SurfProp: A surface-based property prediction framework for antibody developability and screening

Pranav Rao [1] **Henry Isaacson** [1] **Jennifer L. Hofmann** [1] **Darcy Davidson** [1] **Amy Wang** [1] **Andrew M. Watkins** [1] **Richard Bonneau** [1] **Saeed Izadi** [2] **Jae Hyeon Lee** [1]

Abstract

Therapeutic antibodies are an important class of drugs, increasingly used to treat a variety of conditions. Developing antibodies is challenging in part due to issues with viscosity, aggregation, solubility and more *developability* properties that pose issues for manufacturing and delivery. As these properties require large-scale experiments to measure, *in silico* biophysical molecular descriptors are often considered by proxy, and serve as a basis for screening and optimization. In this work, we introduce SurfProp, a surface-based, differentiable property prediction framework aimed at improving the antibody developability workflow in two ways. (1) The *insilico* arm of SurfProp predicts electrostatics and computes hydrophobicity with a significant speedup over traditional methods, facilitating higher throughput *in silico* property screening. (2) the *experimental* arm of SurfProp uses the pre-trained model from the *in silico* task to predict experimental developability properties; here we demonstrate the ability of the pre-trained model to learn hydrophobic interaction chromatography (HIC) more effectively than a model trained from scratch.

1. Introduction

Antibody therapeutics represent the fastest growing class of biological drugs and have been crucial in the treatment of several challenging disease conditions (Carter & Rajpal, 2022; Sharma et al., 2023). However, developing a candidate antibody into a successful drug involves considering several *developability* criteria– including stability and low viscosity at high concentration (Zarzar et al., 2023; Jain et al., 2017). A candidate molecule that passes these criteria is much more likely to be manufacturable, and formulated for safe and efficient delivery for patients. Therefore, assessing these properties is of critical importance to antibody discovery and design.

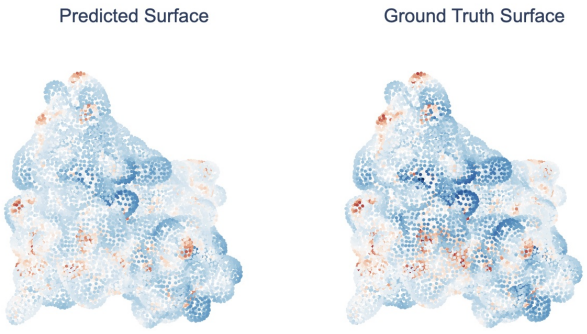

Figure 1. Predicted electrostatic potential for the antibody PDB: `4hs8`. Blue points indicate positive electrostatic potential and red points indicate negative values in the scale $[-6, 6]$. SurfProp predicts the potential with $\rho_{\text{pearson}} = .94$ for this example.

Assessing developability properties directly requires material-intensive and time consuming experiments, making it difficult to influence early stages of antibody discovery, with thousands of candidates or more, with direct experimental signals. This problem becomes even more pronounced in the age of generative models, which hold promise to improve lead identification and optimization of antibodies, but produce large amounts of diverse candidates in need of principled filtering before experimental verification. In light of this, biophysical properties – which can be calculated *in silico*– serve as a surrogate for experimental developability properties in many cases (Park & Izadi, 2024; Raybould & Deane, 2021).

Here we introduce SurfProp, a surface-based property prediction framework that enables developability considerations to influence and enhance earlier stages of the antibody

[1]Prescient Design, Genentech, South San Francisco, CA, USA [2]Pharmaceutical Development, Genentech, South San Francisco, CA, USA. Correspondence to: Pranav Rao <rao.pranav@gene.com>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

drug discovery process. The first way we do this is by speeding up the computation of biophysical properties, which currently takes on the order of minutes for each molecule and frequently relies on molecular dynamics simulations.

SurfProp-*insilico* is able to speed up biophysical calculation to the order of seconds while maintaining good accuracy to ground truth values, including a correlation $\rho_{\text{pearson}} = .90$ between predicted and ground truth potential values at the surface-point level and a 98% recovery of red (extreme) electrostatic risk flags. SurfProp-*insilico* is also able to compute surface hydrophobicity (Figure 2) directly orders of magnitude faster than traditional methods.

This work facilitates larger scale screening of biophysical properties – up to hundreds of thousands of molecules – that would be intractable with previous methods. This is especially relevant for properties annotation of high throughput experimental assays or ML generated designs. As all components of SurfProp are differentiable, the method gives promise to guide generative models based on biophysical properties and can serve as a pre-training foundation for property prediction of low-data regime experimental developability properties, both which are inaccessible with traditional biophysical property calculation methods.

We explore the latter direction through the second part of this work, which we call SurfProp-*expt*. We use hydrophobic interaction chromatography (HIC) as a testbed for experimental property prediction, aided by pretrained representations from SurfProp-*insilico*.

2. Methods

2.1. Dataset preparation & simulations

2.1.1. Molecular dynamics simulations

We utilize a large set of conventional molecular dynamics simulations for training the *in silico* portion of SurfProp. The database was curated from a combination of SabDab (Dunbar et al., 2014), TheraSabDab (Raybould et al., 2020), and diverse subselections of the observed antibody space (OAS) (Olsen et al., 2022) (see Appendix B).

Initial Fab structures were folded using ESMFold, where each variable fragment Fv sequence in the database was given the constant region (CH1, CL) of Herceptin to stabilize simulations. We simulated for 200 ns with explicit water (TIP3P) and neutralizing ions using the Amber 14 force field at a temperature of 298 K (Maier et al., 2015).

2.1.2. *In silico* dataset preparation

To create a dataset for the *in silico* task of learning the surface electrostatic potential for antibodies, we first removed waters and constant domains from the database of conventional MD simulations, totaling 13,667 trajectories from a diverse sequence space.

We used PDB2PQR to protonate 10 evenly spaced frame snapshots from each trajectory, and then utilized the Adaptive Poisson Boltzmann Solver (APBS) to solve for the grid-level electrostatic potential (Dolinsky et al., 2004; Jurrus et al., 2018). We implemented a simple interpolation function to take the grid potential values to the surfaces generated in this work, which are described in the next section.

To create training splits, we utilize MMSeqs2 (Steinegger & Söding, 2017) at 85% sequence identity to cluster the Fv sequences in our dataset, and group clusters into train, validation and test sets based on a 90 / 5 / 5 split at the sequence-cluster level.

2.1.3. Experimental dataset preparation

For the experimental part of the paper, we use a set of hydrophobic interaction chromatography (HIC) datapoints generated for 137 clinical mAbs from (Jain et al., 2017). HIC is the literature standard for experimental quantification of antibody hydrophobicity (Hebditch et al., 2019; Wang et al., 2016). Extreme surface hydrophobicity is a well-established predictor of undesirable antibody behaviors, including self-aggregation (Das et al., 2022) and fast clearance (Lyon et al., 2015).

Despite the small size of the dataset, the represented antibodies cover a diverse sequence space, and we use MMSeqs2 at 85% sequence identity to cluster and generate splits to develop a model that can generalize to unseen sequences.

2.2. Fast surface generation

We use a GPU-accelerated, point-cloud based surface generation method implemented in PyTorch (Paszke, 2019) (implementation details can be found in Appendix A).

To ensure compatibility with established surface representations used for representing biophysical properties, we tuned our surface generation parameters to match those of Nanoshaper (Decherchi & Rocchia, 2013), a widely adopted tool for computing electrostatic and hydrophobic properties on protein surfaces (Jurrus et al., 2018; Park & Izadi, 2024). In Table B, we determined that sampling 60 surface vertices per atom at a 1.0 Å resolution minimizes the distance between Nanoshaper vertices and our point-cloud surface points. This parameter selection shows our point-cloud representation sufficiently reproduces traditional surface geometry, validating its use for biophysical property calculations.

2.3. Multi-scale model architecture

We use a multi-scale model architecture that models atom and surface point-clouds in a joint representation. We first

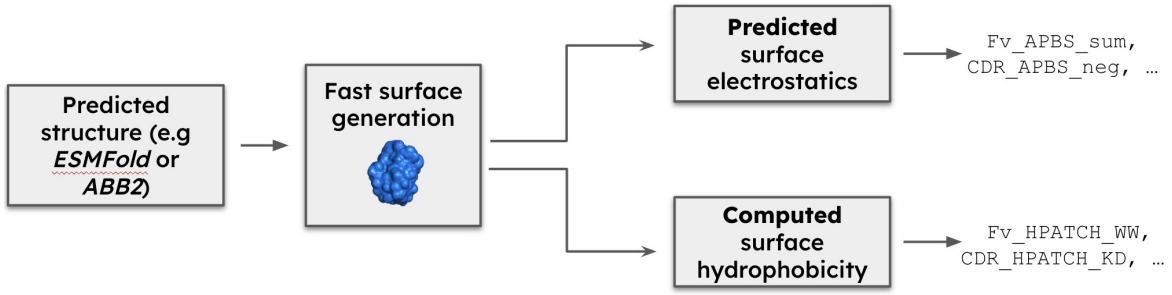

Figure 2. SurfProp-*insilico* workflow: predicted structures are featurized into surfaces, hydrophobicity is computed directly and electrostatics is predicted. Region-level descriptors are computed and compared to risk flags defined by therapeutic antibodies.

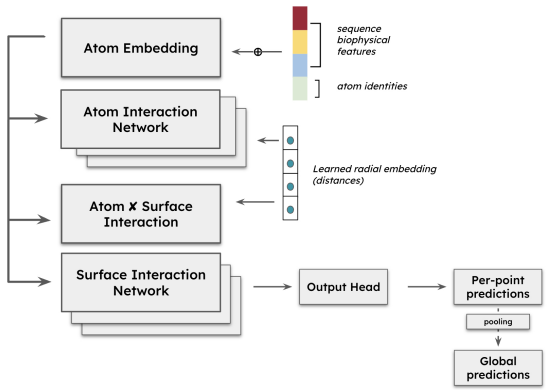

Figure 3. Multi-scale model architecture with surface interaction network for property prediction.

generate an atomic embedding using atom identity and concatenate sequence-level biophysical properties onto the atomic representation. Next, we pass the atomic representations into an atomic interaction network, which is a geometric vector perceptron (`GVP`) (Jing et al., 2020) defined on a radial graph parameterized by inter-atom distances.

The atomic representations are then projected to the surface. For each surface point, the closest atoms' features along with encoded atom-to-surface Euclidean distances– encoded through a learned radial embedding (Gao et al., 2022)– are aggregated with a `scatter` operation and then passed through an `MLP`. The resulting surface representations are passed into an interaction network, which is either another `GVP` or a Geometric Convolution (`GeomConv`) (Sverrisson et al., 2021).

We employ several regularization strategies, including a non-contrastive self supervised loss function (Maser et al., 2023) as well as an $L2$ penalty.

For the *in silico* electrostatics task, we employ an L_p loss with $p = 3$, as we found this better captured outliers (as compared with an L_2 loss) in the electrostatic potential distribution that are important for developability assessment. For the *experimental* HIC prediction task, we use L_1 loss due to the size of the dataset and less emphasis on outlier detection.

3. SurfProp-*insilico*

The in silico portion of SurfProp addresses a bottleneck preventing antibody developability assessment at scale by attempting to improve and speed up biophysical property calculation. We *predict* electrostatic potential and *compute* hydrophobicity allowing for a significantly faster and differentiable version of surface-based biophysical property calculation.

3.1. Electrostatics: surface potential prediction

3.1.1. Per-point regression task performance

We use the per-point predictions of SurfProp to learn the potential values on the antibody surface on the *in silico* dataset from dynamics simulations and the APBS package.

SurfProp is able to effectively learn the per-point potential values at the surface with a correlation of $\rho_{\text{pearson}} = 0.90$ on the test set (see Figure 4). The model serves as a conservative predictor, with errors primarily arising from underprediction of extreme values which is ideal for risk assessment.

3.1.2. Predicted risk flags: classification task

To assess the practical utility of the predicted surface electrostatics, we look at the ability of predicted values to recapitulate risk flags that identify problematic electrostatic properties based on therapeutic antibody distributions.

Risk flags are defined for different antibody regions, calculating electrostatic descriptors separately for the full Fv domain and individual complementarity determining regions (CDRs). Similar to MOLDESK, we define risk flags for each descriptor based on extreme values in the distribution

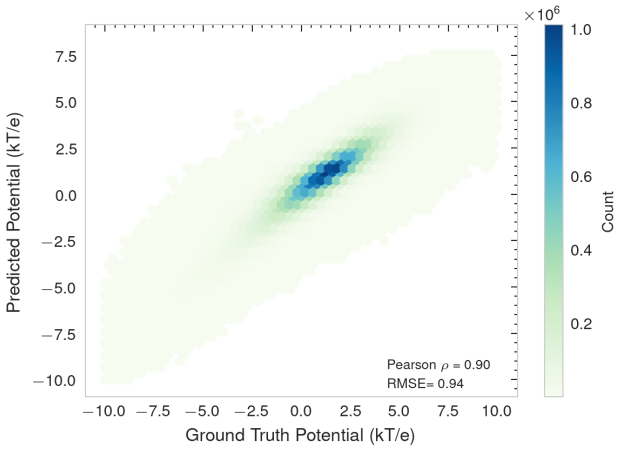

Figure 4. Per-point regression results for learning surface electrostatic potential. The model is a conservative predictor, with error arising from under-prediction.

of the corresponding descriptor over the set of therapeutic antibodies from TheraSabDab. Specifically we define:

$$\text{Red flag} > 95\text{th percentile}$$
$$\text{Amber flag} > 90\text{th percentile}$$
$$\text{Yellow flag} > 75\text{th percentile}$$

The recovery metric represents the fraction of true positives correctly flagged by any predicted risk flag level 3.1.2. When using original thresholds, SurfProp demonstrates strong recovery of red flags (93.39%) and (84.17%) with lower sensitivity to capture yellow (62.92%) flags. After refitting thresholds to predicted values, recovery improves substantially across all categories (98.71% for red, 94.08% for amber, and 77.67% for yellow).

The additional yellow risk flag (not seen in MOLDESK or TAP) allows us to identify risk across more portions of the predicted distribution while maintaining conservative predictions with few false positives. For practical screening, we recommend completely filtering out candidates that trigger red or amber flags, while examining yellow-flagged candidates on a case-by-case basis.

3.2. Hydrophobicity: accelerated calculation

Hydrophobicity is another key biophysical property that can influence developability criteria. Existing methods such as MOLDESK (Park & Izadi, 2024) and TAP (Raybould & Deane, 2021). Both methods report surface hydrophobicity, HPATCH in the former method and patches of surface hydrophobicity (PSH) in the latter. These methods inform sequence level hydrophobicity scales with surface geometry, but come with significant computational cost. In the case of MOLDESK, computing HPATCH takes about 60% of the overall computation time (see Table 3).

To address this, we accelerate the computation of HPATCH with vectorized GPU operations. These scores are calculated by assigning residue-level hydrophobicity scales to both individual atoms and the closest surface points to each atom. Subsequently the hydrophobicity values are averaged within 10 Å neighborhoods, informing the HPATCH score of the surface geometry. This significantly speeds up the computation of hydrophobicity, bringing down the time from around 28 seconds to .03 seconds in the representative cases we show in Table 3. SurfProp also calculates region-level hydrophobicity descriptors and risk flags. Since we calculate these descriptors directly from sequence and surface, a classification-type task as in Sec. 3.1.2 does not apply here.

3.3. Correlation to experimental quantities

We also demonstrate the ability of SurfProp to retain correlations between predicted electrostatics and experimental properties.

We first examine the correlation between electrostatics and Heparin retention time for clinical antibodies (Jain et al., 2023). Heparin is negatively charged and interacts with positive surface patches translating to greater retention time in chromatography for stronger interactions. Our predicted electrostatics with SurfProp captures this phenomena with a strong correlation overall and on held out examples ($\rho_{\text{pearson}} = 0.79$). For this evaluation, we used structures from our MD database for evaluation, taking ensemble averages of model outputs.

Next, we examine high concentration viscosity driven by electrostatics in the Apgar dataset, which contains viscosity measurements of 38 anti-PDGF antibodies at 150 mg/ml (Apgar et al., 2020). The antibodies in this set are outside of the *in silico* dataset, so we use ABodyBuilder2 (Abanades et al., 2023) predicted Fv structures as a starting point for SurfProp evaluation. SurfProp has strong performance, capturing the correlations between negative CDR electrostatic patches and (log) viscosity ($\rho_{\text{pearson}} = 0.84$), despite being evaluated on predicted structures rather than trajectory snapshots, an important test of robustness. This demonstrates that SurfProp can be applied successfully with a variety of more quickly accessible inputs like predicted structures.

We also evaluate SurfProp's hydrophobicity calculations on the Dai dataset (Dai et al., 2024), which contains viscosity measurements for a different set of antibodies. SurfProp captures the correlation between computed surface hydrophobicity and experimental viscosity ($\rho_{\text{pearson}} = 0.48$), showing that our accelerated hydrophobicity calculations retain meaningful correlations with experimental properties.

Predicted (electrostatics) risk flag evaluation						
Metric	Original Thresholds			Re-fit Thresholds		
	Red	Amber	Yellow	Red	Amber	Yellow
Mean Recovery (%)	93.39	84.17	62.92	**98.71**	**94.08**	**77.67**
Mean Precision (%)	**77.88**	**85.02**	**83.78**	69.74	72.60	75.72
Mean Accuracy (%)	31.97	43.54	62.92	**59.32**	**67.32**	**77.67**

Table 1. Performance of SurfProp predicted surface electrostatics on risk flagging in comparison to ground truth risk flags. Recovery is defined as the percentage of ground truth flags at a particular level that are flagged by *any* level of predicted flag. Note that for yellow risk flags, recovery and accuracy are equivalent.

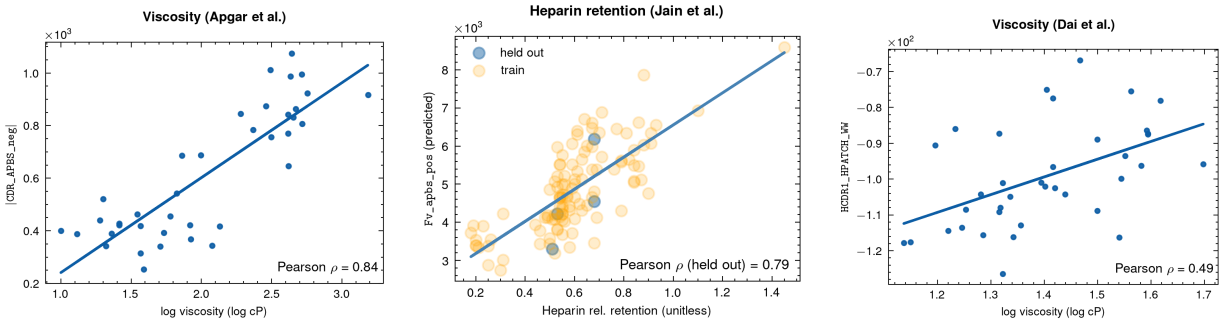

Figure 5. Left: Correlation between predicted `CDR_apbs_neg` and log (base 10) experimental viscosity (Apgar dataset). Middle: Correlation between predicted `Fv_apbs_pos` and relative heparin retention time. Right: Correlation between computed surface hydrophobicity `HCDR1_HPATCH_WW` and log (base 10) experimental viscosity (Dai dataset).

3.3.1. Computational Efficiency

For electrostatics, the average time for one forward pass at inference time is 40 ms on one A100 GPU, while the hydrophobicity calculation takes on average 30 ms per structure. Combined with surface generation, which around 200 ms for our test set on average, the total time for the core evaluation of surface-level is a fast 270 ms. We include pre and post processing steps in Table 3 (for example `PDB2PQR` protonation and `ANARCI` sequence alignment) and show a concrete comparison to MOLDESK for two representative structures.

These steps could be optimized further, for example using `ANARCII` (Greenshields-Watson et al., 2025) for sequence alignment. Even the most conservative time estimate shows a significant speedup of SurfProp over traditional biophysical calculations, allowing users to access property calculations at unprecedented scales and opening up the possibility for guidance of generative models (Watson et al., 2023) and multi-property optimization (Gruver et al., 2023).

3.4. SurfProp-*expt*

In this section, we turn to experimental data relating to antibody developability, focusing specifically on hydrophobic interaction chromatography data, which measures surface hydrophobicity.

In previous work, it was found that the prediction of HIC from sequence hydrophobicity scales varies greatly based on which scale is used (Waibl et al., 2022), and overall correlations are still low (Park & Izadi, 2024) suggesting that existing scales are not sufficient for modeling the hydrophobic interactions of antibodies.

Motivated by this, we use our pretrained model from the previous section and finetune it to learn HIC, adding an additional mean pooling layer to the per-point predictions generated by the model. We also train an additional SurfProp-*insilico* model to predict hydrophobicity (with the Black & Mould scale) specifically for pretraining, and evaluate both versions of *in silico* pretraining. The benefit of pre-training here is that the model will have seen a large, diverse set of antibodies before finetuning on the limited HIC dataset.

We see in Table 3.4 that the model pretrained on the *in silico* task significantly outperforms a model trained from scratch. To our knowledge, SurfProp is able to achieve state of the art performance ($\rho_{\text{pearson}} = 0.86$) on the Jain et. al. dataset (Hebditch & Warwicker, 2019) and is a proof of concept for the utility of pretrained surface representations for experimental developability prediction.

Method	Viscosity (Apgar)	Viscosity (Dai)	Heparin Retention
MOLDESK	0.76 CDR_APBS_neg	0.63 CDR_HPATCH_WW	0.77 CDR_APBS_pos
SurfProp	0.70 CDR_APBS_neg	0.48 HCDR1_HPATCH_WW	0.79 Fv_APBS_pos

Table 2. Comparison of Pearson correlations between MOLDESK and SurfProp descriptors with experimental biophysical properties. Specific descriptors used for each correlation are listed below the correlation values. Correlations are listed here with viscosity (not log viscosity as in Figure 5). Overall, SurfProp is able to capture the biophysical drivers of the experimental quantities considered at speeds much faster than existing methods.

Table 3. Performance Time Comparison between MOLDESK and SurfProp

Method	PDB	Electrostatics[1] (APBS)		Hydrophobicity (HPATCH)		Total Pipeline	
		Time (s)	Speedup	Time (s)	Speedup	Time (s)	Speedup
MOLDESK	4CNI	13.80	–	27.25	–	44.34	–
	8SMT	17.97	–	28.62	–	49.5	–
SurfProp	4CNI	.72	×20	.03	×908	1.58	×28
	8SMT	.89	×20	.03	×954	1.55	×32

[1] We subtract the time for PDB2PQR from the electrostatics portion of *both* methods and subtract the time for ANARCI sequence alignment from the total results for both methods as well.

4. Discussion

We introduce SurfProp, a surface-based property prediction and calculation framework aimed at improving antibody drug discovery by integrating developability considerations at earlier stages. With SurfProp-*insilico*, we offer a significantly faster alternative to biophysical property calculations. This offers the opportunity to screen much larger libraries of sequences or generated designs, and has further promise to be incorporated into multi-property optimization or guidance frameworks, making SurfProp-*insilico* a natural complement to advances in generative modeling for antibody discovery (Frey et al., 2025; Bennett et al., 2024; Wang et al.).

On the other hand, SurfProp-*expt* shows another area of promise directly predicting properties in the low data regime of experimental developability data. By leveraging representations learned by SurfProp-*insilico*, we achieve strong predictive performance for HIC retention ($\rho_p = 0.89$), despite the limited size of the experimental dataset. This transfer learning approach offers a promising direction for building predictive models of other developability properties that similarly suffer from data scarcity.

5. Limitations

We mention a few limitations of our approach. While SurfProp-*insilico* demonstrates robust performance on Fv structures, it needs to be further evaluated to test performance on other diverse formats (e.g. full Fabs, bispecifics, and antibody fragments). Furthermore, it is likely that the model would not generalize to general proteins. We plan on addressing this in the near future by training the model on a wide variety of protein families. Nonetheless, this highlights the continuing need for biophysical calulators such as MOLDESK or TAP as principled and manifestly general methods.

More practically, the benefits of our approach become clear with access to GPU (we utilized an NVIDIA A100 GPU for timing evaluation). In some cases, it could be advantageous to parallelize calculations with many CPU cores and traditional biophysical calculators. However, we find the benefit of our approach lends itself to different regimes, and in tandem with generative models, either for screening or guidance.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Abanades, B., Wong, W. K., Boyles, F., Georges, G., Bujotzek, A., and Deane, C. M. Immunebuilder: Deep-learning models for predicting the structures of immune proteins. *Communications Biology*, 6(1):575, 2023.

Pretraining	Pearson Correlation	Spearman Correlation
Electrostatics	0.86	0.81
None	0.77	0.75

Table 4. Regression performance on the public HIC dataset from (Jain et al., 2017), metrics computed on the test set.

Apgar, J. R., Tam, A. S., Sorm, R., Moesta, S., King, A. C., Yang, H., Kelleher, K., Murphy, D., D'Antona, A. M., Yan, G., et al. Modeling and mitigation of high-concentration antibody viscosity through structure-based computer-aided protein design. *PLoS One*, 15(5): e0232713, 2020.

Bennett, N. R., Watson, J. L., Ragotte, R. J., Borst, A. J., See, D. L., Weidle, C., Biswas, R., Shrock, E. L., Leung, P. J., Huang, B., et al. Atomically accurate de novo design of single-domain antibodies. *biorxiv*, 2024.

Carter, P. J. and Rajpal, A. Designing antibodies as therapeutics. *Cell*, 185(15):2789–2805, 2022.

Dai, J., Izadi, S., Zarzar, J., Wu, P., Oh, A., and Carter, P. J. Variable domain mutational analysis to probe the molecular mechanisms of high viscosity of an igg1 antibody. In *MAbs*, volume 16, pp. 2304282. Taylor & Francis, 2024.

Das, T. K., Chou, D. K., Jiskoot, W., and Arosio, P. Nucleation in protein aggregation in biotherapeutic development: a look into the heart of the event. *Journal of Pharmaceutical Sciences*, 111(4):951–959, 2022.

Decherchi, S. and Rocchia, W. A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PloS one*, 8(4):e59744, 2013.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. Pdb2pqr: an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic acids research*, 32(suppl_2):W665–W667, 2004.

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

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

Feydy, J., Glaunès, A., Charlier, B., and Bronstein, M. Fast geometric learning with symbolic matrices. *Advances in Neural Information Processing Systems*, 33:14448–14462, 2020.

Frey, N. C., Hötzel, I., Stanton, S. D., Kelly, R., Alberstein, R. G., Makowski, E., Martinkus, K., Berenberg, D., Bevers III, J., Bryson, T., et al. Lab-in-the-loop therapeutic antibody design with deep learning. *bioRxiv*, pp. 2025–02, 2025.

Gao, X., Gao, W., Xiao, W., Wang, Z., Wang, C., and Xiang, L. Learning regularized positional encoding for molecular prediction. *arXiv preprint arXiv:2211.12773*, 2022.

Greenshields-Watson, A., Agarwal, P., Robinson, S. A., Williams, B. H., Gordon, G. L., Capel, H. L., Li, Y., Spoendlin, F. C., Boyles, F., and Deane, C. M. Anarcii: A generalised language model for antigen receptor numbering. *bioRxiv*, pp. 2025–04, 2025.

Gruver, N., Stanton, S., Frey, N., Rudner, T. G., Hotzel, I., Lafrance-Vanasse, J., Rajpal, A., Cho, K., and Wilson, A. G. Protein design with guided discrete diffusion. *Advances in neural information processing systems*, 36: 12489–12517, 2023.

Hebditch, M. and Warwicker, J. Charge and hydrophobicity are key features in sequence-trained machine learning models for predicting the biophysical properties of clinical-stage antibodies. *PeerJ*, 7:e8199, 2019.

Hebditch, M., Roche, A., Curtis, R. A., and Warwicker, J. Models for antibody behavior in hydrophobic interaction chromatography and in self-association. *Journal of pharmaceutical sciences*, 108(4):1434–1441, 2019.

Jain, T., Sun, T., Durand, S., Hall, A., Houston, N. R., Nett, J. H., Sharkey, B., Bobrowicz, B., Caffry, I., Yu, Y., et al. Biophysical properties of the clinical-stage antibody landscape. *Proceedings of the National Academy of Sciences*, 114(5):944–949, 2017.

Jain, T., Boland, T., and Vásquez, M. Identifying developability risks for clinical progression of antibodies using high-throughput in vitro and in silico approaches. In *MAbs*, volume 15, pp. 2200540. Taylor & Francis, 2023.

Jing, B., Eismann, S., Suriana, P., Townshend, R. J., and Dror, R. Learning from protein structure with geometric vector perceptrons. *arXiv preprint arXiv:2009.01411*, 2020.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., et al. Improvements to the apbs biomolecular solvation software suite. *Protein science*, 27(1):112–128, 2018.

Lyon, R. P., Bovee, T. D., Doronina, S. O., Burke, P. J., Hunter, J. H., Neff-LaFord, H. D., Jonas, M., Anderson, M. E., Setter, J. R., and Senter, P. D. Reducing hydrophobicity of homogeneous antibody-drug conjugates improves pharmacokinetics and therapeutic index. *Nature biotechnology*, 33(7):733–735, 2015.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

Maser, M., Park, J. W., Lin, J. Y.-Y., Lee, J. H., Frey, N. C., and Watkins, A. Supsiam: Non-contrastive auxiliary loss for learning from molecular conformers. *arXiv preprint arXiv:2302.07754*, 2023.

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

Park, E. and Izadi, S. Molecular surface descriptors to predict antibody developability: sensitivity to parameters, structure models, and conformational sampling. In *MAbs*, volume 16, pp. 2362788. Taylor & Francis, 2024.

Paszke, A. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

Raybould, M. I. and Deane, C. M. The therapeutic antibody profiler for computational developability assessment. In *Therapeutic Antibodies: Methods and Protocols*, pp. 115–125. Springer, 2021.

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

Raybould, M. I., Marks, C., Lewis, A. P., Shi, J., Bujotzek, A., Taddese, B., and Deane, C. M. Thera-sabdab: the therapeutic structural antibody database. *Nucleic acids research*, 48(D1):D383–D388, 2020.

Rekker, R. F. The hydrophobic fragmental constant; an estimation to a 100 data point set. *Eur. J. Med. Chem.*, 14:478–488, 1979.

Sharma, P., Joshi, R. V., Pritchard, R., Xu, K., and Eicher, M. A. Therapeutic antibodies in medicine. *Molecules*, 28 (18):6438, 2023.

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

Sverrisson, F., Feydy, J., Correia, B. E., and Bronstein, M. M. Fast end-to-end learning on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15272–15281, 2021.

Waibl, F., Fernández-Quintero, M. L., Wedl, F. S., Kettenberger, H., Georges, G., and Liedl, K. R. Comparison of hydrophobicity scales for predicting biophysical properties of antibodies. *Frontiers in Molecular Biosciences*, 9: 960194, 2022.

Wang, A., Sang, Z., Stanton, S. D., Hofmann, J. L., Izadi, S., Park, E., Ludwiczak, J., Kirchmeyer, M., Davidson, D., Maier, A., et al. A guided design framework for the optimization of therapeutic-like antibodies. In *ICLR 2025 Workshop on Generative and Experimental Perspectives for Biomolecular Design*.

Wang, G., Hahn, T., and Hubbuch, J. Water on hydrophobic surfaces: Mechanistic modeling of hydrophobic interaction chromatography. *Journal of Chromatography A*, 1465:71–78, 2016.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976): 1089–1100, 2023.

Zarzar, J., Khan, T., Bhagawati, M., Weiche, B., Sydow-Andersen, J., and Sreedhara, A. High concentration formulation developability approaches and considerations. In *MAbs*, volume 15, pp. 2211185. Taylor & Francis, 2023.

A. Surface generation

In this section we describe the approach to generate pointcloud surfaces efficiently on a GPU, without the need for specialized packages such as KeOps (Feydy et al., 2020), as opposed to (Sverrisson et al., 2021) which was the inspiration for this style of surface generation. We also make the estimation of curvature optional, which saves additional compute time.

A.1. Steps

The surface generation scheme consists of the following steps:

1. **Initial Sampling:** Generate B random points around each atom center using a normal distribution:

$$\text{samples} = \mathcal{N}(0, 1) \cdot \text{spread} + \text{atom\_positions} \tag{1}$$

   where spread controls the initial distribution width.

2. **Level Set Optimization:** Apply gradient descent to move points toward the level set of a soft distance function $\phi(\mathbf{x})$:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \eta \nabla_{\mathbf{x}} \frac{1}{2} (\phi(\mathbf{x}) - R_0)^2 \tag{2}$$

   where R_0 is the target level set value (typically 1.05), and $\phi(\mathbf{x})$ is defined as:

$$\phi(\mathbf{x}) = -\frac{\sum_i r_i e^{-\|\mathbf{x}-\mathbf{a}_i\|}}{\sum_i e^{-\|\mathbf{x}-\mathbf{a}_i\|}} \cdot \log \sum_i e^{-\frac{\|\mathbf{x}-\mathbf{a}_i\|}{r_i}} \tag{3}$$

   with $\mathbf{a}_i$ being atom positions and r_i their corresponding radii.

3. **Surface Point Selection:** Filter points to retain only those that are:
   - Near the level set: $|\phi(\mathbf{x}) - R_0| < \text{variance} \cdot R_0$
   - Not trapped inside the protein: $\phi(\mathbf{x} + \mathbf{n}(\mathbf{x})) - R_0 - 0.5\sqrt{R_0} > 0$

   where $\mathbf{n}(\mathbf{x}) = \frac{\nabla\phi(\mathbf{x})}{\|\nabla\phi(\mathbf{x})\|}$ is the normalized gradient.

4. **Uniform Sampling:** Apply grid-based clustering to achieve an approximately uniform distribution of points:

$$\text{labels} = \text{grid\_cluster}(\mathbf{x}, \text{resolution}) \tag{4}$$

   where points within cells of size resolution are aggregated through averaging.

5. **Normal Computation:** Calculate normal vectors as the normalized gradients of the soft distance function:

$$\mathbf{n}(\mathbf{x}) = \frac{\nabla\phi(\mathbf{x})}{\|\nabla\phi(\mathbf{x})\|} \tag{5}$$

6. **Tangent Basis Construction (Optional):** Construct orthogonal vectors $\mathbf{u}$ and $\mathbf{v}$ that span the tangent plane at each point:

$$\mathbf{u} = \text{normalize}([1 + s \cdot n_x^2 \cdot a, s \cdot b, -s \cdot n_x]^T) \tag{6}$$

$$\mathbf{v} = \text{normalize}([b, s + n_y^2 \cdot a, -n_y]^T) \tag{7}$$

   where $s = \text{sign}(n_z)$, $a = -1/(s + n_z)$, and $b = n_x \cdot n_y \cdot a$.

7. **Multi-scale Curvature Computation (Optional):** Calculate curvature features at multiple scales using:

$$\text{PPt} = \sum_{j \in \mathcal{N}(i)} w_{ij} \mathbf{P}_{ij} \mathbf{P}_{ij}^T \tag{8}$$

$$\text{PQt} = \sum_{j \in \mathcal{N}(i)} w_{ij} \mathbf{P}_{ij} \mathbf{Q}_{ij}^T \tag{9}$$

$$\mathbf{S} = \text{PPt}^{-1} \cdot \text{PQt} \tag{10}$$

where $w_{ij} = \exp(-d_{ij}^2/(2\sigma^2))$ are Gaussian weights, $\mathbf{P}_{ij}$ are projections of displacement vectors onto the tangent plane, $\mathbf{Q}_{ij}$ are projections of normal differences, and $\mathbf{S}$ is the shape operator. The mean and Gaussian curvatures are then:

$$H = \frac{1}{2}(\mathbf{S}_{00} + \mathbf{S}_{11}) \tag{11}$$

$$K = \mathbf{S}_{00}\mathbf{S}_{11} - \mathbf{S}_{01}\mathbf{S}_{10} \tag{12}$$

A.2. Key Parameters

- R_0: Level set value (default: 1.05)
- spread: Initial distribution width (default: 10.5)
- resolution (`res`): Subsampling grid size (default: 1.0)
- variance: Level set tolerance (default: 0.1)
- sup_sampling (`B`): Points per atom (default: 20)
- scales: Curvature scales (default: $[1.0, 3.0, 5.0, 7.0, 9.0]$)

B. Subselection of OAS for *insilico* dataset

We use two diverse subselections of sequences from OAS to add to those from SabDab and TheraSabDab for simulations. The first set focused on diversity from the lens of biophysical properties (specifically electrostatics and hydrophobicity) while the second focused on sequence diversity. For the former, we used MOLDESK (Park & Izadi, 2024) to featurize the set of OAS sequences from (Raybould et al., 2019) (specifically with `CDR_apbs_sum` and `CDR_HPATCH_WW`), and selected in 1,774 sequences in the extremes of the joint distribution of hydrophobicity and electrostatics. For the second, we sampled roughly 7,500 sequences from the paired OAS: we first removed sequences that failed aHo alignment using `ANARCI` (Dunbar & Deane, 2016), filtered sequences with lower than 90 amino acids in either heavy or light chain, and finally filtered those with heavy chain CDR3 lengths greater than 15. Of the remaining roughly 1.85 M sequences, we picked 7,500 sequences with unique V-gene combinations of heavy and light chain.

PDB	Parameters	PC Distance to nearest NS point (Å)				Memory (MB)
		Mean	Q75	Q95	Q99	
4cni	$\texttt{B} = 20, \texttt{res} = 2.5$ Å	0.441	0.534	0.774	0.963	0.393
	$\texttt{B} = 20, \texttt{res} = 1.0$ Å	0.451	0.540	0.644	0.710	1.100
	$\texttt{B} = 60, \texttt{res} = 1.0$ Å	0.448	0.538	0.640	0.707	1.300
	$\texttt{B} = 20, \texttt{res} = 0.8$ Å	0.454	0.541	0.647	0.717	1.400
8smt	$\texttt{B} = 20, \texttt{res} = 2.5$ Å	0.446	0.552	0.756	1.010	0.393
	$\texttt{B} = 20, \texttt{res} = 1.0$ Å	0.449	0.538	0.644	0.707	1.100
	$\texttt{B} = 60, \texttt{res} = 1.0$ Å	0.444	0.535	0.636	0.697	1.200
	$\texttt{B} = 20, \texttt{res} = 0.8$ Å	0.453	0.542	0.645	0.714	1.400

Table 5. Surface matching pointcloud (PC) and Nanoshaper (NS) generated surfaces for representative examples 8smt and 4cni. Lowering the PC surface resolution and number of points randomly sampled per atom (B) keeps the mean distance between PC and NS the same but lowers outliers. We choose the configuration B = 20, res = 1.0 Å to balance memory and reduction of outliers.

Term	Principle
Black-Mould	Rekker coefficients (Rekker, 1979)
Eisenberg	Consensus of five scales
Kyte-Doolittle	Consensus of ΔG (water–vapor) and surface accessibility
Wimley-White	ΔG (water–lipid bilayer)

Table 6. Summary of selected hydrophobicity scales, reproduced from (Waibl et al., 2022)