LLMs Enable Large Scale Design of Nanobodies[1]

Gerard Boxó
Center For Genomic Regulation

Raul Cavalcante Dinardi
University of São Paulo, IME-USP

With
Apart Research

Abstract

In the coming decades, AI will be one major contributor to the offense-defense balance. If we want to steer the future towards a high-robustness state, it is important to develop in advance a portfolio of defensive technologies that directly benefit from economic growth and technological progress.
What would a general defensive technology against biological threats look like? As a motivating example, we investigate whether present-day LLMs are capable of autonomous nanobody design.
Using computational tools [2] as a harness for protein design, we enable a generalist LLM to design and test nanobodies. Examination of performance trends across LLMs and reasoning efforts reveals performance improvements in the designed antibodies, demonstrating a scaling trend relevant for safety.
We argue that theoretical and empirical facts about LLMs make them exceptionally well suited for this type of hyperparameter search task [8][9]. Moreover, this type of evaluation enables both a new paradigm for capabilities assessment and potential for scaling defensive technologies.

Keywords: AIxBio, LLM Search, Evaluations

INTERACTIVE WEBSITE

[1] Research conducted at the def_acc hackathon 2025

1. Introduction

We want to build technology that protects us from the biggest threats we will face as a society in the coming decades. To achieve this with high robustness it is important to maintain a balanced portfolio of defensive technologies that both leverage scaling factors like compute, power, and industrial capacity, while maintaining uncorrelated components that function independently of those general improvements.

Resilience against threats of biological origin is especially relevant due to direct impact on public health. This has been made apparent with the recent concerns about AI-enabled bio-threats [6][7] . But, what if we could use the power of AI for the benefit of defense?

To investigate this question, we focus on potential application of AI for Bio Defense, the automated design of pharmacological compounds. Concretely we focus on a type of protein: the nanobodies. Nanobodies (also known as Single Domain Antibodies) are a class of proteins with the ability to bind to specific antigens, much like antibodies, but substantially smaller and simpler.

Nanobodies offer potential advantages that are highly relevant for situations in which a fast reaction is necessary (like a natural or man-made pandemic). For example, nanobodies have very short development times and don't require specialized equipment to be developed, which is especially relevant when dealing with rapidly evolving strains. However, they also have many drawbacks that would make large-scale deployment very challenging.

There is growing interest in using nanobodies for pharmaceutical applications [5][10][11].

In recent years, the tools available to protein engineers have been dramatically expanded by ML-based biological models [12][13]. This is also true for nanobody design [2]. However despite the ease of use of these tools, many aspects of protein design are still poorly understood and rely on diffuse and tacit knowledge that only researchers have.

This prompts our central question: "**Can we leverage the biological knowledge present in state-of-the-art LLMs to automate the design of nanobodies?**"

To answer this question we evaluate several LLMs on the autonomous design of nanobodies for Nipah virus using RFantibody as a harness. Several in-silico metrics [4] will be used to assess the quality of the designed nanobodies.

2. Methods

To operationalize the nanobody design task, we provide the agent with a specification sheet containing information about the Nipah virus and the G protein (PDB ID 2VSM), which is responsible for binding to Ephrin B2 and initiating the

infectious process. We request values for hyperparameters related to design, such as the length of the Complementarity Determining Regions and hotspots.

Experimental Conditions:

- **Baseline (No Aids):** The agent had access to the internet and terminal but no specific protein design tooling.
- **Documentation Aids:** The model was additionally provided with documentation about the RFantibody harness.
- **Few-shot Aids:** The model was provided with n examples of previous designs and their corresponding metrics.

We primarily report model performance under each condition using the ipSAE metric. While ipSAE does not fully capture nanobody quality, it has the benefit of being comparable with a human baseline of expert protein designs [1]. Results will be released on 01/26.

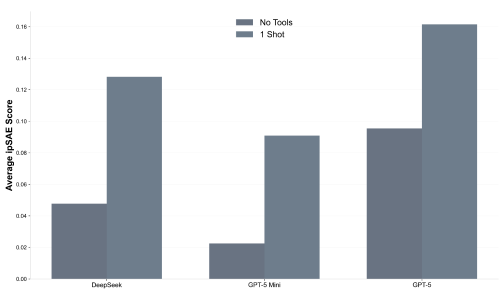

Figure 1 – Comparison of ipSAE across models with and without the Documentation Tool

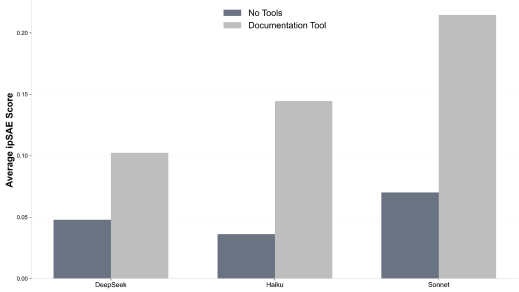

Figure 2 – Comparison of ipSAE across models with and without the 1-Shot Tool

3. Results

More capable models are better at designing nanobodies. Regardless of the affordances provided, a clear trend emerges between design quality and general model capabilities. The worst performing model is GPT-5-Mini, followed by Deepseek R1, Claude Haiku 4.5, GPT-5, and Claude Sonnet 4 Sonnet. If this relationship between general capabilities and safety-relevant capabilities holds across several orders of magnitude, it would provide strong evidence that AI plays an important role in defensive capabilities.

Models struggle with sequential information aggregation. As shown in Figures 2 and 3, providing one example of a previous design improves performance; however,

when more data points are provided, models struggle to aggregate this information, resulting in steep performance degradation.

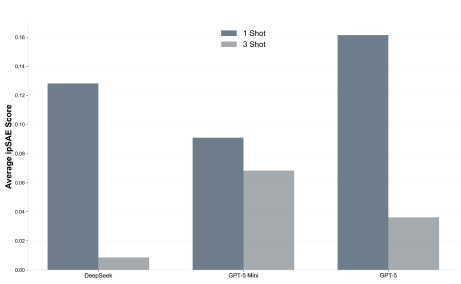

Figure 3 – Comparison of ipSAE across models with different amounts of Serial Information

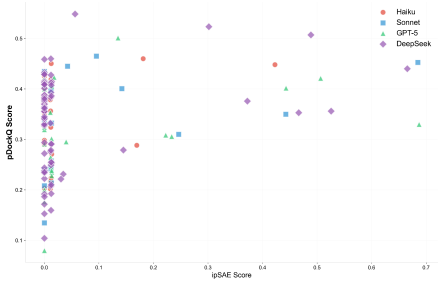

Figure 4 – Scatter plot of ipSAE and pDockQ across models

The best-performing models seem to possess more biological knowledge. We deliberately limited the scope of experiments to empirical metrics; however, inspection of the transcripts in the top right corner of Figure 4 reveals that the best model correctly identified the geometric importance of the H3 loop for proper docking..

4. Conclusion and Limitations

Current generation models can design nanobody libraries in an automated and efficient manner when provided with appropriate tools. Results demonstrate a clear relationship between underlying capabilities and nanobody design performance. Given the versatility of nanobodies and their potential importance for bio-threat defense, improved design capabilities through scaling are encouraging.

However, these experiments are preliminary and we should not jump to conclusions. The experiments were limited to a single use case (Nipah virus), and we focus on coarse-level metrics that often do not correspond with real-world experimental results.

Despite these limitations, if safety-relevant capabilities scale with general capability development, this would provide strong evidence supporting the creation of AI native defensive technologies.

5. References

1. **Adaptyv Bio.** (2025, November 7). *The protein design competition is back!* Adaptyv Bio Substack. Link
2. **Bennett, N. R.,** Watson, J. L., Ragotte, R. J., Borst, A. J., See, D. L., Weidle, C., ... & Baker, D. (2025). Atomically accurate de novo design of antibodies with RFdiffusion. *Nature*, *638*, 320–327. doi:10.1038/s41586-025-09721-5
3. **Bryant, P.,** Pozzati, G., & Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications*, *13*, Article 1664. doi:10.1038/s41467-022-28865-w
4. **Dunbrack Lab.** (2025). *IPSAE: Scoring function for interprotein interactions in AlphaFold2 and AlphaFold3*[Computer software]. GitHub. Link
5. **Lim, H. E.,** Kim, S., & Yoo, H. (2025). Regulation of cellular states via targeted phosphorylation of p53 using a nanobody-coupled kinase system. *Cell Death Discovery*, *11*, Article 82. doi:10.1038/s41420-025-02821-1
6. **Manheim, D.,** Williams, A., Aveggio, C., & Berke, A. (2025). *Understanding the theoretical limits of AI-enabled pathogen design* (Research Report No. RRA4087-1). RAND Corporation. Link
7. **Norwood, S. A.** (2025, April 15). *Understanding AI-enabled biological threats: Hype, hazard, and governance*. Cambridge Engineering Biology. Link
8. **Novikov, A.,** Vu, N., Eisenberger, M., Dupont, E., Huang, P. S.,Wagner, A. Z., ... & Kohli, P. (2025). *AlphaEvolve: A coding agent for scientific and algorithmic discovery*. arXiv. doi:10.48550/arXiv.2506.13131
9. **Romera-Paredes, B.,** Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., ... & Fawzi, A. (2024). Mathematical discoveries from program search with large language models. *Nature*, *625*, 468–475. doi:10.1038/s41586-023-06924-6
10. **Wang, Z.,** Wu, H., He, W., Wei, S., Wei, X., Wei, C., Wang, Y., & Huang, A. (2025). Protective effect of nanobodies targeting Sip protein against *Streptococcus agalactiae* infection in Tilapia. *Animals*, *15*(21), Article 3207. doi:10.3390/ani15213207
11. **Wu, X.,** Zhu, L., Liang, S., Chang, Z., Wang, Y., Zou, Q., ... & Li, M. (2025). A rationally designed cocktail of nanobodies elicited by heterologous vaccination confers protection against SFTSV. *Science Translational Medicine*, *17*(825). doi:10.1126/scitranslmed.ady9025
12. **Dauparas, J.,** Anishchenko, I., Bennett, N. R., Bai, H., Ragotte, R. J., Milles, L. F., ... & Baker, D. (2022). Robust deep learning–based protein sequence design using ProteinMPNN. *Science*, *378*(6615), 49–56.doi:10.1126/science.add2187
13. **Jumper, J.,** Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, with AlphaFold. *Nature*, *596*(7873), 583–589.doi:10.1038/s41586-021-03819-2