De Novo Nanobody Design via Joint Optimization of AlphaFold and an Antibody Language Model

Santiago Mille [*1] **John Wang** [*2 3] **Talal Widatalla** [2 4] **Claudia Driscoll** [2 5] **Xiaowei Zhang** [1] **Xiaojing Gao** [5] **Brian Hie** [2 5]

Abstract

Advances in machine learning have enabled *de novo* design of protein binders with high experimental success rates. Antibodies represent a popular class of protein therapeutics due to their high binding specificity, but machine learning methods struggle to design experimentally validated antibodies due to the hypervariable nature of their binding domains. Here, we introduce a protein design pipeline that frames the nanobody design problem as a joint optimization task between a structure prediction model (AlphaFold-Multimer) and an antibody-specific protein language model (IgLM). In addition, we demonstrate that additional loss terms can effectively "program" the structure of nanobodies to adopt more realistic binding poses. Using hallucination-based design, we generate single-domain antibodies (nanobodies) that score highly across multiple *in silico* metrics. We show that the pipeline can generate nanobodies for a wide range of therapeutically relevant target proteins. Preliminary experimental results confirm that designed nanobodies can express in mammalian cells and further experiments to confirm binding are ongoing.

1. Introduction

Antibodies are proteins produced by the immune system to recognize and bind foreign substances, typically via specific domains known as complementarity determining regions (CDRs) (Bennett et al., 2024). Nanobodies are single domain antibodies – naturally found in camelids and sharks – that retain the stability and binding capacity of conventional antibodies. Traditional methods for developing these molecules rely on techniques such as animal immunization, grafting, or display screening of large antibody libraries (Lu et al., 2020). These processes are slow and often limit the search space for new antibodies to a few mutations from existing ones. Recently, deep learning methods such as BindCraft and RFdiffusion have enabled the design of *de novo* proteins that can successfully bind to specified targets with nanomolar affinity and success rates above 10% (Pacesa et al., 2025; Watson et al., 2023). Furthermore, methods have been developed to design experimentally validated *de novo* nanobodies (Shanehsazzadeh et al., 2024b; Nabla Bio & Biswas, 2025; Bennett et al., 2024). However, these methods suffer from low experimental success rates and typically still require screening thousands of designs.

We aim to develop a framework that generates *de novo* nanobodies with state-of-the-art (SOTA) experimental success rates across a wide range of targets. We summarize our key contributions as follows:

- We modify an existing hallucination-based pipeline to enable nanobody design by introducing a position-specific residue bias and several structural loss terms.
- We implement a method to jointly optimize for multiple models (AlphaFold-Multimer and IgLM).
- We construct a set of *in silico* filters and metrics to efficiently parse thousands of nanobody designs.
- We show that this pipeline generates realistic nanobodies that display strong *in silico* metrics and *in vitro* expression.

2. Related Work

2.1. *De Novo* Protein Design

The *de novo* protein design task seeks to create functional proteins with little to no homology to existing ones. Traditional methods for *de novo* protein design such as Rosetta utilize physics-based energy functions to iteratively search through the space of possible proteins, but these methods are computationally expensive and are limited by inaccuracies in explicitly modeling protein physics (Winnifrith

[*]Equal contribution [1]Department of Bioengineering, Stanford University [2]Arc Institute [3]Department of Computer Science, Stanford University [4]Department of Biophysics, Stanford University [5]Department of Chemical Engineering, Stanford University. Correspondence to: John Wang <jwang003@stanford.edu>.

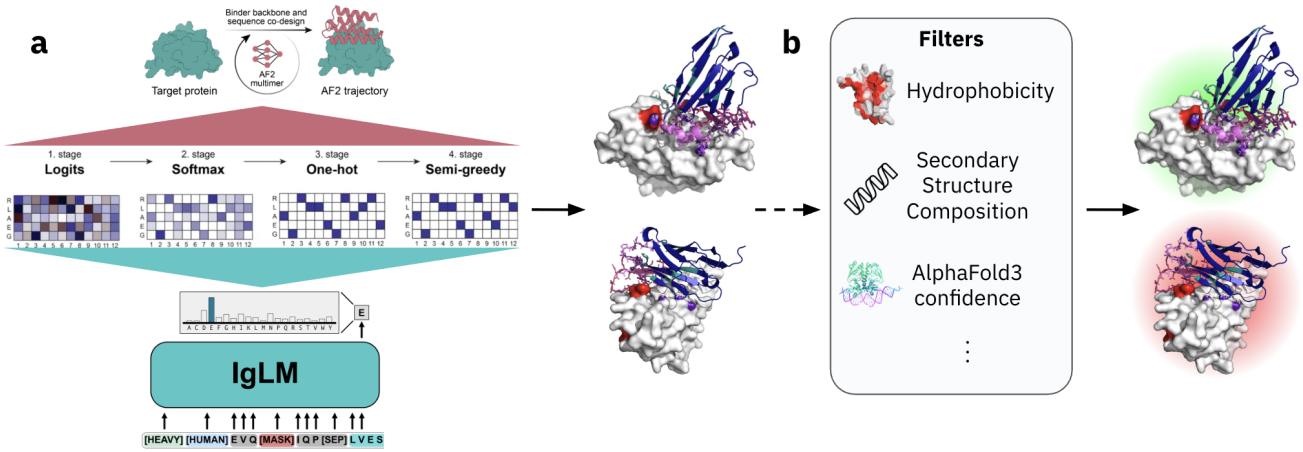

Figure 1. The nanobody design pipeline consists of two main steps. (a) The hallucination process begins by representing the nanobody sequence as a position-specific score matrix (PSSM). The sequence is updated at each iteration via gradients from AlphaFold-Multimer and IgLM while simultaneously being annealed towards a one-hot sequence representation (Evans et al., 2022; Shuai et al., 2023; Pacesa et al., 2025). (b) Following sequence hallucination, the predicted structure is obtined from AlphaFold3 (AF3) and evaluated against several filters to determine passing designs (Abramson et al., 2024).

et al., 2023). Recently, machine learning models trained on massive sequence and structure datasets have demonstrated the ability to learn effective representations of proteins, translating to improved performance on tasks such as protein structure prediction (Abramson et al., 2024). These models have demonstrated the ability to design functional sequences with low homology to known proteins and can operate on sequences, structures, or both (Ruffolo et al., 2024; Frank et al., 2024; Hayes et al., 2024). In particular, structure-based models have shown promise in designing proteins that can bind to a given input target protein (Pacesa et al., 2025). The two main approaches in this paradigm are diffusion-based and hallucination-based models.

Diffusion models are generative models that learn a data distribution by progressively de-noising corrupted samples, effectively learning to interpolate between a noisy distribution (e.g., Gaussian) to a desired distribution (functional proteins). For protein structures, these models are additionally constrained by geometric invariance to translations and rotations in atomic coordinates (Yim et al., 2024). Notable diffusion models such as RFdiffusion and AlphaProteo have been used to generate *de novo* binders for targets such as IL-7Rα, PD-L1, and TrkA with success rates up to 25% (Watson et al., 2023; Zambaldi et al., 2024).

Alternatively, another popular approach to binder design leverages predictive models to "hallucinate" sequences that optimize for certain properties. These hallucination-based approaches typically use gradient descent to iteratively refine sequences based on a loss function for binding. For example, BindCraft uses the structure prediction model AlphaFold-Multimer (AF-M) to hallucinate a binder, starting from a uniform sequence representation and a desired target (Pacesa et al., 2025; Evans et al., 2022). At each iteration, AF-M predicts the structure of the input sequence representation and calculates a loss that optimizes for the probability of binding to the target. This loss updates the input sequence representation via gradient descent. Over several annealing stages, the input converges from a uniform representation to a discrete protein sequence that encodes the desired binder structure. Specifically, given a $L \times 20$ PSSM of raw values representing a sequence s_{logits}, BindCraft uses the following annealing scheme to update the designed sequence s at each stage:

1. At iteration i of n total iterations in the stage, $\lambda = \frac{i+1}{n}$ and $s = (1-\lambda)s_{\text{logits}} + \lambda * \texttt{softmax}(s_{\text{logits}})$.
2. $s = \texttt{softmax}(\frac{s_{\text{logits}}}{\tau}), \quad \tau = 0.01 + 0.99 * (1-\lambda)^2$
3. $s = s_{\text{soft}} + \texttt{stop\_grad}(s_{\text{hard}} - s_{\text{soft}})$
   $s_{\text{soft}} = \texttt{softmax}(s_{\text{logits}}), s_{\text{hard}} = \texttt{argmax}(s_{\text{soft}})$
4. $s \in 0, 1^{L \times 20}, \sum_{j=1}^{20} s_{ij} = 1 \quad \forall i \in \{1, ..., L\}$

Using this pipeline, BindCraft has demonstrated experimental success rates up to 100% and is the current SOTA for designing *de novo* protein binders.

2.2. Deep Learning-based Antibody Modeling

Antibody modeling represents a popular application for deep learning models. Specifically, the loop-like struc-

ture of CDRs has proven challenging for general structure prediction models, while the massive volume of antibody sequences has enabled the training of antibody-specific language models (LMs) (Hitawala & Gray, 2024; Olsen et al., 2022).

Figure 2. Nanobodies consist of three CDRs that comprise the binding domain, while the remainder of the protein, known as the framework region, is highly conserved. Shown PDB: 3EAK

To this end, antibody-specific machine learning models have been developed for almost every common protein design subtask. IgFold is a structure prediction model built on an antibody LM, while AbMPNN is an inverse folding model fine-tuned on antibody structures (Ruffolo et al., 2023; Dreyer et al., 2023). Several antibody LMs have also been developed from pretraining on the observable antibody space (OAS) (Shuai et al., 2023; Olsen et al., 2022; Kenlay et al., 2024). These models can effectively infill CDRs and recover native antibody sequences more effectively than general PLMs. Furthermore, the simple combination of a general protein model - ProteinMPNN - and an antibody LM can recover binding in antibodies with masked CDRs, although each model cannot do so alone (Alamo et al., 2025).

De Novo antibody design generally refers to the construction of novel CDR sequences. Recently, deep learning pipelines for this task have demonstrated the successful generation of experimentally validated nanobodies and scFvs (Nabla Bio & Biswas, 2025). RFantibody uses a finetuned diffusion model to generate novel antibodies, whereas IgDesign utilizes an inverse folding based approach to generate novel CDRs (Bennett et al., 2024; Shanehsazzadeh et al., 2024a). However, a common drawback of all these workflows is their low experimental success rate ($< 2\%$). All existing pipelines for de novo nanobody generation have required library construction and expensive display methods to screen thousands of candidate designs.

2.3. Multi-Objective Optimization

A simple strategy for optimizing over multiple objectives is to perform a weighted sum of loss terms and backpropagate the composite gradient. Although computationally cheap, static weights force can bias solutions away from the Pareto front (Boyd & Vandenberghe, 2023). To alleviate this, adaptive weighting methods such as GradNorm rescale task losses so that their gradient magnitudes maintain a target ratio (Chen et al., 2018). Other methods also operate directly on gradients from different objectives. The Multi-Gradient Descent Algorithm (MGDA) frames multi-objective optimization as finding a common descent direction that maximally decreases all tasks; an exact solution is obtained by solving a small quadratic program at every iteration (Sener & Koltun, 2019). Projection-based variants such as PCGrad detect pairwise gradient conflicts via cosine similarity and subtract the offending component (Yu et al., 2020). More recent approaches such as GradVac alternate between a common descent direction and individual objective gradients (Wang et al., 2020).

3. Motivation

Hallucination-based approaches have achieved high experimental success rates for binder design, but they have not yet been successfully applied for nanobody design. Because models typically used for binder hallucination display poor benchmark results on antibody-antigen complex prediction and existing methods for *de novo* nanobody design suffer from low experimental success rates, we hypothesize that using only general models for protein modeling or antibody-specific models is not sufficient for effective *de novo* nanobody design. Instead, we reason that joint optimization of a general binding objective and an antibody-specific objective will improve experimental success rates. Additionally, we observe that enforcing realistic nanobody binding poses (i.e. CDR binding) is a necessary constraint on the hallucination pipeline. These findings and hypotheses motivate the design of our final pipeline as shown in Fig.1.

4. Methods

4.1. AlphaFold-based Hallucination

We adapt the hallucination pipeline described in BindCraft for nanobody design (Pacesa et al., 2025). Running BindCraft out of the box generates proteins that do not resemble nanobodies. We make the following changes to the pipeline:

4.1.1. Position specific residue bias

We initialize specific positions in the starting sequence PSSM s_{logits} to a desired value x. For nanobody design, we introduce this bias to framework residues to prevent large numbers of mutations in the framework region. No bias is used for CDR positions, allowing free hallucination and design. In practice, we use $x \in (4, 10)$.

4.1.2. Framework contact loss term

We construct a loss term that penalizes close contacts between the framework region of the nanobody and any re-

gion of the target protein. This loss term encourages realistic nanobody binding via CDR regions. We derive probable contacts from the output distogram $D \in \mathbb{R}^{N\times N\times d_{bins}}$ where N is the length of the entire complex and d_{bins} is the number of bins in the distogram. For each residue in the target protein, we calculate a loss by averaging the mean predicted distance of the top k contacts. To calculate the total framework loss $\mathcal{L}_{fw}$, we average the losses of each residue in the target protein.

4.1.3. CDR-specific secondary structure loss terms

These loss terms penalize secondary structure in the CDR regions to promote realistic CDR loops. Namely, we define two losses $\mathcal{L}_\alpha$ and $\mathcal{L}_\beta$ that penalize detected helices and beta strands in the CDR positions. To efficiently detect secondary structure, we measure the distance between every CDR residue i and $i+3$. For helices, we define the helix probability p_α as the sum of the distogram bins between 2 and 6.2 angstroms. Then, we calculate $\mathcal{L}_\alpha = -log(p_\alpha)$. Similarly, we define the distance range for beta sheets as $(9.75, 11.5)$ angstroms and calculate the loss equivalently.

Other changes are detailed in A.x. For all experiments, we pass in the structure of the target protein, a desired epitope or hotspot to bind, and a starting nanobody template derived from 3EAK (Vincke et al., 2009).

4.2. Multiple Gradient Optimization

We frame joint model optimization as a gradient combination task. Given two models AF-M and IgLM, we define two losses to optimize. For AF-M, the loss comprises confidence metrics such as pLDDT, pAE, and iPTM, in addition to the structure specific loss terms described above. Specific weighting of the losses were found through hyperparameter sweeps. For IgLM, the loss is simply the negative log-likelihood of the nanobody sequence. At every iteration, these losses are calculated via a forward pass of the respective models. By freezing the weights of each model, we calculate $\frac{\partial \mathcal{L}}{\partial s_{\text{logits}}} \in \mathbb{R}^{L\times 20}$ for both models. Denote G_{af} as the gradient update from Af-M and G_{iglm} as the gradient update from IgLM. We experiment with different approaches to calculate the final update for s_{logits} based on G_{af} and G_{iglm}.

4.2.1. Normalized Scaling

The simplest method we implement normalizes G_{iglm} to G_{af} before applying a pre-defined scale. Additionally, we scale the final gradient norm by the number of non-zero gradient positions in order to maintain roughly constant per-position update magnitude throughout the design process. The update calculation is shown in Alg. 1.

Algorithm 1 Normalized Scaling Update

Input:
  $G_{af} \in \mathbb{R}^{L\times A}$: Gradient from AF-M.
  $G_{iglm} \in \mathbb{R}^{L\times A}$: Gradient from IgLM.
  $\lambda \in \mathbb{R}$: IgLM scaling factor.
Output: Final gradient $G_{\text{final}} \in \mathbb{R}^{L\times 20}$
$L_{\text{eff}} = \sum_{l=1}^{L} \mathbb{1}\left(\sum_{a=1}^{20} (G_{af})_{la}^2 > \epsilon\right)$
$G'_{iglm} = G_{iglm}$
$s_{\text{norm}} = \frac{\|G_{af}\|}{\|G_{iglm}\|+\epsilon}$
$G'_{iglm} = G_{iglm} \cdot s_{\text{norm}}$
$G_{\text{combined}} = G_{af} + \lambda \cdot G'_{iglm}$
$G_{\text{final}} = G_{\text{combined}} \cdot \frac{\sqrt{L_{\text{eff}}}}{\|G_{\text{combined}}\|_F+\epsilon}$
return G_{final} =0

4.2.2. Loss-Scaled Dynamic Weighting

Normalized scaling offers a simple way to prioritize the dual objectives, but it lacks the ability to change weighting throughout the design process. For example, one may wish to first prioritize the Af-M objective before focusing on IgLM. We implement a simple method to scale the gradient weights by the ratio of each objective's loss. Given two loss values $\mathcal{L}_{af}$ and $\mathcal{L}iglm$, we further scale G'_{iglm} by $\frac{\mathcal{L}iglm}{\mathcal{L}_{af}}$.

Additionally, we experiment with solely applying G_{iglm} to CDR positions in order to encourage updates in the CDR region as opposed to the framework. In this case, we simply apply a zero mask to all framework positions in G_{iglm}.

4.2.3. MGDA

We implement the multi-gradient optimization algorithm from (Sener & Koltun, 2019). Formally, we compute a Pareto-stationary update via a regularized quadratic program. Denote their vectorized forms by $g_1, g_2 \in \mathbb{R}^d$ with $d = 20L$ and set the Gram matrix $G_{ij} = g_i^\top g_j$. Solving

$$\min_{w_1+w_2=1,\ w\geq 0} \tfrac{1}{2}\, w^\top (G + \varepsilon I)\, w$$

yields convex weights $w^\star = (w_1^\star, w_2^\star)$ that minimize the combined squared norm. The MGDA search direction is then $g_{\text{MGDA}} = w_1^\star g_{\text{AF}} + w_2^\star g_{\text{IgLM}}$, which we rescale analogously to Alg. 1.

Algorithm 2 MGDA

Input: $g_{\text{af}},\ g_{\text{iglm}} \in \mathbb{R}^{20d}$, $\varepsilon = 1e-8$
$G \leftarrow \begin{bmatrix} g_{\text{AF}}^\top g_{\text{AF}} & g_{\text{AF}}^\top g_{\text{IgLM}} \\ g_{\text{IgLM}}^\top g_{\text{AF}} & g_{\text{IgLM}}^\top g_{\text{IgLM}} \end{bmatrix}$ {Gram matrix}
$G \leftarrow G + \varepsilon I_2$
Solve $\min_{w\in\mathbb{R}^2} \frac{1}{2} w^\top G w$ s.t. $\mathbf{1}^\top w = 1,\ w \geq 0$
$g_{\text{combined}} \leftarrow w_1 g_{\text{af}} + w_2 g_{\text{iglm}}$
complete as Alg. 1
return g_{final} =0

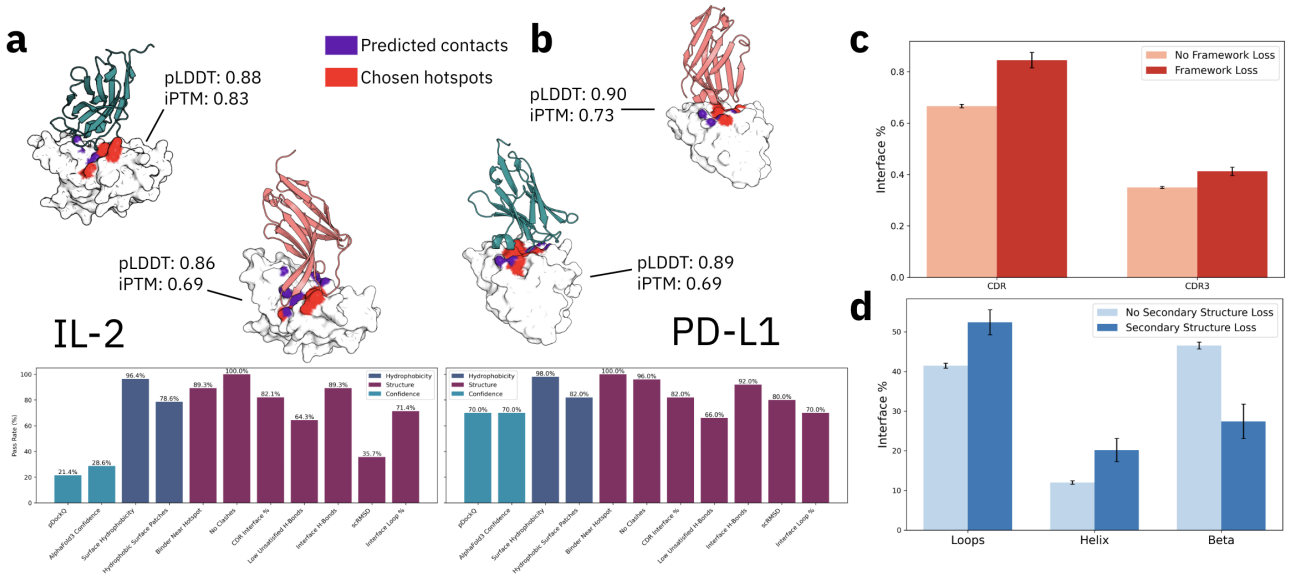

Figure 3. Overview of sampling runs for IL-2 and PD-L1. (a) Randomly selected nanobodies designed against IL-2 shown with the desired epitope in red, actual predicted contacts in purple, and AF3 predicted confidence scores. Filter pass rates for IL-2 samples indicate failures mainly occur due to a low confidence AF3 structure. (b) Analogous metrics for PD-L1 are shown. Designed nanobodies for PD-L1 typically have a higher success rate again *in silico* filters compared to IL-2. (c) Effect of framework contact loss term on the interface CDR percentage for IL-2 nanobody designs. (d) Effect of secondary structure loss terms on nanobody secondary structure composition at the interface.

4.3. Filter Construction

To select nanobodies designed by our computational pipeline for experimental characterization, we develop a set of *in silico* metrics and filters to determine promising sequences (Fig. 1). Specifically, given n generated nanobody designs, we construct a set of filters and a ranking methodology to select the top k designs for experimental validation. Since the pipeline inherently hallucinates proteins with high AF-M confidence, it is susceptible to adversarial attacks, or optimization of specific features that yield high model confidence but low or detrimental biological utility. We therefore re-evaluate every design with an orthogonal structure predictor — AlphaFold3 (AF3) — and base all downstream criteria on AF3 predicted structures. In the first stage, we enforce strict filters that designs must pass to be considered for selection. These filters are summarized in three main classes:

- **Confidence–based metrics**. Designs must satisfy pLDDT ≥ 0.80, iPTM ≥ 0.60, and pDockQ ≥ 0.23 (Bryant et al., 2022).
- **Structural realism**. *Self-consistency:* the AF-M and AF3 models must predict similar poses (RMSD < 5 Å). *Epitope engagement:* at least 70% of the interface must be CDRs, and $\geq 50\%$ of CDR residues must adopt loop conformations.
- **Expressibility and stability**. Two hydrophobicity-based tests are applied: (i) the fraction of hydrophobic residues must be below 0.4, and (ii) structural patches of hydrophobic residues are not permitted.

To avoid over-sampling redundant solutions, we cluster the remaining designs with FOLDSEEK on backbone geometry (van Kempen et al., 2024). If a cluster contains m designs, we allocate the number of designs for *in vitro* characterization propotional to $\lceil \sqrt{m} \rceil$, ensuring representation of both large and small clusters. Within each cluster, candidates are ranked by a weighted composite score that combines ProteinMPNN and IgLM log-likelihoods with pDockQ and auxiliary AF3 metrics. The top-ranked sequences across all clusters constitute the final set of k designs selected for *in vitro* characterization.

5. Results

5.1. *In Silico* Results from Large Scale Sampling

We performed large-scale sampling runs to generate candidate nanobodies against IL-2 and PD-L1, generating roughly 3,000 designs for each target. Success rates for PD-L1 were generally much higher than IL-2, particularly for metrics based on AF3 structure prediction. Nonetheless, the pipeline was able to generate nanobody designs that passed all filters,

including AF3 confidence and predicted docking (Fig.3a, b). Designs were consistently predicted to bind at the specified epitope, and the general structure and binding pose of designed nanobodies mirror those seen in the literature.

We also evaluate the effectiveness of our structural loss terms in constraining designed proteins to resemble realistic nanobodies. We show that the framework contact loss effectively increases both the percentage of the total interface comprised of CDR residues and the percentage of the interface comprised of CDR3 residues, which mainly facilitate binding in natural nanobodies (Fig.3c). Furthermore, we show that loss terms that bias against secondary structure (alpha helices, beta strands) effectively increase the percentage of loops at the binding interface (Fig.3d). We observe a strong negative correlation between beta strand percentage and both loop and helix percentage. However, there is relatively little correlation between loop and helix percentage. Due to this, we prioritize biasing against beta strands in the CDRs, which leads to improved loop percentage at the cost of a slight increase in helix percentage.

Furthermore, we experimentally characterized 38 IL-2 nanobodies and 46 PD-L1 nanobodies with a HiBit assay to measure expression and a split luciferase assay to measure binding. A small fraction – 10 for IL-2 – of the designs expressed *in vitro*, and no designs showed any binding activity. To re-evaluate our prioritization of *in silico* metrics, we plot the correlation between nanobody expression and over 40 logged metrics.

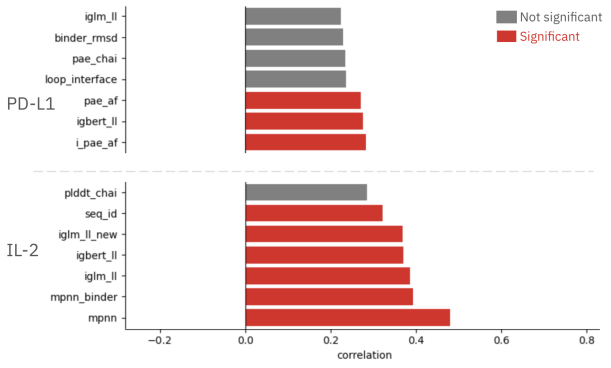

Figure 4. Correlation between various logged metrics during design and *in vitro* expression levels. Only the most highly correlated metrics are shown.

While no singular metric is strongly correlated with expression, we consistently find that external model likelihoods from ProteinMPNN ($r = 0.21, 0.48$) and antibody LMs such as IgLM ($r = 0.26, 0.39$) and IgBERT ($r = 0.33, 0.37$) are among the most correlated metrics (Fig.4).

5.2. Multi-Objective Optimization

In addition to sampling, we conduct experiments on how well the hallucination process optimizes for the AF-M and IgLM objectives by tracking the loss for both models for each iteration across many individual runs (Fig.5).

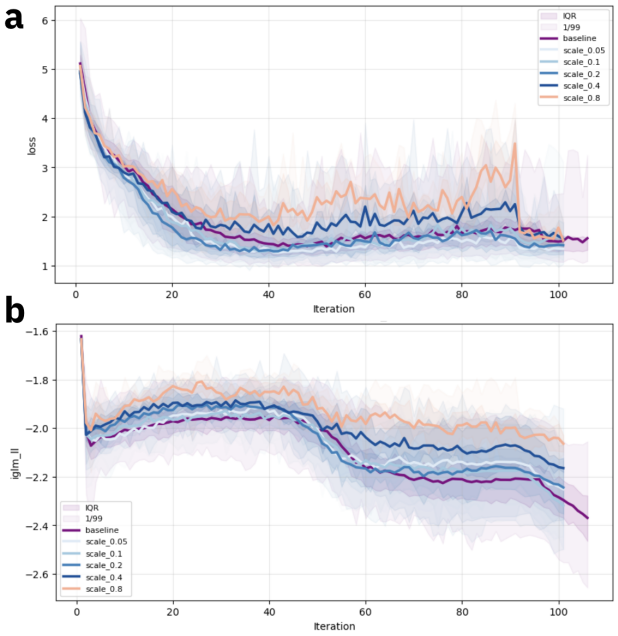

Figure 5. Loss for (a) AF-M and (b) IgLM tracked at each iteration across 50 sampling runs for each plotted configuration. Each line represents the average value for runs using the specified λ IgLM weight under the normalized scaling method described in section 4.2.1

.

As expected, we observe that a higher scaling weight λ increases the IgLM log-likelihood. This improvement in the IgLM likelihood is reflected in the AF-M objective, as increasing λ consistently increases the AF-M loss. Although not shown, we observe a similar trend for the loss-scaled dynamic weighting described in section 4.2.2. Additionally, high λ values logically decrease the hallucination success rate, which is governed by AF-M confidence scores. Experiments for MGDA are ongoing.

6. Discussion

We present an end-to-end pipeline for *de novo* nanobody generation by adapting proven hallucination methods for nanobody design. We introduce specific structural losses that guide model hallucination towards realistic nanobody conformations, while coupling the structural hallucination update with updates from an antibody LM that shows one of the strongest correlations to *in vitro* expression readouts. Nanobodies designed by the pipeline to target IL-2 and

PD-L1 displayed strong *in silico* metrics and realistic binding poses. However, experimentally characterized designs failed to bind to their respective targets and many failed to express in mammalian cells. We hypothesize that allowing mutations in the highly conserved framework may be the cause of low expression among many designs. In upcoming experiments, we aim to test whether completely restricting framework mutations improves expression or yields binding.

Systematic sweeps revealed that naively favoring the sequence prior ($\lambda \gtrsim 0.4$) compromises structural confidence and hallucination success, whereas under-weighting IgLM yields sequences that score poorly on external language-model metrics correlated with expression. In further experiments, we aim to comprehensively test more complex methods for gradient mixing.

Taken together, our results underscore both the promise and the challenges of multi-objective protein design. While designed nanobodies show promise *in silico*, they have failed to yield experimental validation thus far. Future work will focus on bridging the gap between computational and experimental results by further examining which factors influence experimental readouts and identifying the *in silico* metrics that most strongly predict experimental success.

Software

Code for the project will be available after paper submission at https://github.com/jwang307/IgHallucination.

Acknowledgements

We thank members of the Hie and Gao labs for helpful discussions during the project.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016): 493–500, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://www.nature.com/articles/s41586-024-07487-w. Publisher: Nature Publishing Group.

Alamo, D. d., Frick, R., Truan, D., and Karpiak, J. Adapting ProteinMPNN for antibody design without retraining, May 2025. URL https://www.biorxiv.org/content/10.1101/2025.05.09.653228v1. Pages: 2025.05.09.653228 Section: New Results.

Bennett, N. R., Watson, J. L., Ragotte, R. J., Borst, A. J., See, D. L., et al. Atomically accurate de novo design of single-domain antibodies, March 2024. URL https://www.biorxiv.org/content/10.1101/2024.03.14.585103v1. Pages: 2024.03.14.585103 Section: New Results.

Boyd, S. P. and Vandenberghe, L. *Convex optimization*. Cambridge University Press, Cambridge New York Melbourne New Delhi Singapore, version 29 edition, 2023. ISBN 978-0-521-83378-3.

Bryant, P., Pozzati, G., and Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications*, 13(1):1265, March 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28865-w. URL https://doi.org/10.1038/s41467-022-28865-w.

Chen, Z., Badrinarayanan, V., Lee, C.-Y., and Rabinovich, A. GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks, June 2018. URL http://arxiv.org/abs/1711.02257. arXiv:1711.02257 [cs].

Dreyer, F. A., Cutting, D., Schneider, C., Kenlay, H., and Deane, C. M. Inverse folding for antibody sequence design using deep learning, October 2023. URL http://arxiv.org/abs/2310.19513. arXiv:2310.19513 [q-bio].

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., et al. Protein complex prediction with AlphaFold-Multimer, March 2022. URL https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2. Pages: 2021.10.04.463034 Section: New Results.

Frank, C., Khoshouei, A., Fu, L., Schiwietz, D., Putz, D., et al. Scalable protein design using optimization in a relaxed sequence space. *Science*, 386(6720):439–445, 2024. doi: 10.1126/science.adq1741. URL https://www.science.org/doi/abs/10.1126/science.adq1741.

Hayes, T., Rao, R., Akin, H., Sofroniew, N. J., Oktay, D., Lin, Z., et al. Simulating 500 million years of evolution with a language model, July 2024. URL https://www.biorxiv.org/content/10.1101/2024.07.01.600583v1. Pages: 2024.07.01.600583 Section: New Results.

Hitawala, F. N. and Gray, J. J. What has AlphaFold3 learned about antibody and nanobody docking, and what remains unsolved?, September 2024. URL https://www.biorxiv.org/content/10.1101/2024.09.21.614257v1. Pages: 2024.09.21.614257 Section: New Results.

Kenlay, H., Dreyer, F. A., Kovaltsuk, A., Miketa, D., Pires, D., and Deane, C. M. Large scale paired antibody language models, March 2024. URL http://arxiv.org/abs/2403.17889. arXiv:2403.17889 [q-bio].

Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., Tsai, H.-Z., et al. Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, 27 (1):1, January 2020. ISSN 1423-0127. doi: 10.1186/s12929-019-0592-z. URL https://doi.org/10.1186/s12929-019-0592-z.

Nabla Bio and Biswas, S. De novo design of epitope-specific antibodies against soluble and multipass membrane proteins with high specificity, developability, and function, January 2025. URL http://biorxiv.org/lookup/doi/10.1101/2025.01.21.633066.

Olsen, T. H., Moal, I. H., and Deane, C. M. AbLang: An antibody language model for completing antibody sequences, January 2022. URL https://www.biorxiv.org/content/10.1101/2022.01.20.477061v1. Pages: 2022.01.20.477061 Section: New Results.

Pacesa, M., Nickel, L., Schellhaas, C., Schmidt, J., Pyatova, E., et al. BindCraft: one-shot design of functional protein binders, April 2025. URL https://www.biorxiv.org/content/10.1101/2024.09.30.615802v3. Pages: 2024.09.30.615802 Section: New Results.

Ruffolo, J. A., Chu, L.-S., Mahajan, S. P., and Gray, J. J. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications*, 14(1):2389, April 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-38063-x. URL https://www.nature.com/articles/s41467-023-38063-x. Publisher: Nature Publishing Group.

Ruffolo, J. A., Nayfach, S., Gallagher, J., Bhatnagar, A., et al. Design of highly functional genome editors by modeling the universe of CRISPR-Cas sequences, April 2024. URL https://www.biorxiv.org/content/10.1101/2024.04.22.590591v1. Pages: 2024.04.22.590591 Section: New Results.

Sener, O. and Koltun, V. Multi-Task Learning as Multi-Objective Optimization, January 2019. URL http://arxiv.org/abs/1810.04650. arXiv:1810.04650 [cs].

Shanehsazzadeh, A., Alverio, J., Kasun, G., Levine, S., Calman, I., Khan, J. A., et al. IgDesign: In vitro validated antibody design against multiple therapeutic antigens using inverse folding, December 2024a. URL https://www.biorxiv.org/content/10.1101/2023.12.08.570889v2. Pages: 2023.12.08.570889 Section: New Results.

Shanehsazzadeh, A., Alverio, J., Kasun, G., Levine, S., Calman, I., et al. IgDesign: In vitro validated antibody design against multiple therapeutic antigens using inverse folding, December 2024b. URL https://www.biorxiv.org/content/10.1101/2023.12.08.570889v2. Pages: 2023.12.08.570889 Section: New Results.

Shuai, R. W., Ruffolo, J. A., and Gray, J. J. IgLM: Infilling language modeling for antibody sequence design. *Cell Systems*, 14(11):979–989.e4, November 2023. ISSN 24054712. doi: 10.1016/j.cels.2023.10.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2405471223002715.

van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., and Steinegger, M. 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. URL https://doi.org/10.1038/s41587-023-01773-0.

Vincke, C., Loris, R., Saerens, D., Martinez-Rodriguez, S., Muyldermans, S., and Conrath, K. General Strategy to Humanize a Camelid Single-domain Antibody and Identification of a Universal Humanized Nanobody Scaffold. *Journal of Biological Chemistry*, 284(5):3273–3284, January 2009. ISSN 00219258. doi: 10.1074/jbc.M806889200. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925819818914.

Wang, Z., Tsvetkov, Y., Firat, O., and Cao, Y. Gradient Vaccine: Investigating and Improving Multi-task Optimization in Massively Multilingual Models, October 2020. URL http://arxiv.org/abs/2010.05874. arXiv:2010.05874 [cs].

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, August 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL https://www.nature.com/articles/s41586-023-06415-8. Publisher: Nature Publishing Group.

Winnifrith, A., Outeiral, C., and Hie, B. Generative artificial intelligence for de novo protein design, October 2023. URL http://arxiv.org/abs/2310.09685. arXiv:2310.09685 [cs].

Yim, J., Stärk, H., Corso, G., Jing, B., Barzilay, R., and Jaakkola, T. S. Diffusion models in protein structure and docking. *WIREs Computational Molecular Science*, 14(2):e1711, 2024. ISSN 1759-0884. doi: 10.1002/wcms.1711. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1711. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1711.

Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., and Finn, C. Gradient Surgery for Multi-Task Learning, December 2020. URL http://arxiv.org/abs/2001.06782. arXiv:2001.06782 [cs].

Zambaldi, V., La, D., Chu, A. E., Patani, H., Danson, A. E., et al. De novo design of high-affinity protein binders with AlphaProteo, September 2024. URL http://arxiv.org/abs/2409.08022. arXiv:2409.08022 [q-bio].