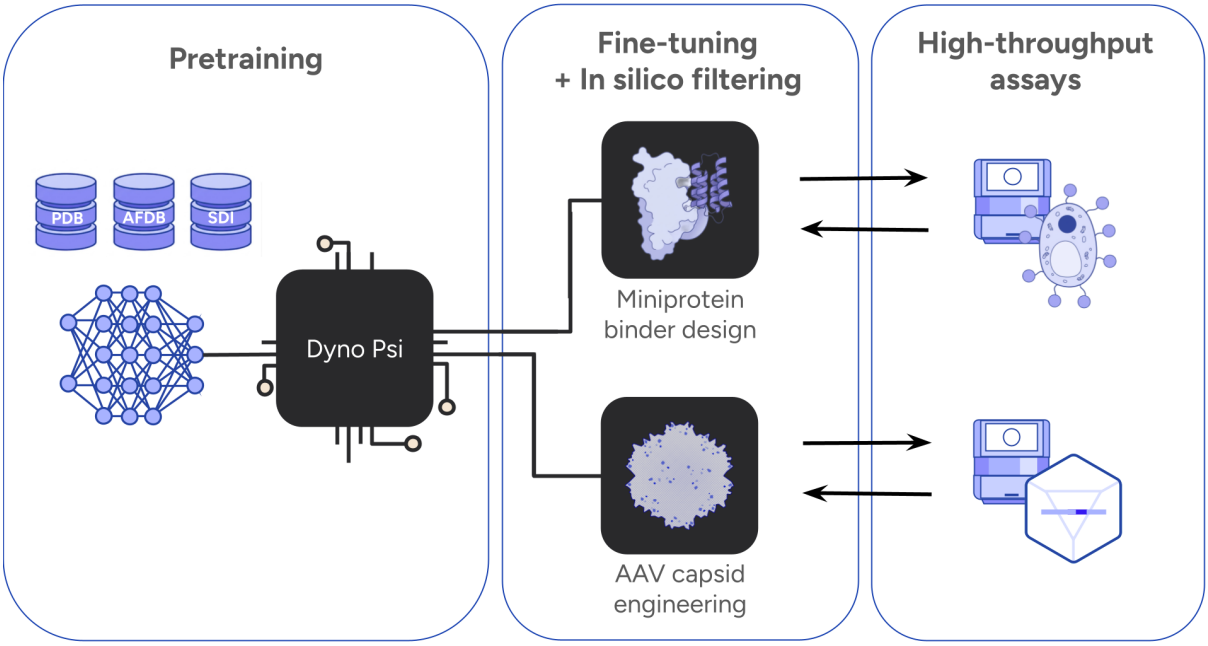
Dyno Psi-1: Flexible de novo design with test-time scaling

Dyno AI Lab
with support from Nvidia Healthcare
March 18, 2026



We introduce Dyno Psi, a family of generative models that harnesses recent advances in protein design to accelerate development for therapeutically relevant targets and modalities. Dyno Psi is a versatile component of our overall approach to enable scalable and experimentally-grounded sequence design. Our approach tightly couples generative design with high-fidelity, high-throughput assays to tailor general-purpose models and *in silico* screening pipelines to for the problems that matter most to drug developers. We start with Dyno Psi-0, a foundation model for protein design that is trained on a diverse set of generative tasks on Dyno SDI (Synthetic Domain Interfaces), a dataset of over 4 million real-world and synthetic multimer examples. We then show that Dyno Psi-0 can be fine-tuned for benchmarking tasks such as miniprotein binder design and match the performance of other open-sourced state-of-the-art models. We release model weights and inference code for this miniprotein binder design model, Dyno Psi-1, along with the training data for Dyno Psi-0. In conjunction, we develop a quantitative yeast display binding assay that recapitulates gold-standard surface plasmon resonance (SPR) binding affinity measurements but at orders-of-magnitude higher scale. With this assay, we benchmark Dyno Psi-1 against BoltzGen head-to-head in the same experiment and see that Dyno Psi-1 achieves superior experimental performance, which scales robustly with more computational screening.

1 Introduction

Advances in structure-based generative protein design have resulted in the rapid development and improvement of general-purpose protein design models. Models like Chroma (1) and RFdiffusion (2) demonstrated how diffusion-based approaches could be successfully applied to protein design. RFdiffusion (2) further showed that these models could generate binders *de novo* to therapeutically-relevant targets and achieve unprecedented experimental hit rates. Experimental success rates for binder design have only improved since then with the release of subsequent models from many academic and industry groups (3–7). This foundation is paving the way for task-specific models that optimize performance for specific targets and/or modalities, which is necessary for applying generative protein design to real-world drug development problems. Though groups have seen improved performance on antibody design by focusing on antibody-specific tasks (8–11), this approach remains largely unexplored for most other modalities. The success of these models also varies widely by target, and there has been limited focus on how design pipelines can be optimized for specific targets.

We needed to bridge this gap on both the model side and the data generation side in order to apply generative protein design to therapeutically-relevant challenges at Dyno,including Adeno-associated virus (AAV) capsid engineering. To support our aim to develop fine-tuned generative models for specific applications, we first developed (1) a flexible base model that learns generalizable features and a diverse repertoire of potential structures, and (2) a platform for quantitative high-throughput measurements that enables accelerated design campaigns grounded in real data. This flexible base model can then be fine-tuned for specific design tasks and calibrated in a tight feedback loop with the most relevant experimental outcomes.

Related studies have typically relied on low-throughput binding measurement methods, such as biolayer interferometry (BLI) or surface plasmon resonance (SPR), to measure binding affinities of designs towards the desired target. Building on previous work using yeast display to quantify miniprotein binder affinity in high-throughput (12), we developed a quantitative yeast display assay that measures binding affinities of tens of thousands of designed miniproteins in a pooled format. We applied this assay to characterize far more candidates than is practical with low-throughput methods such as BLI or SPR, generating large-scale experimental data to validate and benchmark the capacity for generative design models to scale their performance with more samples.

At this scale, we were able to screen Dyno Psi designs without passing those designs through stringent *in silico* filters and found that Dyno Psi designs have a higher rate of experimental validation compared to a benchmark model and that experimental success rate scales with more sampling. These results are in line with the intuition that while *in silico* filtering is useful for improving experimental validation rates, *in silico* metrics are still a crude proxy for experimental success. They also come with the cost of collapsing the design space around only designs with acceptable structure prediction confidence metrics. High-throughput screening allows for a different explore-exploit tradeoff which results in a more diverse set of candidates that progress to subsequent drug development stages.

2 Dyno Psi-1

2.1. Overview.

Our design approach combines several recent innovations in the field. The first is our extension of the scalable flow-matching approach used by the Proteína family of models (7, 13–15) to enable a wide range of protein binder design tasks. This required a way to specify arbitrary combinations of target chains, new chains to be designed, and/or scaffold chains to be partially re-designed to bind to the target. In the case of partially re-designing a scaffold, which is particularly relevant for engineering antibodies and viral capsids, the model needs to learn to

both inpaint a binder interface and dock the binder to the target at the same time.

Drawing inspiration from a solution to the pose estimation problem in computer vision (16), we allow the model to denoise the target structure concurrently with the binder, while providing a distogram of the target as conditioning input. This allows the model to reconstruct the target structure while dynamically adjusting its position to accommodate the binding interface. A similar strategy was also used by BoltzGen (3). See Appendix A1 for more details on model architecture.

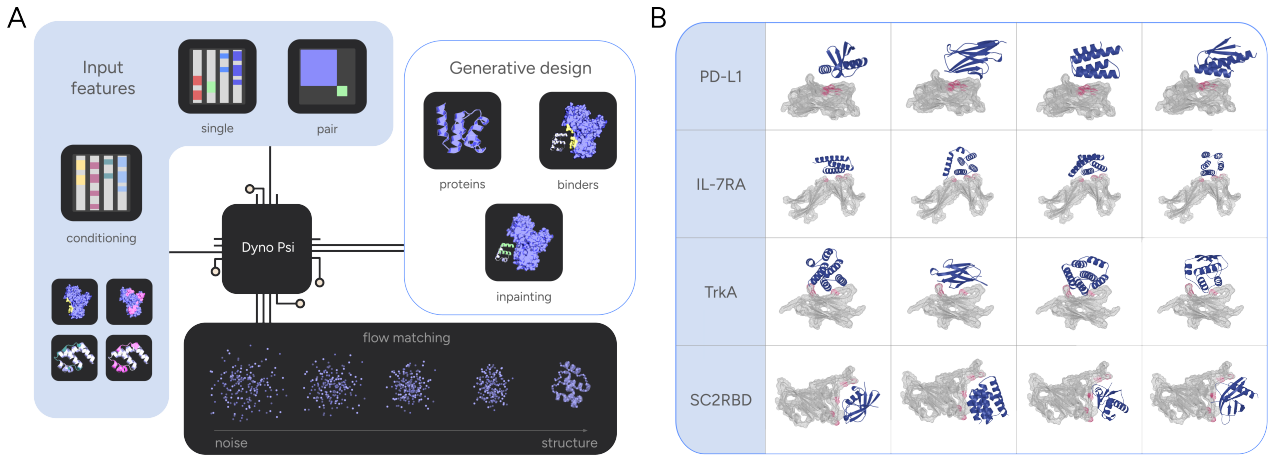


Figure 1. Dyno Psi-0 applies flow-matching to diverse protein design tasks. A) Dyno Psi-0 is a flow-matching model trained on unconditional protein design, as well as binder design and binder inpainting. B) Generated examples from Dyno Psi-1, which was fine-tuned on the miniprotein design task.

Model predictions are converted into new binders via discrete simulation of a stochastic differential equation (SDE) that initializes the binder backbone coordinates from Gaussian noise. Consistent with other diffusion and flow matching approaches, we find that SDE simulation yields improved sample quality relative to a deterministic ordinary differential equation. We adapt the Proteína (13) sampling approach, which is designed for monomer generation, to the binder design setting by providing an initial estimate of the relative positioning of the binder relative to the target. Because our model was trained on both unconditional protein generation and target-conditioned binder design, we are also able to optionally apply classifier-free guidance (CFG) (17) to further optimize for binding to specific targets. Appendix A2 contains details on specific parameter choices for sampling, though these parameters can be tuned depending on the desired balance between designability and diversity.

2.2. Training Data and Strategy.

Developing a model for *de novo* protein binder generation necessitates a large and structurally diverse dataset of protein-protein interactions (PPIs). While the Protein Data Bank (PDB) (18) contains approximately 200,000 entries, the subset of complexes containing multiple interacting protein chains suitable for training is considerably smaller (estimated between 50,000 and 100,000, depending on filtering criteria). This number is further reduced to an estimated 20,000 when clustered based on structural similarity. In contrast, the AlphaFold Protein Structure Database (AFDB) (19) offers an order of magnitude more synthetic structures; however, it was limited to protein monomers when we began this project, and a comparable resource for synthetic multimeric structures did not exist.

Inspired by the approach taken by the Proteína-Complexa team (7), we curated a dataset of inter-domain interactions between TED-annotated (20) domains present within synthetic monomer structures from the AFDB. This process yielded approximately four million "dimer" examples that represent real PPIs, which served as a

pretraining dataset. These examples were clustered into roughly one million clusters using Foldseek (21) and subsequently sampled alongside PDB complex clusters during Dyno Psi-0 pretraining. We refer to this dataset as the Dyno Synthetic Domain Interfaces (SDI) dataset.

Phase	Task(s)	Dataset	Training summary
Dyno Psi-0 pretraining phase 1	Monomer generation	PDB protein monomers AFDB50 cluster reps	256 max tokens; 200M parameters
Dyno Psi-0 pretraining phase 2	Monomer generation, Binder design, Binder inpainting, Target docking	PDB protein monomers AFDB50 cluster reps Dyno SDI dataset	512 max tokens; 200M parameters
Dyno Psi-1 binder design fine-tuning	Binder design	PDB multimers; Dyno SDI dataset (refined)	512 max tokens; 200M parameters

Table 1. Dyno Psi training phases and datasets.

To maximize the generalizability of our model, we chose a multi-phase training procedure (Table 1) that begins with a pretraining phase on a monomer generation task, trained on PDB and AFDB monomers. This was followed by another pretraining phase on a combination of monomer generation, binder inpainting, binder design, and target docking docking using the full combination of PDB and synthetic datasets. The resulting pretrained model, Dyno Psi-0, serves as the foundation for fine-tuning models for specific applications. We opted for a near uniform time sampler in the pretraining phase, followed by a time sampler that weights later times (i.e. closer to the data distribution) higher during fine-tuning (Appendix A1). This approach leverages the diversity of the full dataset to establish global structure in earlier time steps, while allowing the task-specific dataset to dominate the resolution of finer structural details.

The first task-specific model we trained was Dyno Psi-1, which was trained on the miniprotein binder design task using a combination of PDB examples and a more stringently-curated set of synthetic dimers. For this curated set, we used AlphaFold2 (22) to refold all ~1M cluster representatives from Dyno SDI and applied strict interface confidence filters to yield 32,253 high-confidence and diverse synthetic "dimers", which we refined with PyRosetta FastRelax (23) to yield the final synthetic dataset. See Appendix A3 for more details. All model training was conducted using 96 NVIDIA H200 GPUs.

3 Results

3.1. *In silico* benchmarking.

To evaluate our approach, we compared Dyno Psi-1 against two prominent, publicly available generative binder design models: RFdiffusion3 (4) and BoltzGen (3). We selected 5 targets and generated an equal number of samples from each model, evenly divided into length bins from 60 to 120 residues (see Appendix A4.1 for details):

- Common benchmarking targets: IL-7RA, PD-L1, TrkA, SARS-Cov2 RBD
- Dyno Target A: target of interest for an internal program

For each set of samples we used the AlphaFold2 monomer "initial guess" protocol as previously described (2, 24) to calculate pass rates for filters: binder RMSD < 1 Å, binder pLDDT > 80, interface pAE < 10. For targets with native binders in the miniprotein size range, we additionally filtered out designs that were similar to those native binders. We found that Dyno Psi-1 achieves similar filter pass rates to RFdiffusion3 and BoltzGen, with Dyno Psi-1 achieving the best pass rates for 2/5 targets (Figure 2) while generating diverse designs (Figure A4). Interestingly, we find that each model has different filter pass rates for different lengths (Figure A3). BoltzGen tends to have pass rates that are high for shorter binders and falls off for longer lengths, RFdiffusion3 has fairly consistent pass rates regardless of length, and Dyno Psi-1 has a variable pattern that is target-dependent. This suggests that the different models could be used to complement each other in a design campaign, depending on the target and therapeutic strategy.

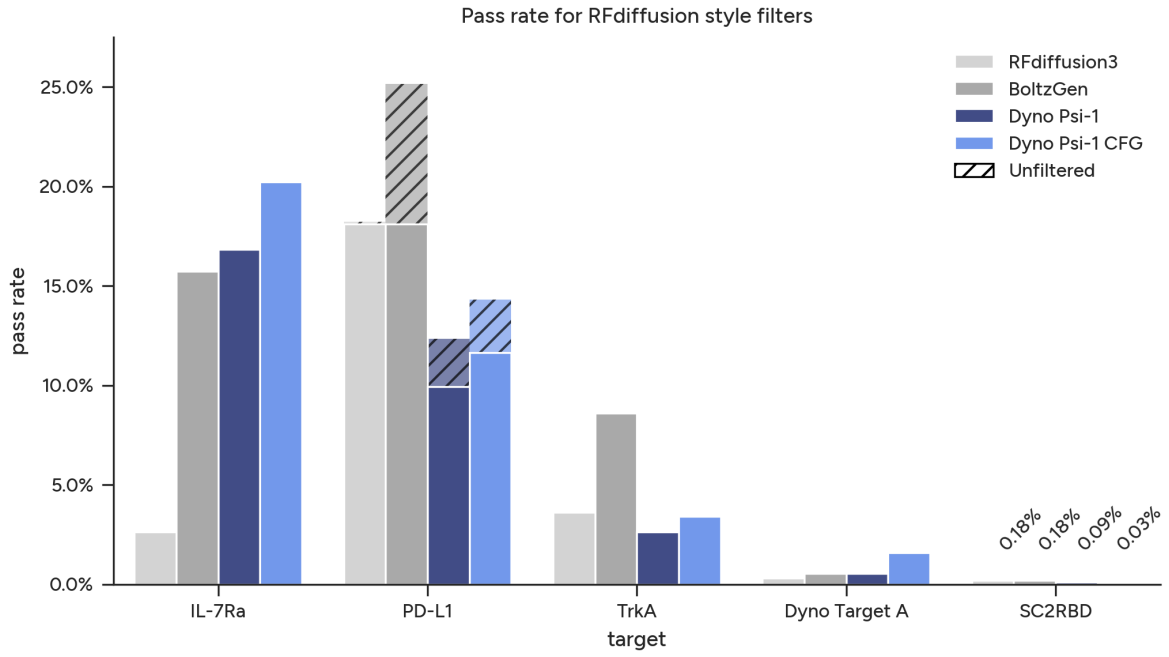


Figure 2. Dyno Psi-1 matches state-of-the-art performance on *in silico* metrics. Four inverse folding sequences were generated for each method according to recommended settings and pass rates reflect the number of backbones for which at least one sequence passes AlphaFold2 "initial guess" filters. PD-L1, IL-7RA, and TrkA backbones were additionally filtered for those with a TM-score < 0.6 to PD-1, IL-7, and nerve growth factor (NGF), respectively (unshaded).

3.2. Yeast display allows for target-specific model validation and demonstrate test-time scaling dynamics for Dyno Psi-1.

In silico filters are indispensable for improving experimental hit rates for small-scale experiments, but the development of quantitative high-throughput assays enables a different paradigm. When the experimental budget for a design campaign is is in the thousands rather than tens, it is better to relax filters and generate more diverse samples so that the final set of candidates that advances to downstream screening steps is as diversified as possible. To test this theory, we generated a library of miniprotein designs targeting IL-7RA, a common modeling benchmark and clinically relevant receptor, from an early version of Dyno Psi-1 (referred to as Dyno Psi-0.5) and BoltzGen (3). We used the full BoltzGen design pipeline, including their filtering pipeline to generate 125 designs from an initial sampling budget of ~12k. For Dyno Psi-0.5 designs, we applied a lenient filter (binder RMSD < 3.5Å, binder pLDDT < 0.7, ipSAE > 0.1). To provide a fair comparison between Dyno Psi-0.5 and BoltzGen, we randomly subsampled the Dyno Psi-0.5 designs such that they came from the same initial sampling budget as BoltzGen. In these sample-matched comparisons, Dyno Psi-0.5 reliably achieved higher experimental success rates and generated the stronger binder (Figure 3A).

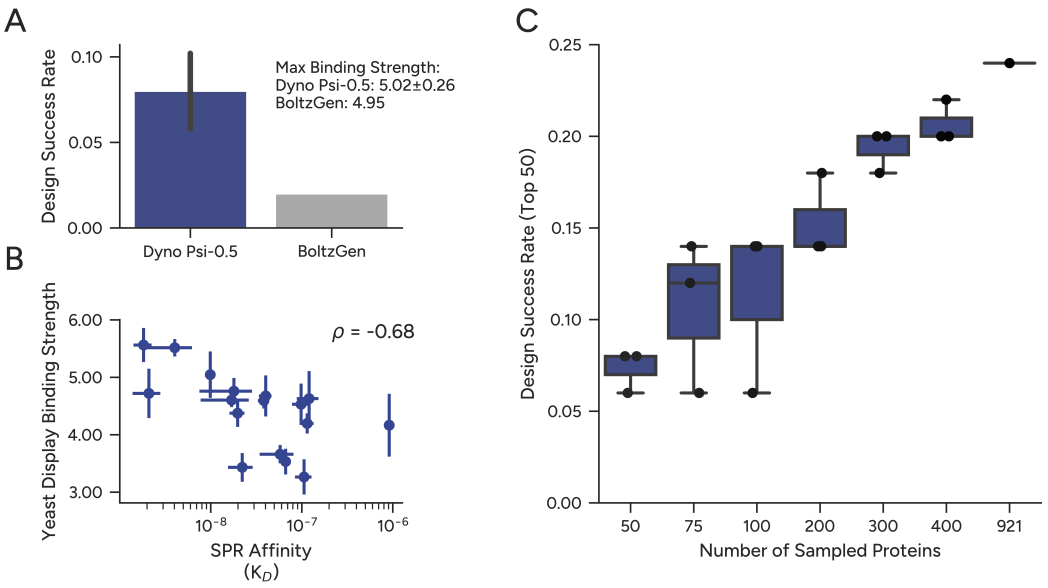


Figure 3. Dyno Psi-0.5 achieved superior IL-7RA design success rate and affinity compared to other methods. A) Binder design success by model as measured with yeast display: Design Success Rate as measured by the fraction of IL-7RA binders identified by yeast display ($\log_2$ fold enrichment in IL-7RA condition over input library > 2.5, BH FDR-adjusted p-value < 0.05) out of the top 100 designs ranked by ipSAE is displayed for each model. Inset shows the highest yeast display binding strength ($\log_2$ fold enrichment in IL-7RA condition over input library) in these sets for each model. B) Yeast display binding correlates well with gold-standard SPR: Yeast display binding strength ($\log_2$ fold enrichment in IL-7RA condition over input library) for a subset of Dyno Psi-1-designed miniproteins is shown on the y-axis, with affinity measurements (KD) of the same miniproteins by SPR on the x-axis . Inset shows the Spearman correlation ($\rho = -0.68$). C) Efficient test-time scaling leads to increased binder design success rates: When selecting the top 50 proteins using ipSAE from a number of randomly selected sequences in the library, the rate of binder design success increases with increasing numbers of proteins selected to rank.

In parallel, we selected a subset of designs for SPR validation. Binding measurements from yeast display correlated with affinities (KD) measured by SPR (Figure 3B), and there were no false positives, indicating that our yeast display assay provides a quantitative and reliable readout of binding affinity. We next evaluated whether the scale of our assay enables simulation of test-time scaling effects. To test this, we simulated increasing design time by randomly sampling progressively larger sets of Dyno Psi-0.5 designed proteins from the yeast display dataset and ranking them by ipSAE scores. Selecting the top 50 sequences from each sampled set resulted in progressively higher design success rates as the number of evaluated designs increased, without apparent saturation within the tested range (Figure 3C). This indicates that more investment in compute with Dyno Psi-0.5 will translate to meaningfully improved experimental success.

4 Discussion

We have introduced a coupled approach to protein discovery in which advances in generation are developed together with advances in experimental measurement and validation. Our generative binder design models are paired with a high-fidelity, high-throughput experimental platform that offers validation at a scale that low-throughput techniques cannot match.

On the generative side, we demonstrated that our protein design foundation model, Dyno Psi-0, serves as a powerful base for us to develop fine-tuned models geared towards specific discovery problems, such as miniprotein binder design. The resulting miniprotein binder design model, Dyno Psi-1, produces high-quality, diverse candidate binders across both standard benchmarking targets and a challenging, therapeutically-relevant internal

target. In carefully controlled head-to-head benchmarking with matched design budgets and filtering criteria, Dyno Psi-1 is competitive with state-of-the-art external methods. We have simultaneously released the model weights and inference code for Dyno Psi-1 as well as the Dyno Psi training data containing ~4 million synthetic dimer interactions, enabling easy adoption of our generative component for miniprotein binders.

This generative capability is coupled with an experimental framework that enables high-fidelity, quantitative assessment of binding at scale. Our pilot study showed a higher rate of experimental validation for Dyno Psi candidates than those produced using existing binder generation models. Our high-throughput yeast display assay is both reliable compared to gold-standard low-throughput measurements, and can be leveraged to understand the effects of inference-time scaling. Looking ahead, we will scale this framework to characterize Dyno Psi-1 designs towards a diverse set of therapeutically-relevant targets. We will also leverage this framework to systematically evaluate the predictive power of *in silico* metrics against experimental outcomes.

Taken together, we have presented both a strong generative system and the experimental framework needed to evaluate it at a meaningful scale. Scaling these experimental campaigns to larger, more diverse sets of proteins will enable both faster validation cycles and tighter feedback loops for improving and fine-tuning generative models. We also aim to apply both Dyno Psi and our experimental platform to a broader set of design challenges beyond miniproteins, advancing both the computational tools for complex design tasks and the high-throughput screening approaches for therapeutically-relevant targets and modalities.

5 Model and Data Availability

Inference code for the binder design version of Dyno Psi is available https://github.com/dynotx/dynopsi and as a pypi package https://pypi.org/project/dynopsi. Model weights can be downloaded on huggingface at https://huggingface.co/dynotx/dynopsi and the synthetic multimer data can be downloaded at https://huggingface.co/datasets/dynotx/synthetic_dimers.

6 Contributions

Dyno AI Lab
Lead technical contributors: Akosua Busia, Josh Wilde, Jesse Weller, Priya Veeraraghavan, Patrick Stock
Experimental contributors: Viktoria Betin, Alex Kern, Meghan Lambie, Jordan Majka
Engineering support: David Levy-Booth, Danny Dannaher, Bobby Ryterski, Ina Chen
Project Leadership: Kathy Lin, Sam Sinai, Eric Kelsic

Nvidia Healthcare
Technical and engineering support: Roy Tal Dew, Sohil Nadajaran, Kushal Shah, Zoey Zhang, Anthony Costa

7 Acknowledgements

We thank David Brookes, Jeffrey Chan, Leo Chen, Oberon Dixon-Luinenburg, Abhishaike Mahajan, Stephen Malina, Jakub Otwinowski, Neha Prasad, Theo Sternlieb, Gian-Marco Visani for their technical contributions to this project.

8 References

1. John B. Ingraham, Max Baranov, Zak Costello, Karl W. Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M. Lord, Christopher Ng-Thow-Hing, Erik R. Van Vlack, Shan Tie, Vincent Xue, Sarah C. Cowles, Alan Leung, João V. Rodrigues, Claudio L. Morales-Perez, Alex M. Ayoub, Robin Green, Katherine Puentes, Frank Oplinger, Nishant V. Panwar, Fritz Obermeyer, Adam R. Root, Andrew L. Beam, Frank J. Poelwijk, and Gevorg Grigoryan. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, 2023. ISSN 0028-0836. doi: 10.1038/s41586-023-06728-8.
2. Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023. ISSN 0028-0836. doi: 10.1038/s41586-023-06415-8.
3. Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, Mateo Reveiz, Roman Bushuiev, Tomáš Pluskal, Josef Sivic, Karsten Kreis, Arash Vahdat, Shamayeeta Ray, Jonathan T. Goldstein, Andrew Savinov, Jacob A. Hambalek, Anshika Gupta, Diego A. Taquiri-Diaz, Yaotian Zhang, A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, Chang C. Liu, Gene-Wei Li, Denes Hnisz, David M. Sabatini, William F. DeGrado, Jeremy Wohlwend, Gabriele Corso, Regina Barzilay, and Tommi Jaakkola. BoltzGen: Toward universal binder design. *bioRxiv*, page 2025.11.20.689494, 2025. doi: 10.1101/2025.11.20.689494.
4. Jasper Butcher, Rohith Krishna, Raktim Mitra, Rafael I. Brent, Yanjing Li, Nathaniel Corley, Paul T. Kim, Jonathan Funk, Simon Mathis, Saman Salike, Aiko Muraishi, Helen Eisenach, Tuscan Rock Thompson, Jie Chen, Yuliya Politanska, Enisha Sehgal, Brian Coventry, Odin Zhang, Bo Qiang, Kieran Didi, Max Kazman, Frank DiMaio, and David Baker. De novo design of all-atom biomolecular interactions with RFdiffusion3. *bioRxiv*, page 2025.09.18.676967, 2025. doi: 10.1101/2025.09.18.676967.
5. Vinicius Zambaldi, David La, Alexander E Chu, Harshnira Patani, Amy E Danson, Tristan O C Kwan, Thomas Frerix, Rosalia G Schneider, David Saxton, Ashok Thillaisundaram, Zachary Wu, Isabel Moraes, Oskar Lange, Eliseo Papa, Gabriella Stanton, Victor Martin, Sukhdeep Singh, Lai H Wong, Russ Bates, Simon A Kohl, Josh Abramson, Andrew W Senior, Yilmaz Alguel, Mary Y Wu, Irene M Aspalter, Katie Bentley, David L V Bauer, Peter Cherepanov, Demis Hassabis, Pushmeet Kohli, Rob Fergus, and Jue Wang. De novo design of high-affinity protein binders with AlphaProteo. *arXiv*, 2024. doi: 10.48550/arxiv.2409.08022.
6. Latent Labs Team, Alex Bridgland, Jonathan Crabbé, Henry Kenlay, Daniella Pretorius, Sebastian M Schmon, Agrin Hilmkil, Rebecca Bartke-Croughan, Robin Rombach, Michael Flashman, Tomas Matteson, Simon Mathis, Alexander W R Nelson, David Yuan, Annette Obika, and Simon A A Kohl. Latent-x: An atom-level frontier model for de novo protein binder design. *arXiv*, 2025. doi: 10.48550/arxiv.2507.19375.
7. Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M. Bronstein, Martin Steinegger, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *The Fourteenth International Conference on Learning Representations*, 2026.
8. Nathaniel R. Bennett, Joseph L. Watson, Robert J. Ragotte, Andrew J. Borst, DéJenaé L. See, Connor Weidle, Riti Biswas, Yutong Yu, Ellen L. Shrock, Russell Ault, Philip J. Y. Leung, Buwei Huang, Inna Goreshnik, John Tam, Kenneth D. Carr, Benedikt Singer, Cameron Criswell, Basile I. M. Wicky, Dionne Vafeados, Mariana Garcia Sanchez, Ho Min Kim, Susana Vázquez Torres, Sidney Chan, Shirley M. Sun, Timothy T. Spear, Yi Sun, Keelan O'Reilly, John M. Maris, Nikolaos G. Sgourakis, Roman A. Melnyk, Chang C. Liu, and David Baker. Atomically accurate de novo design of antibodies with RFdiffusion. *Nature*, 649(8095):183–193, 2026. ISSN 0028-0836. doi: 10.1038/s41586-025-09721-5.
9. Chai Discovery Team, Jacques Boitreaud, Jack Dent, Danny Geisz, Matthew McPartlon, Joshua Meier, Zhuo-

ran Qiao, Alex Rogozhnikov, Nathan Rollins, Paul Wollenhaupt, and Kevin Wu. Zero-shot antibody design in a 24-well plate. *bioRxiv*, page 2025.07.05.663018, 2025. doi: 10.1101/2025.07.05.663018.

10. Nabla Bio and Surojit Biswas. De novo design of epitope-specific antibodies against soluble and multipass membrane proteins with high specificity, developability, and function. *bioRxiv*, page 2025.01.21.633066, 2025. doi: 10.1101/2025.01.21.633066.
11. Daniel Cutting, Frédéric A. Dreyer, David Errington, Constantin Schneider, and Charlotte M. Deane. De novo antibody design with se(3) diffusion. *Journal of Computational Biology*, 32(4):351–361, 2025. doi: 10.1089/cmb.2024.0768.
12. Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M. Jude, Iva Marković, Rameshwar U. Kadam, Koen H. G. Verschueren, Kenneth Verstraete, Scott Thomas Russell Walsh, Nathaniel Bennett, Ashish Phal, Aerin Yang, Lisa Kozodoy, Michelle DeWitt, Lora Picton, Lauren Miller, Eva-Maria Strauch, Nicholas D. DeBouver, Allison Pires, Asim K. Bera, Samer Halabiya, Bradley Hammerson, Wei Yang, Steffen Bernard, Lance Stewart, Ian A. Wilson, Hannele Ruohola-Baker, Joseph Schlessinger, Sangwon Lee, Savvas N. Savvides, K. Christopher Garcia, and David Baker. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, 2022. ISSN 0028-0836. doi: 10.1038/s41586-022-04654-9.
13. Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Proteina: Scaling flow-based protein structure generative models. *arXiv*, 2025. doi: 10.48550/arxiv.2503.00710.
14. Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-proteina: Atomistic protein generation via partially latent flow matching. *arXiv*, 2025. doi: 10.48550/arxiv.2507.09466.
15. Danny Reidenbach, Zhonglin Cao, Zuobai Zhang, Kieran Didi, Tomas Geffner, Guoqing Zhou, Jian Tang, Christian Dallago, Arash Vahdat, Emine Kucukbenli, and Karsten Kreis. Consistent synthetic sequences unlock structural diversity in fully atomistic de novo protein design. *arXiv*, 2025. doi: 10.48550/arxiv.2512.01976.
16. Tao Sun, Liyuan Zhu, Shengyu Huang, Shuran Song, and Iro Armeni. Rectified point flow: Generic point cloud pose estimation. *arXiv*, 2025. doi: 10.48550/arxiv.2506.05282.
17. Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv*, 2022. doi: 10.48550/arxiv.2207.12598.
18. Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.235.
19. Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, 2021. ISSN 0305-1048. doi: 10.1093/nar/gkab1061.
20. Andy M. Lau, Nicola Bordin, Shaun M. Kandathil, Ian Sillitoe, Vaishali P. Waman, Jude Wells, Christine A. Orengo, and David T. Jones. Exploring structural diversity across the protein universe with the encyclopedia of domains. *Science*, 386(6721):eadq4946, 2024. ISSN 0036-8075. doi: 10.1126/science.adq4946.
21. Michel van Kempen, Stephanie S Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L M Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein structure search with foldseek. *bioRxiv*, page 2022.02.07.479398, 2022. doi: 10.1101/2022.02.07.479398.
22. John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas

Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. ISSN 0028-0836. doi: 10.1038/s41586-021-03819-2.

23. Sidhartha Chaudhury, Sergey Lyskov, and Jeffrey J. Gray. PyRosetta: a script-based interface for implementing molecular modeling algorithms using rosetta. *Bioinformatics*, 26(5):689–691, 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq007.
24. Nathaniel R. Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, Frank DiMaio, Steven De Munck, Savvas N. Savvides, and David Baker. Improving de novo protein binder design with deep learning. *Nature Communications*, 14 (1):2625, 2023. doi: 10.1038/s41467-023-38328-5.
25. Michael S Albergo, Mark Goldstein, Nicholas M Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. *arXiv*, 2023. doi: 10.48550/arxiv.2310.03725.
26. Jason Yim, Andrew Campbell, Andrew YK Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S Veeling, Regina Barzilay, Tommi Jaakkola, et al. Fast protein backbone generation with se (3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023.
27. Avishek Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. Se(3)-stochastic flow matching for protein backbone generation. *arXiv*, 2023. doi: 10.48550/arxiv.2310.02391.
28. Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. ISSN 0036-8075. doi: 10.1126/science.ade2574.
29. Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014. ISSN 1465-6906. doi: 10.1186/s13059-014-0550-8.

A1 Dyno Psi model details

For the backbone generation model, we adopt a flow-matching approach with a scalable protein transformer architecture and triangle updates, similar to Proteína (13). In this setup, a flexible transformer is trained as a denoiser model to approximate the conditional vector field that transports a given noisy sample $x_t, t \in [0,1)$ to a clean sample $\mathbf{x}_1 \sim p_D(x)$ from the data distribution. As in Proteína, we use a simple-to-sample Gaussian prior as the initial noise distribution p_0 and a training loss that combines a flow loss on the conditional vector field with an auxiliary cross-entropy distogram loss. In order to flexibly extend this setup to a diverse set of protein design tasks, we also integrate ideas from the stochastic interpolants perspective on flow matching (25), which introduces position-specific noising and denoising. Specifically, for a protein complex with L residues, we introduce $\xi \in \{0,1\}^L$, a noise mask indicating the designed (as opposed to globally fixed) positions. This results in the training objective:

$$\min_\theta \mathbb{E}_{\mathbf{x}_1 \sim p_D(\mathbf{x}), \mathbf{x}_0 \sim p_0, t \sim p(t)} \Bigg[\frac{1}{3L} \underbrace{\left\| \mathbf{v}_t^\theta(\mathbf{x}_t, t, \hat{\mathbf{x}}(\mathbf{x}_t), \xi, \mathbf{c}) - \xi \cdot (\mathbf{x}_1 - \mathbf{x}_0) \right\|_2^2}_{\text{conditional flow-matching loss}} - \frac{\mathbf{1}(t > 0.3)}{9L^2} \sum_{i,j} \sum_{b=1}^{64} \underbrace{D_{b,ij}(\mathbf{x}_1) \log p_{b,ij}^\theta(\mathbf{x}_t, t, \hat{\mathbf{x}}(\mathbf{x}_t), \xi, \mathbf{c})}_{\text{auxiliary binned distogram loss}} \Bigg].$$

where $p(t)$ is the distribution over times $t \in [0,1)$, $\mathbf{x}_t = (1-\xi) \cdot t\mathbf{x}_1 + \xi \cdot (1-t)\mathbf{x}_0$ is the linear stochastic interpolant, $\mathbf{v}_t^\theta$ is the denoising model with parameters θ, $\hat{\mathbf{x}}(\mathbf{x}_t)$ is the (optional) self-conditioning prediction, c are additional conditioning inputs, and $D_{b,ij}(\mathbf{x})$ are the binned pairwise distance between residue atoms i and j. We adapt the time sampling distribution from Proteína (13), $p(t) = (1-\gamma)U(0,1) + \gamma\beta(1.9, 1.0)$, where β is a beta distribution and U is a uniform distribution. During training, we use $\gamma = 0.98$ for pretraining phase 1 and fine-tuning and $\gamma = 0.49$ for pretraining phase 2.

Note that our training objective is defined over all three backbone atoms. Other approaches tend to either denoise a Cα-only backbone structure and design the sequences with ProteinMPNN or attempt to denoise the backbone and sequence at the same time, either by modeling all atoms or by using a latent representation over sequences and side chains. We chose an approach that is most similar to FrameFlow (26), where we denoise the 3 backbone atom coordinates (analogous to FrameFlow's frames, except in Cartesian coordinates) and design the sequence with ProteinMPNN. While we did experiment with denoising the sequence and side chains along with the backbone structure, we found that this approach did not yield better results while being much more expensive to run. This is bolstered by the observation that many prominent generative approaches in the field also choose to re-design the sequences of binders with an inverse folding model, even if they initially generate sequences (3, 4).

The model architecture of $\mathbf{v}_t^\theta$ is comprised of three tracks: (1) the residue track, which creates a per-residue representation of the protein; (2) the pair track, which creates a representation for each pair of residues; and (3) the residue conditioning track, which encodes additional information about the input protein and design task (Figure A1). While Proteína enables conditioning on a scalar class value per protein, our architecture features modified conditioning and pair tracks that embed rich single and pairwise residue conditioning information, thereby unlocking a wide range of design tasks that incorporate elements of both rigid body docking and partial redesign (Figure A2), as well as enabling a way to incorporate crucial information about multi-chain complexes, such as chain breaks, gaps between resolved residues, or known protein pockets. These architecture modifications integrate the key ideas of position-specific noising (25) and pose estimation techniques in computer vision (16), which add context about relative positioning. This enables the model to leverage both global and relative backbone position information.

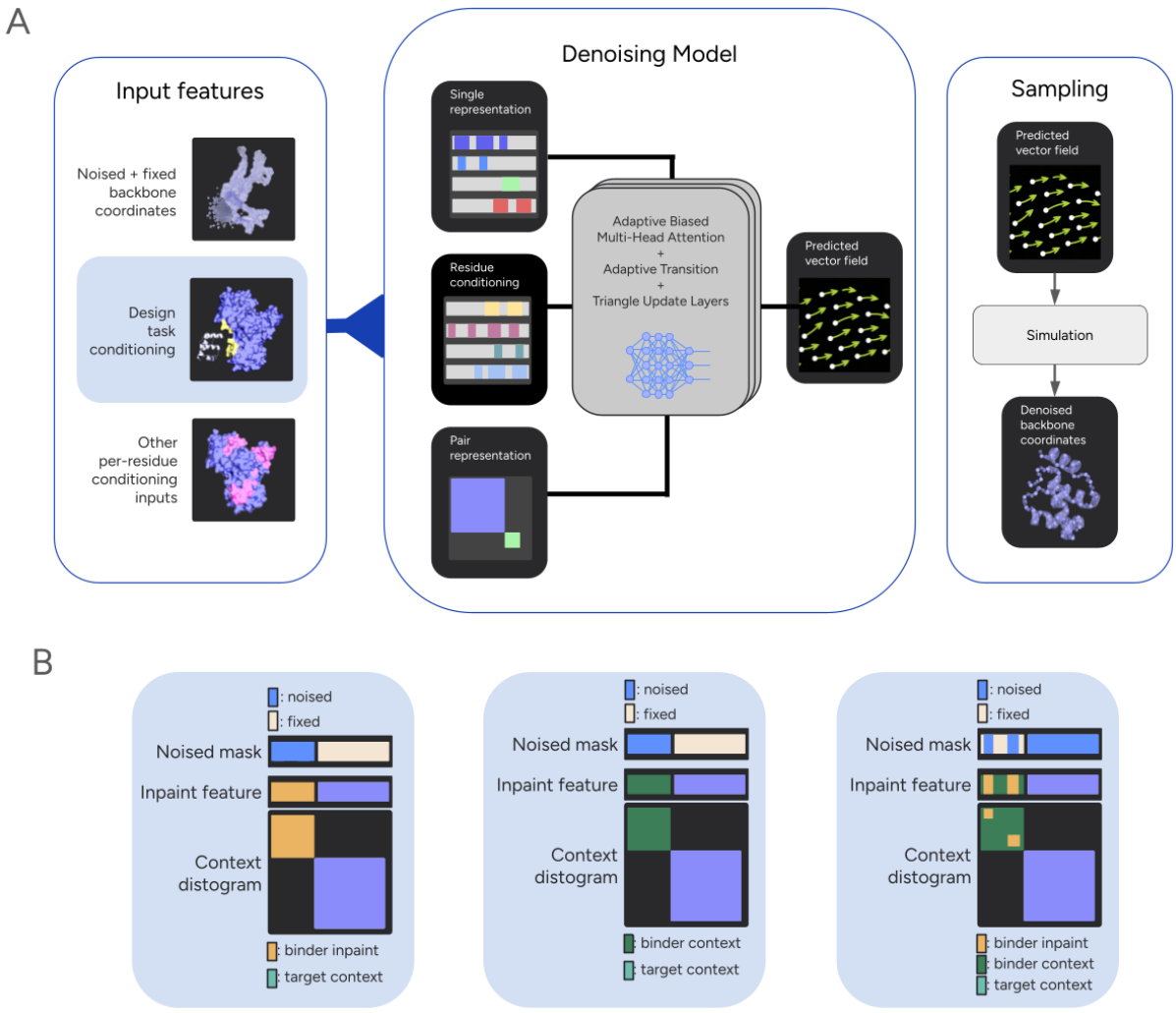


Figure A1. Dyno Psi model architecture. A) Overview of the denoising model $\mathbf{v}_t^\theta$, which operates on three interconnected tracks: a residue track (single representation), a residue conditioning track, and a pair track. The model ingests both global and relative structural information, including noised backbone coordinates and pairwise context distogram features, and predicts a vector field used for denoising during sampling. B) Additional examples of conditioning features used to encode different design tasks. A noised mask indicates which residues are noised and updated by the learned flow, while inpainting features and context distograms encode pairwise relative positioning. Different configurations of these inputs enable a range of tasks including (from left to right) binder design, rigid-body docking, and binder inpainting, thereby enabling a unified framework for diverse protein design tasks.

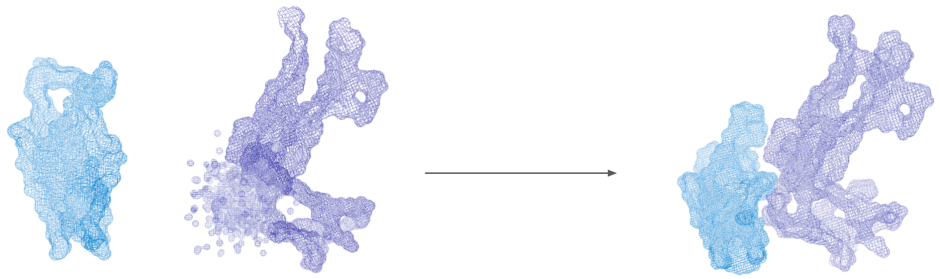


Figure A2. Illustration of simultaneous rigid-body docking and interface redesign. An example of an inpainting design task where a binder (blue) is docked onto a target protein (purple) while the residues at the target interface are simultaneously redesigned. Dyno Psi jointly denoises the global positioning of the binder and refines the target interface, leveraging both the known pairwise relative positioning of the binder residues and the fixed target context, and noisy global coordinates at the designed residues.

Global information is provided through 3D atomic coordinates in the sequence track and a mask in the conditioning track that distinguishes residues with fixed positions from those moved by the learned flow. Relative positions are provided through context distograms in the pair track and an inpainting feature represented in both conditioning and pair tracks. The inpainting feature encodes which pairwise relative distances involve designed residues moved by the learned flow and which are fixed. Figure A1 displays the way the binder design task, used for fine-tuning Dyno Psi-1, is encoded through these model inputs; the model intakes known global and relative position information for the target and fully designs a binder chain. A small change to the inputs enables rigid body docking between two chains: we provide relative intrachain positioning information for both chains (instead of only one) in the context distogram and inpainting feature, and the model learns to denoise the rigid body positioning of the noised chain (Figure A1B). Overall, the updated transformer architecture allows Dyno Psi models to simultaneously handle a broad variety of design tasks, from fully unconditional protein design to partial re-design of complex interfaces (Figures A1B), all encoded through different combinations of pairwise and per-residue context and conditioning inputs.

To permit the broader range of design tasks in Dyno Psi-0, we developed a cohesive strategy for centering and data augmentation that works across monomer design, binder design, inpainting and partial re-design. Specifically, we first center the global frame on chains that include designed residues; for binder design, this means centering on the binder. During training, we then apply a random global rotation, to induce learned rotational equivariance, and small random global translations, to reduce leakage of the true center.

In order to force the model to reason over reason over binder-target docking, we introduce chain-specific priors with random translations in the prior centers $\delta \sim p_0 = \mathcal{N}(\delta \mid 0, r^2)$, similar to (7). For binder design, the requires the model to reason over not only the positions of atoms within the binder, but also over relative positioning of the binder with respect to the target. In this setup, the interpolant is

$$\mathbf{x}_t = \xi\left[t\mathbf{x}_1 + (1-t)(\mathbf{x}_0 + \delta)\right] + (1-\xi)\mathbf{x}_1$$

where $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, $\sigma = 10$ Å. To improve robustness to errors in the predicted binder center of mass at inference time — particularly for larger binders — we scale r as a function of binder length. We reasoned that errors in placing the binder center of mass would scale with its radius of gyration. Since the radius of gyration follows a power law relationship with chain length, we fit a relationship empirically using our PDB training data: $r(L) = a \cdot L^b + c$, where $a = 2.1$, $b = 0.4$, and $c = -10.0$.

A2 Sampling Details

A2.1. SDE simulation.

We adapt the stochastic differential equation that Proteína describes (13) to handle fixed target coordinates in the binder design setting:

$$d\mathbf{x}_t = \mathbf{v}_t^\theta(\mathbf{x}_t)d_t + g(t,\xi)\mathbf{s}_t^\theta(\mathbf{x}_t)dt + \sqrt{2g(t,\xi)\gamma}dW_t$$

for $t \in [0,1)$, where $\mathbf{v}_t^\theta$ is the vector field prediction from our model, $\mathbf{s}_t^\theta$ is the score of the intermediate marginal distribution, W_t is a Weiner process, $g(t,M)$ is a time-dependent Langevin scaling term that affects the level of stochasticity in the simulation, and γ is a noise scaling term that modulates sampling temperature. Since Dyno Psi-1 models Gaussian flows, we are able to obtain the score $\mathbf{s}_t^\theta(x_t)$ directly from the vector field model predictions.

With the exception of the very beginning ($t < 0.06$) and the very end ($t > 0.9$) of the simulation, a positive Langevin scaling term was highly beneficial to the quality of our binder design samples. We set the scaling term according to this schedule:

$$g(t) = \begin{cases} 0 & t < 0.06 \\ \frac{1}{t+0.01} \cdot \xi & 0.06 \le t \le 0.90 \\ 0 & t > 0.90 \end{cases}$$

Consistent with Proteína's results, we found that emphasizing Langevin dynamics terms at earlier time steps during formation of the global binder structure produced the best *in silico* results, but switching to a deterministic ODE was required to maintain simulation stability at the end of the trajectory. For this time period we utilize a low temperature, score-scaled ODE that again progresses more aggressively toward high-density (more designable) regions of the intermediate marginal distributions compared to a standard ODE but without any noise injection. An initial "warmup" period of deterministic ODE simulation without score scaling improved sample quality and tempered initial volatility in early-trajectory denoised coordinate predictions.

Consistent with other flow matching and diffusion approaches (13, 27), we find that a low-temperature noise schedule improves the designability of the Dyno Psi-1 samples and set $\gamma = 0.3$ for all in-silico benchmarking results. We use a linear time schedule that discretizes the simulation into 100 uniformly-spaced steps.

A2.2. Chain-specific prior distributions.

The Dyno Psi-1 sampling procedure is initialized with an estimated position of the binder center of mass relative to the target chains. We experimented with algorithmic strategies for determining these relative positions, such as the hotspot-based center-of-mass conditioning described in RFdiffusion3 (4), but we found these approaches susceptible to pathologies for certain targets. While Dyno Psi-1 is trained to be robust to misspecification of this relative positioning (see Appendix A1), we find that very large errors can degrade sample quality. Instead, we find manual estimation of the binder center-of-mass to be a more reliable strategy across our benchmarked targets that also generalizes to other design tasks, such as binder inpainting. We set the standard deviation of each prior chain to 10 Å to match our training setup.

A2.3. Target-conditioned classifier-free guidance.

Classifier-free guidance (CFG) is an common technique for generating samples from diffusion and flow matching models conditional on auxiliary information. Other biomolecular design models have applied this technique to generate monomers with particular fold class annotations (13) or to control aspects of small-molecule binding interactions (4).

When sampling with CFG, each denoising step in the simulation is computed via a weighted difference between a conditional and unconditional predictions:

$$v_t^{CFG} = \omega v_t^C + (1 - \omega) v_t^U$$

where v_t^C is the conditional prediction, v_t^U is the unconditional prediction, and ω, known as the guidance scale and typically set to $\omega > 1$, governs the the strength of the guidance. CFG trajectories more explicitly emphasize the difference between the conditional and unconditional states compared to a simulation governed only by conditional predictions.

We hypothesized that CFG could be utilized in the binder design setting to encourage the formation of structural features that are more common in protein-protein interactions than in unconditional monomer structures. Specifically, our conditional predictions are the standard predictions from the model with fixed target features. For our unconditional predictions, we alter the input features to remove the target information and generate the "binder" chain in a target-free context. We set $\omega = 1.5$.

A3 Dyno Synthetic Inter-domain Dimers dataset curation

Following the approach established by Proteína-Complexa (7) for curating their Teddymer dataset, we started with the ~52M AlphaFold DB entries corresponding to AFDB50 cluster representatives and enumerated all pairs of TED-annotated domains with at least a CAT-level domain classification, yielding ~14M dimer pairs. For each pair, we determined interface residues by calculating how many Cα atoms in one domain were within 10Å of a Cα atom in the other domain. Here is where our pipeline diverges from the Teddymer procedure: we filter to domain pairs for which each domain has at least 8 interface residues, whereas Teddymer's cutoff is 4 at this stage. We additionally filtered out domain pairs that participate in a 3-way interaction because we did not want any of the final dimer examples to have a "ghost" third domain stabilizing the interaction. Finally, we filtered for domain pairs with interface pAE < 10Å and interface pLDDT > 70. Overall, 4,070,959 domain pairs survived this filtering step.

We then clustered all the dimer pairs with FoldSeek using the same settings that were used to cluster Teddymer examples. This yielded 1,063,207 clusters, including 724,966 singletons and 338,241 non-singletons. All of the ~4M dimers are included in the data release, along with metadata on which clusters they belong to, and how many interface residues were estimated.

To curate the final synthetic dataset to augment the PDB multimer dataset for fine-tuning binder design, we used AlphaFold2 to refold each of the ~1M cluster representatives to determine if AlphaFold2 would predict that the chains would form a complex, even when not tethered together. We refolded with AlphaFold2 monomer models 1 and 2 with templates turned on for both chains and filtered to dimers with interface pTM > 0.5 and interface pAE < 12.4Å for both models. We also aligned the refolded structures to the original "dimers" by one chain and calculated the RMSD of the other chain and filtered out any dimers where that RMSD was greater than 6Å. We then refolded each domain in the "dimer" with ESMfold (28) and filtered for "dimers" where both participating domains had a reconstruction RMSD < 3.5Å and pLDDT > 70. Finally, we used PyRosetta FastRelax

(23) to refine the refolded dimer structures (AlphaFold2 monomer model 1). A total of 32,253 "dimers" passed this set of stringent interface filters.

A4 *In silico* benchmarking

A4.1. Target details.

Table A1. *In silico* benchmarking targets with the structural crops and hotspot residues that specify the binder design task.

Target	PDB ID	Chain and residue indices	Hotspot residues	Hotspot amino acid identities
IL-7RA	3DI3	B17–209	A58, A80, A139	V, L, Y
PD-L1	5O45	A19–132	A54, A56, A66, A115	I, Y, Q, M
SARS-CoV2-RBD	6M0J	E333–526	E485, E489, E494, E500, E505	G, Y, S, T, Y
TrkA	1WWW	X282–382	X294, X296, X333	V, M, L

A4.2. Generating benchmarking samples.

We generated 16,900 miniprotein binder designs using each of Dyno Psi-1, BoltzGen, and RFdiffusion3. The designs are distributed evenly across five targets (four benchmarking targets (Table A1) and Dyno Target A) and thirteen binder lengths $[60, 65, 70, \cdots, 115, 120]$. We used the default inference settings for both RFdiffusion3 and BoltzGen-1 Diffusion, saving all designs prior to any filtering or ranking. For each generated structure four sequences were designed using an inverse folding model. For Dyno Psi-1 and RFdiffusion3 sequences we used ProteinMPNN with temperature 0.0001 and fixed target positions. For BoltzGen sequences we used BoltzIF with default settings. Designs from all three methods were then passed through the same complex refolding and filtering pipeline.

A4.3. Pass rate calculation.

All designed binder sequences (see Appendix A4.2) were refolded in complex with the corresponding target following the AlphaFold2 initial guess protocol (24). For each method we computed per target pass rates using the following filter thresholds:

1. Interface pAE < 10
2. Binder pLDDT > 80
3. Binder RMSD < 1 Å

Binder RMSD is calculated using the designed binder aligned with the refolded complex binder. These thresholds were selected to match those used in external benchmarking (24).

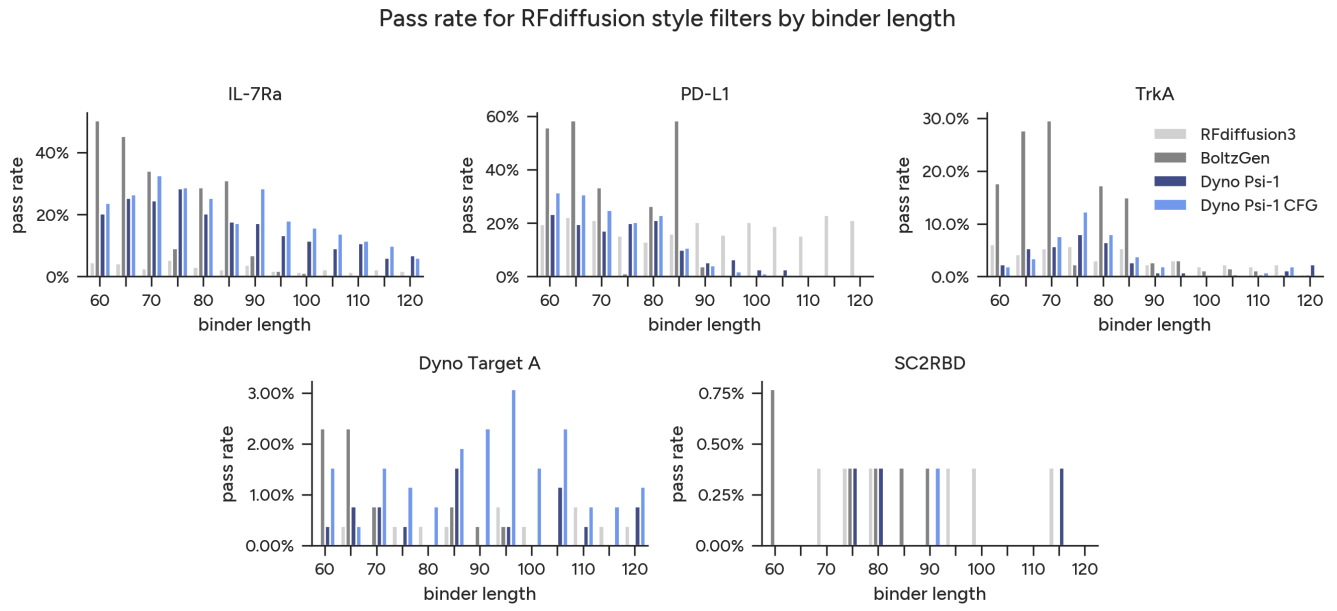


Figure A3. ***In silico*** **pass rates by binder length.** Pass rates from the benchmarking results are shown for each model separated by target and binder length. Pass rates reflect the number of backbones for which at least one sequence passes AlphaFold2 "initial guess" filters. PD-L1, IL-7RA, and TrkA backbones were additionally filtered for those with a TM-score < 0.6 to PD-1, IL-7 and NGF, respectively.

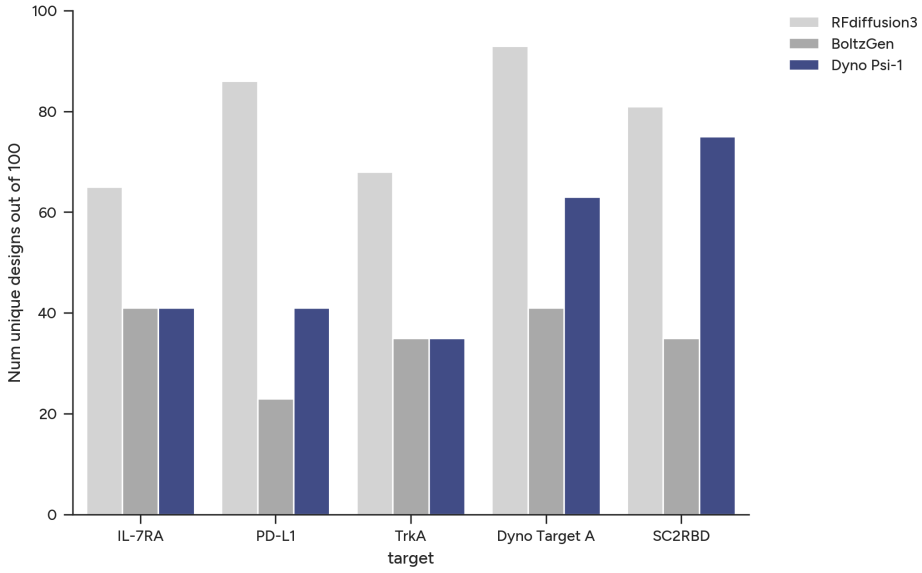


Figure A4. Diversity of generated samples. Number of unique clusters for 100 random samples generated from RFdiffusion3, BoltzGen, and Dyno Psi-1. Samples were clustered using a greedy clustering approach with a TM-score cutoff of 0.6.

A4.4. Supplementary in silico results.

A5 Yeast Display

DNA encoding miniprotein designs were synthesized commercially (Twist Biosciences) with golden gate restriction sites for cloning into a plasmid backbone for yeast expression and surface display via fusion to Aga2p. Yeast strain EBY100 was transformed with the plasmid library by electroporation and grown in minimal media selective for the yeast strain (-ura) and the transforming plasmid (-trp). Expression was induced by growth in 2% galactose for 24h, and surface expression of miniproteins was confirmed by staining cells with anti-Myc-biotin (R&D systems; BAM3696) and strep-PE (Abcam; ab239759) antibodies and Flow sorting on the Sony MA900

instrument.

For binding assays, the induced yeast library was spun down, washed with PBSATE (PBS with 0.5% BSA, 0.5 mM EDTA, 0.02% Tween-20) and normalized to an input of 1E9 cells per reaction. Each binding reaction was performed in triplicate. Yeast cells were incubated for 2 hours with 1000μM biotinylated IL7RA (SinoBiological; 10975-H49H-B), washed once with PBSATE, incubated for 2 hours with 240μL streptavidin beads (NEB; S1420S), and washed three times with PBSATE to remove unbound yeast. To determine which miniproteins bound IL7RA, plasmids encoding the miniproteins were purified from the starting library and the IL7RA-bead-bound yeast using the Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research; D2004). Extracted plasmids were indexed for sequencing with primers containing Illumina p5 and p7 barcodes. NGS libraries were quantified on a Tapestation4200 and normalized for sequencing on an Illumina Nextseq 2000 instrument. Each sequencing library received a minimum of 27 million reads.

Sequenced reads were aligned to a reference using custom alignment tools. Differential enrichment analysis was performed by DESeq2 (29) using default parameters and without shrinkage, to determine a Log2FoldChange enrichment value for each variant in the library relative to the input sample. Binding hit calling was determined by a Log2FoldChange measurement above a threshold of 2.5 along with a significant Benjamini-Hochberg adjusted p-value <0.05. Variants with low overall abundance (raw reads <100) were excluded from the analysis due to high noise and low confidence.

Dyno Psi-0.5 designs included in the yeast display assay were the top ranking designs by AlphaFold2 ipSAE from among a loosely pre-filtered pool of all generated designs that met all of the following *in silico* criteria: Binder RMSD < 3.5 Å and binder pLDDT > 0.7 when refolded using ESMFold ipSAE > 0.1 for at least one of the AlphaFold2 monomer models “model_1_ptm” and “model_2_ptm” RMSD between the binder in the designed complex and refolded complex is at most 20 Å when the complexes are aligned on the target TM scores of the binder against ubiquitin, IL7, and PD1 are at most 0.6

BoltzGen designs included in the yeast display assay were generated with the standard BoltzGen design pipeline, which ranks designs. The top ranking sequences were selected for further validation. Model comparison statistics: We generated different numbers of designs per model (Table A2), so we used a sample-size matched strategy for comparison. Design success rates were computed by randomly sampling sets of 100 measured miniproteins from each model and calculating the fraction that bound to IL-7RA. The BoltzGen IL-7RA success rate observed here is lower than previously reported (3), likely because we generated fewer designs *in silico* relative to the number of designs that were experimentally tested.

Table A2. Counts per model of designs that were generated, tested experimentally, and found to bind IL-7RA in a yeast display assay.

Model	Total designs generated	Designs tested with yeast display	Percentage of designs measured	Designs identified as binders
Dyno Psi-0.5	181,000	921	0.51%	74
BoltzGen	12,164	122	1.02%	2

A5.1. SPR.

Experiments were carried out by Adaptyv Bio.
Surface plasmon resonance (SPR) binding assays were performed on a Carterra LSA XT instrument in running buffer consisting of 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, and 0.05% Tween-20 (pH 7.4). Twin-Strep-tagged miniproteins were produced by cell-free expression from codon-optimized DNA constructs (Twist Bio-

science) assembled using NEBuilder HiFi DNA Assembly (NEB) and expressed in 8 µL in vitro translation reactions containing 4 nM DNA at 37 °C for 8 h, followed by normalization using an affinity-based quantification assay. Carboxymethylated SPR chips were functionalized with Strep-Tactin XT (IBA Lifesciences) by conditioning the surface with 50 mM NaOH, activating with EDC/NHS, coupling Strep-Tactin XT (50 µg mL^{-1} in 10 mM sodium acetate, pH 4.5), quenching with 1 M ethanolamine (pH ~8.5), and washing with 0.1 M sodium borate and 1 M NaCl (pH 9.0). The Twin-Strep-tagged miniproteins were captured using a 96-channel multichannel head with bidirectional flow for 750 s followed by a 600 s baseline in running buffer. IL7-RA protein (Cat. No. IL7-H5258, Acrobiosystems) was injected in a dilution series (10 nM, 100 nM, 1000 nM) at 50 µL min^{-1} in a single-cycle kinetic format without intermediate regeneration, with each concentration consisting of a 60 s baseline, 300 s association, and 600 s dissociation phase. Surfaces were regenerated after the full series using 10 mM glycine-HCl (pH 1.5) for 5 min followed by a 20 min buffer wash. Sensorgrams were preprocessed by trimming kinetic phases, correcting signal discontinuities, aligning association and dissociation segments, and subtracting reference and baseline signals, and were globally fit to a 1:1 Langmuir binding model using fitting software to estimate kinetic parameters (k_{on}, k_{off}, and k_D). Binding was classified based on the presence of quantifiable kinetic fits or significant association-phase responses relative to negative controls.