ABDD: EXPERIMENTALLY VALIDATED ANTIBODY DEVELOPABILITY OPTIMIZATION VIA DISCRETE DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Therapeutic antibody development often fails due to poor developability issues identified late in discovery, such as aggregation, polyspecificity, poor expression, or low solubility. Elevated hydrophobicity is a common liability, contributing to aggregation and high viscosity. Here, we introduce AbDD, a 350M-parameter antibody-specific discrete diffusion model that jointly models amino acid sequences and Foldseek structural tokens. By combining AbDD with Reward-Guided Evolutionary Refinement in Diffusion (RERD), we provide a flexible, training-free framework for optimizing antibodies against arbitrary black-box property predictors while constraining structural deviation from the parent antibody. We experimentally validated this approach by optimising two clinical antibodies with known hydrophobicity liabilities, Galiximab and Rilotumumab. Across nine validated variants, we achieved an 89% success rate in reducing hydrophobic interaction chromatography (HIC) retention times, with top variants reaching the therapeutically acceptable range with only two mutations. Orthogonal experimental assays (SEC, BVP, VIBE) confirmed no introduction of major new liabilities.

1 INTRODUCTION

Antibodies are among the fastest-growing therapeutic modalities, with the potential to target virtually any biomolecule with high affinity and specificity. However, antibody discovery is extremely expensive, estimated at over $2 billion per approved therapeutic, largely due to high attrition rates (DiMasi et al., 2016). A significant cause of these failures is developability issues driven by antibody biophysical properties, such as aggregation, polyspecificity, poor expression, immunogenicity and low solubility (Raybould et al., 2024; Mieczkowski et al., 2023). These failure modes are often identified late in the discovery process, requiring expensive downstream engineering or abandoning a promising lead. Furthermore, current *in vitro* and *in vivo* discovery techniques often lead to antibodies with high affinity and specificity, but poor biophysical developability properties (see Svilenov et al. (2023); Shehata et al. (2019) for a consideration of this trade-off).

Discrete diffusion models have emerged as a promising paradigm for protein sequence generation, formulating generation as iterative denoising from masked sequences (Austin et al., 2023). This enables both *de novo* generation and sequence optimization through infilling. Several models have been developed for general proteins, including EvoDiff (Alamdari et al., 2024), DPLM (Wang et al., 2024a), and ESM3 (Hayes et al., 2024), with recent work incorporating discrete structural representations alongside sequence (Wang et al., 2024b). Importantly, diffusion models are amenable to inference-time conditioning, allowing optimization for arbitrary properties without retraining through derivative-free sampling methods (Li et al., 2024; Uehara et al., 2025). These approaches have been applied to property-conditioned protein and antibody de novo generation (Zhao et al., 2025; Uehara et al., 2025), but have not been applied to antibody optimization or been experimentally validated.

Here, we introduce AbDD, a 350M-parameter antibody-specific discrete diffusion model that jointly models amino acid sequences and Foldseek structural tokens (van Kempen et al., 2024), and demonstrate its combination with Reward-Guided Evolutionary Refinement in Diffusion (RERD) (Uehara

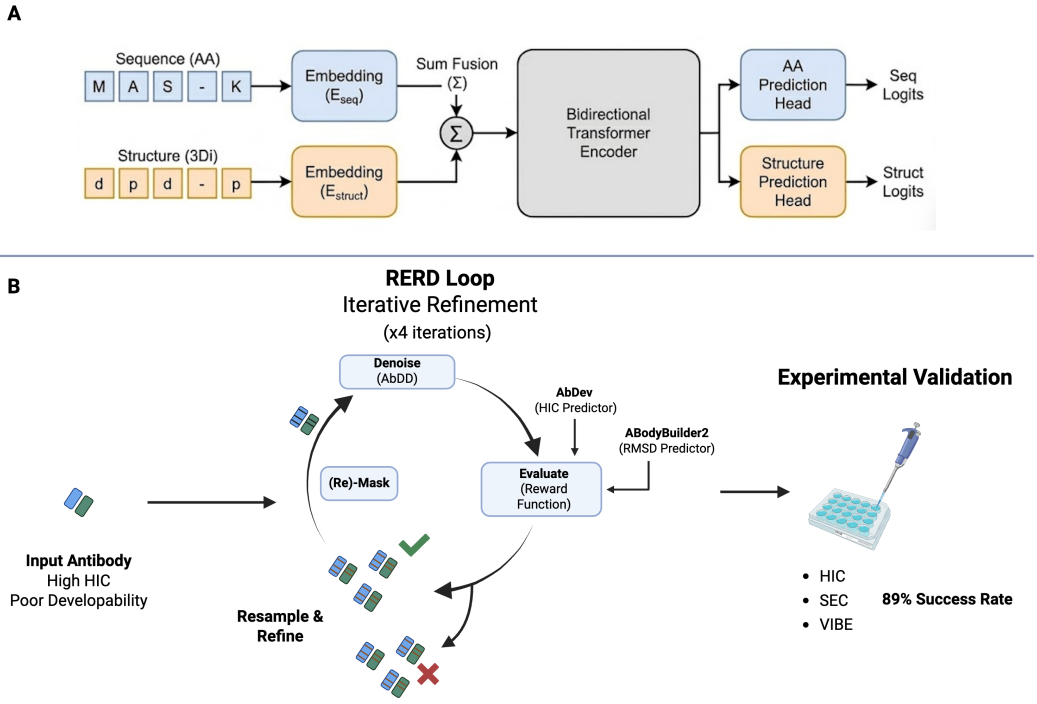

Figure 1: **Overview of AbDD and the RERD optimization framework.** (A) AbDD architecture: antibody sequences are aligned to IMGT numbering, structures predicted with ABodyBuilder2, and converted to Foldseek 3Di tokens. Sequence and structure tokens are independently embedded and combined via sum fusion before processing by a bidirectional transformer encoder. Separate prediction heads output sequence and structure logits. (B) RERD iterative refinement loop: starting from a parent antibody with poor developability, a subset of positions are masked and candidates are iteratively refined through cycles of denoising (AbDD), reward evaluation (combining AbDev-predicted HIC with RMSD to parent structure), resampling (discarding low-reward candidates), and remasking. Selected variants are experimentally expressed and characterized using orthogonal biophysical assays.

et al., 2025) for developability optimization. We experimentally validate optimized variants of two clinical antibodies with known hydrophobicity concerns, achieving an 89% success rate in reducing HIC values while maintaining structural similarity to the parent structure and not introducing new liabilities. This framework can incorporate any differentiable or black-box property predictor, highlighting broad applicability across antibody engineering tasks.

2 METHODS

2.1 ABDD: ANTIBODY DISCRETE DIFFUSION MODEL

AbDD is a 350M-parameter discrete diffusion model for paired antibody generation that jointly models amino acid sequences and Foldseek structural tokens (van Kempen et al., 2024). The model uses a transformer backbone with separate embeddings for sequence and structure tokens, combined via sum fusion (Figure 1A). The complete optimization framework, combining AbDD with RERD for iterative refinement, is illustrated in Figure 1B. Full architecture details are provided in Appendix A.1.

We trained on approximately 2.4 million paired heavy-light chain antibody sequences from OAS (Olsen et al., 2022), with structures predicted using ABodyBuilder2 (Abanades et al., 2023) and converted to discrete Foldseek 3Di tokens. All sequences were aligned to IMGT numbering using

ANARCI (Dunbar & Deane, 2015), yielding a fixed 288-position representation per antibody (157 heavy chain, 131 light chain positions).

The model was trained using the discrete diffusion formulation of Zheng et al. (2024), which generalises discrete diffusion to masked language modelling with variable masking rate. Training used AdamW with learning rate 6×10^{-4} for 20,000 steps with effective batch size 2048.

2.2 Inference-Time Conditioning

We use Reward-Guided Evolutionary Refinement in Diffusion (RERD) (Uehara et al., 2025), a sampling-based method for conditioning diffusion models on arbitrary reward functions at inference time. RERD improves over single-shot sampling methods by introducing an iterative refinement loop that alternates between: (1) denoising with importance sampling to bias token selection toward higher rewards, and (2) global resampling across candidate sequences, discarding low-reward samples and duplicating high-reward ones. Selected sequences are then partially remasked and refined further, allowing correction of suboptimal tokens from previous steps.

The reward for intermediate, partially masked sequences is approximated using a lookahead step: the clean sequence is predicted by selecting the most likely token for all masked positions, and the reward function is evaluated on this predicted sequence. Uehara et al. (2025) show this formulation ensures final sequences are sampled from the target distribution $p(x) \propto \exp(r(x)/\alpha)p^{\text{pre}}(x)$, balancing optimisation of the reward $r(x)$ while remaining close to the pretrained distribution $p^{\text{pre}}(x)$.

2.3 Reward Design

We optimise antibodies with known developability issues while minimising structural deviation from the parental backbone, which serves as a proxy for preserving binding function; mutations that significantly alter backbone geometry are more likely to disrupt paratope-antigen interactions. We chose RMSD over direct affinity prediction due to computational tractability (each optimization run evaluates ∼10,880 candidates, prohibitive for structure-based affinity models), and furthermore current affinity predictors lack sufficient accuracy and sensitivity to point mutations to provide reliable optimization signal (Hummer et al., 2025). We use AbDev (Wu et al., 2025), a suite of machine learning models trained on biophysical measurements from 137 clinical-stage antibodies (Jain et al., 2017) to predict 12 developability-related properties from sequence, as a developability oracle. In our case study, we targeted reducing hydrophobicity, which is implicated in aggregation and viscosity developability risks. We used AbDev predictions of hydrophobic interaction chromatography (HIC) and standup monolayer adsorption chromatography (SMAC), two *in vitro* assays of antibody hydrophobicity, as reward signals. Notably, RERD is derivative-free, enabling the use of non-differentiable black-box reward models.

The aggregate reward combines developability metrics with a structural penalty:

$$R_{\text{agg}}(x) = \sum_i R_i(x) - \lambda_{\text{struct}} \cdot \text{RMSD}(x) \tag{1}$$

where R_i are normalised developability rewards, and $\lambda_{\text{struct}} = 1.0$ weights the structural constraint, chosen empirically to balance property improvement against structural deviation given typical candidate RMSD values of 0.1–0.3 Å. Each developability metric is min-max normalised to [0, 1] using observed ranges from AbDev predictions on the Jain et al. (2017) clinical-stage therapeutic dataset (HIC: 9–13, SMAC: −0.5–0.8). For HIC and SMAC, where lower assay values indicate better developability, normalised rewards are inverted. RMSD measures backbone deviation between each candidate's predicted structure (ABodyBuilder2) and the parent structure.

We used the following RERD parameters: population size $R = 20$, candidates per edit $C = 20$, iterations $T = 4$, edit budget per iteration $m \approx 9$ positions (3% of sequence), selection strength $\alpha = 5$. Each optimization run evaluates approximately 10,880 candidates.

2.4 Experimental Validation

We experimentally validated optimised variants to ensure: (1) variants could be productively expressed and purified, (2) hydrophobicity was successfully reduced, and (3) optimization did not

introduce other developability concerns. We used HIC to validate hydrophobicity changes, and assessed standard developability assays, including baculovirus particle (BVP) ELISA for polyreactivity and size exclusion chromatography (SEC) for aggregation (monomer percentage). To further characterize the antibodies, we employed the Variations of Interfacial Behaviour and its Evolution (VIBE) assay. The assay measures interfacial wave phenomena generated when a protein sample stimulates a liquid interface held near a thermodynamic phase transition. The resulting biophysical fingerprint provides a multiplexed view of the sample's behaviour in solution (John et al., 2025). VIBE can identify developability risk orthogonal to conventional assays, providing a multi-property assessment of unfavourable changes.

Expression and biophysical characterisation were performed by Sino Biological. Antibody variants were expressed via transient transfection in HEK293 cells and purified using one-step Protein A affinity chromatography. Purity exceeded 90% for all variants.

3 RESULTS

3.1 DEVELOPABILITY OPTIMIZATION

We selected two clinical antibodies with known hydrophobicity liabilities as case studies: Galiximab, a discontinued anti-CD80 therapeutic, and Rilotumumab, an anti-HGF antibody. Both exhibit elevated HIC retention times relative to the distribution of approved therapeutics (Jain et al., 2017; Arsiwala et al., 2025). We jointly optimised for lower AbDev-predicted HIC and minimal structural deviation from the parent structure. For Galiximab, we also performed a second optimization run jointly targeting both HIC and SMAC, while still minimizing structural deviation.

RERD was run for four iterations, generating a final batch of 20 variants per run. We selected variants from the Pareto frontier of each run, balancing predicted property improvement, structural deviation (RMSD), and mutation count to provide diverse candidates for experimental testing. In total, nine variants were expressed and experimentally characterized: seven for Galiximab (three from HIC-only, four from HIC+SMAC optimization) and two for Rilotumumab (Table 1).

Across both antibodies, 8/9 variants (89%) showed improved HIC retention time. For Galiximab, the most improved variant achieved a 2.58-unit decrease in HIC to 11.21 (19% improvement) with only two mutations, bringing it within the 'safe' HIC range defined by Jain et al. (2017). For Rilotumumab, the best variant achieved a 2.45-unit decrease (19% improvement) with only two mutations, and both variants were rescued to within the 'safe' range. The dual optimisation run for Galiximab (HIC + SMAC) produced the best performing variant and showed 100% success rate (4/4 variants improved) compared to 67% (2/3) for HIC-only optimization, suggesting that optimising for multiple correlated properties may provide beneficial regularisation. However, the small sample sizes preclude statistical comparisons of efficacy.

One Galiximab variant (APH6-8) exhibited anomalous behaviour: although predicted to have the lowest HIC, it showed the highest experimental value (14.45, above wild-type). This failure may reflect limitations in the AbDev predictor for certain sequence contexts, or epistatic interactions between mutations that the model fails to capture. Overall correlation between predicted and experimental HIC for Galiximab was weak (Pearson $r = 0.22$, $n = 7$); excluding this outlier improved correlation substantially ($r = 0.74$, $p = 0.055$, $n = 6$). Despite mixed ranking performance, direction accuracy—correctly predicting improvement versus wild-type—was 89% across both antibodies (8/9 variants).

To assess whether HIC optimization introduced other biophysical liabilities, we evaluated variants using orthogonal assays (see Appendix Table 2). All variants maintained SEC monomer content $\geq$ 97%, which may be due to the RMSD constraint preventing destabilizing mutations by keeping variants close to the parent structure. BVP (baculovirus particle) ELISA measures polyreactivity risk; elevated BVP is associated with faster clearance and reduced efficacy. Most variants showed modest BVP changes: for Galiximab, values ranged from 0.09–0.39 (WT: 0.15), and the best-performing variant APH6-10 showed decreased BVP (0.10, Δ=-0.05). For Rilotumumab, variants showed slight BVP increases (0.22–0.35 vs WT: 0.18) but remained within acceptable ranges. Overall, BVP values remained within acceptable ranges, as defined by Jain et al. (2017).

Table 1: Experimental validation results for Galiximab and Rilotumumab variants. Best HIC improvement per antibody in bold. All variants maintained SEC $\geq 97\%$. VIBE assesses orthogonal developability properties.

Antibody	Opt.	Mut.	HIC	ΔHIC	VIBE	ΔVIBE
Galiximab	—	0	13.79	—	1.38	—
Galiximab	HIC	8	12.02	-1.77	1.31	-0.07
Galiximab	HIC	4	14.45	$+0.66$	1.60	+0.22
Galiximab	HIC	4	11.84	-1.95	1.09	-0.29
Galiximab	**HIC+SMAC**	**2**	**11.21**	$\mathbf{-2.58}$	1.55	+0.17
Galiximab	HIC+SMAC	4	11.83	-1.97	1.23	-0.16
Galiximab	HIC+SMAC	4	13.65	-0.14	1.25	-0.13
Galiximab	HIC+SMAC	3	13.71	-0.08	1.24	-0.15
Rilotumumab	—	0	12.63	—	0.15	—
Rilotumumab	**HIC**	2	**10.18**	$\mathbf{-2.45}$	-0.02	-0.17
Rilotumumab	HIC	1	11.07	-1.56	-0.08	-0.23

We additionally assessed variants using VIBE, a recently developed interfacial wave assay that captures developability risk orthogonal to conventional hydrophobicity measurements. VIBE changes were modest across variants, with no catastrophic increases observed, and the variant with the largest VIBE increase (+0.22) occurred in the HIC failure case. However, the Galiximab variant with the lowest HIC value showed a 12% increase in VIBE, indicating it may be capturing further information beyond hydrophobicity. Overall, these results confirm that targeted HIC optimization did not introduce major orthogonal liabilities, with either a decrease or modest increase to VIBE scores, though they highlight the value of multi-property monitoring.

4 Conclusions

By combining an antibody-specific discrete diffusion model (AbDD) with training-free inference-time conditioning (RERD), we have demonstrated experimentally validated developability optimization of therapeutic antibodies. Across two antibodies with distinct sequences and targets, we achieved an 89% success rate (8/9 variants improved), with the best variants showing HIC reductions of 2.58 units (Galiximab) and 2.45 units (Rilotumumab). Both antibodies were rescued from elevated HIC values to within the therapeutically acceptable range, with the best variants for both requiring only two mutations. Orthogonal assays (SEC, BVP, VIBE) confirmed that no major new liabilities were introduced.

While our approach achieved high direction accuracy (89%), the poor ranking correlation for Galiximab ($r = 0.22$) and one clear failure case highlight limitations of current developability predictors. The AbDev HIC model provides sufficient signal for optimization, identifying variants likely to improve over wild-type, but cannot reliably rank variants among themselves. We used RMSD as a proxy for maintaining binding function, but this does not guarantee preserved affinity, and we did not experimentally assess binding. Ideally, an antibody-antigen affinity predictor would be integrated as an additional reward, though current models lack sufficient accuracy (Hummer et al., 2025). Additionally, larger-scale validation would strengthen conclusions about success rates.

The flexibility of this framework, requiring no model retraining and accepting any differentiable or black-box reward function, makes it readily applicable to diverse antibody engineering objectives, including affinity maturation, immunogenicity reduction, and multi-property optimization as improved predictors become available. While predictor accuracy remains a limiting factor, the high directional accuracy suggests this approach can meaningfully accelerate therapeutic antibody development when combined with targeted experimental validation.

References

Brennan Abanades, Wing Ki Wong, Fergus Boyles, Guy Georges, Alexander Bujotzek, and Charlotte M. Deane. ImmuneBuilder: Deep-learning models for predicting the structures of

immune proteins. *Communications Biology*, 6(1):1–8, May 2023. ISSN 2399-3642. doi: 10.1038/s42003-023-04927-7.

Sarah Alamdari, Nitya Thakkar, Rianne van den Berg, Neil Tenenholtz, Robert Strome, Alan M. Moses, Alex X. Lu, Nicolò Fusi, Ava P. Amini, and Kevin K. Yang. Protein generation with evolutionary diffusion: Sequence is all you need, November 2024.

Ammar Arsiwala, Rebecca Bhatt, Lood van Niekerk, Porfirio Quintero-Cadena, Xiang Ao, Adam Rosenbaum, Aanal Bhatt, Alexander Smith, Yaoyu Yang, K. C. Anderson, Lucia Grippo, Xing Cao, Rich Cohen, Jay Patel, Joshua Moller, Olga Allen, Ali Faraj, Anisha Nandy, Jason Hocking, Ayla Ergun, Berk Tural, Sara Salvador, Joe Jacobowitz, Kristin Schaven, Mark Sherman, Sanjiv Shah, Peter M. Tessier, and David W. Borhani. A high-throughput platform for biophysical antibody developability assessment to enable AI/ML model training. *mAbs*, 17(1):2593055, December 2025. ISSN 1942-0870. doi: 10.1080/19420862.2025.2593055.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces, February 2023.

Joseph A. DiMasi, Henry G. Grabowski, and Ronald W. Hansen. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, 47:20–33, May 2016. ISSN 1879-1646. doi: 10.1016/j.jhealeco.2016.01.012.

James Dunbar and Charlotte M. Deane. ANARCI: Antigen receptor numbering and receptor classification. *Bioinformatics*, pp. btv552, September 2015. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btv552.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J. Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q. Tran, Jonathan Deaton, Marius Wiggert, Rohil Badkundri, Irhum Shafkat, Jun Gong, Alexander Derry, Raul S. Molina, Neil Thomas, Yousuf Khan, Chetan Mishra, Carolyn Kim, Liam J. Bartie, Matthew Nemeth, Patrick D. Hsu, Tom Sercu, Salvatore Candido, and Alexander Rives. Simulating 500 million years of evolution with a language model, July 2024.

Alissa M. Hummer, Constantin Schneider, Lewis Chinery, and Charlotte M. Deane. Investigating the volume and diversity of data needed for generalizable antibody–antigen $\Delta\Delta$G prediction. *Nature Computational Science*, 5(8):635–647, August 2025. ISSN 2662-8457. doi: 10.1038/s43588-025-00823-8.

Tushar Jain, Tingwan Sun, Stéphanie Durand, Amy Hall, Nga Rewa Houston, Juergen H. Nett, Beth Sharkey, Beata Bobrowicz, Isabelle Caffry, Yao Yu, Yuan Cao, Heather Lynaugh, Michael Brown, Hemanta Baruah, Laura T. Gray, Eric M. Krauland, Yingda Xu, Maximiliano Vásquez, and K. Dane Wittrup. Biophysical properties of the clinical-stage antibody landscape. *Proceedings of the National Academy of Sciences*, 114(5):944–949, January 2017. doi: 10.1073/pnas.1616408114.

Alexander Nicholas St John, Alec Nelson Thomas, Piergiorgio Caramazza, Alireza Meghdadi, Joe Bailey, Sejeong Lee, Vivek Ramakrishna, Alexandra Leigh Boehmke Amoruso, Elizabeth Suk-Hang Lam, Natesh Ganesh, and Shamit Shrivastava. A High-Dimensional Interfacial Wave Assay for Early Biophysical Profiling of Therapeutic Antibodies, May 2025.

Xiner Li, Yulai Zhao, Chenyu Wang, Gabriele Scalia, Gokcen Eraslan, Surag Nair, Tommaso Biancalani, Shuiwang Ji, Aviv Regev, Sergey Levine, and Masatoshi Uehara. Derivative-free guidance in continuous and discrete diffusion models with soft value-based decoding, October 2024.

Carl Mieczkowski, Xuejin Zhang, Dana Lee, Khanh Nguyen, Wei Lv, Yanling Wang, Yue Zhang, Jackie Way, and Jean-Michel Gries. Blueprint for antibody biologics developability. *mAbs*, 15 (1):2185924, December 2023. ISSN 1942-0862. doi: 10.1080/19420862.2023.2185924.

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. *Protein Science*, 31(1):141–146, 2022. ISSN 1469-896X. doi: 10.1002/pro.4205.

Matthew I. J. Raybould, Oliver M. Turnbull, Annabel Suter, Bora Guloglu, and Charlotte M. Deane. Contextualising the developability risk of antibodies with lambda light chains using enhanced therapeutic antibody profiling. *Communications Biology*, 7(1):1–13, January 2024. ISSN 2399-3642. doi: 10.1038/s42003-023-05744-8.

Laila Shehata, Daniel P. Maurer, Anna Z. Wec, Asparouh Lilov, Elizabeth Champney, Tingwan Sun, Kimberly Archambault, Irina Burnina, Heather Lynaugh, Xiaoyong Zhi, Yingda Xu, and Laura M. Walker. Affinity maturation enhances antibody specificity but compromises conformational stability. *Cell Reports*, 28(13):3300–3308.e4, September 2019. ISSN 2211-1247. doi: 10.1016/j.celrep.2019.08.056.

Hristo L. Svilenov, Paolo Arosio, Tim Menzen, Peter Tessier, and Pietro Sormanni. Approaches to expand the conventional toolbox for discovery and selection of antibodies with drug-like physicochemical properties. *mAbs*, 15(1):2164459, December 2023. ISSN 1942-0862. doi: 10.1080/19420862.2022.2164459.

Masatoshi Uehara, Xingyu Su, Yulai Zhao, Xiner Li, Aviv Regev, Shuiwang Ji, Sergey Levine, and Tommaso Biancalani. Reward-guided iterative refinement in diffusion models at test-time with applications to protein and DNA design, February 2025.

Michel van Kempen, Stephanie S. Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L. M. Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein structure search with foldseek. *Nature Biotechnology*, 42(2):243–246, February 2024. ISSN 1546-1696. doi: 10.1038/s41587-023-01773-0.

Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Diffusion language models are versatile protein learners, February 2024a.

Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. DPLM-2: A multimodal diffusion protein language model, October 2024b.

I-En Wu, Lateefat Kalejaye, and Pin-Kuang Lai. Machine learning models for predicting monoclonal antibody biophysical properties from molecular dynamics simulations and deep learning-based surface descriptors. *Molecular Pharmaceutics*, 22(1):142–153, January 2025. ISSN 1543-8384. doi: 10.1021/acs.molpharmaceut.4c00804.

Siqi Zhao, Joshua Moller, Porfi Quintero-Cadena, and Lood van Niekerk. Guided generation for developable antibodies, July 2025.

Lin Zheng, Jianbo Yuan, Lei Yu, and Lingpeng Kong. A reparameterized discrete diffusion model for text generation, February 2024.

A Appendix

A.1 Model Architecture Details

The 350M AbDD model has the following configuration: hidden dimension 960, transformer layers 32, attention heads 15, feed-forward dimension 3840, dropout 0.1. The model uses RoPE positional embeddings and separate prediction heads for amino acid and structure tokens. Figure 2 provides an overview of the model architecture.

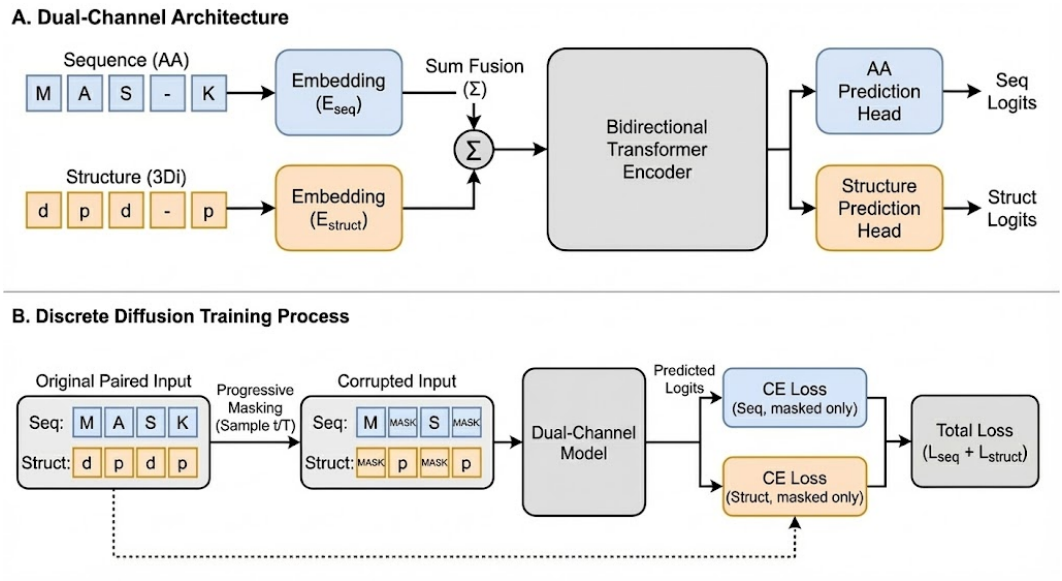

Figure 2: **A.** The Dual-Channel Architecture. Aligned amino acid (AA) sequences and Foldseek structural (3Di) sequences are independently embedded and combined via sum fusion before being processed by a bidirectional transformer encoder. Two independent prediction heads output logits for sequence and structure. **B.** The Discrete Diffusion Training Process. During training, paired inputs are corrupted via progressive masking based on a sampled timestep t. The model is trained to predict the original tokens for the masked positions, optimising the sum of cross-entropy (CE) losses for both channels.

A.2 Validation Results

Table 2: Experimental validation results for Galiximab and Rilotumumab variants. Best HIC improvement per antibody in bold. All variants maintained SEC $\geq$ 97%. BVP and VIBE assess orthogonal developability properties.

Antibody	Opt.	Mut.	RMSD	HIC	ΔHIC	SEC%	BVP	ΔBVP	VIBE	ΔVIBE
Galiximab	—	0	0.00	13.79	—	99.3	0.15	—	1.38	—
Galiximab	HIC	8	0.11	12.02	−1.77	98.5	0.21	+0.06	1.31	−0.07
Galiximab	HIC	4	0.14	14.45	+0.66	99.1	0.19	+0.04	1.60	+0.22
Galiximab	HIC	4	0.07	11.84	−1.95	98.6	0.39	+0.24	1.09	−0.29
Galiximab	**HIC+SMAC**	**2**	**0.12**	**11.21**	**−2.58**	99.2	0.10	−0.05	1.55	+0.17
Galiximab	HIC+SMAC	4	0.09	11.83	−1.97	98.4	0.11	−0.04	1.23	−0.16
Galiximab	HIC+SMAC	4	0.08	13.65	−0.14	98.8	0.13	−0.02	1.25	−0.13
Galiximab	HIC+SMAC	3	0.23	13.71	−0.08	98.4	0.09	−0.06	1.24	−0.15
Rilotumumab	—	0	0.00	12.63	—	99.0	0.18	—	0.15	—
Rilotumumab	**HIC**	2	0.04	**10.18**	**−2.45**	97.9	0.35	+0.17	−0.02	−0.17
Rilotumumab	HIC	1	0.04	11.07	−1.56	97.2	0.22	+0.04	−0.08	−0.23

A.3 GLOSSARY OF ABBREVIATIONS

Abbreviation	**Definition**
AbDD	Antibody Discrete Diffusion model
BVP	Baculovirus Particle ELISA (polyreactivity assay)
RMSD	Root Mean Square Deviation
HIC	Hydrophobic Interaction Chromatography
IMGT	ImMunoGeneTics numbering scheme
OAS	Observed Antibody Space database
RERD	Reward-Guided Evolutionary Refinement in Diffusion
SEC	Size Exclusion Chromatography
SMAC	Standup Monolayer Adsorption Chromatography
VIBE	Variations in Interfacial Behaviour and its Evolution