Probabilistic Modeling of Antibody Structural Dynamics

Eva Smorodina
Department of Immunology, University of Oslo and Oslo University Hospital, Oslo, Norway
eva.smorodina@medisin.uio.no

Montader Ali
Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry,
University of Cambridge, Lensfield road, Cambridge CB2 1EW, UK
ma986@cam.ac.uk

Rahmad Akbar
InSilico Antibody Discovery, AI and Digital Innovation, Novo Nordisk Park 2, 2760, Måløv, Denmark
rdak@novonordisk.com

Victor Greiff
Department of Immunology, University of Oslo and Oslo University Hospital, Oslo, Norway
Imprint Labs, LLC, New York, NY, USA
victor.greiff@medisin.uio.no

Abstract

Antibody function depends on a distribution of structural conformations, with different states exhibiting distinct binding properties. Accurately characterizing these ensembles is critical for therapeutic design, yet current experimental data are sparse and noisy, and molecular simulations are computationally expensive. This frames antibody dynamics as a posterior inference and sampling problem: efficiently exploring high-dimensional conformational space and estimating probable structures with quantified uncertainty. Solving this problem would transform antibody design, advance machine learning for high-dimensional, multimodal inference, and bridge theory and experiment with rigorous structural confidence estimates.

1 Scientific Context

Antibodies are flexible proteins whose biological function (binding disease or autoimmune targets called antigens [1]) depends not on a single static structure, but on a distribution of conformations [5]. This conformational ensemble is naturally expressed as a probability distribution over structures in high-dimensional space. Different conformational states can exhibit different binding characteristics [3, 2, 4], which have profound implications for antibody therapeutic engineering. Indeed, antibody therapeutics are among the most successful modern drugs, with applications in oncology, infectious disease, and autoimmune disorders [5]. However, the relationship between conformational dynamics and binding function remains expensive to simulate and challenging to represent with AI [8, 10]. To design effective therapeutics, it is essential to infer and sample from the full ensemble of antibody structures, rather than relying on single-point predictions [9], raising a fundamental research question:

Given sparse and noisy experimental data (e.g., cryogenic electron microscopy (cryo-EM) maps) and computationally expensive simulations (e.g., molecular

Preprint.

dynamics (MD) trajectories), our goal is to infer a distribution (ensemble) of plausible antibody conformations over millions of structures, with quantified uncertainty and tractable computational cost. In particular, we seek methods that either approximate the full conformational ensemble or, at minimum, recover a functionally meaningful subset of conformations sufficient to guide the engineering of surface-dependent properties (i.e. solubility, binding, etc.).

This makes antibody structural dynamics a problem in sampling and probabilistic inference from and inferring complex posteriors:

- **Sampling**: Exploring the high-dimensional space of antibody conformations (the posterior distribution)
- **Inference**: Estimating which conformations are probable given data (e.g., Bayesian updating)

A model that manage to incorporate both sampling and inference of antibody dynamics would enable the discovery of realistic, meaningful, conformations, in addition to the possibility of covering the entire conformational landscape. In turn, such data would boost the success rate in both practical and theoretical approaches of drug design.

2 Limitations of the Current Approaches

Available methods include physics-based and data-driven techniques,such as:

- **Molecular simulations (classical MD)**: Produces samples from the Boltzmann distribution but is inefficient in exploring rare conformations
- **Enhanced Sampling (e.g., metadynamics, replica exchange)**: Attempts to accelerate sampling but lacks principled uncertainty quantification
- **AI-based Predictors (Bioemu [7], AlphaFlow [6])**: Provide single-point estimates rather than full posterior distributions
  - generate structures that fall outside any meaningful posterior over conformations
  - can contain clear structural violations or energetically implausible geometries
  - in some cases simply yield oversimplified, nonphysical, or biologically unrealistic states

These approaches are limited due to one of the following issues: 1) fail to provide a well-characterized posterior over conformations, 2) produce samples without clear inferential guarantees, or 3) lack explainable and benchmarked biological, chemical, and physical outputs.

Key computational bottlenecks include:

1. **Multimodal Posteriors**: Antibody energy landscapes contain multiple basins (e.g., distinct antibody binding loop conformations). Classical samplers get trapped in local modes
2. **Rare Event Inference**: Functionally relevant states correspond to low-probability but biologically critical regions of the posterior
3. **Likelihood Evaluations**: Evaluating Boltzmann weights via force fields is computationally expensive, making naive Markov chain Monte Carlo (MCMC) impractical
4. **Integration of Heterogeneous Evidence**: Observational data (e.g., cryo-EM maps, nuclear magnetic resonance (NMR) restraints) are partial and noisy, requiring Bayesian inference to combine with simulations

Current limitations reinforce why posterior inference remains crucial: without a principled uncertainty model, we cannot reliably distinguish meaningful conformations from artifacts. In addition, they give rise to another problem of filtering predicted conformations afterwards.

We believe that any successful model must discover realistic, biophysically plausible conformations, not just aim for exhaustive exploration of the energy landscape. This way we highlight both the theoretical ambition (full posterior) and the practical utility (usable, physically grounded ensembles).

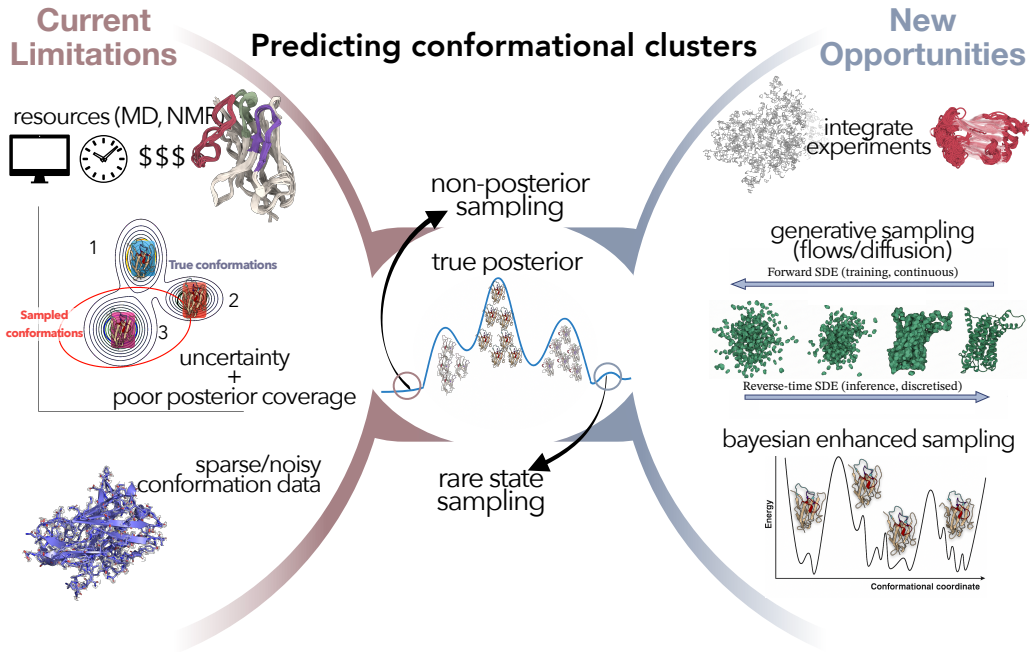

Figure 1: **Limitations and opportunities of accurate conformational sampling**

3 Opportunities for Progress

Several directions hold promise for advancing computational antibody modeling and improving ensemble sampling efficiency (Table 1):

1. **Scaling Laws for Ensemble Sampling**: Understanding how sampling efficiency scales with compute, dataset size, and model capacity will clarify what's required to robustly capture antibody ensembles
2. **Simulation–Experiment Integration**: Scaling inference methods to combine MD with cryo-EM, NMR, and mutational scans across thousands of antibodies would transform therapeutic design from case-by-case modeling to systematic prediction
3. **Generative Samplers**: Flows and diffusion models enable efficient proposals, providing up to 100$\times$ faster conformational exploration
4. **Bayesian Enhanced Sampling**: Adaptive methods such as sequential Monte Carlo (SMC) or reinforcement-learning-guided MD improve coverage of rare but critical states
5. **Approximate Inference**: Variational approaches provide tractable posteriors and allow real-time updates as experimental data are collected
6. **Hierarchical Data Fusion**: Unified probabilistic models generate principled ensembles from noisy, heterogeneous experimental and computational data

Appendix

A Probabilistic Inference Tasks

Table 1 summarizes key tasks in probabilistic inference for antibody structural dynamics, outlining their goals, primary challenges, commonly used methods, and potential impact on therapeutic design and structural biology.

Table 1: Antibody structural dynamics probabilistic inference tasks.

Task	Goal	Challenges	Methods	Impact
Posterior Approximation	Estimate distribution of plausible conformations (posterior)	High dimensionality, multi-modality	Variational Inference (VI), Normalizing Flows	Avoids overconfident predictions; captures uncertainty
Rare-Event Sampling	Efficiently sample rare states (e.g., binding-competent conformations)	Metastable states, high energy barriers	Metadynamics, SMC, Diffusion Models	Critical states may determine biological function (e.g., antigen binding)
Uncertainty Propagation	Quantify confidence in predictions (e.g., per-residue uncertainty)	Noisy/sparse data (cryo-EM, NMR), force-field errors	Bayesian Neural Networks, Posterior Calibration	Enables risk-aware antibody design
Data Integration	Combine MD, cryo-EM, and AI (Bioemu [7], AlphaFlow [6]), and similar) into a unified model	Conflicting data, heterogeneous noise/resolution	Hierarchical Bayesian Models, Likelihood-Free Inference	Maximizes information from expensive experiments

References

[1] Rahmad Akbar, Habib Bashour, Puneet Rawat, Philippe A. Robert, Eva Smorodina, Tudor-Stefan Cotet, Karine Flem-Karlsen, Robert Frank, Brij Bhushan Mehta, Mai Ha Vu, Talip Zengin, Jose Gutierrez-Marcos, Fridtjof Lund-Johansen, Jan Terje Andersen, and Victor Greiff. Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies. *mAbs*, 14(1):2008790, 2022. doi: 10.1080/19420862.2021.2008790. URL https://doi.org/10.1080/19420862.2021.2008790. PMID: 35293269.

[2] Zahra Bahrami Dizicheh, I-Ling Chen, and Patrick Koenig. Vhh cdr-h3 conformation is determined by vh germline usage. 6(1):864, 2023. ISSN 2399-3642. doi: 10.1038/s42003-023-05241-y. URL https://doi.org/10.1038/s42003-023-05241-y.

[3] Monica L. Fernández-Quintero, Johannes R. Loeffler, Johannes Kraml, Ursula Kahler, Anna S. Kamenik, and Klaus R. Liedl. Characterizing the diversity of the cdr-h3 loop conformational ensembles in relationship to antibody binding properties. *Frontiers in Immunology*, Volume 9 - 2018, 2019. ISSN 1664-3224. doi: 10.3389/fimmu.2018.03065. URL https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2018.03065.

[4] Monica L. Fernández-Quintero, Nancy D. Pomarici, Barbara A. Math, Katharina B. Kroell, Franz Waibl, Alexander Bujotzek, Guy Georges, and Klaus R. Liedl. Antibodies exhibit multiple paratope states influencing vh–vl domain orientations. 3(1):589, 2020. ISSN 2399-3642. doi: 10.1038/s42003-020-01319-z. URL https://doi.org/10.1038/s42003-020-01319-z.

[5] Monica L. Fernández-Quintero, Nancy D. Pomarici, Anna-Lena M. Fischer, Valentin J. Hoerschinger, Katharina B. Kroell, Jakob R. Riccabona, Anna S. Kamenik, Johannes R. Loeffler, James A. Ferguson, Hailee R. Perrett, Klaus R. Liedl, Julianna Han, and Andrew B. Ward. Structure and dynamics guiding design of antibody therapeutics and vaccines. *Antibodies*, 12(4), 2023. ISSN 2073-4468. doi: 10.3390/antib12040067. URL https://www.mdpi.com/2073-4468/12/4/67.

[6] Bowen Jing, Bonnie Berger, and Tommi Jaakkola. Alphafold meets flow matching for generating protein ensembles, 2024. URL https://arxiv.org/abs/2402.04845.

[7] Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew Foong, Victor Satorras, Osama Abdin, Bastiaan Veeling, Iryna Zaporozhets, Yaoyi Chen, Soojung Yang, Adam Foster, Arne Schneuing, Jigyasa Nigam, Federico Barbero, Vincent Stimper, Andrew Campbell, Jason Yim, and Frank Noé. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389, 07 2025. doi: 10.1126/science.adv9817.

[8] Prerna Sharma, Rahul V. Joshi, Robert Pritchard, Kevin Xu, and Maya A. Eicher. Therapeutic antibodies in medicine. *Molecules*, 28(18), 2023. ISSN 1420-3049. doi: 10.3390/molecules28186438. URL https://www.mdpi.com/1420-3049/28/18/6438.

[9] Eva Smorodina, Oliver Crook, Johannes R. Loeffler, Monica Lisa Fernandez Quintero, Lucas Matthias Weissenborn, Hannah L Turner, Aleksandar Antanasijevic, Rahmad Akbar, Puneet Rawat, Khang Lê Quý, Brij Bhushan Mehta, Ole Magnus Fløgstad, Dario Segura-Peña, Nikolina Sekulić, Andrew B. Ward, Fridtjof Lund-Johansen, Jan Terje Andersen, and Victor Greiff. Structural modeling of antibody variant epitope specificity with complementary experimental and computational techniques. In *ICLR 2025 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2025. URL https://openreview.net/forum?id=d6ExTOu00A.

[10] William R Strohl. Structure and function of therapeutic antibodies approved by the us fda in 2024. *Antibody Therapeutics*, 8(3):197–237, 06 2025. ISSN 2516-4236. doi: 10.1093/abt/tbaf014. URL https://doi.org/10.1093/abt/tbaf014.