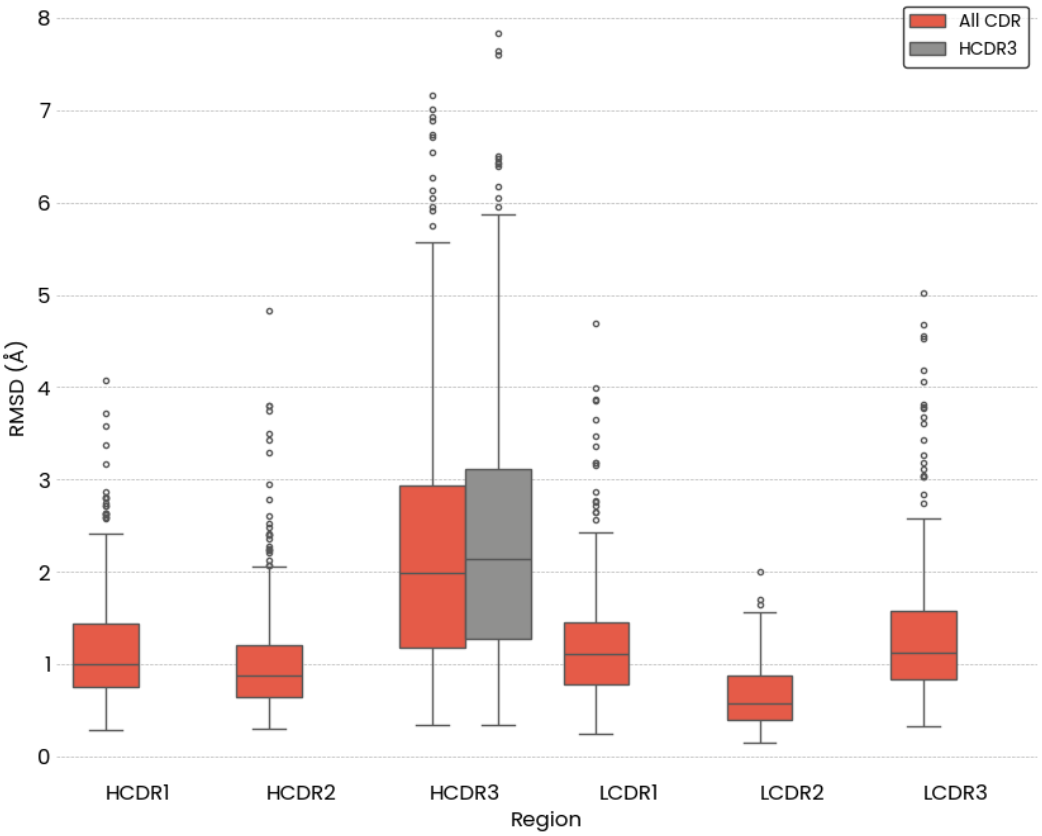

Figure 5: **RMSD on a set of 50 SAbDab antibody structures held out from a conditional IgFlow variant on two tasks: design of all 6 CDRs (orange) and design of just the HCDR3 (grey)** . For each task, 8 CDR designed structures were sampled. As expected, the HCDR3 RMSD is the highest. The median HCDR3 RMSD for the all-CDR design task is 1.98 Å while median HCDR3 RMSD for the HCDR3 design task is 2.13 Å. The conditional IgFlow model was trained to recapitulate all 6 CDRs given framework features and frames, which may explain the higher median HCDR3 RMSD when IgFlow is tasked with solely designing the HCDR3 region at inference time. Unexpectedly, the distribution of heavy and light chain predictions suggests that LCDR1 and LCDR3 design are difficult for IgFlow.

Folding Model	Model	HCDR3
ABB2	IgFlow	0.95
	IgDiff	0.95
ESMFold	IgFlow	0.72
	IgDiff	0.78

Table 4: **HCDR3 scRMSD success rates for IgDiff and IgFlow on the HCDR3 conditional designability task**. IgFlow and IgDiff are equally performant on this task with IgDiff performing slightly better under ESMFold.

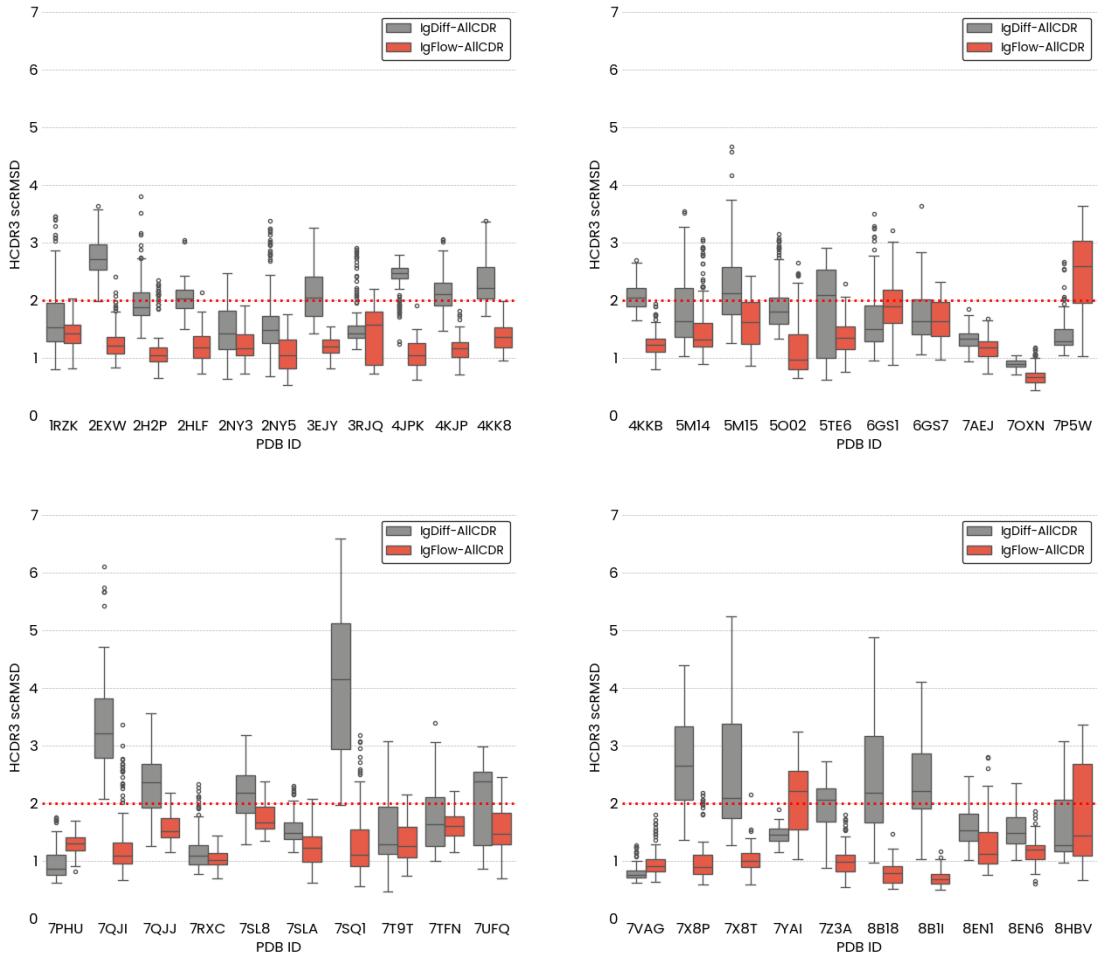

Figure 6: **HCDR3 scRMSD on the all-CDR conditional design task between IgDiff (grey) and IgFlow (orange) using ESMFold as the folding model**. 41 unique antibody structures in SAbDab of various formats (i.e. nanobodies and Fabs) from our conditional benchmark are visualized here. Each structure has 160 datapoints, representing 8 designs sampled from each given model and the associated 20 ESMFold predicted structures per designed sample. IgFlow is generally more performative than IgDiff in terms of HCDR3 loop design with $34/41 \approx 82.9\%$ of targets having lower IgFlow median HCDR3 scRMSD than IgDiff. The red line indicates the typical 2 Å cutoff for measuring designability with the scRMSD metric. Almost 40% $(16/41)$ of the targets have no IgFlow designs failing the designability threshold. This highlights the usefulness of IgFlow for CDR design in antibody engineering tasks.

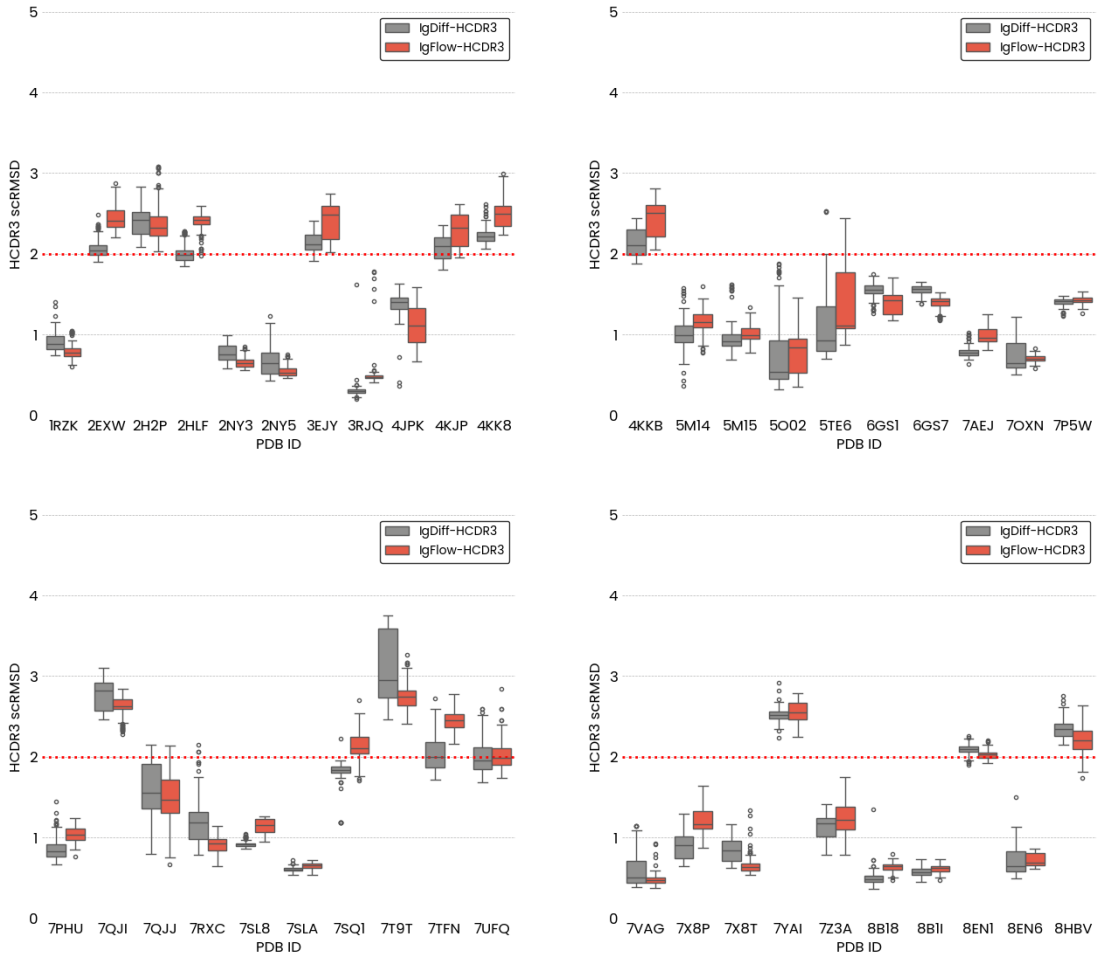

Figure 7: **HCDR3 scRMSD on the HCDR3-only conditional design task between IgDiff (grey) and IgFlow (orange) using ESMFold as the folding model**. The same protocol used in the all-CDR design task is followed to generate predicted structures for each of the 41 references. IgDiff performs slightly better on a per-target basis with $26/41 = 63.4\%$ having lower median HCDR3 scRMSD than IgFlow. For $24/41 = 58.5\%$ of targets, IgFlow designs have no structures with greater than 2 Å scRMSD, while $23/41 = 56.1\%$ of IgDiff designs meet the same threshold for designability. We see examples of successes with VHHs (i.e. 7SL8/7SLA) and Fabs (i.e. 4JPK), which indicates that both models can be useful across antibody formats. Additionally, we see designability success on both medium (8EN6 with a 13 residue HCDR3) and long HCDR3 lengths (1RZK with a 21 residue HCDR3).