Structure-Guided Antibody Design Through Multimodal Generative Models and Epitope-Specific Conditioning

Pandu Mandasari
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
mandasari493@unh.edu

Majken Bergstrom
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
majkenb@uc.edu

Aron Langevoort
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
alangevoort@ku.edu

Tobiasz Losiak
Department of Computer Science, University of Houston, Houston, TX, USA.
tobiaszl@uh.edu

Abstract

Antibody design is shifting from expert-guided sequence optimization toward structure-aware generative systems that integrate molecular representation, immunological context, and deployability constraints. This paper examines structure-guided antibody design through multimodal generative models and epitope-specific conditioning as a socio-technical infrastructure problem rather than a purely algorithmic one. It argues that recent advances in protein structure prediction, diffusion-based design, language-model-guided sequence generation, and all-atom generative modeling create new scientific possibilities while intensifying governance, robustness, fairness, and sustainability challenges. The discussion traces how multimodal architectures combine sequence, structure, surface, and interaction signals to support antibody candidates that are not only binding-competent but also developable, manufacturable, and clinically translatable. Epitope-specific conditioning is analyzed as a control layer that constrains generative search toward functionally relevant regions, yet it also raises concerns about dataset bias, pathogen diversity, and overfitting to dominant epitope classes. The paper then considers deployment architecture, including model serving, experimental feedback loops, laboratory information systems, and reproducibility infrastructure. It further evaluates fairness in access to generative design capabilities, biosafety and dual-use risks, and the environmental footprint of large-scale training and inference. A final section addresses evaluation and clinical translation, emphasizing prospective validation, human oversight, and regulatory evidence. The conclusion recommends layered governance, open benchmarking, uncertainty-aware interfaces, and lifecycle accounting as conditions for responsible adoption. The central claim is that epitope-conditioned multimodal antibody design can become a durable public-interest platform only if

model performance is coupled with institutional accountability, infrastructure equity, and sustained empirical validation.

Keywords

antibody design; multimodal generative models; epitope-specific conditioning; protein structure prediction; computational immunology; AI governance; research infrastructure.

1. Introduction

Antibody design has become a focal point for artificial intelligence because it combines precise molecular recognition with enormous therapeutic and diagnostic value. The field has moved rapidly from sequence-based heuristics and physics-inspired docking toward learned representations that capture protein structure, evolutionary constraints, and interaction interfaces. Landmark structure prediction systems demonstrated that deep learning can approximate folding and complex formation with unprecedented accuracy, creating a foundation for downstream design tasks [1]. Complementary three-track architectures further expanded the ability to predict protein structures and interactions, including antibody-antigen complexes, thereby enabling more reliable computational starting points for engineering campaigns [2]. These developments did not eliminate experimental uncertainty, but they changed the design cycle by making structure-guided hypothesis generation faster, cheaper, and more scalable.

The next phase is defined by generative models that propose novel antibodies rather than merely evaluate existing ones. Diffusion models for protein structure and function have shown that de novo design can be framed as a controllable generation problem over geometric representations [3]. Sequence design methods such as graph-based and message-passing models have improved the recovery of sequences compatible with a target backbone, which is central to antibody engineering because small sequence changes can alter affinity, stability, and immunogenicity [4]. At the same time, protein language models have been adapted to antibody repertoires, enabling efficient exploration of mutational landscapes and humanization trajectories [5]. Classical computational antibody design frameworks remain important because they encode domain knowledge about complementarity-determining regions, framework regions, and interface packing [13]. The emerging synthesis is therefore not a replacement of prior methods but an integration of learned and mechanistic approaches.

This paper advances a system-level argument. Structure-guided antibody design through multimodal generative models and epitope-specific conditioning should be understood as a layered infrastructure that connects data curation, model training, inference services, laboratory automation, clinical evaluation, and governance. The technical promise is substantial, but the risks are distributed across the lifecycle. Poorly curated training data can encode narrow pathogen coverage, biased germline representation, or hidden assay conditions. Inference systems can produce confident but non-developable molecules. Deployment in commercial or public health settings can widen inequalities between well-resourced and under-resourced institutions. Governance frameworks that focus only on model release may miss upstream data provenance and downstream clinical translation. The paper therefore analyzes architecture, conditioning, robustness, fairness, sustainability, and policy as interconnected concerns.

The central claim is that epitope-conditioned multimodal antibody design can become a durable public-interest platform only when model performance is coupled with institutional accountability, infrastructure equity, and sustained empirical validation. The sections that

follow examine the foundations of structure-guided design, the architecture of multimodal generative systems, the role of epitope-specific conditioning, deployment infrastructure, robustness and fairness, sustainability, and clinical translation. Throughout, the emphasis is on structural trade-offs rather than isolated benchmarks. A system that generates high-affinity binders but cannot be manufactured, validated, or governed at scale is not a complete solution. Conversely, a governance framework without capable generative tools may fail to address urgent immunological needs. The productive path is an integrated sociotechnical design agenda.

2. Structure-Guided Antibody Design as a Socio-Technical System

Structure-guided antibody design is often presented as a computational pipeline, but it is more accurately a socio-technical system. The pipeline begins with target selection, antigen preparation, epitope definition, and dataset curation. It then moves through model training, candidate generation, filtering, experimental characterization, and iterative optimization. Each stage depends on human expertise, institutional priorities, and material infrastructure. The accuracy of structure prediction systems has improved dramatically, yet the choice of antigen conformation, glycosylation state, and oligomeric context remains a scientific judgment that models cannot fully automate [1]. Similarly, interaction prediction methods can guide design, but they operate within the limits of their training distributions and cannot guarantee biologically meaningful epitope recognition in vivo [2].

The antibody molecule itself imposes distinctive constraints. It must bind a target with sufficient affinity and specificity while maintaining stability, solubility, low immunogenicity, and manufacturability. Computational methods that optimize binding alone may produce candidates that aggregate, express poorly, or require costly process development. Methods such as automated interface design and solubility-aware mutation modeling illustrate the importance of multi-objective reasoning [6]. Solubility prediction tools and developability heuristics have become essential because a molecule that performs well in silico can fail at the formulation stage [14]. The design system therefore must represent not only geometric complementarity but also biophysical liabilities, post-translational modifications, and expression-system compatibility.

A second socio-technical dimension is the feedback loop between computation and experimentation. Generative models are often trained on public structures and sequences, but the most valuable data are generated in proprietary laboratories through binding assays, stability measurements, and functional screens. This creates a data asymmetry: institutions with high-throughput experimentation can fine-tune models on privileged feedback, while others rely on stale public datasets. The result is not merely a performance gap but a governance gap. If the best generative systems are concentrated in a few organizations, then agenda-setting power over antibody design shifts toward those organizations. Open benchmarks and data-sharing agreements can partially mitigate this, but they require sustained funding and clear intellectual property arrangements.

A third dimension is the regulatory and ethical context. Antibodies are used in oncology, autoimmunity, infectious disease, and diagnostics, and each domain carries different risk profiles. A broadly neutralizing antiviral antibody may have public health implications, while an engineered therapeutic antibody may raise immunogenicity and safety concerns. The dual-use potential of generative protein design has been recognized in broader AI safety discussions, where capability gains can outpace oversight mechanisms [19]. Structure-guided antibody design is not exempt from these concerns. The same models that accelerate

pandemic preparedness could be misused to enhance harmful properties. A socio-technical framing therefore treats governance not as an external constraint but as part of the design architecture.

3. Multimodal Generative Models and Molecular Representation

Multimodal generative models for antibody design aim to learn joint distributions over sequence, structure, surface properties, and interaction context. Early generative approaches often focused on one modality, such as sequence generation with autoregressive models [7] or structure generation with equivariant diffusion processes [8]. These methods demonstrated that deep generative modeling could produce plausible protein-like outputs, but they also exposed the difficulty of enforcing cross-modal consistency. A generated backbone must be compatible with a generated sequence, and that sequence must fold into the intended structure under physiological conditions. Multimodal architectures attempt to address this by sharing representations across modalities or by alternating between sequence and structure refinement.

Diffusion-based methods have been particularly influential because they can generate geometric objects while conditioning on desired functions or interfaces [3]. In antibody design, diffusion models can be adapted to generate complementarity-determining regions, full variable domains, or antibody-antigen complexes. Some approaches focus on full-atom representations to capture side-chain packing and hydrogen bonding, which are critical for affinity and specificity [10]. Others use antigen-specific conditioning to steer generation toward a particular epitope or surface patch [11]. Iterative refinement graph neural networks further show that sequence and structure can be co-designed rather than treated as separate stages [12]. These developments suggest a trend toward unified models that reason across molecular scales.

A unified multimodal Bayesian flow framework illustrates how sequence, structure, and all-atom representations can be jointly modeled within a single generative process [9]. The significance is not only architectural but epistemic: the model can represent uncertainty across modalities and update beliefs as new conditioning information becomes available. In antibody design, this matters because the target epitope may be partially characterized, the antigen may adopt multiple conformations, and experimental feedback may arrive incrementally. A multimodal generative system that maintains uncertainty can support decision-making under incomplete knowledge rather than forcing premature commitment to a single candidate. This is a departure from one-shot design pipelines and a step toward interactive molecular engineering.

However, multimodality also introduces new failure modes. Models may learn spurious correlations between modalities when training data are biased toward certain antibody families or antigen classes. They may generate structures that satisfy geometric constraints but violate biochemical rules. They may overfit to crystallographic artifacts or to computational labels rather than experimental measurements. Evaluation therefore cannot rely on a single reconstruction metric. It must assess binding affinity, specificity, stability, expression, immunogenicity, and cross-reactivity across diverse conditions. The architecture of a multimodal generative model is inseparable from the evaluation infrastructure that surrounds it. Without strong validation, multimodal fluency can create an illusion of design competence.

4. Epitope-Specific Conditioning and Control

Epitope-specific conditioning is the process of constraining generative antibody design toward a defined region on an antigen. It is a control layer that transforms a broad generative problem into a targeted search over paratope sequences and structures. Conditioning can be implemented through masks, distance constraints, surface descriptors, or learned embeddings of the epitope. The goal is to increase the probability that generated antibodies bind the intended site while avoiding off-target interactions. This is especially important for infectious disease, where neutralizing epitopes may be conserved across variants, and for oncology, where selectivity between tumor and normal tissue is critical. Antigen-specific diffusion models have shown that conditioning can improve the relevance of generated candidates [11].

Yet epitope conditioning is not a neutral technical operation. It depends on how the epitope is defined, which structures are available, and which conformations are considered representative. Many antigens are flexible, glycosylated, or membrane-embedded, and their epitopes may be discontinuous. A conditioning signal derived from a single crystal structure may miss alternative states that are relevant for antibody recognition. Full-atom models can capture more chemical detail, but they also require accurate side-chain conformations and protonation states [10]. If the conditioning representation is too coarse, the model may generate antibodies that bind a similar but non-identical surface. If it is too rigid, the model may fail to explore productive binding modes. The design of conditioning signals is therefore a modeling choice with biological consequences.

There is also the risk of overfitting to dominant epitope classes. Public databases and published studies overrepresent certain viruses, cancer antigens, and antibody formats. A model trained predominantly on these examples may perform well on benchmark epitopes but generalize poorly to neglected pathogens or rare autoimmune targets. This is a fairness issue in computational immunology: diseases that attract less research funding may also attract less data, leading to weaker generative support. Epitope-conditioned design could therefore amplify existing disparities unless training corpora and benchmarks are deliberately diversified. Governance mechanisms should require documentation of epitope coverage, antigen sources, and known blind spots.

From a systems perspective, epitope conditioning creates a trade-off between controllability and creativity. Strong conditioning narrows the search space and improves interpretability, but it may exclude unconventional solutions that could overcome resistance or cross-reactivity. Weak conditioning allows broader exploration but increases the burden on downstream screening. The optimal balance depends on the application. In pandemic response, speed and conserved epitope targeting may favor strong conditioning. In therapeutic antibody discovery, broader exploration may reveal unexpected binding modes. Adaptive systems that adjust conditioning strength based on experimental feedback could offer a middle path, but they require robust uncertainty estimation and human oversight.

5. Architecture, Infrastructure, and Deployment

Deploying multimodal generative antibody design at scale requires more than a trained model. It requires an architecture that connects data ingestion, preprocessing, model serving, candidate ranking, experimental design, and laboratory information management. These components are often developed independently, which creates integration debt. The machine learning systems literature has long warned that hidden technical debt accumulates when data dependencies, configuration, and feedback loops are not treated as first-class engineering concerns [18]. Antibody design platforms are especially vulnerable because they combine

sensitive biological data, proprietary assay results, and complex computational pipelines. A model that works in a notebook may fail in a regulated production environment.

A practical deployment architecture typically separates training, inference, and orchestration. Training requires high-performance computing, curated datasets, and reproducible experiment tracking. Inference requires low-latency services that can generate and score candidates for interactive design sessions. Orchestration connects computational predictions to robotic liquid handlers, plate readers, and sequencing instruments. Each layer has different reliability, security, and audit requirements. For clinical or public health use, the system must also support traceability from generated sequence to experimental evidence to regulatory submission. This is not merely an information technology problem; it shapes what kinds of scientific questions can be asked. If feedback loops are slow or opaque, the generative model cannot learn effectively from experiments.

Cloud and hybrid deployments offer scalability, but they raise concerns about data sovereignty and access. Many biological datasets are subject to institutional review, national regulations, or commercial agreements. Moving them to centralized cloud infrastructure may simplify computation but complicate governance. Federated learning and secure enclaves are potential alternatives, yet they introduce communication overhead and coordination challenges. Open-source model weights can democratize access, but they can also lower barriers to misuse. The deployment architecture therefore must align with the governance model. A system intended for global public health may prioritize open standards and equitable compute access, while a commercial therapeutic platform may prioritize intellectual property protection and regulated validation.

Sustainability is another infrastructure consideration. Large multimodal models require substantial energy for training and inference, and repeated design cycles can multiply the computational footprint. The environmental cost is not limited to electricity; it includes water use for cooling, hardware manufacturing, and electronic waste. Institutions should account for these costs when choosing model size, retraining frequency, and deployment location. Efficiency techniques such as distillation, quantization, and caching can reduce inference burden, but they may affect accuracy and uncertainty calibration. The system-level goal is not maximal model scale but fit-for-purpose capability. A smaller, well-calibrated model that supports routine design may be more sustainable than a massive model used sporadically.

6. Robustness, Fairness, and Governance

Robustness in antibody design means maintaining performance across distribution shifts, noisy inputs, and adversarial conditions. A model trained on one set of antigens may fail on newly emerging variants. A model fine-tuned on one assay platform may not transfer to another. A model conditioned on a particular epitope may generate candidates that bind but do not neutralize. These failures are not merely technical; they can have clinical and public health consequences. Robustness testing should therefore include out-of-distribution antigens, sequence mutations, structural perturbations, and assay variability. Uncertainty quantification is essential because it allows human experts to distinguish confident predictions from speculative ones. Systems that present all outputs with equal confidence are dangerous in high-stakes settings.

Fairness in this domain has multiple meanings. At the data level, fairness concerns the representation of pathogens, populations, and antibody germlines. At the model level, it concerns whether generative performance is equitable across target classes and institutions. At

the deployment level, it concerns who can access the technology and who benefits from it. If generative antibody design is available only to well-funded pharmaceutical companies and elite research centers, then public health priorities may be neglected. If training data overrepresent certain populations, then antibody therapies may be less effective for underrepresented groups. Datasheets for datasets and model cards for model reporting provide useful transparency mechanisms, but they must be enforced through procurement and publication standards [16]. Model cards can document intended use, limitations, and evaluation conditions, helping downstream users make informed decisions [17].

Governance must be layered because risks emerge at different stages. Upstream, funders and data providers should require provenance documentation and ethical review for sensitive sequences. Midstream, model developers should implement access controls, misuse monitoring, and red-teaming for dual-use potential. Downstream, clinical and public health institutions should validate candidates prospectively and maintain pharmacovigilance. Broader AI safety principles, such as avoiding reward hacking and ensuring corrigibility, are relevant when generative systems are coupled to automated laboratories [19]. An antibody design platform that optimizes a narrow binding metric without human oversight could produce candidates that are potent but unsafe. Governance should therefore combine technical safeguards, institutional review, and regulatory oversight.

Policy implications extend beyond individual projects. National and international bodies may need to classify certain generative protein design capabilities as sensitive. Export controls, licensing regimes, and codes of conduct could shape who can train and deploy models. At the same time, overly restrictive policies could impede legitimate research and pandemic preparedness. The challenge is to design governance that is proportionate, adaptive, and internationally coordinated. Public-private partnerships, open benchmarks, and shared safety evaluations can help align incentives. The goal is not to slow innovation but to ensure that innovation is directed toward socially beneficial ends and that risks are identified before harm occurs.

7. Sustainability and Economic Considerations

The economic sustainability of generative antibody design depends on the total cost of ownership across the lifecycle. Training a large multimodal model requires compute, data engineering, and expert labor. Inference adds ongoing costs, especially if design campaigns involve thousands of candidates. Wet-lab validation is often the largest expense, and it scales with the number of candidates that must be synthesized and tested. A generative model that reduces the number of failed candidates can create substantial savings, but only if its predictions are reliable enough to guide resource allocation. The economic value proposition is therefore tied to accuracy, calibration, and integration with experimental workflows.

There is also a risk of overinvestment in model scale at the expense of data quality and process integration. A smaller model trained on high-quality, standardized experimental data may outperform a larger model trained on noisy public data. Institutions should evaluate models using decision-relevant metrics, such as the number of validated binders per unit cost, rather than benchmark scores alone. This requires close collaboration between computational teams and experimental scientists. Economic sustainability also depends on maintaining skilled personnel who understand both machine learning and immunology. Training such hybrid experts is an infrastructure investment that cannot be replaced by automated pipelines.

From a global equity perspective, the economics of generative design are uneven. High-income countries and large corporations can afford the necessary compute and laboratory infrastructure, while low- and middle-income countries may struggle to participate. This can lead to a situation where globally relevant pathogens are understudied because they do not represent profitable markets. Public funding, technology transfer, and regional manufacturing capacity are essential to counterbalance these tendencies. Open-source models and shared data repositories can lower barriers, but they require sustainable funding models. The governance challenge is to create incentives for both innovation and equitable access.

Environmental sustainability intersects with economic sustainability. Energy-intensive training and inference contribute to carbon emissions, and repeated wet-lab experiments generate plastic and chemical waste. Lifecycle accounting should include computational and experimental footprints. Institutions can adopt greener computing practices, such as scheduling training in low-carbon regions and using efficient hardware. They can also design experiments to maximize information gain per sample. These choices are not merely operational; they reflect the values embedded in the research system. A responsible antibody design platform should aim for scientific impact with minimal ecological burden.

8. Evaluation and Clinical Translation

Evaluation is the bridge between generative design and clinical translation. In silico metrics such as predicted binding affinity, structure confidence, and sequence likelihood are necessary but insufficient. Experimental validation must assess binding, neutralization, stability, expression, and immunogenicity. Prospective studies should compare generated candidates against standard discovery methods, not just against other models. Clinical translation further requires pharmacokinetics, toxicology, and manufacturing consistency. The history of AI in medicine shows that promising computational results often face challenges in real-world deployment, including workflow integration, clinician trust, and regulatory evidence [20]. Antibody design will face similar hurdles.

Clinical decision support systems provide a useful analogy. Reinforcement learning and other AI methods have been proposed for treatment recommendation, but their adoption depends on interpretability, safety, and integration with clinical workflows [15]. Antibody design is not a bedside decision tool, but it shares the need for human oversight and clear accountability. Computational predictions should be presented with uncertainty ranges and supporting evidence, not as definitive answers. Experimental scientists and clinicians should be able to interrogate the rationale behind a candidate. This requires models that can provide attribution, counterfactuals, or at least confidence estimates. Without such interfaces, generative design risks becoming a black box that experts cannot effectively supervise.

Regulatory pathways for computationally designed antibodies are still evolving. Agencies may require detailed documentation of design rationale, model versioning, training data, and validation experiments. Reproducibility is essential: independent groups should be able to recreate the computational pipeline and verify the results. This necessitates standardized data formats, open benchmarks, and version-controlled models. It also requires collaboration between regulators, developers, and clinical researchers. The goal is not to burden innovation but to build trust. A transparent evaluation ecosystem can accelerate adoption by reducing uncertainty for investors, regulators, and patients.

Translation also depends on manufacturing and distribution. A designed antibody must be expressible in a suitable cell line, purifiable, stable during storage, and affordable at scale.

Generative models can incorporate developability constraints, but they cannot replace process development. Early collaboration with manufacturing experts can prevent late-stage failures. Similarly, distribution and access require health system capacity, cold chain logistics, and pricing policies. These are not peripheral concerns; they determine whether a scientifically elegant design becomes a useful therapy. Evaluation should therefore be end-to-end, from sequence generation to patient access.

9. Conclusion

Structure-guided antibody design through multimodal generative models and epitope-specific conditioning represents a significant convergence of artificial intelligence, structural biology, and immunology. The technical foundations are strong: structure prediction, diffusion-based design, language-model-guided sequence generation, and all-atom multimodal modeling have expanded the space of designable antibodies. Epitope-specific conditioning offers a powerful control mechanism for directing generation toward functionally relevant surfaces. However, the success of these systems depends on more than model architecture. It depends on data provenance, experimental feedback loops, deployment infrastructure, robustness testing, fairness safeguards, sustainability accounting, and regulatory translation.

The central argument of this paper is that these concerns are not externalities but constitutive elements of the design system. A model that generates high-affinity binders but cannot be manufactured or governed is incomplete. A platform that performs well on benchmark epitopes but neglects neglected pathogens is inequitable. A deployment that concentrates capability in a few institutions may undermine public health priorities. Addressing these challenges requires layered governance, open benchmarking, uncertainty-aware interfaces, equitable compute access, and lifecycle sustainability. It also requires sustained collaboration among computational scientists, immunologists, clinicians, policymakers, and community stakeholders.

Future work should focus on adaptive conditioning that responds to experimental feedback, robust evaluation across diverse antigen classes, and governance frameworks that anticipate dual-use risks without stifling beneficial research. The field should invest in shared data infrastructure, reproducible pipelines, and training for hybrid experts who can bridge machine learning and immunology. If these conditions are met, epitope-conditioned multimodal antibody design can become a durable public-interest platform. If they are neglected, the field may produce impressive demonstrations that fail to translate into equitable, safe, and sustainable therapies. The choice is not between innovation and caution but between fragmented advancement and integrated responsibility.

References

1. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

2. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-

track neural network. Science, 373(6557), 871-876. https://doi.org/10.1126/science.abj8754

3. 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., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., ... Baker, D. (2023). De novo design of protein structure and function with RFdiffusion. Nature, 620(7976), 1089-1100. https://doi.org/10.1038/s41586-023-06415-8

4. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Courbet, A., de Haas, R. J., Bethel, N., Leung, P. J. Y., Huddy, T. F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., ... Baker, D. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. Science, 378(6615), 49-56. https://doi.org/10.1126/science.add2187

5. Hie, B. L., Candido, S., Lin, Z., Kabeli, O., Rao, R., Smetanin, N., Sercu, T., Rives, A., & Kim, P. S. (2024). Efficient evolution of human antibodies from general protein language models. Nature Biotechnology, 42(2), 275-283. https://doi.org/10.1038/s41587-023-01763-2

6. Marks, C., Hummer, A. M., Chin, M., & Deane, C. M. (2021). Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. PLOS Computational Biology, 17(8), e1009338. https://doi.org/10.1371/journal.pcbi.1009338

7. Shin, J. E., Riesselman, A. J., Kollasch, A. W., McMahon, C., Simon, E., Sander, C., Manglik, A., Kruse, A. C., & Marks, D. S. (2021). Protein design and variant prediction using autoregressive generative models. Nature Communications, 12, 2403. https://doi.org/10.1038/s41467-021-22732-w

8. Anand, N., & Achim, T. (2022). Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. arXiv. https://arxiv.org/abs/2205.15019

9. Wu, H., Song, Y., Zhang, Z., Zhang, Z., Zhou, H., Ma, W. Y., & Liu, J. (2026). Rationalized All-Atom Protein Design with Unified Multi-Modal Bayesian Flow. Advances in Neural Information Processing Systems, 38, 3019-3050.

10. Kong, X., Huang, W., & Liu, Y. (2023). End-to-end full-atom antibody design with diffusion models. arXiv. https://arxiv.org/abs/2302.02277

11. Luo, S., Su, Y., Peng, X., Wang, S., Peng, J., & Ma, J. (2022). Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. Advances in Neural Information Processing Systems, 35, 9754-9767.

12. Jin, W., Wohlwend, J., Barzilay, R., & Jaakkola, T. (2022). Iterative refinement graph neural network for antibody sequence-structure co-design. International Conference on Learning Representations.

13. Adolf-Bryfogle, J., Kalyuzhniy, O., Kubitz, M., Weitzner, B. D., Hu, X., Adachi, Y., Schief, W. R., & Dunbrack, R. L. (2018). RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. PLOS Computational Biology, 14(4), e1006112. https://doi.org/10.1371/journal.pcbi.1006112

14. Sormanni, P., Aprile, F. A., & Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. Journal of Molecular Biology, 427(2), 478-490. https://doi.org/10.1016/j.jmb.2014.09.026

15. Barata, C., Rotemberg, V., Codella, N. C. F., Tschandl, P., Rinner, C., Akay, B. N., Puig, S., Ceylan, C., & Kittler, H. (2021). A reinforcement learning model for AI-based decision support in skin cancer. Nature Medicine, 27, 1941-1946. https://doi.org/10.1038/s41591-021-01497-1

16. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

17. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229. https://doi.org/10.1145/3287560.3287596

18. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J. F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems, 28, 2503-2511.

19. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv. https://arxiv.org/abs/1606.06565

20. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7