BoltzGen:
Toward Universal Binder Design

Hannes Stark*[1], Felix Faltings†[1], MinGyu Choi†[1], Yuxin Xie†[1], Eunsu Hur†[1], Timothy O'Donnell†[3], Anton Bushuiev†[4], Talip Uçar†[2], Saro Passaro[2], Weian Mao[1], Mateo Reveiz[1], Roman Bushuiev[4,5], Tomáš Pluskal[5], Josef Sivic[4], Karsten Kreis[6], Arash Vahdat[6], Shamayeeta Ray[7], Jonathan T. Goldstein[7], Andrew Savinov[1], Jacob A. Hambalek[8], Anshika Gupta[8], Diego A. Taquiri-Diaz[8], Yaotian Zhang[9], A. Katherine Hatstat[10], Angelika Arada[10], Nam Hyeong Kim[10], Ethel Tackie-Yarboi[10], Dylan Boselli[10], Lee Schnaider[10], Chang C. Liu[8], Gene-Wei Li[1,11], Denes Hnisz[9], David M. Sabatini[5,7], William F. DeGrado[10], Jeremy Wohlwend[2], Gabriele Corso[2], Regina Barzilay[1,12], Tommi Jaakkola[1]

Correspondence to `hstark@csail.mit.edu`

Abstract

We introduce *BoltzGen*, an all-atom generative model for designing proteins and peptides across all modalities to bind a wide range of biomolecular targets. BoltzGen builds strong structural reasoning capabilities about target-binder interactions into its generative design process. This is achieved by unifying design and structure prediction, resulting in a single model that also reaches state-of-the-art folding performance. BoltzGen's generation process can be controlled with a flexible design specification language over covalent bonds, structure constraints, binding sites, and more. We experimentally validate these capabilities in a total of eight diverse wetlab design campaigns with functional and affinity readouts across 26 targets. The experiments span binder modalities from nanobodies to disulfide-bonded peptides and include targets ranging from disordered proteins to small molecules. For instance, we test 15 nanobody and protein binder designs against each of nine novel targets with low similarity to any protein with a known bound structure. For both binder modalities, this yields nanomolar binders for 66% of targets. We release model weights, data, and both inference and training code at: https://github.com/HannesStark/boltzgen.

*Interned at Boltz for a part of the project, †Equal core computational contributors, [1]MIT, [2]Boltz, [3]Open Athena, [4]CTU Prague, [5]IOCB Prague, [6]NVIDIA, [7]IOCB Boston, [8]UC Irvine, [9]MPI, [10]UCSF, [11]HHMI, [12]Jameel Clinic

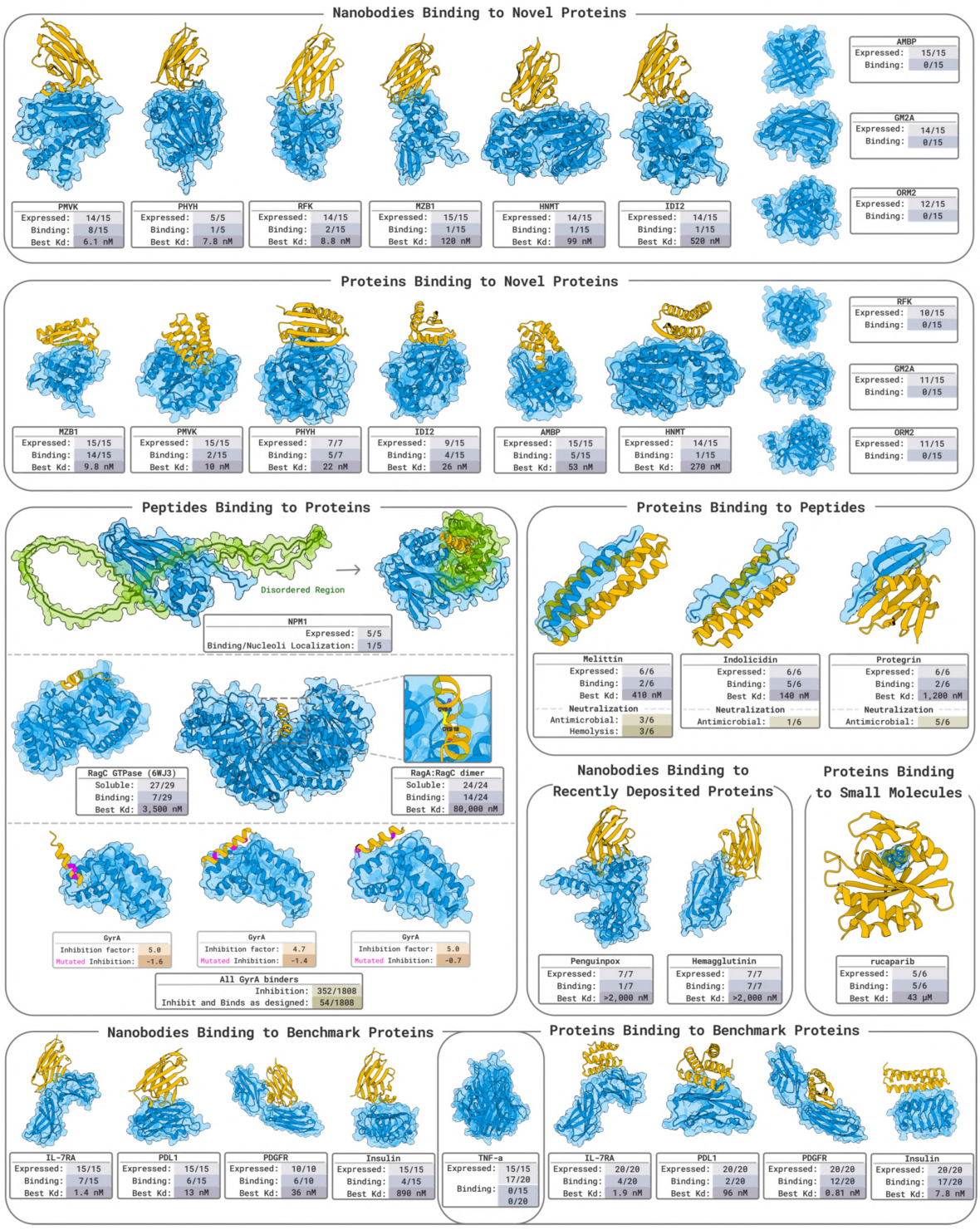


Figure 1: **BoltzGen Wetlab Validation.** We validate BoltzGen in collaboration with leading wet labs working on high-impact biological problems. These independently test designs for their specific applications. Additionally, we validate on 9 "novel targets" meaning that there are no proteins in a bound context with more than 30% sequence identity in the entire PDB.

Main Text Table of Contents

1 Introduction

De-novo binder design offers considerable potential for automating drug discovery. A number of previous techniques have been proposed to address parts of this challenge, including [Watson et al., 2023, Pacesa et al., 2024, Bennett et al., 2025, Mille-Fragoso et al., 2025]. Several key limitations remain, however. For example, many of the techniques are tailored to specific classes of biomolecules such as nanobodies or peptides. As models learn to emulate physics primarily through examples provided, we believe expanding the generality of the method further improves its design capabilities for specific classes as well. Another key limitation has to do with evaluation as methods are often tested on targets that have closely related complexes in the training data. The potential of de-novo binder design comes precisely from its presumed ability to extrapolate beyond easy targets. We believe design methods should be tested accordingly. Moreover, in real-world discovery campaigns, a number of additional requirements and constraints govern successful designs. It is important to be able to control the design process in a flexible manner.

Here we introduce BoltzGen, a binder design algorithm that addresses the above desiderata. At its core, the BoltzGen pipeline uses a single, all-atom generative model that unifies design and structure prediction. A purely geometry-based representation of designed residue types enables scalable training on both tasks simultaneously. As a result, unlike any previous design model, BoltzGen matches the performance of state-of-the-art folding models (Figure 20). BoltzGen's structure-based reasoning about target-binder interactions supports design of high-affinity binders to novel targets, unrelated to complexes seen during training (Figure 1). We also provide a design specification language that allows for constraining binders across a variety of possible requirements, such as selecting a desired binding site or a (partial) structure for the target and covalent bonds or residue identity constraints in the design. The method is described in detail in Section 3.

Wetlab Validation. We experimentally validate our designs in a large-scale distributed effort involving multiple wetlabs. Each group selected targets and output modalities, relevant to their specific application, and then independently validated BoltzGen designs. To measure generalization capacity, we explicitly focus on targets that are highly dissimilar to any proteins for which bound structures exist.

Here we report the experimental results available to date; additional validation is ongoing. Some data are temporarily confidential at collaborators' request, and we will update this work as further results become available.

1. Section 2.2: We design nanobodies against 9 novel targets for which there are no proteins with >30% sequence identity in a bound context in the entire PDB. Experimentally validating 15 or fewer designs against each target yields nM binders for 66% of them. The analogous experiment for protein binder designs results in the same success rate of nM binders against 66% of targets.
2. Section 2.3: When designing proteins to bind 3 bioactive peptides with diverse structures, we obtain nM binders for 2 and μM binders for the other while only testing 6 binders per target.
3. Section 2.4: We generate and test 5 designs for binding the disordered region of NPM1 and obtain evidence of de-novo designs binding disordered proteins in live cells.
4. Section 2.5: When designing linear peptides to bind RagC we obtain 7 binders after testing 29 with the highest affinity being 3.5 μM.
5. Section 2.6: Similarly, designing disulfide-bonded peptides to bind the RagA:RagC dimer yields binders for 14 of 28 tested designs.
6. Section 2.7: We find hits when testing 7 nanobody designs against each of 2 novel targets in a yeast display assay.
7. Section 2.8: We obtain weak binders against two small molecules.
8. Section 2.9: Our campaign to design antimicrobial peptides binding to GyrA results in 19.5% of designs inhibiting cell growth by more than 4×.
9. Section 2.10: When testing at most 20 designed proteins and 15 nanobodies against 5 benchmark targets we obtain nM binders against 80% of them with both modalities.

Open Source Release. We release training code, inference code, model weights, and all designs under the MIT License: `https://github.com/HannesStark/boltzgen`. The design pipeline is freely

available, with easy-to-use interface to specify a binder-design problem and run BoltzGen, producing a filtered, ranked, diversity-optimized set of designs ready for experimental validation. We hope that fully open-sourcing the project puts state-of-the-art biomolecular design capabilities in the hands of any researcher and enables the community to build upon BoltzGen or contribute to BoltzGen-2.

2 Wetlab Results Summary

This section provides a summary of the wetlab results. Each subsection contains a figure that illustrates the best designs against each target. For the same set of experiments, more detailed descriptions are in Section 4, and their wetlab methodology is laid out in Appendix D. Unless mentioned otherwise, we provide the structure of the targets as input to BoltzGen.

2.1 Interpreting Affinities and Expression Numbers

Expression. To test a designed binder, one typically first produces the DNA that encodes the design. If the DNA is introduced into an environment with molecular machinery that translates DNA into proteins, the protein is produced and the design is being *expressed*. Expression can fail for various reasons. For instance, a protein could fail to fold as intended, or a design could contain a large hydrophobic patch that binds to itself, causing aggregation (the proteins "clump up"). Usually, more stable and more soluble proteins have a higher chance of expressing well.

Affinity. Binding affinity describes how tightly two molecules stick to each other. It is often quantified via their dissociation constant K_d, defined as the binder's *off-rate* (how often they fall apart) divided by its *on-rate* (how often they come together). A smaller K_d indicates that the molecules stay bound longer and interact more strongly. Natural protein–protein interactions cover a broad range of affinities: transient signaling complexes typically bind in the μM range, whereas stable complexes such as enzyme–inhibitor pairs or antibody–antigen assemblies can reach nM affinities. In contrast, therapeutic binders, such as monoclonal antibodies, engineered nanobodies, and peptide drugs, are often optimized to achieve tighter binding. Antibodies and nanobodies frequently reach picomolar or low-nanomolar affinities, while therapeutic peptides typically bind in the nanomolar range, depending on their size and conformational rigidity. Representative affinities for selected therapeutic binders are summarized in Table 1. Importantly, a high affinity is only the first step toward an effective therapeutic. It indicates that a molecule can recognize and stably engage its target, but not whether it will reach the target in the body, remain stable, avoid off-target interactions, or produce the desired biological effect.

Drug	Target	K_d (nM)
Caplacizumab	vWF	8.5
Brolucizumab	VEGF-A	0.03
Ozoralizumab	TNFα	0.02
Degarelix	GnRH-R	1.68
Desmopressin	AVPR2	0.76
Tirzepatide	GIPR	0.13

Table 1: Reported binding affinities (K_d) of therapeutic antibodies and peptides.

2.2 Designing Nanobodies and Proteins against 9 Novel Targets

Experiments carried out by Adaptyv Bio.

Targets. The majority of prior experimental validation of binder design models is carried out on targets that appear in their training data in complex with existing binders. In contrast to this, we choose 9 targets that are highly dissimilar to any other protein in PDB with an existing binder. For all 9 targets, we enforce that they are monomers and that there is no protein appearing in a bound structure in PDB with a sequence identity greater than 30%. Thus, it is possible that some of the targets do not even have a surface patch that allows for high-affinity protein-protein or nanobody-protein binding. We detail the potential therapeutic relevance of the targets in Section 4.1.

We evaluate BoltzGen's ability to design both nanobodies and general proteins against these targets. Designing high-affinity nanobodies is generally more challenging, as it involves additional structural constraints that restrict the diversity of sequences and structures the model can generate. Nevertheless,

nanobodies are often preferred as therapeutic modalities due to their favorable developability characteristics, including high solubility, robust expression yields, low aggregation propensity, good thermal and chemical stability, and ease of engineering [Jovčevska and Muyldermans, 2020].

Nanobody Designs. We use BoltzGen to generate 60 000 nanobodies against each of the 9 targets *without* specifying any binding site. BoltzGen randomly samples from its 4 default nanobody scaffolds for each design (see Figure 9). Given a selected scaffold, we fix the structure and sequence of the framework region, but replace the 3 CDR regions with loops of random length.

Protein Designs. We use BoltzGen to generate 60 000 proteins of lengths 80-140 against each of the 9 targets *without* specifying any binding site.

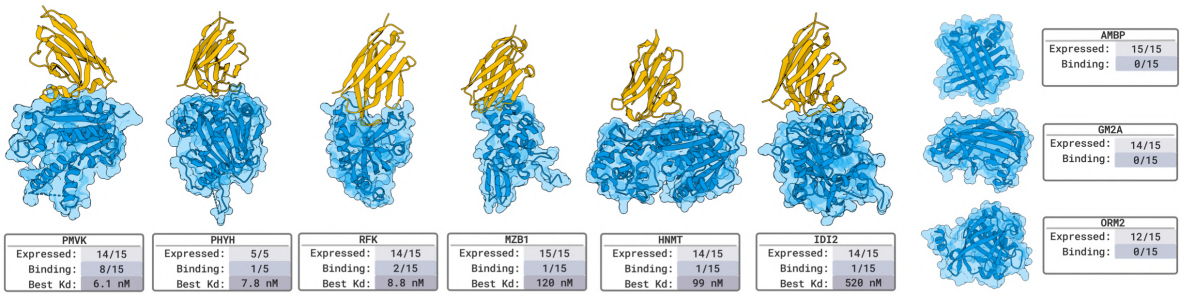


Figure 2: **Nanobody binders for 9 novel targets.**

Nanobody Design Results in Figure 2. For each target, we select 15 nanobodies for experimental validation. Surface plasmon resonance (SPR) and biolayer interferometry (BLI) affinity assays confirm that we obtain nM-affinity binders for 6 out of 9 targets. This represents a 66% success rate against *novel* targets, none of which have similar proteins in a bound context in all of PDB.

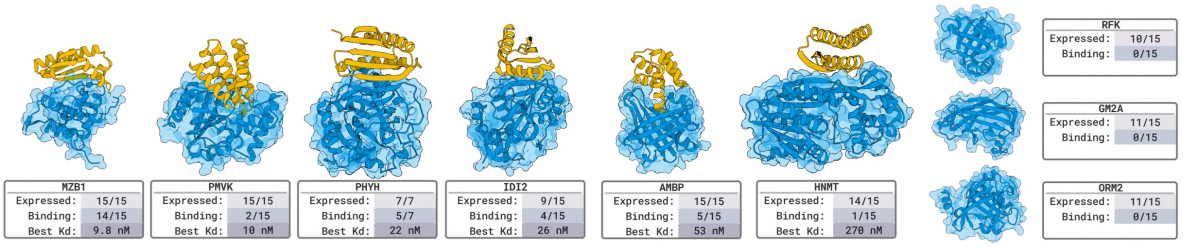


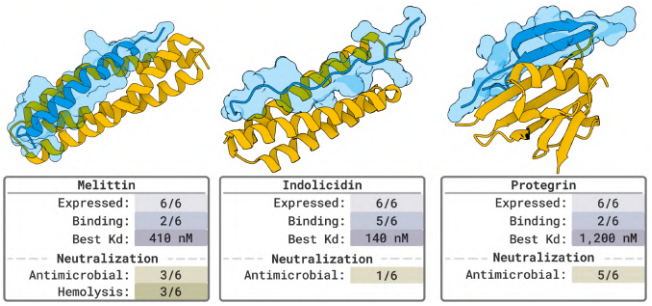
Figure 3: **Protein binders for 9 novel targets.**

Protein Design Results in Figure 3. For each target, we evaluate a set of 15 designed protein binders. Using SPR and BLI, we detect nM binders for 6 out of 9 targets. These results represent a 66% success rate on *novel* targets without any similar protein in all of PDB that is in a bound structure.

2.3 Designing Proteins to Bind Bioactive Peptides with Diverse Structures

Experiments by A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, and William F. DeGrado.

Targets. We design proteins to bind to three antimicrobial peptides and cytotoxic peptides as a class of biologically important compounds. We targeted multiple structural classes, including: 1) protegrin, a disulfide-rich beta hairpin; 2) melittin, which is intrinsically unfolded in dilute aqueous solution but forms a helix when bound to membranes; 3) indolicidin, which forms a polyproline II or amphipathic conformation in the presence of bilayers.



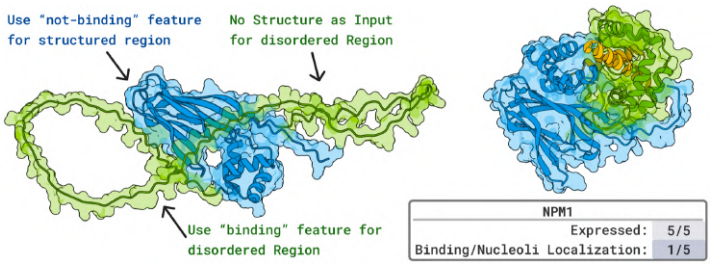
Designs and Results. All selected designs were first screened for peptide binding in vitro via changes in intrinsic tryptophan fluorescence and/or by surface plasmon resonance (details in 4.2). As the target

peptides are cytotoxic and antimicrobial peptides, we also assessed binders for their ability to neutralize antimicrobial activity and, where relevant, hemolysis. For each target, at least one binder design had single-digit μM affinity (for protegrin) or nM affinity (for indolicidin and melittin) and neutralized antimicrobial and, for melittin, hemolytic activity.

2.4 Designing Peptides to Bind the Disordered Region of NPM1.

Experiments by Yaotian Zhang, and Denes Hnisz.

Target. The NPM1-c mutant of NPM1 is a known driver of Acute Myeloid Leukemia. We aimed to design peptides that bind to the disordered region of NPM1. NPM1 is particularly appealing as a target due to its intrinsic disorder and cellular localization: it naturally accumulates in the nucleoli, and peptides that bind to it are expected to co-localize there. Therefore, nucleolar localization of a designed peptide can be a proxy for assessing its binding to NPM1 in live cells.



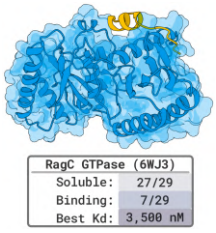
Designs. We generate 20 000 peptide designs, each 40-80 residues in length, targeting the disordered region of NPM1. To guide the design process, we leverage BoltzGen's binding site conditioning feature, explicitly directing the model to target the disordered region while avoiding interaction with the structured β-sheet region through its "not-binding" constraint. Additionally, we provide the structure of the ordered region and leave the disordered region flexible. Thus, tasking BoltzGen to model how the disordered region will fold and become structured in the presence of the designed peptide that is simultaneously being designed.

2.5 Designing Peptides to Bind a Specific Site of RagC GTPase

Experiments by Shamayeeta Ray, Jonathan T. Goldstein, and David M. Sabatini.

Target. RagC GTPase is a central part of a cellular pathway for sensing nutrients and regulating cell growth, protein synthesis, and other metabolic processes. There are no existing peptide binders against RagC.

Designs and Results. With one of RagC's interaction surfaces as binding-site input for BoltzGen, we generate 10 000 ranked designs of length 5-20. We test 29 in a binding affinity assay (SPR), and find 7 binders. The highest affinity is 3.5 μM and the second highest 60 μM.

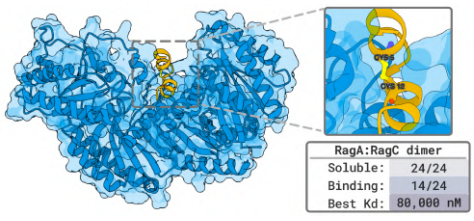


2.6 Designing Disulfide Bonded Cyclic Peptides to Bind a Specific Site of RagA:RagC

Experiments by Shamayeeta Ray, Jonathan T. Goldstein, and David M. Sabatini.

Target. The RagA:RagC dimer is part of a cellular pathway responsible for sensing nutrients and regulating cell growth, protein synthesis, and other metabolic processes. There are no existing peptide binders against the RagA:RagC dimer.

Designs. We use BoltzGen to design 50 000 ranked disulfide-cyclized peptides of size 10-18 against the RagA:RagC dimer with one of its interaction surfaces specified as binding site. The aim of introducing a disulfide bond between two residues of the peptide is to stabilize it and reduce its flexibility (rigidity reduces entropy loss during binding, thus potentially aiding stronger binding). To achieve this with BoltzGen, we specify the design to contain two cysteines that are covalently bonded. The cysteines are separated by six designed residues, with an additional one to five designed residues flanking either side of this eight-residue segment.



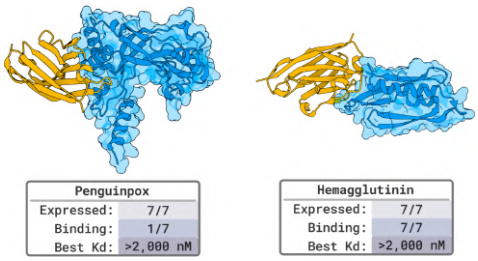
Results. We test 24 designs in a binding affinity assay and find 14 binders. For 8 of those, we resolved their affinities (SPR) and obtained 80 μM as the highest 164 μM as the second highest affinity.

Results. We experimentally test the top five highest-ranked designs in live human cells using fluorescence readouts based on GFP attached to the designs (see Section 4.3 for details). One design reliably localized to the nucleoli, suggesting successful binding to NPM1. This, we obtain a *de-novo* designed protein with *in vivo* evidence of binding to a disordered protein. Importantly, this in-cell assay provides insight beyond binding affinity. It also reflects functional viability, including the design's selectivity for NPM1 over potential off-targets that do not localize to the nucleoli. However, this experiment does not definitely confirm that the binding occurs specifically at the disordered region - the main evidence for this comes from the BoltzGen-generated structure and structure predictions within the BoltzGen pipeline.

2.7 Designing Nanobodies that Bind Penguinpox and Hemagglutinin

Experiments by Jacob A. Hambalek, Anshika Gupta, Diego Taquiri Diaz, and Chang C. Liu.

Targets. We choose two monomer targets that were recently deposited in PDB. The first target is the cyclic GMP-AMP phosphodiesterase of Penguinpox (cGAMP PDE), a protein known to inhibit host STING signaling by degrading cyclic dinucleotides [Hobbs et al., 2024]. The second is Filamentous Hemagglutinin (FhaB), an adhesion protein expressed by the pathogen Bordetella, which allows the pathogen to colonize and infect hosts [Costello et al., 2025].

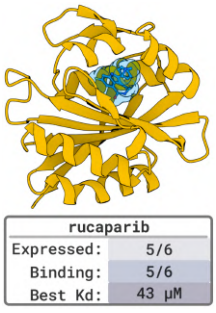


Designs and Results. We generate 60 000 nanobodies against each target in the fashion described in Section 2.2 and select 7 per target for experimental characterization. A yeast surface display assay shows binding signal for a nanobody to bind Penguinpox and for 7 to bind Hemagglutinin. The assay does not allow for computing a binding affinity but indicates that it is at best 2 μM. We note that we carried out similar experiments on a set of designs from a previous version of BoltzGen that suffered from a serious flaw resulting in close-to-random ranking and filtering (Details in C). For those designs no hits were found.

2.8 Designing Proteins that Bind to Small Molecules

Experiments by Nam Hyeong Kim and William F. DeGrado.

Targets. We evaluate BoltzGen's ability to design binders against two small molecules: rucaparib and a rhodamine derivative. Binders to these targets could serve as components in biosensors, delivery systems, or detoxification agents.



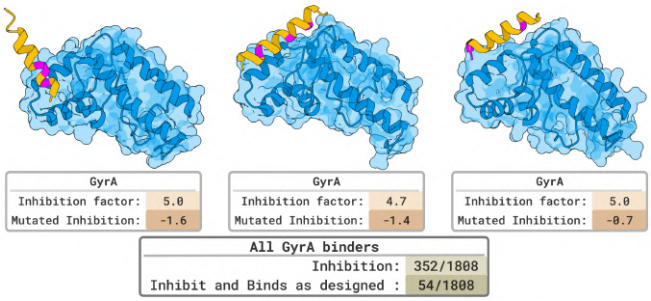
Designs and Results. We generate 10 000 protein designs targeting rucaparib and 20 000 targeting the rhodamine derivative with design lengths ranging from 140 to 180 residues. We select six designs against rucaparib for experimental validation, five of which show binding with affinities between 50 and 150 μM. For the rhodamine derivative, four designs were tested experimentally, all showing weak binding with affinities between 30 and 250 μM.

Previous computational work [Lu et al., 2024] designed a low-nanomolar binder to rucaparib through a specialized, expert-guided approach that involved identifying specific chemical groups on the small molecule. In contrast, our work demonstrates that a general-purpose design model, BoltzGen, can generate diverse scaffolds with moderate binding affinity and simpler customization.

2.9 Designing Antimicrobial Peptides that Inhibit the GyrA to GyrA Interaction

Experiments by Andrew Savinov, and Gene-Wei Li.

Target. We used BoltzGen to design inhibitory peptides of the essential bacterial protein DNA gyrase subunit A (GyrA), a target of interest for developing antibiotics. For its function, GyrA needs to interact with a copy of itself. When disrupting this interaction in bacteria, they die.



Designs and Results. We specify the surface where GyrA interacts with a copy of itself as the binding site when generating peptides of length 10-50. We select 1 808 designs for experimental validation in a growth inhibition assay. Of these, 352 (19.5%) were found to inhibit *E. coli* growth by more than 4×. In a second experiment, we replace the design's three closest residues to the target (pink in the Figure) with alanines to verify whether they bind as intended. 54 (3.0% of total) of the growth inhibitors lose their activity after introducing these mutations. The inhibitory effects of the 54 successful designs were, in most cases, strong enough to completely eliminate the cell populations in which they were expressed.

2.10 Designing Nanobodies and Proteins against 5 Benchmark Targets

Experiments carried out by Adaptyv Bio.

Targets. We designed binders against PD-L1, TNFα, PDGFR, IL-7Rα, and InsulinR, which were considered in previous work [Chai-Discovery et al., 2025, Zambaldi et al., 2024]. All these targets have known binders that are included in the training data of most previous models and in our training data.

1. **TNFα** There are 18 matching PDB complexes, i.e. entries with more than 1 protein entity and with over 90% sequence identity to the structure used to design binders. See, for example, PDB 5M2I, released in 2017, showing TNFα in complex with picomolar-affinity nanobodies [Beirnaert et al., 2017].
2. **PD-L1** There are 37 matching PDB complexes. See, for example, PDB 5JDS, released in 2017, showing PD-L1 in complex with a 3.0-nanomolar-affinity nanobody [Zhang et al., 2017].
3. **PDGFR** The PDB entry 3MJG used for making designs, released in 2010, shows PDGFR in complex with its binding partner PDGF, and previous work has reported *de novo* mini protein binders with nanomolar affinities [Cao et al., 2022].
4. **IL-7Rα** There are 6 matching PDB complexes, such as PDB entry 6P50, released in 2019, which shows IL-7Rα in complex with a 1-nanomolar-affinity Fab [Hixon et al., 2020].
5. **InsulinR** There are 74 matching PDB complexes. The PDB entry 4ZXB used for design, released in 2016, shows the target in complex with four Fabs (83-7 and 83-14) [Croll et al., 2016].

Nanobody Designs. We generate 60 000 nanobodies (except for TNFα, where we only generate 30 000) against each of the 5 targets while specifying the binding sites listed in Chai-Discovery et al. [2025]. BoltzGen randomly chooses between its 4 default scaffolds for each design (see Figure 9). For a selected scaffold, we fix the structure and sequence of the framework region, but replace the 3 CDR regions with loops of random length.

Protein Designs. We use BoltzGen to generate 60 000 proteins (except for TNFα where we only generate 30 000) of lengths 80-120 against each of the 5 targets while specifying the binding sites listed in Chai-Discovery et al. [2025].

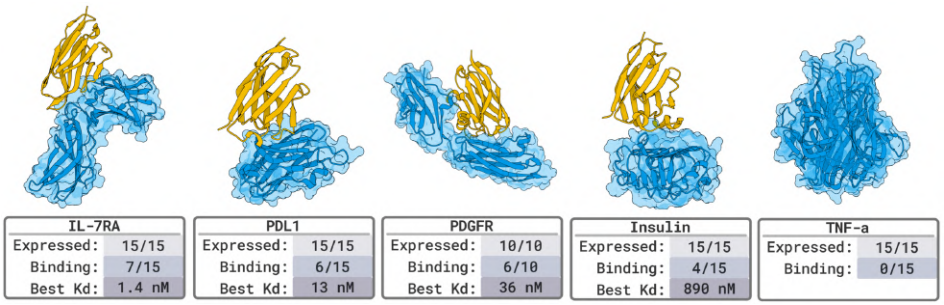


Figure 4: **Nanobody binders targeting 5 Benchmark Proteins.**

Nanobody Binder Results in Figure 4. For each target, we evaluate a set of 15 or fewer designed protein binders. SPR and BLI assays confirm nM-affinity binders for 4 out of 5 targets (an 80% success rate).

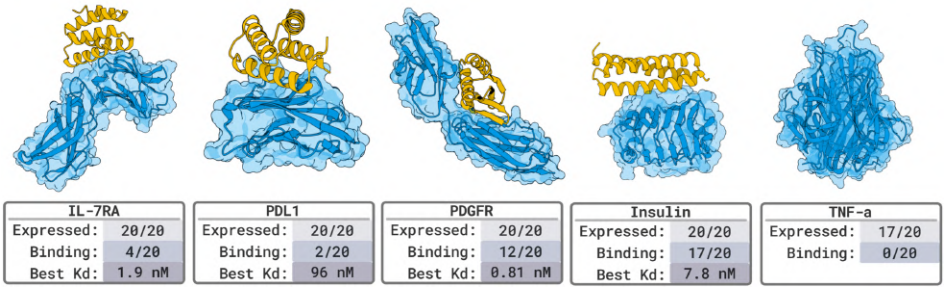


Figure 5: **Protein binders targeting 5 Benchmark Proteins.**

Protein Binder Results in Figure 5. We tested 20 designed protein binders per target. SPR and BLI assays identified binders for 4 out of 5 targets, including pM hits on PDGFR, yielding an 80% success rate. The missing target is TNFα for which previous work [Chai-Discovery et al., 2025] successfully designed.

We note that these benchmark targets include cases where high-affinity binders are already present in public datasets. In such settings, a model may succeed by reusing or recombining interaction motifs encountered during training, rather than by generating truly novel binding solutions. Consequently, results on these benchmark targets provide marginal evidence toward a model's ability to generalize to discovery scenarios involving targets without known binders.

3 Method

BoltzGen is a single all-atom diffusion model capable of performing both structure prediction and protein design. The model takes a set of molecular entities as input and outputs their all-atom three-dimensional structure. Molecular entities include small molecules, RNA, DNA, or protein sequences, along with any post-translational modifications and covalent bonds. New proteins are sampled by specifying *design residues*, for which the model generates both the all-atom structure and amino acid identities. Structure prediction and design capabilities can be exercised in tandem; for example, when generating a binder given only the sequence of the target, the model simultaneously folds the target and designs the binder's atomic structure, producing a bound complex.

Unified Design and Structure Prediction. The joint all-atom sequence and structure sampling ability of the model and its scalable training come from its purely geometry-based representation of designed amino-acid types. This representation encodes residue identities according to the position of the "left-over" atoms of side chains (see Figure 7). This change helps maintain a scalable architecture and its associated diffusion training process, similar to state-of-the-art biomolecular structure prediction methods.

Design Specification Language. Predictions can be conditioned on a broad set of additional inputs. These include standard annotations such as desired residue types and covalent bonds, as well as secondary structure identity, binding site location, and structure templates. The conditions can be

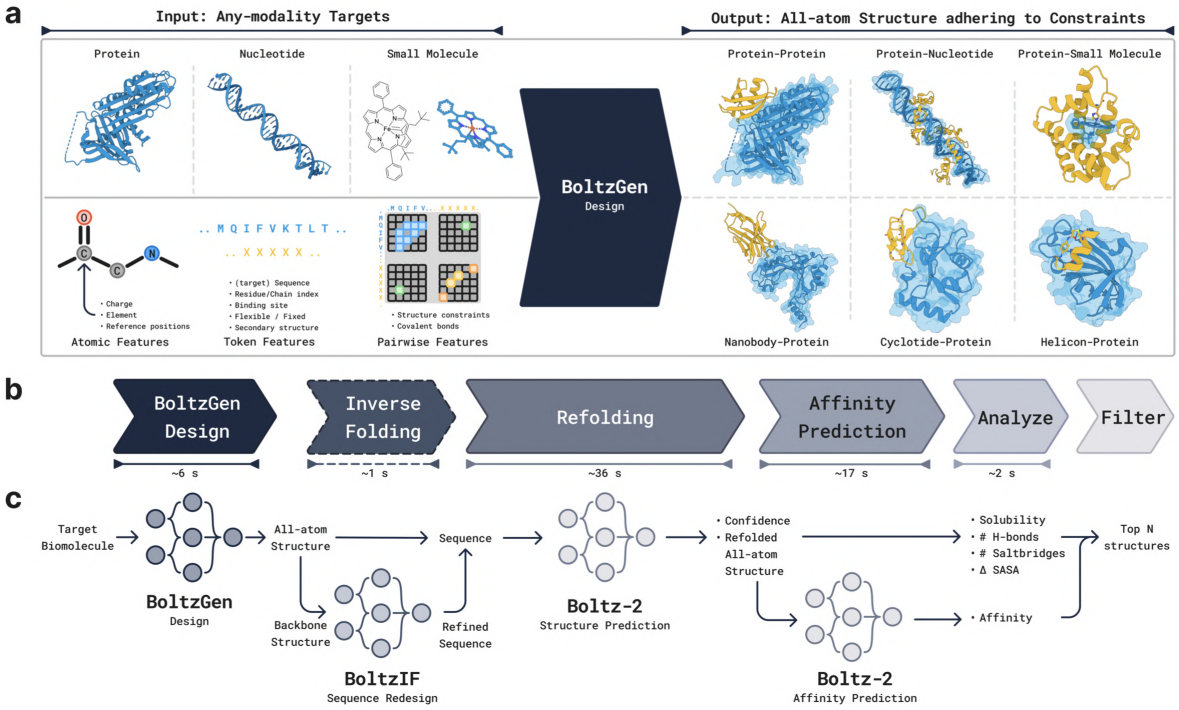


Figure 6: **Overview of BoltzGen Pipeline:** (**a**) The overall flow from target specification to binder generation. The generative model designs binders for arbitrary targets, including proteins, nucleic acids, and small molecules. It supports various constraints such as covalent bonds, structural motifs, and specific binding sites. The generated designs are passed through a filtering and ranking pipeline to produce a small, diverse set suitable for experimental validation. (**b**) Runtime per stage for a representative case with 200 total residues across the binder and target, illustrating the model's efficiency. (**c**) A detailed breakdown of each pipeline stage, showing intermediate inputs and outputs.

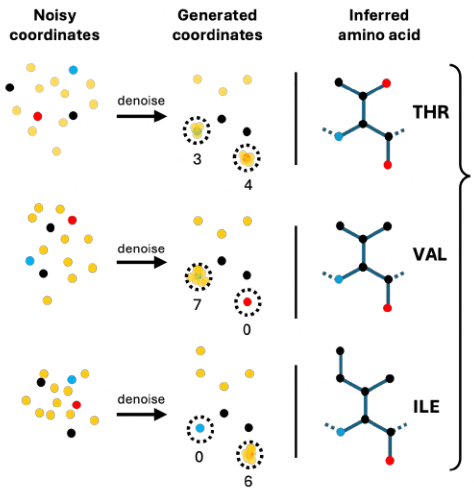
incorporated with the help of a rich design specification language, allowing us to support the needs of our experimental collaborators. As a result, we can address a wide range of design challenges, including diverse modalities such as nanobodies, cyclic peptides with various cyclizations, helicons, or cyclotides.

In addition to the core generative model, we provide a comprehensive design pipeline that includes: (1) the initial generation of candidate designs, (2) optional sequence redesign through inverse folding, (3) evaluation of refolding quality and, for small molecule targets, affinity estimation, (4) ranking and filtering of designs, and (5) selection of a final candidate pool with diversity optimization.

3.1 All-atom Generative Model Formulation

Model Representations. All non-designed molecular entities provided as input to the model are represented at the atomic level, including atom positions, element types, and charge. These atoms are grouped into tokens, such as residues for proteins or nucleotides for RNA and DNA. Designed residues, which do not have a specified amino acid type during generation, use a fixed-size representation consisting of exactly 14 atoms, some of which serve as *virtual* placeholders. Once the model determines the final residue types, these virtual atoms are discarded.

Geometric Encoding of Residue Type. Instead of generating a discrete residue label, the model encodes the residue identity through the geometry of the 14-atom representation (see Figure 7). Specifically, it learns to place the virtual atoms on top of designated backbone atoms to signal the intended residue type. The first four atoms in each designed residue are fixed as the backbone N, Cα, C, and O atoms, in that order. As a result, residue types can be inferred directly from the generated structure by counting how many atoms are placed within 0.5 Å of each backbone atom. Any remaining atoms that are not superposed onto the backbone are interpreted as the side chain. For example, proline



RES	1	2	3	4	5	6	7	8	9	10	11	12	13	14	#N	#O
GLY	N	CA	C	O											0	10
ALA	N	CA	C	O	CB										0	9
CYS	N	CA	C	O	CB	SG									0	8
SER	N	CA	C	O	CB	OG									8	0
PRO	N	CA	C	O	CB	CG	CD								0	7
THR	**N**	**CA**	**C**	**O**	**CB**	**OG1**	**CG2**								**3**	**4**
VAL	**N**	**CA**	**C**	**O**	**CB**	**CG1**	**CG2**								**7**	**0**
ILE	**N**	**CA**	**C**	**O**	**CB**	**CG1**	**CG2**	**CD1**							**0**	**6**
ASN	N	CA	C	O	CB	CG	OD1	ND2							1	5
ASP	N	CA	C	O	CB	CG	OD1	OD2							2	4
LEU	N	CA	C	O	CB	CG	CD1	CD2							4	2
MET	N	CA	C	O	CB	CG	SD	CE							6	0
GLN	N	CA	C	O	CB	CG	CD	OE1	NE2						0	5
GLU	N	CA	C	O	CB	CG	CD	OE1	OE2						2	3
LYS	N	CA	C	O	CB	CG	CD	CE	NZ						5	0
HIS	N	CA	C	O	CB	CG	ND1	CD2	CE1	NE2					0	4
PHE	N	CA	C	O	CB	CG	CD1	CD2	CE1	CE2	CZ				0	3
ARG	N	CA	C	O	CB	CG	CD	NE	CZ	NH1	NH2				3	0
TYR	N	CA	C	O	CB	CG	CD1	CD2	CE1	CE2	CZ	OH			0	2
TRP	N	CA	C	O	CB	CG	CD1	CD2	NE1	CE2	CE3	CZ2	CZ3	CH2	0	0

Figure 7: **Residue Type Encoding** In BoltzGen, each designed residue is represented using 14 atoms. To determine the residue type, the model is trained to superpose a subset of these atoms onto specific backbone atoms. These auxiliary atoms act as markers and are discarded after decoding. For example, placing three atoms on the backbone nitrogen and four on the backbone oxygen is interpreted as a threonine. The atoms in positions 5, 6, and 7 are then assigned as the threonine CB, OG1, and CG2 atoms, respectively.

is encoded by placing 7 atoms on the backbone oxygen, while threonine is represented by placing 3 atoms on the nitrogen and 4 on the oxygen.

This geometric encoding allows the model to operate entirely in a continuous space, avoiding the need to mix discrete and continuous representations. As a result, it enables efficient joint training for both structure prediction and design tasks.

Diffusion Process. Because the model operates on a continuous space, we can use the same diffusion process as Boltz-2 [Passaro et al., 2025]. The only difference is that the data samples now contain additional virtual atoms for the designed residues.

Formally, let $X \sim p_{\text{data}}$, $X \in \mathbb{R}^{N\times 3}$ the 3D atomic coordinates of a training sample and let X_t follow the forward diffusion process

$$dX_t = \sqrt{2t}dB_t, \tag{1}$$

with initial condition $X_0 \sim p_{\text{data}}$. For large T, X_T will be approximately Gaussian with variance T^2. This process can be reversed to obtain samples from the data distribution starting from Gaussian noise. Reversing the diffusion process requires a denoiser $D_\theta(x, t; z)$, parameterized by learnable weights θ. The goal of the denoiser is to approximate the posterior mean

$$D_\theta(x, t; z) \approx \mu_t(x) = \mathbb{E}_{X_0|X_t=x}[X_0], \tag{2}$$

conditioned on trunk features z. The model is trained using a standard denoising loss

$$\mathcal{L}(\theta) = \mathbb{E}_t\mathbb{E}_{X_0,X_t}\left[w(t) \cdot \|D_\theta(X_t, t; z) - X_0\|^2\right], \tag{3}$$

where $w(t)$ is a weighting function. In the absence of parametric constraints, the minimizer of this loss is μ_t, the posterior means.

3.2 Architecture

The model preserves the Boltz-2 architecture with some modifications to include additional conditioning inputs, as shown in Fig. 8. The model is split into two parts. The larger *Trunk* produces token and pairwise representations used to condition the *Diffusion Module*, which generates the structure. The trunk is only run once, while the diffusion module is run many times to progressively denoise the 3D coordinates of all atoms.

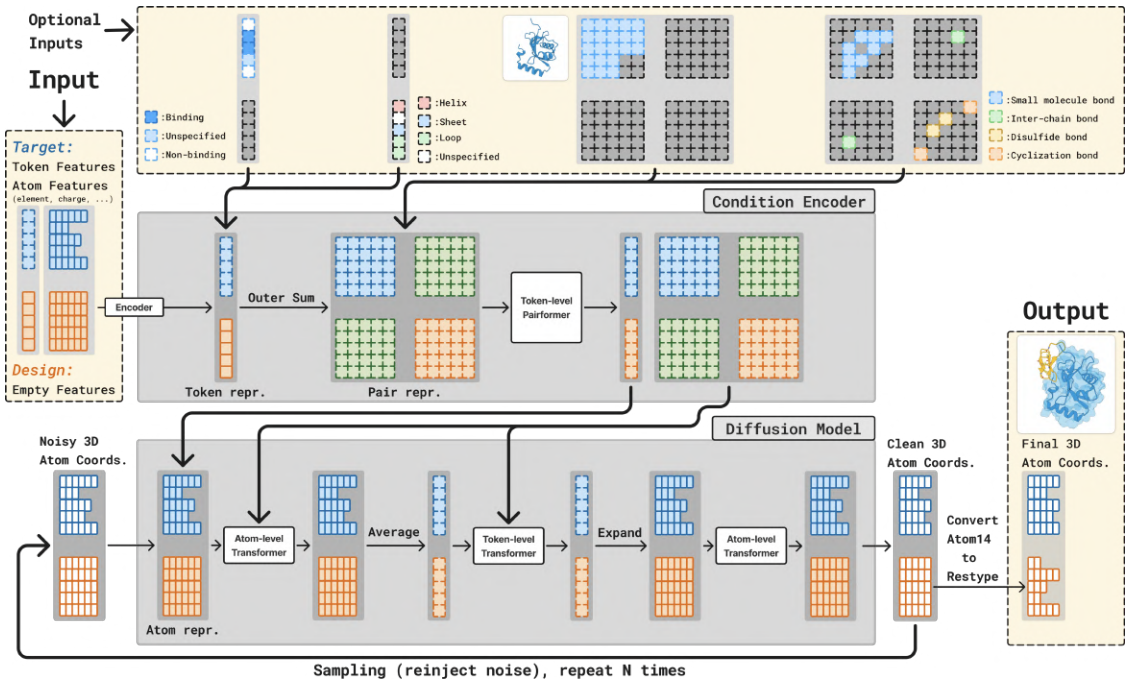


Figure 8: **Model Architecture** The architecture preserves the main components of the AlphaFold3 and Boltz-2 architectures, including the condition encoder (trunk) and diffusion model. The main difference lies in the inclusion of design tokens and additional inputs such as binding site and target structure.

Trunk. The trunk operates on tokenized structures, where proteins are tokenized into amino acids, RNA and DNA into nucleotides, and small molecules into atoms. Each token consists of a group of atoms with associated features such as charge and element type, along with token-level attributes including residue index, amino acid type, and a flag indicating whether the residue is designed. All features are encoded into vector representations. Atom-level embeddings are averaged to produce token-level vectors, and a pair representation is constructed using an outer-sum of the token embeddings. Both token and pair representations are then passed through a PairFormer stack. Each PairFormer block includes triangle multiplicative and triangle attention layers that update the pair representations, along with a transformer layer that updates the token representations using the pair features as the attention bias.

Diffusion Module. The diffusion module takes noisy 3D atomic coordinates as input and predicts denoised coordinates. It uses a standard transformer architecture that operates on both atom and token levels, consisting of 3 atom-level layers, followed by 24 token-level layers, and concluding with another 3 atom-level layers. The atom-level layers utilize sequence-local attention. Transitions between atom and token levels are handled by averaging or expanding the representations. Conditioning information from the trunk is incorporated by adding expanded token-level features to the atom-level input representations, and by biasing the attention layers based on the pair representation.

As in Boltz-2, AlphaFold3 and EDM [Karras et al., 2022], the diffusion module is preconditioned to parameterize the denoiser D_θ,

$$D_\theta(x, t; z) = \frac{\sigma_{\text{data}}^2}{\sigma_{\text{data}}^2 + t^2} x + \frac{t \cdot \sigma_{\text{data}}}{\sqrt{\sigma_{\text{data}}^2 + t^2}} F_\theta \left(\frac{x}{\sqrt{\sigma_{\text{data}}^2 + t^2}}, \frac{1}{4} \log(t); z \right), \tag{4}$$

where F_θ is the diffusion module. This parameterization is derived in [Karras et al., 2022] so that (1) the inputs to the network F_θ have unit variance (2) the effective training target of F_θ has unit variance and (3) the scalar multiplier of F_θ in Eqn. 4 is minimal.

Controllability. Several additional inputs can optionally be provided to BoltzGen to steer the generation process according to user-specified requirements. Figure 8 shows where these inputs are integrated into the architecture, and Figure 9 illustrates the resulting expressive design specification language.

1. **Covalent Bonds.** Covalent bonds can be specified between individual atoms, in which case the identity of the residues that contain the bound atoms must be specified.
2. **Structure Conditioning.** Parts of the structure can be specified to the model via pairwise distances, for example to perform motif scaffolding. All structures are specified in *structure groups*, where all pairwise distances between atoms in the same structure group are fixed, but not across groups. For example, during nanobody design, the nanobody framework structure and the target structure can be fixed but their relative positions to each other can be left unconstrained by putting them in different structure groups.
3. **Binding Site.** Residues can be specified as binding, or not-binding. The model will try to place designed residues close to binding residues and away from non-binding residues.
4. **Secondary Structure.** Designed residues can also be specified to be part of alpha-helices, beta-sheets, or coils.

Covalent bonds, specified as a matrix of pairwise bonds, and pairwise distances for structure conditioning are both encoded and added to the trunk's input pair representation. Binding and secondary structure labels are incorporated into the input token representation.

These conditioning options can be used to control the model and address a variety of design tasks. For example,

1. **Cyclic Peptides** can be designed by specifying a covalent bond. This includes disulfide-stapled peptides, head-to-tail cyclic peptides, and any other type of cyclization.
2. **Helicons** can be designed by including a staple molecule in the model's input and enforcing covalent bonds between the staple and the sulfurs of two cysteines in the design.
3. **Nanobodies** can be designed by enforcing the design to adhere to a given template, allowing the model to generate the CDRs as well as place the scaffold in relation to the target.

Figure 9 also illustrates how to realize these examples using the BoltzGen design interface.

3.3 Training

The model is trained with a diffusion objective with a mixture of experimental and self-distilled biomolecular structures. This data is then randomly cropped and employed in a diverse set of tasks by randomly selecting parts of the structure to be designed or conditioned on. The procedure is described in Figure 10. This conditioning sampling process, combined with the diffusion objective, supervises BoltzGen to simultaneously learn folding, binder design, motif scaffolding, and more, resulting in a universal binder design model that maximally extracts information from the data.

Training Data. Our data pipeline mostly retains the datasets used in Boltz-2 [Passaro et al., 2025], while adapting the sampling procedure for the task of biomolecular design. Specifically, we use experimental structures from the Protein Data Bank (PDB) [Berman et al., 2000], as well as self-distilled structures from AlphaFold2 (AFDB), as well as Boltz-1 (for protein-ligand, RNA, and DNA-protein structures) (see Appendix A.1 for dataset details). Our data also differs from Boltz-2 by not including the upsampled antibody and TCR datasets, since including them reduces generation diversity.

Cropping. The crop size for folding is up to 768 residues, as done in [Abramson et al., 2024, Passaro et al., 2025]. Crop size for generative tasks (binder design, motif scaffolding, and unconditional design) is reduced to 512 residues, to accommodate augmented fake atom representations (Appendix A.2).

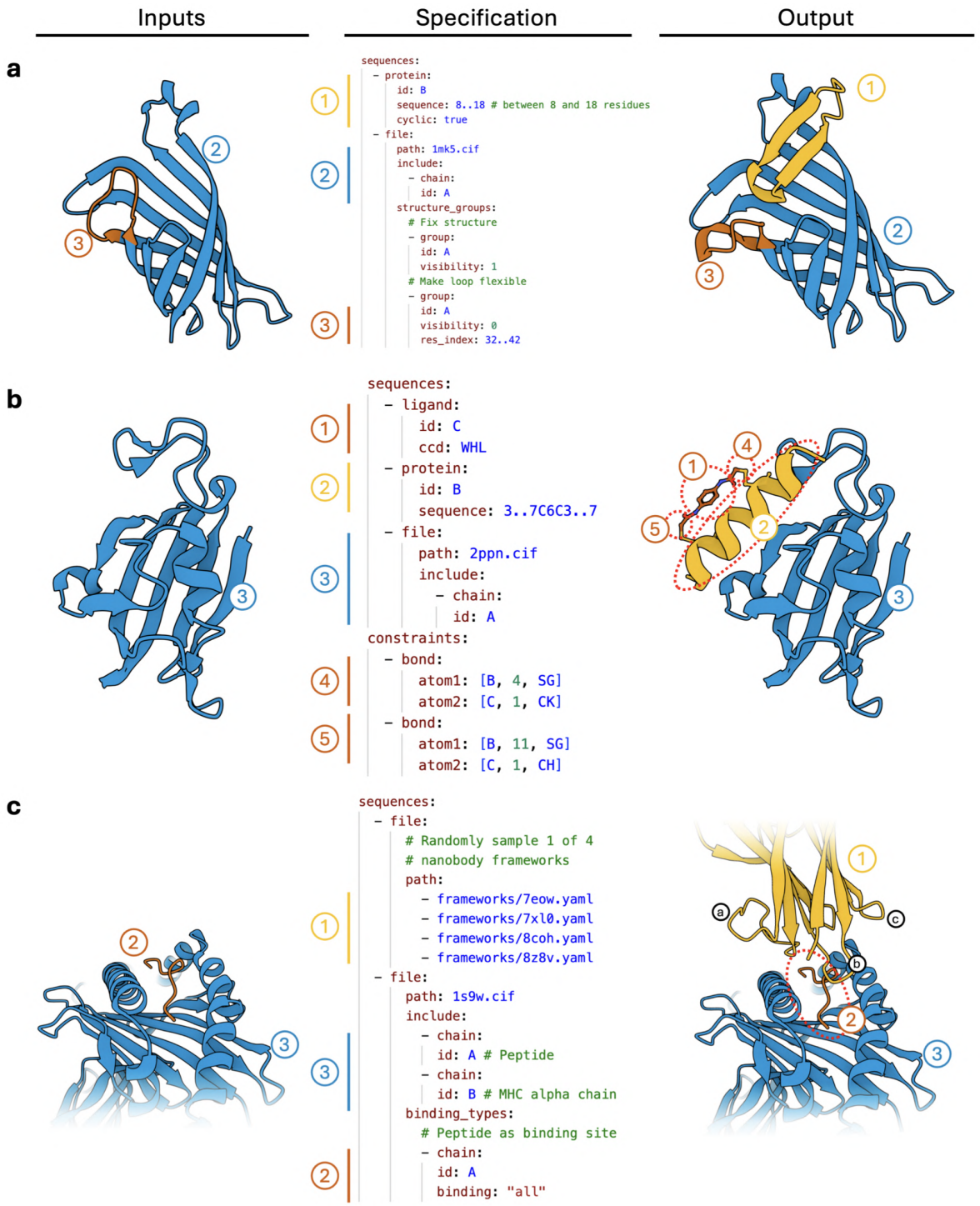


Figure 9: **Design Specification Language.** 3 examples of how the BoltzGen design specification can be used to solve different design tasks. (**a**) Designing a cyclic peptide against streptavidin. Part of the target structure is left flexible (3, orange) and is predicted to change conformation upon binding the design (1, yellow). (**b**) Designing a helicon binder. The helicon is created by including the staple molecule (1, WHL) and specifying covalent bonds between two cysteines in the design and the staple (4,5, orange). (**c**) Designing a nanobody against a peptide-MHC complex. The peptide (2, orange) is specified as a binding site, and the designed regions are limited to the 3 CDR loops (a,b,c) of the nanobody (1, yellow). The nanobody frameworks are themselves yaml files that specify the PDB structure to use and which are parts to design. With multiple specifications, a random one is sampled for each design.

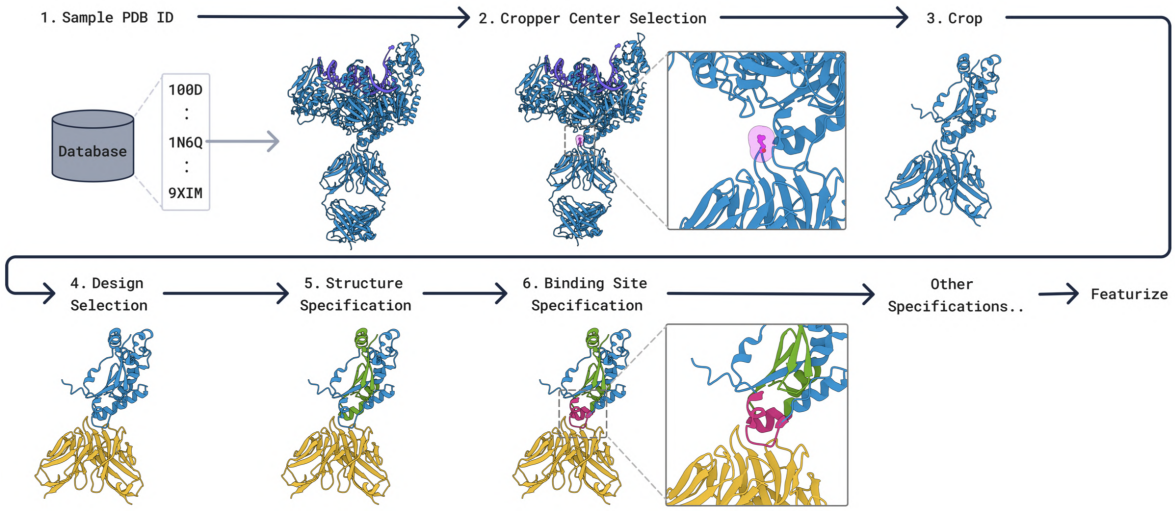


Figure 10: **Crop, Selection, and Specification** During training, we sample one PDB structure from the database and crop it using our cropping algorithm 4. On top of this cropped structure, we optionally select chains that will be designed (yellow). Diverse conditions, such as fixed target substructures (green), binding sites (red), and more, are also optionally specified. Cropped, selected, and specified structure is then featurized to be supplied to the model.

Diffusion Objective. The loss used to train the model is

$$\mathcal{L}(\theta) = \mathbb{E}_{\substack{X\sim p_{\text{data}} \\ t\sim p_{\text{noise}} \\ \epsilon\sim\mathcal{N}}} \left[w(t) \left(\mathcal{L}_{\text{MSE}}(\theta; X, t, \epsilon) + \mathcal{L}_{\text{bond}}(\theta; X, t, \epsilon) \right) + \mathcal{L}_{\text{smooth\_lDDT}}(\theta; X, t, \epsilon) \right], \tag{5}$$

where p_{noise} is described below, ϵ is standard isotropic Gaussian noise, σ^2_{data} is the variance of the data, $\mathcal{L}_{\text{MSE}}, \mathcal{L}_{\text{bond}}$, and $\mathcal{L}_{\text{smooth\_lDDT}}$ are the three components of the loss described below and the weighting is defined as

$$w(t) = \frac{t^2 + \sigma^2_{\text{data}}}{(t \cdot \sigma_{\text{data}})^2}. \tag{6}$$

Let $\hat{X} = D_\theta(X + t\epsilon, t; z)$ be the output of the denoiser. The MSE loss is a weighted version of the denoising loss including a rigid alignment of the denoiser output and target,

$$\mathcal{L}_{\text{MSE}}(\theta; X, t, \epsilon) = \frac{1}{3} \sum_l w_l \|\hat{X}_l - X^{\text{aligned}}_l\|^2_w, \tag{7}$$

where l iterates over atoms in the structure, X^{aligned} is rigidly aligned to $\hat{X}$, and w_l upweighs nucleotide and ligand atoms,

$$w_l = \begin{cases} 1 & \text{if protein} \\ 6 & \text{if DNA or RNA} \\ 11 & \text{if ligand} \end{cases}. \tag{8}$$

The bond loss encourages bond length correctness,

$$\mathcal{L}_{\text{bond}}(\theta; X, t, \epsilon) = \frac{1}{|\mathcal{B}|} \sum_{l,m\in\mathcal{B}} \left(\|\hat{X}_l - \hat{X}_m\| - \|X^{\text{aligned}}_l - X^{\text{aligned}}_m\| \right)^2, \tag{9}$$

Finally, $\mathcal{L}_{\text{smooth\_lDDT}}$ is described in [Abramson et al., 2024] and is a smooth version of the lDDT which can be directly optimized.

The noise sampling distribution p_{noise} used during training is

$$\sigma_{\text{data}} \cdot \exp(-1.2 + 1.5 \cdot \mathcal{N}(0, 1)). \tag{10}$$

Training Tasks. In order to train the model to use the various forms of conditioning above, we construct a number of tasks from our training data, which we describe below (Appendix A.3 for more details). Each task is distinguished by which parts of a cropped structure are chosen to be designed. These selected parts are represented using the fixed-size representation with virtual atoms described above, and their residue and atom types are masked.

1. **Folding.** No design residues are specified, corresponding to structure prediction.
2. **Binder Design.** One protein chain in the target structure is chosen to be designed. This corresponds to binder design against whatever other biomolecules are in the target structure, which could be a small molecule, DNA, RNA, or another protein. Sometimes, only the interface of the protein chain with the rest of the structure is designed, corresponding to binder design with a supplied scaffold.
3. **Motif Scaffolding.** Either a crop of the target structure is chosen to be designed, or everything but the crop is designed. This corresponds to completing a scaffold in the first case, and scaffolding a motif in the second case.
4. **Unconditional Design.** Everything is chosen to be designed, corresponding to unconditional protein generation.

Each task is sampled with some probability during training as long as the target structure is appropriate for the task. For example, the binder design tasks are not sampled for structures that contain a single protein monomer.

In addition to the tasks, we also sample different conditioning to supply the model:

1. **Structure Conditioning.** We randomly choose portions of the target structure to supply as input to the model based on crops. These are also randomly grouped together.
2. **Binding Site.** We randomly annotate residues as being part of a binding site or not, based on proximity to any portions of the training structure selected to be designed.
3. **Secondary Structure.** We annotate random residues by their secondary structure.

See Appendix A.3 for full details on the various schemes we use to construct the tasks and sample the different conditioning inputs to give the model.

3.4 Generation

To sample from the model, the inputs are first passed through the trunk to obtain the token and pair representations z that condition the diffusion module. The diffusion module then generates a structure x using the stochastic sampler from [Karras et al., 2022]. This sampler starts from random noise and alternates between adding noise and denoising.

The sampler makes use of the probability flow ODE, which in the case of our forward process is,

$$x'(t) = -\frac{x - \mu_t(x)}{t} dt, \tag{11}$$

for $t > 0$, where μ_t is the posterior mean in Eqn. 2 that is approximated by the denoiser D_θ. Note that this ODE runs forward (towards noise). Let $F(x, t, s)$, be the associated flow map, i.e. if $x(t)$ is a solution to the above ODE, then

$$F(x(t), t, s) = x(s). \tag{12}$$

The probability flow ODE satisfies that $F(X_t, t, s)$ has the same marginal distribution as X_s. In particular, $F(X_t, t, 0)$ gives samples from the data distribution for any $t > 0$. Rather than simulate the ODE the whole way, the sampler in [Karras et al., 2022] interleaves noising steps to add stochasticity based on the observation that $X_t = X_s + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sqrt{t^2 - s^2})$ for $t > s$. Hence, alternating between adding noise ϵ and applying the flow map F always gives samples with the correct marginals. The

Algorithm 1: EDM Sampler

Input: Times t_i, $\gamma_i > 1$, $\alpha > 0$, $\beta > 0$, D_θ, z, T
Output: Sample trajectory x_i
Sample $x_0 \sim \mathcal{N}(0, T^2\mathbf{I})$;
foreach $i \in \{0, \ldots, N-1\}$ **do**
  Sample $\epsilon_i \sim \mathcal{N}(0, \mathbf{I})$;
  $\hat{t}_i \leftarrow t_i + \gamma_i t_i$;
  // 1) Step toward noise:
  $\hat{x}_i \leftarrow x_i + \beta\sqrt{\hat{t}_i^2 - t_i^2}\epsilon_i$;
  // 2) Step in ODE direction:
  $d_i \leftarrow (\hat{x}_i - D_\theta(\hat{x}_i, \hat{t}_i; z))/\hat{t}_i$;
  $x_{i+1} \leftarrow \hat{x}_i + \alpha(t_{i+1} - \hat{t}_i)d_i$;
end

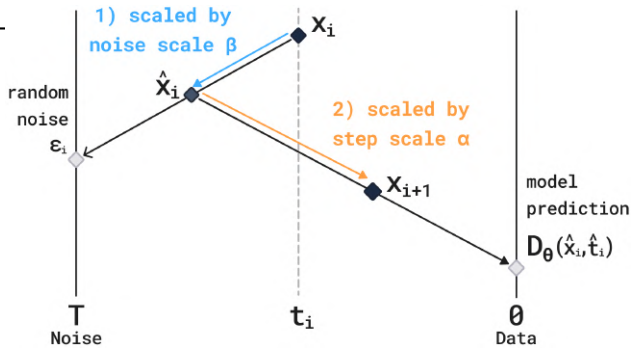


Figure 11: **The EDM Sampler** steps toward noise and then toward the model prediction in each denoising iteration. The magnitude with which to make these steps can be scaled by β (noise) and α (prediction). Using $\alpha \neq 1$, $\beta \neq 1$ no longer samples the training distribution, but is a heuristic scaling to sample more diverse (higher β, lower α) or more designable (lower β, higher α) proteins.

algorithm is given in Alg. 1 which additionally makes use of a step scale α and noise scale β. To sample more high-designability but lower diversity binder structures, we can increase the step scale or decrease the noise scale (and the opposite to obtain more diverse proteins). When drawing many samples, we vary the step scales and noise scales for each generated protein.

Dilated schedule. The time schedule t_i used in AlphaFold3 and Boltz is,

$$t_i = \sigma_{\text{data}} \cdot \left(s_{\max}^{1/p} + \tau_i \cdot (s_{\min}^{1/p} - s_{\max}^{1/p})\right)^p, \tag{13}$$

where $\sigma_{\text{data}} = 16, s_{\min} = 4 \cdot 10^{-4}, s_{\max} = 160, p = 7$, and $\{\tau_i\}_{i=1}^N \in [0, 1]$ is a sequence of steps. By default, $\tau_i = i/N$.

Because of our geometric encoding, we found that the amino acid types of designed residues were determined within a short window, approximately at $\tau_i \in [0.6, 0.8]$. In order to spend more function evaluations when generating the residue types, we therefore use a *dilated* schedule where the interval $[0.6, 0.8]$ is stretched out. Concretely, to dilate an interval $[\tau_s, \tau_e]$ by $1 < \lambda < 1/(\tau_e - \tau_s)$, we map step $\tau \in [0, 1]$ according to

$$\phi(\tau) = \begin{cases} \tau/r, & \text{if } \tau < l \\ (\tau - u)/r + u & \text{if } \tau > u \\ (\tau - l)/\lambda + l & \text{otherwise} \end{cases}, \tag{14}$$

where $r = (1 - \lambda \cdot (\tau_e - \tau_s))/(1 - (\tau_e - \tau_s))$, $l = r \cdot \tau_s$, $u = l + \lambda(\tau_e - \tau_s)$. In practice, we choose $\lambda = 8/3$, $\tau_s = 0.6$, and $\tau_e = 0.8$, with $N = 300$ total function evaluations which we found to work well in experiments.

3.5 BoltzGen Pipeline

On top of the generative model, we run a computational pipeline to filter from the potentially thousands of designs sampled by the model down to the most promising candidates. The pipeline consists of the following stages.

1. **BoltzGen Diffusion (GPU).** Given a design specification, we generate a large number of designed binders with BoltzGen.
2. **Inverse Folding (GPU).** We optionally inverse fold the designed binders. This step tends to create sequences that are more likely to be predicted to refold into the designed structure. It is also likely to improve solubility (the inverse folding model was only trained on soluble proteins). We use BoltzIF (detailed below), which is similar to SolubleMPNN [Goverde et al., 2024].

3. **Folding (GPU).** We predict the structure of the design in complex with the target using Boltz-2 which is provided the template of the target structure (produced in step 1) and no MSAs. We compute the RMSD between the predicted structure and the structure produced by step 1 for later filtering (if the refolded structure is similar to the designed structure, the design is more likely to be a binder). This also produces confidence metrics (pTMs, pAEs) used in filtering.

   In scenarios where we design globular proteins, we perform an additional refolding step where we only predict the structure of the designed binder (in absence of the target), and compute the same RMSD metrics. As detailed in C, we found this necessary to assess whether the designed protein's structure can be achieved in absence of the target, indicating that the design can express well and fold on its own.

4. **Affinity Prediction (GPU).** When designing proteins that bind small-molecules, we predict their affinity using Boltz-2's affinity module.

5. **Analyze (CPU).** We compute physics-based metrics that provide information about the binder-target interaction strength and developability properties of the design. Here we describe those that are used in the filtering step by default. A complete list is in Appendix A.4. We count the hydrogen bonds and salt bridges between design and target and assess the buried surface area (how much contact is there and how much shape complementarity) between the design and target. Additionally, we compute a sequence based solubility metric and the area of the largest hydrophobic patch on the surface of the designed protein (large hydrophobic patches can cause protein aggregation and difficulty during purification or expression).

6. **Filter (CPU).** Based on the metrics computed in the previous stage, we produce a ranking of the designs (detailed below). This ranking is used in a quality-diversity algorithm (detailed below) which returns a final set of filtered (to a desired budget k), ranked, and diversity optimized candidates. This step takes around 20 seconds independent of the number of designs.

Inverse Folding Model. Our inverse folding model is largely similar to SolubleMPNN [Goverde et al., 2024]. The main differences are that (1) we use 6 encoder layers instead of the 3 of SolubleMPNN, (2) we exclude fibril proteins from the training set next to excluding transmembrane proteins and (3) we train on structures cropped to 1024 amino acids rather than excluding larger instances. We verify that our inverse folding model performs similar to ProteinMPNN and SolubleMPNN in Section B.3.

Ranking. To rank designs, we first compute two groups of metrics that we expect to correlate with experimental success: Boltz-2 confidence scores [Passaro et al., 2025] and interaction-type scores such as the number of hydrogen bonds. These are aggregated into a single quality score q by taking the worst rank across all metrics (Algorithm 2). Metrics are given varying importance by weighting their respective ranks, where the weights are calibrated on a benchmark of validated binders (see Appendix B.1). This prioritizes designs with the best worst-case performance across all metrics.

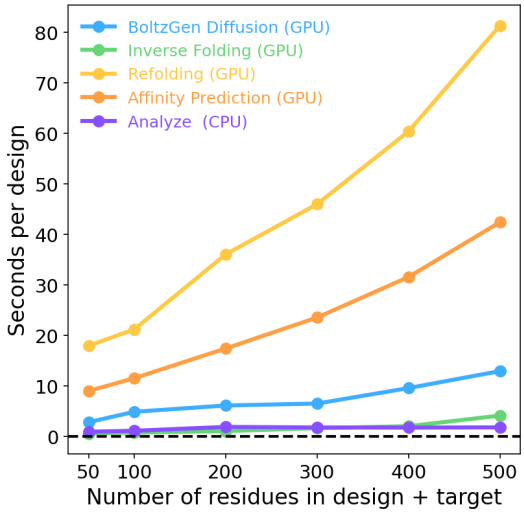


Figure 12: Time consumption of each step in seconds per design for different numbers of residues in the design + target complex (on a single NVIDIA A100 GPU).

Algorithm 2: Top-k binder designs selection by worst weighted metric rank

Input: Data matrix X (rows indexed by i correspond to protein binder designs, columns indexed by j are metrics), metric weights w_j, number of selections k

Output: Indices of top-k rows

```
/* Default metrics and weights: design_iiptm: 1, design_ptm: 2, neg_min_design_to_target_pae:
   1, plip_hbonds_refolded: 2, plip_saltbridge_refolded: 2, and delta_sasa_refolded: 2      */
```

foreach *column* j **do**
  foreach *row* i **do**
    Compute rank $r_{i,j}$ across column; `// Lower is better`
    Compute weighted rank $\tilde{r}_{i,j} \leftarrow r_{i,j}/w_j$;
  end
end
foreach *row* i **do**
  Compute score $s_i \leftarrow \max_j \tilde{r}_{i,j}$;
end
return Indices of k rows with smallest s_i;

Quality-Diversity Selection. In order to select a *diverse* set of high-ranking designs, we use a greedy selection algorithm described in Alg. 3. Each design x produced by the model is ranked according to the aggregated score $q(x)$ introduced above. For a given set of candidates S, we also measure the similarity of design x to all designs in the set S based on a mixture of the TM-score and sequence similarity,

$$\text{Diversity}(x, A) = 1 - (w_{\text{struct}} \cdot \text{StructSim}(x, A) + w_{\text{seq}} \cdot \text{SeqSim}(x, A)), \tag{15}$$

where

$$\text{StructSim}(x, A) = \max_{a \in A} \text{TM-score}(x, a), \tag{16}$$

and

$$\text{SeqSim}(x, A) = \max_{a \in A} \frac{\text{alignment}(x, a)}{\max(\text{len}(x), \text{len}(a))}. \tag{17}$$

The algorithm then proceeds by greedily adding designs that maximize a combination of quality and diversity with respect to the current set of candidates.

Algorithm 3: Greedy Quality-Diversity (QD) Selection

Input: Set of candidate designs $S = \{x_1, x_2, \ldots, x_N\}$, quality scores $q(x)$, trade-off parameter $\alpha \in [0, 1]$, diversity weights $w_{\text{struct}}, w_{\text{seq}}$, and selection budget B.

Output: Selected subset of designs A, with $|A| = B$

$A \leftarrow \{\arg\max_{x \in S} q(x)\}$; `// Initialize selected set`
while $|A| < B$ **do**
  $x^* \leftarrow \arg\max_{x \in S \setminus A} [\alpha \cdot \text{Diversity}(x, A) + (1 - \alpha) \cdot \text{Quality}(x)]$; `// Choose next design` x^*
  $A \leftarrow A \cup \{x^*\}$; `// Update selection set`
end
return A

4 Detailed Wetlab Results

4.1 Designing Nanobodies and Proteins against 9 Novel Targets

Experiments carried out by Adaptyv Bio.

The target choice, design process, and results are described in Section 2.2. Here we provide Tables 2 and 3 to list all attained affinity measurements and describe the targets' therapeutic and translational relevance in Section 4.1.1.

Table 2: **Affinities for Novel Targets - Protein Designs**. Entries in blue correspond to the average of replicate measurements, orange corresponds to single measurements. Affinity K_D in nM. Expressed designs that do not bind are marked as ○; lack of expression is marked as ×.

AMBP (3qkg)	GM2A (1g13)	HNMT (1jqd)	IDI2 (2pny)	MZB1 (7aah)	ORM2 (3apu)	PHYH (2a1x)	PMVK (3ch4)	RFK (1nb0)
53	○	270	26	9.8	○	22	10	○
190	○	○	66	12	○	31	12	○
710	○	○	73	14	○	91	○	○
890	○	○	230	23	○	120	○	○
weak	○	○	○	25	○	160	○	○
○	○	○	○	25	○	○	○	○
○	○	○	○	26	○	○	○	○
○	○	○	○	31	○		○	○
○	○	○	○	31	○		○	○
○	○	○	×	32	○		○	○
○	○	○	×	73	○		○	×
○	×	○	×	160	×		○	×
○	×	○	×	220	×		○	×
○	×	○	×	310	×		○	×
○	×	×	×	○	×		○	×

Table 3: **Affinities for Novel Targets - Nanobody Designs.** Entries in blue correspond to the average of replicate measurements, orange corresponds to single measurements. Affinity K_D in nM. Expressed designs that do not bind are marked as ○; lack of expression is marked as ×.

AMBP (3qkg)	GM2A (1g13)	HNMT (1jqd)	IDI2 (2pny)	MZB1 (7aah)	ORM2 (3apu)	PHYH (2a1x)	PMVK (3ch4)	RFK (1nb0)
○	○	99	520	120	○	7.8	6.1	8.8
○	○	○	○	○	○	○	9.1	18
○	○	○	○	○	○	○	13	○
○	○	○	○	○	○	○	23	○
○	○	○	○	○	○	○	66	○
○	○	○	○	○	○		76	○
○	○	○	○	○	○		200	○
○	○	○	○	○	○		440	○
○	○	○	○	○	○		○	○
○	○	○	○	○	○		○	○
○	○	○	○	○	○		○	○
○	○	○	○	○	○		○	○
○	○	○	○	○	×		○	○
○	○	○	○	○	×		○	○
○	×	×	×	○	×		×	×

4.1.1 Therapeutic and Translational Relevance of 9 Novel Targets

While largely chosen for benchmarking generalization, many of our 9 "hard" targets also play important roles in disease pathways, therapeutic mechanisms, or emerging biotechnological and synthetic biology applications. Designing binders against them could therefore enable new ways to modulate, inhibit, stabilize, or study these proteins in both therapeutic and translational contexts. We should note that, for intracellular targets, we assume that nanobody-based intrabodies, genetically encoded and expressed within cells, offer a promising modality for studying and modulating protein function *in vivo*, including within organelles such as peroxisomes [Wagner and Rothbauer, 2020].

Some of the targets are directly implicated in chronic inflammation and cancer. For instance, orosomucoid-2 (**ORM2**) is an acute-phase glycoprotein that promotes cytokine production in rheumatoid arthritis synovial tissue [Kim et al., 2024]. Binders that block ORM2-mediated immune signaling

could serve as potential anti-inflammatory agents or probes to dissect its role in autoimmune disease. Similarly, **MZB1**, an ER-resident co-chaperone involved in antibody secretion and calcium regulation, is overexpressed in multiple myeloma and chronic lymphocytic leukemia [Chanukuppa et al., 2020]. A binder that perturbs MZB1 folding function could interfere with the secretory machinery of transformed plasma cells and thus represent a novel intervention for B-cell malignancies.

Several of our targets are enzymes whose dysfunction causes inherited or acquired metabolic disorders. Phytanoyl-CoA hydroxylase (**PHYH**) is a peroxisomal enzyme required for phytanic acid breakdown. While mutations in PHYH cause Adult Refsum disease, recent studies have also implicated PHYH in cancer metabolism and other contexts of metabolic dysregulation [Zhengqi et al., 2021]. A highly specific binder could serve as a research tool to selectively modulate or visualize PHYH activity and to study the cellular consequences of impaired alpha-oxidation. Another example is riboflavin kinase (**RFK**), which catalyzes the rate-limiting step in the biosynthesis of FMN and FAD, cofactors essential in redox metabolism. Species-specific differences between human and microbial RFK enzymes support the development of microbial RFK-targeting binders as potential antimicrobial agents [Anoz-Carbonell et al., 2020]. Other enzymes in our panel link directly to cancer metabolism. Phosphomevalonate kinase (**PMVK**) is a critical enzyme in the mevalonate pathway, and recent work shows it drives tumor progression via metabolite-dependent activation of Wnt/β-catenin signaling [Chen et al., 2023]. Designed binders could serve to inhibit this signaling axis or act as pathway-specific probes. **IDI2** is also part of isoprenoid biosynthesis and widely used in engineered microbes to boost terpenoid and carotenoid yields [Farhi et al., 2011]. Tunable binders for IDI2 could thus be applied in enzyme control or stabilization in synthetic biology workflows.

Some of our targets are involved in signaling and regulatory processes relevant to inflammation, neurobiology, and RNA metabolism. **HNMT**, which methylates histamine, plays a central role in histamine clearance and is genetically associated with asthma, allergy, and various neurological traits [Szczepankiewicz et al., 2010, Yoshikawa et al., 2019]. Increasing brain histamine levels through novel HNMT inhibitors could offer therapeutic potential for certain neuropsychiatric conditions [Yoshikawa et al., 2019]. In contrast to HNMT's role in small-molecule metabolism, **METTL16** is an m^6A RNA methyltransferase that regulates MAT2A mRNA splicing and S-adenosylmethionine homeostasis. Its functional roles span both RNA processing and metabolic control. Recent studies highlight its context-dependent behavior in cancer: METTL16 can promote tumor progression in colorectal and gastric cancer [Wei et al., 2023, Wang et al., 2021], while acting as a tumor suppressor in bladder and papillary thyroid cancers, where its downregulation correlates with more aggressive disease [Yu et al., 2024, Li et al., 2024]. This functional duality underscores the potential of selective binders as orthogonal tools to dissect METTL16 pathway wiring and regulatory logic across tissue types [Qi et al., 2023].

Finally, two of our targets highlight distinct modes of translational value. **GM2A**, a lysosomal activator protein required for ganglioside GM2 degradation, is deficient or functionally compromised in the AB variant of GM2 gangliosidosis, which leads to neurodegeneration [Renaud and Brodsky, 2015, Conzelmann and Sandhoff, 1978]. Stabilizing or activating binders could therefore serve therapeutic or diagnostic functions in this rare disease, while inhibitory binders may have utility for other neurodegenerative diseases such as Alzheimer's Disease, where GM2A activity is elevated [Hsieh et al., 2022]. Alpha-1-microglobulin/bikunin precursor (**AMBP**) is a complex plasma glycoprotein that is processed into two distinct, functional proteins: alpha-1-microglobulin (A1M) and bikunin. A1M is a radical-scavenging and heme-binding protein that protects against oxidative stress and tissue injury, while bikunin is a Kunitz-type protease inhibitor involved in extracellular matrix stabilization and inflammation control. Recent work in a preeclampsia mouse model showed that recombinant A1M can alleviate placental and renal oxidative damage, reduce hypertension, and preserve organ function, underscoring its therapeutic potential in pregnancy-related disorders and other oxidative pathologies [Erlandsson et al., 2019]. Binders that selectively enhance or inhibit cleavage, secretion, or the extracellular interactions of AMBP products could be developed as tools to modulate protease activity, track oxidative damage *in vivo*, or regulate post-translational processing of these multifunctional proteins.

In sum, the 9 hard targets selected here are not merely difficult in structural terms — they are underexplored yet promising proteins across therapeutic, diagnostic, and synthetic biology domains. The ability to design functional binders against them suggests new experimental and translational tools that could complement or extend small-molecule and protein binder–based strategies.

4.2 Designing Proteins to Bind Bioactive Peptides with Diverse Structures

Experiments by A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, and William F. DeGrado.

We sought to test the capacity of BoltzGen to generate binders of biologically active peptides with diverse secondary structures and antimicrobial and/or cytotoxic activity. We targeted three peptides with diverse secondary structures: protegrin (disulfide stapled beta-hairpin) [Gidalevitz et al., 2003, Steinberg et al., 1997], melittin (amphipathic helix in the presence of membranes) [DeGrado et al., 1982, Dempsey, 1990, Guha et al., 2021, Hristova et al., 2001, Vogel and Jähnig, 1986], and indolicidin (polyproline II or amphipathic conformation in the presence of bilayers) [Falla et al., 1996, Ladokhin et al., 1997, 1999]. In this campaign, we define a successful design as one that expresses at high levels in Escherichia coli, is monomeric with the desired secondary structure, and binds its desired target with at least single-digit μM affinity. Because all the binding targets are antimicrobial peptides (AMPs), we also screened binders for their ability to neutralize antimicrobial activity against Bacillus subtilis and, where applicable, their ability to inhibit peptide-induced hemolysis. The former also serves as a measure of the designs' proteolytic stability as B. subtilis secretes numerous proteases.

From the top ranked designs produced by BoltzGen, we selected six per peptide through manual inspection, prioritizing those with consistent burial, well-oriented hydrogen bonds, and tightly packed interfaces for experimental characterization.

Target Peptide Melittin. We first evaluated melittin binders for their ability to bind melittin and neutralize its antimicrobial and hemolytic activity. All six selected melittin binder designs (termed mel1–mel6) expressed, and mel1–3 displayed the expected helical structure (Figure 13a,b; $[\Theta]_{222,\text{mel1}} = -29,900\text{deg} * \text{cm}^2/\text{dmol}$; $[\Theta]_{222,\text{mel2}} = -17,200\text{deg} * \text{cm}^2/\text{dmol}$; $[\Theta]_{222,\text{mel3}} = -13,600\text{deg} * \text{cm}^2/\text{dmol}$). Size exclusion chromatography (SEC) showed that mel1 forms a monodisperse species consistent with a monomer, while mel2 and mel3 tended to aggregate (Figure 13c). However, pre-incubation of mel2 and mel3 with melittin before SEC analysis shows that melittin binding acts as a switch, driving mel2 and mel3 from an aggregated state to monodisperse, monomeric species. Mel4–6 show poor folding and form aggregates in SEC (Figure 13b,c).

Mel1–3 were moved forward for assessment of melittin binding through a combination of *in vitro* binding assays and neutralization assays (both antimicrobial and hemolysis assays, as melittin has both antimicrobial activity and cytotoxicity) (Figure 13e,f). All three binders neutralize antimicrobial activity (Figure 13e) and hemolysis (Figure 13f). A single molar equivalent of Mel1 and Mel2 reversed the cytotoxic effect of melittin near its HD50 (1.2 μM) for hemolysis of erythrocytes. Melittin contains a single tryptophan residue that is predicted to be near or fully buried in the peptide:binder interface in the designed complexes. Thus, we complemented neutralization assays by monitoring changes in intrinsic tryptophan fluorescence *in vitro* to quantify design binding affinity. Mel2 and mel3 have low to sub-μM affinity for melittin (0.41 and 4.4 μM K_d for mel2 and mel3, respectively), while mel1 showed limited spectral shift at the excitation and emission wavelengths monitored.

Target Peptide Indolicidin. All selected indolicidin binder designs were predicted to be helical, comprised of 3-6 helices with a peptide binding cleft on the surface of the helical binder (Figure 14a). Three of the six (indo1, indo3, and indo4) designs formed single, homogenous species by analytical size exclusion chromatography, while indo2, indo5, and indo6 tended to aggregate (Figure 14b). All six binder designs had helical character as measured by circular dichroism (Figure 14c; $[\Theta]_{222,\text{indo1}} = -21,700\text{deg} * \text{cm}^2/\text{dmol}$; $[\Theta]_{222,\text{indo2}} = -19,800\text{deg} * \text{cm}^2/\text{dmol}$; $[\Theta]_{222,\text{indo3}} = -8,700\text{deg} * \text{cm}^2/\text{dmol}$; $[\Theta]_{222,\text{indo4}} = -17,700\text{deg} * \text{cm}^2/\text{dmol}$; $[\Theta]_{222,\text{indo5}} = -9,600\text{deg} * \text{cm}^2/\text{dmol}$; $[\Theta]_{222,\text{indo6}} = -22,900\text{deg} * \text{cm}^2/\text{dmol}$). All indolicidin binder designs showed indolicdin binding as monitored by changes in indolicidin tryptophan fluorescence (Figure 14d; $K_{d,\text{indo4}} < K_{d,\text{indo1}} < K_{d,\text{indo5}} < K_{d,\text{indo2}} < K_{d,\text{indo3}} < K_{d,\text{indo6}}$), with indo 1, 3, and 5 exhibiting affinities $<5\mu$M and indo4 exhibiting sub-μM affinity. (Figure 14d). The indo4:indolicidin interaction was further analyzed by surface plasmon resonance, which confirmed that the complex binds with nanomolar affinity (Figure 14e). Despite all six designs having detectable indolicidin binding, only indo4 showed robust neutralization of indolicidin antimicrobial activity (Figure 14f). This may be due to low proteolytic stability of the designs in the presence of B. subtilis, which secretes an array of proteases [van Wely et al., 2001].

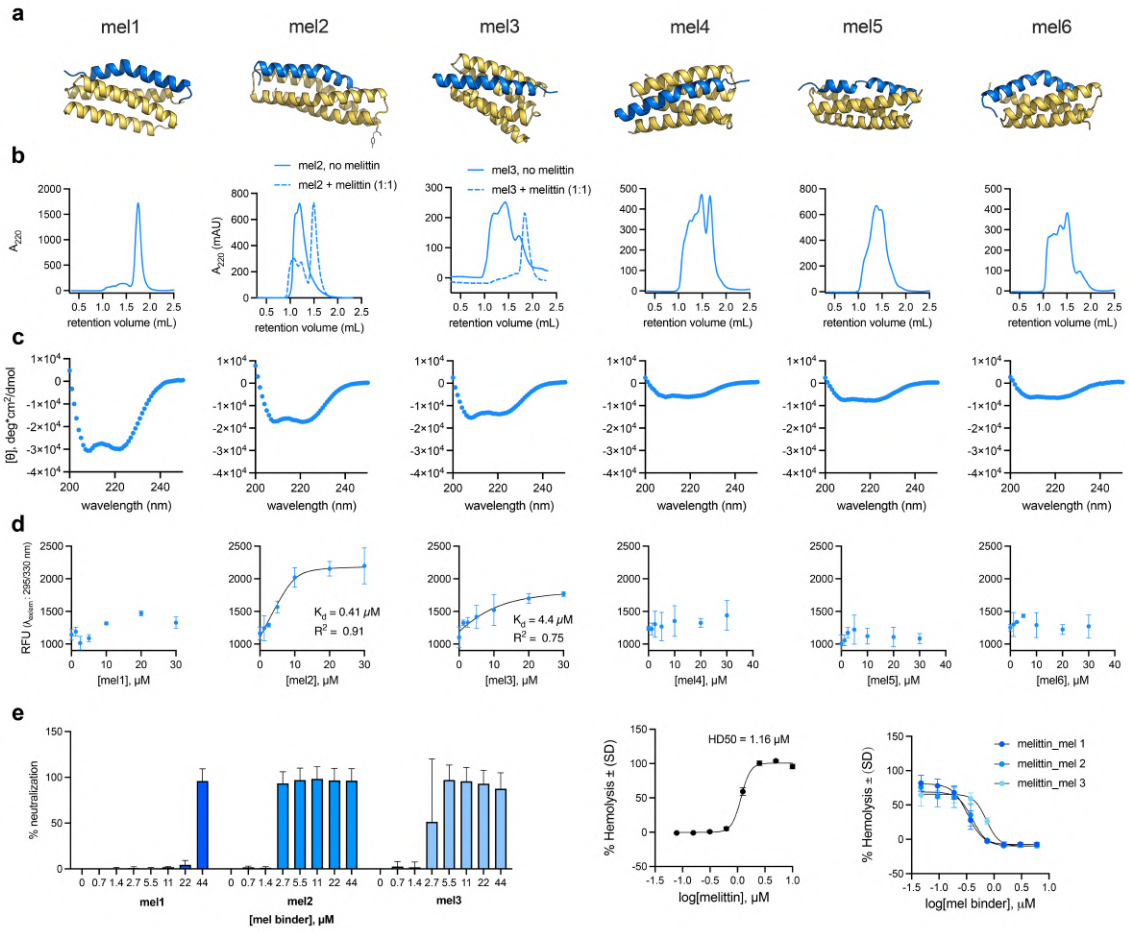


Figure 13: **Experimental characterization of melittin binder designs. (a)** Structure prediction models of all selected melittin binders are predicted to adopt a helical conformation and to bind melittin in its amphipathic helix state. **(b)** Analytical size exclusion chromatography traces of Mel1–Mel6 show varied oligomeric states. **(c)** Circular dichroism shows that Mel1–3 have helical character, while Mel4–6 are only partially folded. **(d)** Mel2 and Mel3 show detectable melittin binding as measured by changes in intrinsic fluorescence of melittin's tryptophan residue. Melittin concentration is held constant at 10 μM while binders, which lack tryptophan residues, are varied from 0–40 μM. **(e)** Mel1–3 neutralize melittin's antimicrobial activity against *B. subtilis*. **(f)** Mel1–3 also neutralize melittin's hemolytic activity.

Target Peptide Protegrin. Finally, we assessed the ability of BoltzGen to generate binders to protegrin, a disulfide-stapled, beta-hairpin antimicrobial peptide. Unlike the previous targets, the generated designs for protegrin binders showed mixed alpha/beta or all beta topologies (Figure 15 a). By SEC, pro1 and pro6 formed single monodisperse species consistent with a monomeric state, while pro2–5 eluted at volumes consistent with higher-order oligomers or mixed oligomeric states (Figure 15 b). All six designs showed the expected secondary structure by circular dichroism (beta or mixed alpha/beta) (Figure 15 c). Two of the six designs, pro1 and pro6, showed detectable binding to protegrin via changes in binder tryptophan fluorescence, with affinities of 7.2 and 1.2 μM, respectively (Figure 15 d). Both pro1 and pro6 neutralized protegrin, with pro6 being a more potent neutralizer. This is consistent with the relative affinities of the two designs as measured by tryptophan fluorescence changes. While pro2, pro4, and pro5 showed little spectral shift in tryptophan fluorescence assays, they all neutralized protegrin's activity against B. subtilis. The trp residues in these designs are not predicted to be fully buried in the bound complex; thus, neutralization assays indicate a binding event is occurring but it may not be detectable by the *in vitro* binding assay used herein (Figure 15 e).

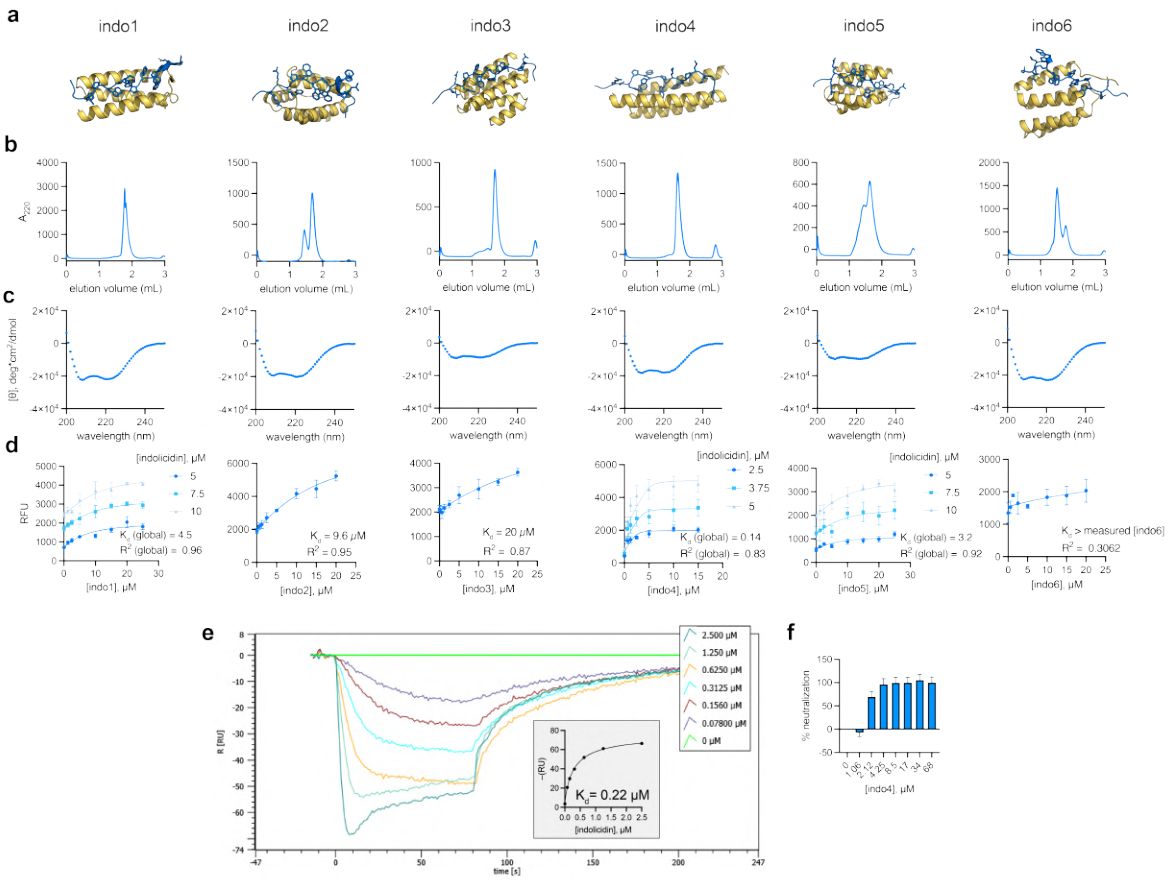


Figure 14: **Experimental characterization of indolicidin binder designs. (a)** Structure prediction models of all selected indolicidin binders are predicted to adopt a helical conformation. **(b)** Analytical size exclusion chromatography traces of Indo1–Indo6 show varied oligomeric states. **(c)** Circular dichroism shows that Indo1–6 all have helical character. **(d)** All Indo binders show detectable indolicidin binding as measured by changes in intrinsic fluorescence of indolicidin's tryptophan residues. Indo1, Indo4, and Indo5 binding was measured with multiple indolicidin concentrations, and K_d was determined via a global fit. **(e)** Only Indo4 neutralizes indolicidin antimicrobial activity against *B. subtilis*. **(f)** For Indo4, binding was validated by surface plasmon resonance, in which Indo4 was immobilized and exposed to varied [indolicidin]. Affinity was determined as a function of response units at the pre-injection stop points of each [indolicidin] (inset).

4.3 Designing Peptides to Bind the Disordered Region of NPM1.

Experiments by Yaotian Zhang, and Denes Hnisz.

Designs. Using BoltzGen, we generate 20,000 designs of size 40-80 to bind NPM1's disordered region and test 5 designs experimentally. We make use of BoltzGen's binding site conditioning and specify that the design should interact with the disordered region and *not* its structured beta-sheet region via our model's "not-binding" feature. Additionally, we provide the structure of the ordered region and leave the disordered region flexible. Thus tasking BoltzGen to model how the disordered region will fold and become structured in presence of the peptide while designing that peptide.

Detailed Results. We assessed the target engagement of the NPM1-binders in live cells. Five selected NPM1-binders were genetically fused to GFP, and the GFP-tagged binders were transiently expressed in human osteosarcoma (U2OS) cells. The subcellular localization of the binders was visualized with GFP fluorescence (Figure 16 A). Among the five tested binders, one binder (NPM1-binder-4) showed localization consistent with enrichment in nucleoli where the endogenous NPM1 protein is known to localize (Figure 16 B). Immunofluorescence staining for NPM1 indeed confirmed the colocalization of

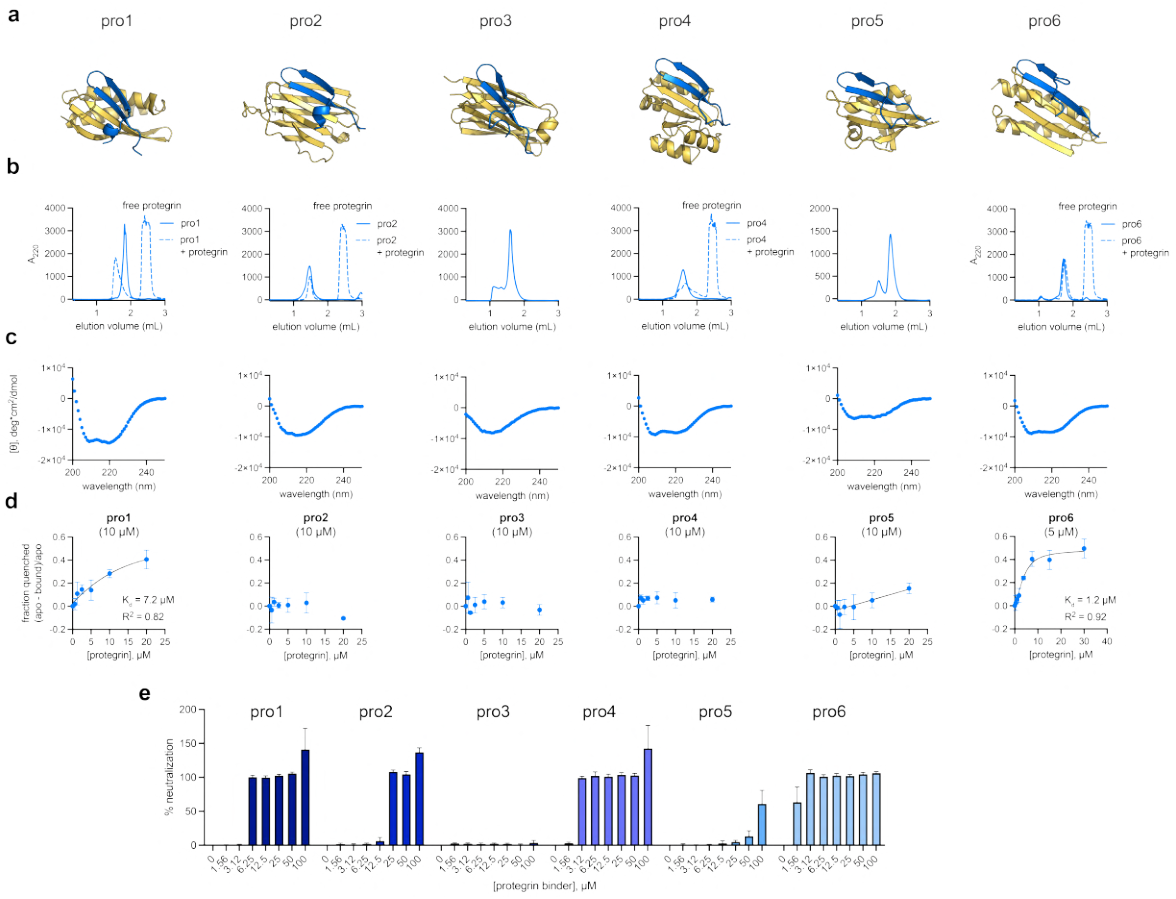


Figure 15: **Experimental characterization of protegrin binder designs. (a)** Structure prediction models of all selected protegrin binders are predicted to adopt a helical conformation. **(b)** Analytical size exclusion chromatography traces of Pro1–Pro6 show varied oligomeric states. **(c)** Circular dichroism shows that Pro1, 4, 5, and 6 have partial helical character, while Pro2 and Pro3 have the expected β-structure in their apo form. **(d)** Pro1, 5, and 6 show detectable indolicidin binding as measured by changes in intrinsic fluorescence of the binders' tryptophan residues. **(e)** Pro1, 2, 4, and 6 fully neutralize protegrin antimicrobial activity against *B. subtilis* at the concentrations tested.

the GFP-tagged NPM-binder-4 with endogenous NPM1 (Figure 16 C), thus providing evidence of a de-novo designed protein binding disordered proteins *in live cells*. As an additional control for nucleolar localization, another well-known protein that also localizes in nucleoli, SURF6, was also visualized (Figure 16 C).

4.4 Designing Peptides to Bind a Specific Site of RagC GTPase

Experiments by Shamayeeta Ray, Jonathan T. Goldstein, and David M. Sabatini.

Target and Designs. All information is already present in the summary Section 2.5.

Results. We test 29 designs in an SPR assay and find 7 binders with affinities ranging from 3.5 μM to 893 μM (Table 4). Additionally, for 4 binders, we randomly permute their residues and re-test binding. Three lose their affinity and 1 shows 10× weaker binding. Furthermore, the permuted version of peptide 23 showed poor binding at lower concentration and displayed uninterpretable sensograms at concentrations greater 25 μM, indicating non-specific binding.

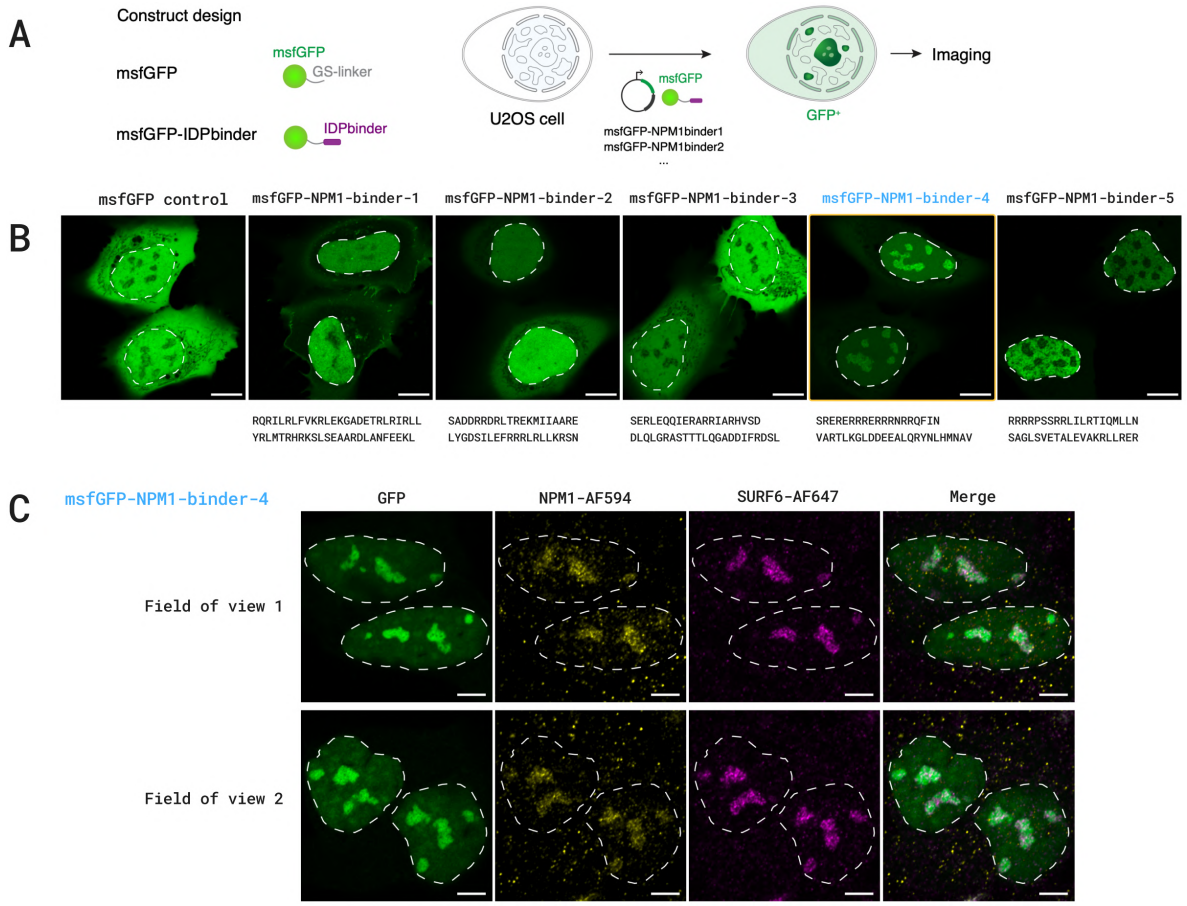


Figure 16: A. Schematic model of the msfGFP-tagged binder design and cellular assay to visualize subcellular localization of the NPM1-binders. B. Live cell fluorescence microscopy images of U2OS cells expressing ectopic msfGFP-NPM1 binders. The cell nucleus is highlighted with a dashed white line contour. Scale bar: 10 μm. The experiment was repeated twice independently with similar results. C. Fixed-cell immunofluorescence of U2OS cells expressing exogenous msfGFP-NPM1-binder-4. Endogenous NPM1 and SURF6 are stained with antibodies. The cell nucleus is highlighted with a dashed white line contour. Scale bar: μ µm. The experiment was repeated twice independently with similar results.

4.5 Designing Disulfide Bonded Cyclic Peptides to Bind a Specific Site of RagA:RagC

Experiments by Shamayeeta Ray, Jonathan T. Goldstein, and David M. Sabatini.

Target and Designs. All information is already present in the summary Section 2.6.

Results. We test 24 peptides and find 14 of them to show specific binding by SPR. For 8 of those we resolved the affinities obtaining values ranging from 80 μM to 1100 μM (Table 5).

4.6 Designing Nanobodies that Bind Penguinpox and Hemagglutinin

Experiments by Jacob A. Hambalek, Anshika Gupta, Diego Taquiri Diaz, and Chang C. Liu.

Targets. All information is already present in the summary Section 2.7.

Designs and Results. We generate 60 000 nanobodies against each target in the fashion described in Section 2.2 and select 7 per target for experimental characterization. Of the 7 designed nanobodies that we tested for each antigen through yeast surface display at a range of antigen concentrations (7 nM–2 µM), we observed a weak binding signal at the highest antigen concentration (1.6–2 µM) for some designs. Specifically, 1 of 7 designs against cGAMP PDE showed a binding signal at 1.6 µM of the

Peptide	Sequence	K_d (μM)	$R_{\max}$ (RU)
6	ICTLHRK	60	375
13	GVKEDCQALRAQSKALRK	117	249
18	IMTLKRFSKNYGEIERLALY	893	505
19	QLYHIRIARSAQRIFKNGG	268	3041
20	HAMSKNMQRFLRKAKAMVIV	3.5	2083
23	MMSDVRQLRTIVRELRRV	268	285
29	DYSAGRQLLRTLKDKLTTS	373	185
Scr20	KWVHIFRALAMKAMRSQKNM	33.5	451

Table 4: **Binding affinities and response maxima of BoltzGen's peptide designs.** 29 peptides were tested in total. $R_{\max}$ is the saturated signal of a sensogram at maximum binding representing complete occupancy of all ligand sites. RU is Response Units. Scr20 denotes a version of peptide 20 with randomly permuted residues.

Peptide	Sequence	K_d (μM)	$R_{\max}$ (RU)
7	RLRERCRLNPLYCL	308	670
8	RRRERCRLNPLYCG	318	2200
11	SRRERCRLNRLLCLL	447	2000
12	RRRELCKLNPLVCG	1100	1750
14	RRRELCRLDRRACL	297	550
16	KRREACARYRTICLH	164	250
18	IAKRCKADPYRCKLLSR	80	1200
22	GCSKDVQKCKLLK	274	290

Table 5: **Binding affinities and response maxima of BoltzGen's disulfide-bonded cyclic peptide designs against the RagA:RagC dimer.** 24 peptides were tested in total, 14 bind, and for 8 we resolved the affinities. $R_{\max}$ denotes the maximum sensogram signal corresponding to full ligand occupancy. RU: Response Units.

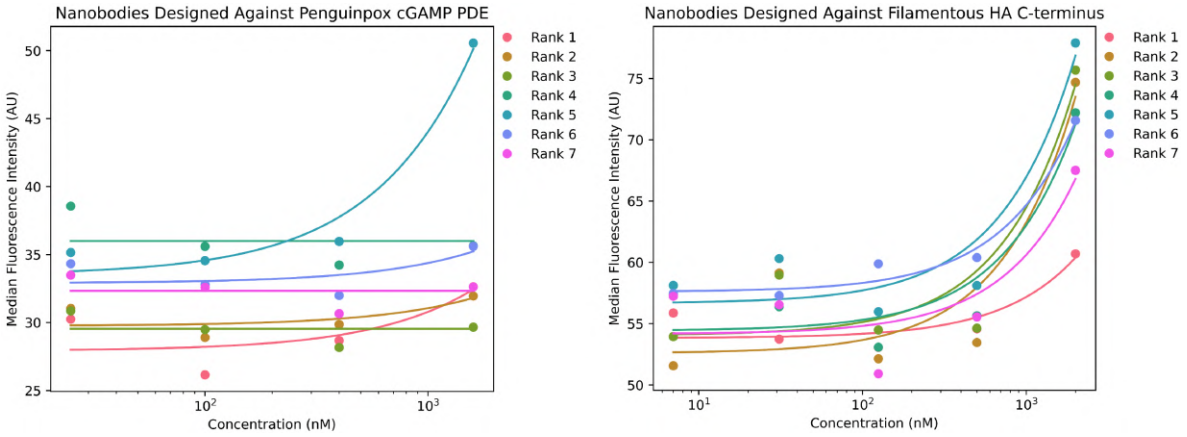


Figure 17: **YSD of nanobodies for feasibility assessment of design optimization.** Median fluorescence intensities above background for cells displaying nanobodies targeting **(left)** cGAMP PDE and **(right)** FhaB. The highest concentration of antigen in both cases was the only test case that demonstrated any binding signal for any nanobody design.

antigen, and 7 of 7 designs against FhaB—where all 7 designs were against the same epitope—showed a binding signal at 2 µM of the antigen. Because 2 µM was the highest antigen concentration used for labeling and was the minimal concentration tested that resulted in any binding signal, we cannot calculate an EC_{50} from these studies. However, we can conclude that the binders exhibiting weak binding signals are at best 2 µM affinity binders.

The plots in Figure 17 represent the median Alexa Fluor 647 fluorescence intensity of nanobody-displaying cells (HA tag label positive) minus the median Alexa Fluor 647 fluorescence intensity of the non-nanobody-displaying cells (HA tag label negative) at different concentrations of antigen. The latter acts as an internal control and is subtracted to remove the signal from nanobody-independent background binding of cells to antigen. Higher fluorescence intensity indicates more binding. The background-subtracted median fluorescence intensities across concentrations for each nanobody design were then fit to a Hill function using `SciPy curve_fit`. For initializing curve fitting, the background value used was the minimum fluorescence intensity observed at 0 antigen concentration and the EC_{50} used was the median antigen concentration tested. The Hill coefficient, n, was explicitly set to 1, as we expect non-cooperative binding of the antigen to the antibody. The fluorescence value, Y, is fit as:

$$Y = B_{\min} + \frac{B_{\max} L}{\tilde{C} + L}$$

where $B_{\min}$ is the best-fit background fluorescence value, $B_{\max}$ is the best-fit maximum binding value, $\tilde{C}$ is the best-fit EC_{50} projection, and L is the concentration of antigen.

4.7 Designing Proteins that Bind to Small Molecules

Experiments by A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, and William F. DeGrado.

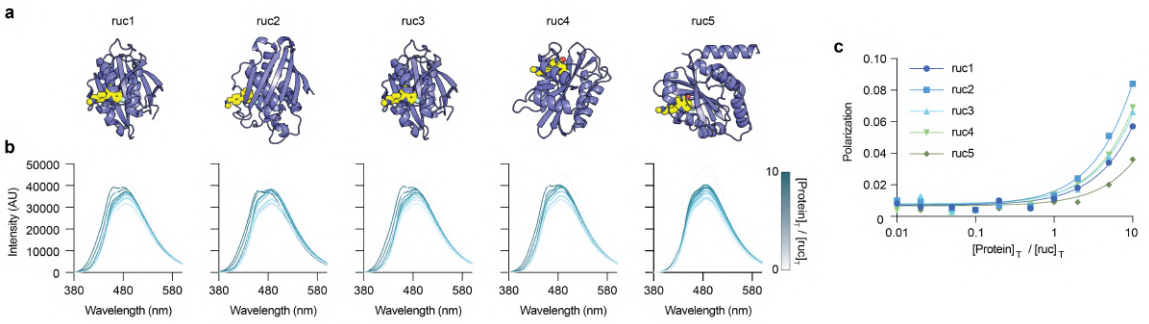


Figure 18: **rucaparib Binder Design a)** Structural model of the designed rucaparib binder (purple) in complex with rucaparib (yellow). **b)** Fluorescence emission spectra of rucaparib (10 μM) upon titration with increasing concentrations of the designed protein (0–10 equivalents). The emission spectrum of rucaparib exhibited a blue shift upon binding, indicating changes in its local environment. **c)** Fluorescence polarization assay of rucaparib binding to the designed protein, measured with an excitation wavelength of 405 nm and emission wavelength of 516 nm. Polarization values were plotted against the molar ratio of protein to rucaparib, and the data were fitted to a one-site binding model using nonlinear regression in GraphPad Prism 10 to determine the dissociation constant (K_d).

We selected rucaparib as a benchmark target, as published reports indicate that high-affinity binders ($K_d < 5$ nM) can be achieved. We also selected a rhodamide derivative related to rucaparib, which we leave undisclosed. Our computational design pipeline generated on the order of 10,000 designs for each target. The initial set was filtered to a highest-confidence subset of 100 designs based on RMSD < 2.5 Å relative to the Boltz-2 refolded models. Within this filtered pool, designs were ranked using a composite metric (interaction score + Boltz score), which effectively integrates predicted structural fidelity with biophysical interaction quality. To identify essential interactions, we fragmented the chemical groups of rucaparib and calculated the number of hydrogen bonds formed with each fragment. We prioritized designs forming hydrogen bonds with the carboxamide chemical groups, as this interaction is considered essential for specific binding. A total of six designs for rucaparib and 4 designs for the rhodamine derivative were selected for experimental validation by binding assay (Figure 18a).

Among the selected rucaparib binder designs (termed ruc1–ruc6), five out of six were expressed in moderate to good yields (15 – 69.5 mg/L). Incubation of each binder with equimolar concentrations of rucaparib led to a marked blue-shift and an increase in the intensity of its fluorescence spectrum, which is characteristic of the rucaparib indole core being bound in a rigid, solvent-inaccessible site (Figure 18b). Fluorescence polarization data showed that ruc1–ruc4 exhibited moderate affinity for rucaparib, with values of 75.9, 43.0, 64.2, and 59.1 μM K_d, respectively. Ruc5 showed the weakest binding affinity (151.5 μM K_d) among the tested candidates (Figure 18c). All designs against the rhodamine derivative expressed in moderate to good yields (15 – 69.5 mg/L), and fluorescence polarization data showed binding affinities of 30.9, 69.2, 144.7 and 252.2 μM. Affinities of all designs against both targets are summarized in Table 6.

	Length	**Expressed**	K_d **(μM)**
rucaparib			
ruc1	173	yes	75.9
ruc2	180	yes	43.0
ruc3	173	yes	64.2
ruc4	180	yes	59.1
ruc5	177	yes	151.5
ruc6	179	no	—
rhodamine deriv.			
rhd1	154	yes	69.2
rhd2	141	yes	252.2
rhd3	158	yes	30.9
rhd4	154	yes	144.7

Table 6: **Small Molecule Binders.** Experimental characterization of designed binders against rucaparib and an undisclosed rhodamine derivative.

4.8 Designing Antimicrobial Peptides that Inhibit the GyrA to GyrA Interaction

Experiments by Andrew Savinov, and Gene-Wei Li

Recent work has demonstrated that peptide fragments of full-length proteins (protein fragments) are generalizable inhibitors of native protein interactions in living cells [Savinov et al., 2022, 2025]. Libraries of tiling protein fragments also reveal specific interfaces prone to such protein fragment-based inhibition [Savinov et al., 2022], identifying good target sites for alternative drug design modalities.

Here, we sought to leverage previous results uncovering potent inhibitory protein fragments of the essential bacterial protein DNA gyrase (subunit A; hereafter, GyrA), where the C-gate closure interaction between two GyrA subunits was identified as a desirable target site [Savinov et al., 2022, 2025]. DNA gyrase is a target of considerable interest for developing novel antibiotics. Existing compounds, such as clinically employed fluoroquinones, function by trapping gyrase in the DNA-cleavage complex [Khan et al., 2018, Bax et al., 2019] and aminocoumarins interfere with the enzyme's ATPase activity [Khan et al., 2018], but interfering with C-gate closure represents a complementary drug modality. Indeed, a previously reported monoclonal antibody against *Mycobacterium tuberculosis* GyrA appears to target this same site [Manjunatha et al., 2005].

We therefore used the inhibitory peak identified by fragment scanning experiments [Savinov et al., 2022, 2025] and massively parallel predictions of fragment binding modes with FragFold [Savinov et al., 2025] to identify target residues in the GyrA C-gate closure complex, and used BoltzGen to design *de novo* peptide binders targeting this same interface. We note that this was a challenging target for *de novo* binder design owing to the small size of the C-gate closure interface (i.e., only 6 residues within 4 Å in the experimental structure of the full gyrase complex (PDB ID 6RKS), and 11 residues within 4 Å in the FragFold model [Savinov et al., 2025]) and we therefore designed binders with diverse design parameters. We then leveraged an established experimental method to measure the inhibitory effects of these designed binders in living *E. coli* cells, taking advantage of the importance of GyrA for cell growth to determine inhibitory function from massively parallel measurements of growth inhibition [Savinov et al., 2022, 2025]; in this way a large library of BoltzGen designs was tested in parallel (Methods).

1,808 designed GyrA inhibitors were tested alongside 1,788 mutants designed to break the designed binding modes (3 alanine substitutions at the binding interface per design) to determine the specificity of inhibitory activity to binding GyrA as desired. These designs were tested alongside 30-aa fragments tiling across GyrA with a 1 aa step size, matching previously published work [Savinov et al., 2022,

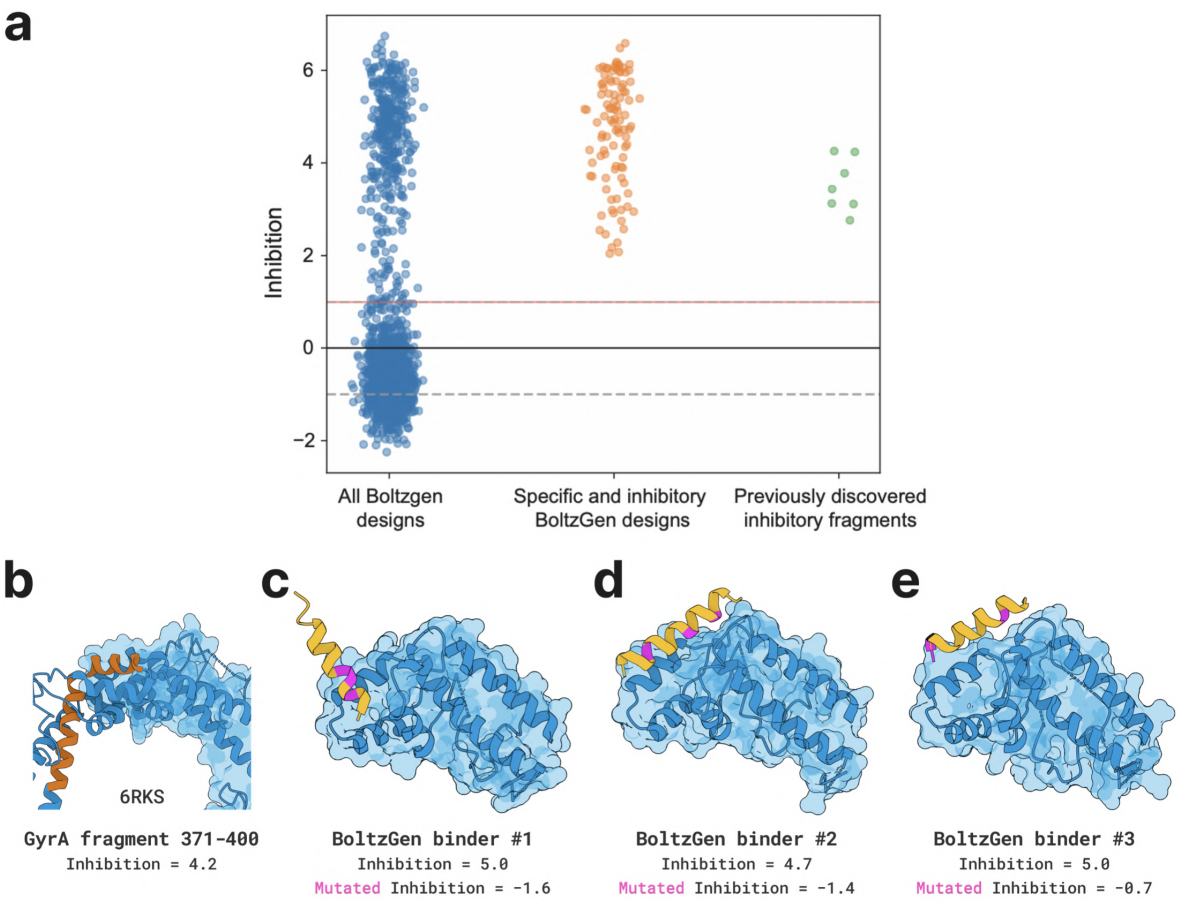


Figure 19: **BoltzGen-designed binders are potent DNA Gyrase inhibitors *in vivo*. (a)** GyrA inhibition from massively parallel in-cell measurements of peptides (Methods) for all designed GyrA binders; designed binders that were inhibitory and specific to the designed binding mode (5.5% of all designs); and inhibitory fragments of GyrA targeting the same site ([Savinov et al., 2022, 2025]). Note that all values for the inhibitory GyrA fragments and 91% of the values for interface-specific designed binders represent lower bounds on inhibitory activity, as these peptides inhibited growth so thoroughly that they completely dropped out of the library when expressed. **(b)** Structure of inhibitory fragment 371-400 of GyrA (orange; [Savinov et al., 2022, 2025]) bound to full-length GyrA (blue) in the context of the full-length GyrA dimer structure (PDB ID 6RKS), with *in vivo* inhibitory effect indicated. **(c) - (e)** Three example BoltzGen-designed binders targeting the same interface as GyrA fragment 371-400, exhibiting high inhibition and specificity *in vivo*. In each case the in-cell inhibitory effects of the designed binder as well as the mutated form with 3 alanine substitutions at the interface (magenta residues) are indicated. In all three cases, inhibitory effects are completely lost upon mutation.

2025], as an internal positive control for GyrA-inhibitory fragments; and also 20-aa fragments tiling (1 aa step size) across enhanced GFP (eGFP), approximately matching the average fragment length of the designs (15 ± 3.4 aa, mean ± s.d.). The eGFP fragments provided an internal negative control in the form of fragments not expected to interact specifically with *E. coli* proteins.

Across all designed GyrA binders tested, 352 (19.5%) were found to substantially inhibit *E. coli* growth, similar to known inhibitory protein fragments of GyrA (Figure 19a-b; Methods). The binders inhibiting growth specifically by binding the GyrA C-gate at the desired interface should generally exhibit loss of activity when residues at the designed interface are mutated (Figure 19c-e), so we calculated the quantity Δ(Inhibition) = Inhibition(design) - Inhibition(mutated design) as a measure of designed GyrA inhibitor specificity. Of all 352 growth-inhibiting designs, 54 designed binders (3.0% of total) were strongly specific for the designed GyrA binding site (Δ(Inhibition) ≥ 2) and 99 designed binders (5.5% of total) were significantly specific under a less stringent requirement (Δ(Inhibition) ≥ 1). Thus, ∼3-6% of designed GyrA inhibitors appear to inhibit growth by binding GyrA at the C-gate site as designed, and almost 20% of designs targeting this essential protein inhibited growth more broadly.

The fraction of successfully designed GyrA inhibitors was comparable to the fraction of tiling fragments

across GyrA producing inhibitory activity (∼9%), reflecting the previously noted widespread inhibitory activity of protein fragments across diverse proteins [Savinov et al., 2022, 2025]. Comparing these results to the effects of expected nonfunctional protein fragments of similar length as represented by the 20 aa eGFP fragment library, we found that 0.45% of eGFP fragments inhibited *E. coli* growth. Therefore, compared to nonspecific control peptides, a ∼43-fold larger fraction of designed GyrA binders inhibit cell growth, and ∼7-12-fold larger fraction specifically inhibit growth dependent on the designed binding interface.

The inhibitory effects of interface-specific designed binders (Inhibition = 4.7 ± 1.2, mean ± s.d.) were comparable to the inhibitory effects of GyrA fragments from the corresponding inhibitory peak (Inhibition = 3.5 ± 0.6, mean ± s.d.) in internal control measurements in the same experiment (Figure 19a). We note as well that all GyrA-inhibitory protein fragments and 91% of interface-specific designed GyrA binders were sufficiently strong inhibitors of this essential protein that they completely dropped out of the cell population when expressed, meaning these values represent a lower bound on inhibitory activity. The successfully designed DNA gyrase inhibitors represent promising novel antimicrobials targeting this essential bacterial protein.

4.9 Designing Nanobodies and Proteins against 5 Benchmark Targets

Experiments carried out by Adaptyv Bio.

The target choice, design process, and results are described in Section 2.10. Here we provide Table 7 to list all attained affinity measurements.

(a) Protein designs

IL-7RA	Insulin	PDGFR	PDL1	TNFα
1.9	7.8	0.81	96	○
66	17	12	310	○
79	25	23	○	○
210	27	110	○	○
○	29	110	○	○
○	33	170	○	○
○	42	260	○	○
○	49	310	○	○
○	54	350	○	○
○	56	350	○	○
○	58	370	○	○
○	140	550	○	○
○	710	○	○	○
○	730	○	○	○
○	1400	○	○	○
○	1400	○	○	○
○	1600	○	○	○
○	○	○	○	×
○	○	○	○	×
○	○	○	○	×

(b) Nanobody designs

IL-7RA	Insulin	PDGFR	PDL1	TNFα
1.4	890	36	13	○
23	1300	70	27	○
26	1400	71	49	○
38	1500	87	54	○
120	○	98	82	○
150	○	120	180	○
240	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○		○	○
○	○		○	○
○	○		○	○
○	○		○	○
○	○		○	○

Table 7: **Affinities for Benchmark Targets**. Protein designs **(a)** and Nanobody designs **(b)**. Entries in blue correspond to the average of replicate measurements, orange corresponds to single measurements. Affinity K_D in nM. Expressed designs that do not bind are marked as ○; lack of expression is marked as ×.

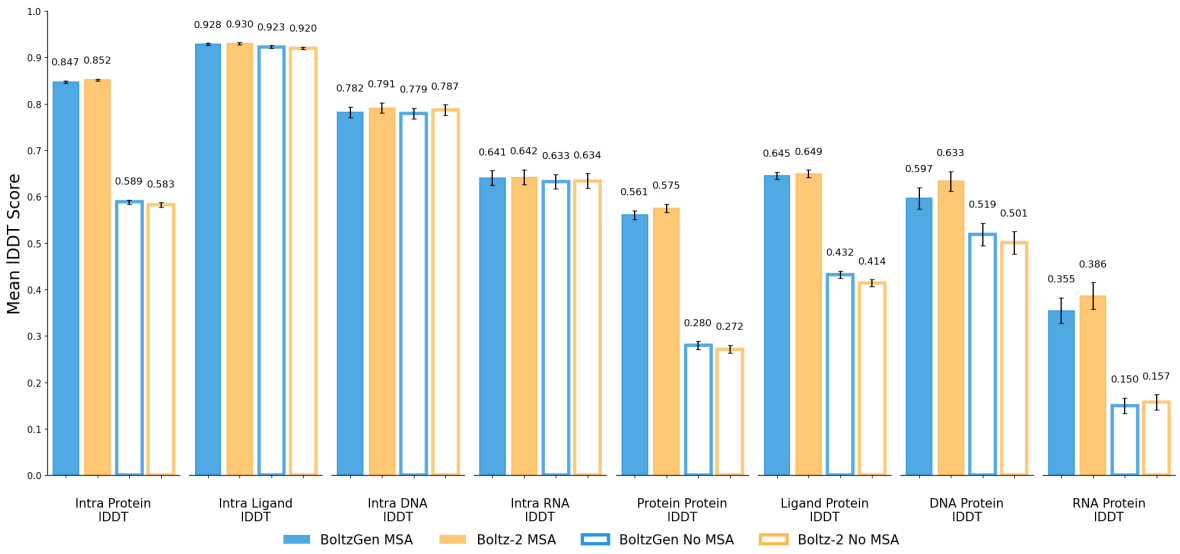


Figure 20: **Structure Prediction Evaluation.** We reason that structure-based binder design requires strong structure prediction and reasoning capability. BoltzGen can perform design *and* folding. Its folding performance matches Boltz-2. Shown is the best lDDT out of 5 diffusion samples for each complex.

5 Computational Results

5.1 Structure Prediction

BoltzGen's development was driven by the hypothesis that designing high-affinity binders requires strong structure prediction and reasoning capabilities. Expressive features that allow for accurate structure prediction are also crucial for design and enable the model to place atoms that tightly interact with the target. Thus, we assess BoltzGen's structure prediction performance.

Figure 20 shows that BoltzGen's folding performance matches Boltz-2 on its test set (minus 187 complexes that did not fit on a 40 GB GPU). This test set is based on clustering sequences with a 40% similarity threshold (details in A.1.2).

5.2 Computational Binder Design

In a comparison with RFdiffusion [Watson et al., 2023] and its extension RFdiffusionAA [Krishna et al., 2024] we aim to assess the degree to which the models ignore the target and produce the same structures independent of their conditioning. To evaluate a model's target dependence, we collect a set of targets and draw one binder design per target and filter that set, only keeping the designs that match the model's generated structure after refolding with Boltz-2 (based on an RMSD threshold of 2.5 Å). We then assess the diversity of the resulting set as the Vendi score [Friedman and Dieng, 2022] with TM-score as the similarity kernel. Methods that tend to produce the same structures irrespective of the target will obtain a lower diversity.

We carry this evaluation out for monomeric proteins and small molecules as targets. The monomer set is selected based on sequence similarity clustering and June 2023 as a date cutoffs (details in A.1). The small molecule set is a random sample from all ligands that satisfy standard drug-like property criteria (Lipinski's rule of five [Lipinski et al., 2001]) and have a Tanimoto similarity of less than 0.35 to any small molecule with the same date cutoff. The comparison in figure 21 shows that BoltzGen "pays more attention" to the target.

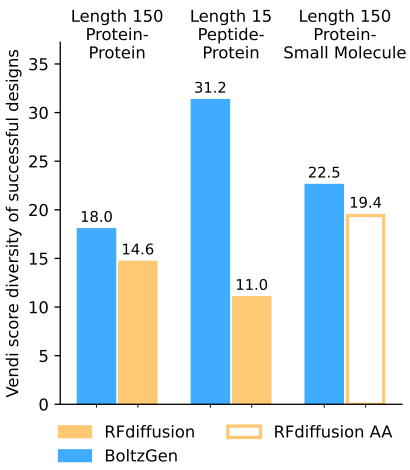


Figure 21: **Target conditioning quantification.** The diversity of successfully refolded complexes when designing a single binder against 110 targets. This assesses the degree to which the models are conditioned on the target instead of generating designs independent of the target.

6 Limitations

For therapeutic development, generating high-affinity binders is only the first step. Whether a design will achieve the intended function depends on a range of additional properties, including selectivity, developability, and the precise characteristics of the target. While predictors for these could be included in BoltzGen's filtering steps, the *selective* binding problem may be particularly ripe for direct integration into the generative process. Classifier-free guidance techniques could be used for combining the scores of BoltzGen with different targets as conditioning to achieve guiding toward one target while steering away from off-targets.

More specific to BoltzGen, we note that there is a memorization issue when designing binders of length 73-76. For certain protein targets, BoltzGen's generation diversity collapses in this length range and it nearly exclusively samples ubiquitin as the binder. For future BoltzGen training runs, ubiquitin should be downsampled (appears more than 900 times in the PDB). More details about this ubiquitin memorization is provided in Appendix B.5.

Lastly, we comment on how there is a tendency in the field to claim that binder design models are "zero-shot" and "plug-and-play" solutions without a chance for failure. We do not make this claim and encourage users to use BoltzGen thoughtfully, carefully inspect the generated structures, and potentially rerun the pipeline multiple times, first at smaller and then larger scales. BoltzGen's rich design specification language provides a large degree of control that should be experimented with for optimal results.

7 Conclusion

BoltzGen is a general-purpose generative model for biomolecular binder design, supporting a broad range of modalities, including proteins, peptides, nanobodies, and related modalities, which can be designed to target virtually any biomolecule such as proteins, small molecules, and nucleic acids. While BoltzGen is fundamentally a diffusion-based generative model, we provide it as part of a broader framework that includes tools for specifying design tasks, generating candidates, and filtering, ranking, and optimizing for diversity. This makes it a practical, end-to-end solution for real-world binder design problems.

Our strongest results to date are in nanobody design against protein targets, where we obtain nanomolar-affinity binders against two-thirds of the tested novel targets with only 15 or fewer designs. We also demonstrate strong results in designing miniproteins and peptides against a wide variety of ordered and disordered regions of proteins.

We release the entire BoltzGen package under the MIT license, including model weights, training and inference code, and a user-friendly pipeline for design and evaluation. By providing a complete and accessible solution, we hope BoltzGen serves as a practical tool and a foundation for future work in general-purpose biomolecular design.

Acknowledgments

We would like to thank Rachel Wu, Sergey Ovchinnikov, Wenxian Shi, Tally Portnoi, Umesh Padia, Peter Mikhael, Edwin Neumann, Christopher Garcia, Kevin Jude, Juno Nam, Cora Wendlandt, Dane Wittrup, Bradley Pentelute, Jigar Patel, Reda Radi, Sarah Fahlberg, Giannis Daras, Sushil Mishra, Mathai Mammen, Donovan Chin, Thomas Schwartz, Jeremie Alexander, Jonathan Stokes, Robert Hurt, Matthew Shoulders, Robbie Wilson, Aneesh Karatt-Vellatt, and Soojung Yang for insightful discussions and feedback. We thank Jiri Damborsky for providing insights into early versions of the model. D.H. thanks Henri Niskanen and Alexandre Magalhaes for consultation on experimental design. We thank Philip J. Kranzusch and Samuel J. Hobbs for providing the penguinpox cGAMP PDE target and Celia W. Goulding and Christine D. Hardy for providing the filamentous hemagglutinin target. A.S. and G-W.L. acknowledge members of the Li lab for helpful discussions.

For the compute resources required to complete the project, we thank Novo Nordisk, the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility using NERSC award ERCAP0035823, Anantha P. Chandrakasan, Chris Hill and the Office of Research Computing and Data (ORCD) via a Seed Fund grant, the EuroHPC Joint Undertaking for access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI consortium through a EuroHPC Regular Access call. We thank The Infrastructure Group (TIG) for managing our MIT computer cluster.

We acknowledge support the Abdul Latif Jameel Clinic from the Machine Learning for Pharmaceutical Discovery and Synthesis (MLPDS) consortium, the DTRA Discovery of Medical Countermeasures Against New and Emerging (DOMANE) threats program, the NSF Expeditions grant (award 1918839) Understanding the World Through Code, and the DSO Singapore grant on next generation techniques for protein ligand binding.

J.S. acknowledges support by the European Union (CLARA (101136607), ERC FRONTIER (101097822), ELIAS (101120237)) and by the Czech Technology Agency (TA CR) projects RETEMED (TN02000122) and TEREP TN02000122/001N). T.P. acknowledges support by the Czech Science Foundation (GA CR) grant 21-11563M and by the European Union's Horizon Europe program (ERC, TerpenCode, 101170268). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. This work was also supported by awards K99 GM148718 (to A.S.) and R35 GM124732 (to. G.-W.L.) from the NIH. G.-W.L. is an investigator of the HHMI. We acknowledge the following: NIH R35GM122603 to W.F.D., NIH K99GM157551 to L.S., NIH K99GM155611 to A.K.H., NIH F32GM150255 to E.T.Y. We thank the Institute for Rapid Antibody Engineering and Evolution, part of the Engineering+Health Initiative of the UCI Samueli School of Engineering, for support.

Supplementary Material

Appendix Table of contents

A Computational Method Details

A.1 Datasets

A.1.1 Training

Our data pipeline builds upon Boltz-2 [Passaro et al., 2025], while adapting the sampling procedure for the task of biomolecular design.

Table 8 summarizes the datasets used for sampling during training, including their sources, sampling cluster types, and associated weights. BoltzGen is primarily trained on entries from the Protein Data Bank (PDB) [Berman et al., 2000] and the AlphaFold Protein Structure Database (AlphaFold DB) [Varadi et al., 2022], while leveraging Boltz-1 distillation [Wohlwend et al., 2025] to enhance performance on underrepresented modalities. For additional information on individual datasets, please see Passaro et al. [2025].

PDB We process every structure in the PDB following a pipeline similar to those previously described in Boltz-2 [Passaro et al., 2025]:

- We use every PDB structure up to the training date cutoff of 06/01/2023. We parse the Biological Assembly 1 from these structures.
- For each polymer chain, we use the reference sequence and align it to the residues available in the structure.
- For ligands, we refer to the CCD dictionary to get the reference ligand and atom composition. We compute up to 10 3D conformers per ligand and sample one at random during training.
- We remove large complexes that are over 7MB or with more than 5000 residues.
- We apply the same filters as AlphaFold3, namely excluding crystallization aids and other non-biologically relevant ligands, removing clashing chains, and filtering out chains with fewer than 4 resolved residues or composed only of unknown residues.
- We compute multiple-sequence alignments for every protein chain (and only protein chains) using ColabFold search. Once monomeric MSAs are produced, we assign a taxonomy ID to every sequence in every MSA using their Uniref100 IDs as reference, if any. The preprocessing of the MSAs is analogous to AlphaFold3.
- We produce template hits for protein chains as described in AlphaFold3, using hmmbuild and hmmsearch on PDB sequences deposited at least 60 days prior to any given query's deposition date.

Distilled datasets We use Boltz-2 distilled datasets, in particular:

- *AlphaFold Database (AFDB) distillation:* In order to construct a protein monomer distillation set, we begin with uniref30 and find the overlap between those sequences and the uniclust multiple sequence alignments provided by OpenFold. We then fetch structures from the AFDB where we impose a minimum global lDDT of 0.5. This procedure results in a monomer distillation of about 5 million proteins.
- *Protein-Ligand distillation*: We construct a dataset of protein-ligand distillation from BindingDB and ChEMBL that were excluded from the main hit-to-lead affinity training set of Boltz-2 [Passaro et al., 2025]. The distillation set was formed by filtering Boltz-1 predictions to examples with a maximum interface predicted distance error (iPDE) ≤ 1.0 and a minimum interface predicted TM-Score (ipTM) ≥ 0.9.

- *RNA distillation*: Following AlphaFold3, we clustered Rfam (v14.9) [Kalvari et al., 2021] using MMSeqs2 [Steinegger and Söding, 2017] with 90% identity and 80% coverage. To form the distillation set, Boltz-1 predictions for cluster representatives are filtered to those where the maximum average predicted distance error (PDE) ≤ 2.0.
- *Protein-DNA distillation*: The protein-DNA distillation data is constructed similarly to AlphaFold3. Using the JASPAR 2024 release (specifically, the CORE collection), we first find transcription factor profiles with matching gene IDs across two high-throughput SELEX datasets [Jolma et al., 2015, Yin et al., 2017]. For each filtered profile, a protein sequence is assigned in two ways: i) using the canonical protein sequence under the profile's Uniprot ID and ii) searching for the sequence in the two SELEX datasets (with matching gene ID) with the highest similarity to the Uniprot sequence. Sequence similarity is calculated using KAlign v2.0, computed as the number of non-gap matches between the two sequences divided by the minimum length of pre-aligned sequences. Unlike AlphaFold3, we did not apply any sequence clustering. To generate binding DNA sequences for each protein sequence, we use the corresponding JASPAR profile's position frequency matrix (PFM) to sample 10 single-stranded motifs. For each distillation example, the inputs include the protein sequence, the single-strand DNA sequence and its corresponding reverse complement. After generating Boltz-1 predictions, we filtered examples to those that satisfied all the following conditions PDE ≤ 2.0, maximum interface predicted distance error (iPDE) ≤ 1.0 and minimum interface predicted TM-Score (ipTM) ≥ 0.7.

Table 8: **Training data composition**

Dataset	Source	Sampling Clusters	Sampling Weight
PDB	experimental	chains & interfaces	0.6
AFDB	AF2 distillation	chains	0.3
Protein-ligand	Boltz-1 distillation	interfaces	0.03
RNA	Boltz-1 distillation	chains	0.04
DNA-protein	Boltz-1 distillation	interfaces	0.03

A.1.2 Structure Prediction Test Set

For structure prediction, the test set in table 20 was constructed following Boltz-2 [Passaro et al., 2025]:

1. Initial release date is between 2024-01-01 and 2024-12-31.
2. Resolution is below 4.5Å.
3. We select all polymer chains that have less than 40% similarity to training polymer chains.
4. We select all interfaces where at least one of the two chains is dissimilar from the training chains.
5. Given these chains and interfaces, we get all the relevant full targets and always predict assembly 1.

We exclude 187 complexes since they do not fit on 40GB GPUs.

A.2 Cropping

Each sampled training entry is randomly cropped. While AlphaFold3 [Abramson et al., 2024] alternates between contiguous, spatial, and interface-spatial cropping, we use the single strategy of Boltz-1 [Wohlwend et al., 2025] tailored to biomolecule design. Algorithm 4 shows how training tokens T are processed: given a center index c determined by the sampling cluster type (Table 8), we iteratively add contiguous fragments of size W from chains nearest to the center until reaching the target crop size L. The maximum crop size is 768 for folding, matching AlphaFold3 and Boltz-2, and 512 for design tasks to accommodate additional memory for fake atoms.

Algorithm 4: CROPNEIGHBORHOOD

Input: tokens T, centre index c, fragment-size W, maximum length L
Output: *cropped* — indices centered on c, grown in W-sized fragments, truncated at L

```
/*order residues by spatial proximity to the center                    */
ordered ← indices of residues sorted by ‖ T[i].x − T[c].x ‖;
cropped ← ∅;
for i ∈ ordered do
    chain_members ← { j | T[j].chain_id = T[i].chain_id };
    if |chain_members| ≤ W then
        block ← chain_members ;                          // short chain - keep all
    else
        block ← contiguous window in chain_members centered on i,
                expand left/right until |block| ≥ W;
    end
    if |cropped| + |block| > L then
        break ;                                          // length budget reached
    end
    cropped ← cropped ∪ block;
end
return cropped (sorted);
```

A.3 Training Tasks

During training, we select different parts of the data sample to be designed. We do this according to different training tasks, which correspond to common use cases such as binder design, or motif scaffolding, as described in Sec. 3.3. The tasks are outlined in Table. 9.

Table 9: **Training Tasks Descriptions**

Design Task	Training Task	Description	Alg.
Folding	Folding	No design residues selected	5
Binder Design	Protein Chains	Select a protein chain to be designed	7
	Protein Interfaces	Select residues in a protein chain at the interface with another protein chain	11
	Non-Protein Interfaces	Select residues in a protein chain at the interface with a non-protein chain	10
Motif Scaffolding	Scaffolding	Select all residues in a crop	8
	Motif	Select all residues except a crop	9
Unconditional Design	Standard Protein	Select all protein residues	6

Each task is sampled with a certain probability depending on the data sample. These sampling probabilities are given in Table 10.

In addition to selecting which residues will be designed, we also sample other conditioning features, such as binding site specifications. The procedures for each feature are listed in Table 11 along with references to algorithmic descriptions.

Table 10: **Training Tasks Distribution**

Task	Condition				
		0 Non-Protein		> 0 Non-Protein	
	0 Protein	1 Protein	> 1 Protein	1 Protein	> 1 Protein
Folding (Alg. 5)	1	0.1	0.05	0.05	0.05
Scaffolding (Alg. 9)	0	0.5	0.2	0.2	0.2
Motif (Alg. 8)	0	0.3	0.15	0.15	0.1
Non-Protein Interface (Alg. 10)	0	0	0	0.2	0.05
Standard Protein (Alg. 6)	0	0.1	0.1	0.4	0.1
Protein Interfaces (Alg. 11)	0	0	0.1	0	0.1
Protein Chains (Alg. 7)	0	0	0.4	0	0.4

Table 11: **Conditioning Inputs Sampling**

Input Feature	Description	Mode	Weight	Alg.

Binding Site	Which residues are part of the binding site, not part of the binding site, or unspecified.	`binding`
		`not_binding`
		`both`
		`none`
			0.15
			0.075
			0.075
			0.70
				Alg. 12

Pairwise Distances	Which pairwise distances are given as input to the model. Used, for example, to specify the structure of the target.	`all`
		`uniform`
		`crops`
			0.40
			0.30
			0.30
				Alg. 13

Secondary Structure	Which residues are part of an alpha helix, beta-sheet, loop, or unspecified.	`all` `uniform`	0.50 0.50	Alg. 14

Algorithm 5: SELECT_NONE

Input: tokens T
Output: updated T with `T.design_mask` (1 = redesign)
`T.design_mask` $\leftarrow 0$
return T;

Algorithm 6: SELECT_STANDARD_PROT

Input: tokens T
Output: updated T with `T.design_mask`
foreach *residue index* i **do**
  if $T[i]$.*is_protein* **and** $T[i]$.*is_standard* *(not modified)* **then**
    `T[i].design_mask` $\leftarrow 1$;
  end
end
return T;

Algorithm 7: SELECT_PROTEIN_CHAINS

Input: tokens T
Output: updated T with `T.design_mask`
$chain\_ids \leftarrow \text{unique}\{\, T[i].chain\_id \mid T[i].\texttt{is\_protein} \text{ and } T[i].\texttt{is\_standard} \,\}$;
$chosen\_ids \leftarrow \text{np.random.choice}(chain\_ids, \text{ size} = \text{np.random.randint}(1, |chain\_ids|))$;
foreach *residue* i *with* $T[i].chain\_id \in chosen\_ids$ **do**
  `T[i].design_mask` $\leftarrow 1$;
end
return T;

Algorithm 8: SELECT_MOTIF

Input: tokens T (one per residue), fragment-size K (desired motif width
Output: updated T with `T.design_mask`

/* canonical protein residues */
$protein\_std \leftarrow \{\, i \mid T[i].\texttt{is\_protein}$ and $T[i].\texttt{is\_standard} \,\}$;
/* decide maximum motif length */
$max\_len \leftarrow \max\big(\text{np.random.randint}(|protein\_std|),\ fragment\_size + 1\big)$
/* choose a central residue and gather a contiguous window around that */
$center\_token \leftarrow \text{np.random.choice}(protein\_std)$;
$crop\_set \leftarrow \textsc{CropNeighborhood}\big(\text{tokens} = T,\ \text{center} = center\_token,\ \text{window} = fragment\_size,\ \text{limit} = max\_len\big)$;
$T[crop\_set].\texttt{design\_mask} \leftarrow 1$;
return T;

Algorithm 9: SELECT_SCAFFOLD

Input: tokens T, fragment-size set K
Output: updated T with `T.design_mask`

/* first select a motif based on SELECT_MOTIF (Alg. 8), then set the remaining standard protein residues as the scaffold to be designed. */
$protein\_std \leftarrow \{\, i \mid T[i].\texttt{is\_protein}$ and $T[i].\texttt{is\_standard} \,\}$;
$max\_len \leftarrow \max\big(\text{np.random.randint}(|protein\_std|),\ fragment\_size + 1\big)$
$center\_token \leftarrow \text{np.random.choice}(protein\_std)$;
$crop\_set \leftarrow \textsc{CropNeighborhood}\big(\text{tokens} = T,\ \text{center} = center\_token,\ \text{window} = fragment\_size,\ \text{limit} = max\_len\big)$;
/* design *scaffold* part: all protein-standard residues *outside* the motif */
$T[\,protein\_std \setminus crop\_set\,].\texttt{design\_mask} \leftarrow 1$;
return T;

Algorithm 10: SELECT_NONPROT_INTERFACE

Input: tokens T
Output: updated T with `T.design_mask`

/* pick 1+ non-protein chains as the target and design the k closest standard protein residues at the interface. */
$nonprot\_ids \leftarrow \text{unique}\big\{\, T[i].chain\_id \mid T[i].\texttt{is\_protein} = \text{False} \,\big\}$
$target\_ids \leftarrow \text{np.random.choice}(nonprot\_ids,\ \text{size} = \text{np.random.randint}(1, |nonprot\_ids|+1)$;
$candidates \leftarrow \{\, i \mid T[i].\texttt{is\_protein}$ and $T[i].\texttt{is\_standard} \,\}$;
foreach $i \in candidates$ **do**
  $d[i] \leftarrow \min_{j:T[j].chain\_id \in target\_ids}\Big(\big\| T[i].\texttt{center\_coords} - T[j].\texttt{center\_coords} \big\|_2 + \mathcal{N}(0,\sigma) \Big)$;
end
$order \leftarrow candidates$ sorted by d;
$k \leftarrow \text{np.random.randint}(1, |order|+1)$;
`T[order[:k]].design_mask` $\leftarrow 1$;
return T;

Algorithm 11: SELECT_PROTEIN_INTERFACES

Input: tokens T
Output: updated T

/* Select 1+ protein chains and mark the k standard residues on those chains that lie closest to other protein chains to be designed. */

$prot\_chain\_ids \leftarrow$ unique $\{T[i].chain\_id \mid T[i].\texttt{is\_protein}$ and $T[i].\texttt{is\_standard}\}$;
$redesign\_chain\_ids \leftarrow$ np.random.choice($prot\_chain\_ids$, size = np.random.randint$(1, |prot_chain_ids|)$);
$Redesign \leftarrow \{\, i \mid T[i].chain\_id \in redesign\_ids$ and $T[i].\texttt{is\_standard}\,\}$;
$Target \leftarrow \{\, i \mid T[i].chain\_id \notin redesign\_ids\,\}$;
foreach $i \in Redesign$ **do**
  $d[i] \leftarrow \min_{j \in \text{Target}} \Big(\left\| T[i].\texttt{center\_coords} - T[j].\texttt{center\_coords} \right\|_2 + \mathcal{N}(0, \sigma) \Big)$;
end
$order \leftarrow Redesign$ sorted by d;
$k \leftarrow$ np.random.randint$(1, |order|+1)$;
`T[order[:k]].design_mask` $\leftarrow 1$;
return T;

Algorithm 12: SPECIFY_BINDING_SITE

Input: tokens T
Output: updated T with T.`binding_type` set on target residues
$design \leftarrow \{\, i \mid T[i].\texttt{design\_mask} = 1\,\}$;
$target \leftarrow \{\, i \mid T[i].\texttt{design\_mask} = 0\,\}$;

/* compute atom-atom contacts between every target token and all design-token atoms */

foreach $i \in target$ **do**
  $is\_atomic\_contact[i] \leftarrow \exists\, a \in \text{atoms}(T[i]),\, b \in \text{atoms}(design)$ s.t. $\|a - b\| < 5\,\text{Å}$;
end
$contact\_targets \leftarrow \{\, i \in target \mid is\_atomic\_contact[i] = 1\,\}$;

$mode \leftarrow$ random.choice([`binding`, `not_binding`, `both`, `none`], p = [0.15, 0.075, 0.075, 0.70]);
if $mode \in \{\texttt{binding}, \texttt{both}\}$ **then**
  $S \leftarrow$ random nonempty subset of $contact\_targets$;
  foreach $i \in S$ **do**
    $T[i].\texttt{binding\_type} \leftarrow$ BINDING;
  end
end
if $mode \in \{\texttt{not\_binding}, \texttt{both}\}$ **then**
  $U \leftarrow$ random nonempty subset of $(target \setminus contact\_targets)$;
  foreach $i \in U$ **do**
    $T[i].\texttt{binding\_type} \leftarrow$ NOT_BINDING;
  end
end
return T;

Algorithm 13: SPECIFY_STRUCTURE_GROUPS (pairwise–distance conditioning)

Input: tokens T
Output: updated T with T.`structure_group` set for target residues

/* we set structure_group to drive pairwise-distance conditioning: pairs with group 0 receive *no* distances; pairs whose residues share a nonzero group (≥ 1) have their pairwise distance specified */

$target \leftarrow \{\, i \mid T[i].\texttt{design\_mask} = 0 \,\}$;
$chain\_ids \leftarrow$ unique $T[i].\texttt{chain\_id}$ for $i \in target$;

/* pick one or more target chains to specify structure group */
$specified \leftarrow$ random subset of $chain\_ids$ of size randint$[1, |chain_ids|]$;

/* for each chosen chain, select all tokens or sub-regions to specify the structure */
$subsets \leftarrow \varnothing$;
foreach $c \in specified$ **do**
  $tokens_c \leftarrow \{\, i \in target \mid T[i].\texttt{chain\_id} = c \,\}$;
  $mode \leftarrow$ random.choice([`all`, `uniform`, `crops`], p $= [0.40, 0.30, 0.30]$);
  if $mode =$ *`all`* **then**
    /* specify the entire set of target tokens on the chain */
    $subsets \leftarrow subsets \cup [\, tokens_c \,]$;
  else if $mode =$ *`uniform`* **then**
    /* split the chain's target tokens into contiguous segments */
    $m \leftarrow$ randint$[1, \min(6, |tokens_c|)]$;
    split $tokens_c$ into m contiguous segments;
    $subsets \leftarrow subsets \cup [\,$each segment$\,]$;
  else `crops`
    /* take several spatially local crops grown around random centres */
    $R \leftarrow$ randint$[2, 4]$;
    for $r = 1$ **to** R **do**
      $center\_token \leftarrow$ random element of tokens not yet chosen in any crop;
      $crop\_set \leftarrow$ CROPNEIGHBORHOOD(tokens $= T$, center $= center\_token$, window $= fragment\_size$, limit $=$ |tokens not yet chosen in any crop|);
      $subsets \leftarrow subsets \cup [\, crop\_set \,]$;
    end
  end
end

/* assign frame IDs: sample num_groups and give each subset a random ID in $\{1, \ldots, num\_groups\}$ (group 0 means no distances) */
$num\_groups \leftarrow$ randint$[1, |subsets|]$;
foreach $S \in subsets$ **do**
  $g \leftarrow$ random choice in $\{1, \ldots, num\_groups\}$;
  $\forall i \in S:\ T[i].\texttt{structure\_group} \leftarrow g$;
end
return T;

Algorithm 14: SPECIFY_SECONDARY_STRUCTURE_MASK

Input: tokens T
Output: updated T with T.`design_ss_mask` set for designed residues

```
/* design_ss_mask controls secondary structure conditioning:  1 means condition on the secondary
   structure for a designed residue                                                          */
```

$designed \leftarrow \{\, i \mid T[i].\texttt{design\_mask} = 1 \,\}$;

```
/* select a mode for how much SS to reveal                                                   */
```

$mode \leftarrow$ random choice $\in \{\texttt{all}, \texttt{uniform}\}$;
if $mode = \texttt{all}$ **then**
  $\forall i \in designed:\ T[i].\texttt{design\_ss\_mask} \leftarrow 1$;
else `uniform`
  `/* partition into contiguous intervals; randomly choose to reveal SS or not for each interval */`
  $m \leftarrow \text{randint}[1, |T|]$;
  split the token index range $[1..|T|]$ into m contiguous intervals $I_1, \ldots, I_m$;
  for $k = 1$ **to** m **do**
    with prob $1/2$ set $T[i].\texttt{design\_ss\_mask} \leftarrow 1$ for all $i \in (designed \cap I_k)$;
  end
end
return T;

A.4 Details about Computed Metrics

Table 12 provides a comprehensive reference for all metrics computed by the BoltzGen pipeline.

Table 12: BoltzGen metrics reference.

Metric Name	Description
1 Design quality metrics	
1.1 Predicted structure quality	
`ptm`	Predicted TM-score, measuring overall structure quality (higher is better).
`iptm`	Predicted TM-score across chain pairs, measuring overall complex stability (higher is better).
`design_ptm`	Predicted TM-score for the designed structure, measuring how well the designed structure folds (higher is better).
`design_iptm`	Predicted TM-score for interactions between the entire design chain and target, measuring overall binding interface quality (higher is better).
`design_to_target_iptm`	Predicted TM-score for interactions between only the designed residues (not the full chain) and target, measuring specific binding interface quality (higher is better).
`design_iiptm`	Predicted TM-score for interactions between design residues that are within 8 Å of target atoms and any target residues, measuring interface interaction quality (higher is better).
`design_ptm>[threshold]`	Binary flag for `design_ptm` above threshold (1 = pass, 0 = fail), where threshold is 0.75 or 0.8.
`design_iptm>[threshold]`	Binary flag for design `design_iptm` above threshold (1 = pass, 0 = fail), where threshold is 0.5, 0.6, 0.7, or 0.8.
`interaction_pae`	Predicted aligned error for all design–target interactions (lower is better).

Continued on next page

Metric Name	Description
`min_design_to_target_pae`	Minimum Predicted Aligned Error (PAE) between any design and target residue pair, indicating the most confidently predicted contact (lower is better).
`neg_min_design_to_target_pae`	Negative of `min_design_to_target_pae` for ranking purposes (higher is better).
1.2 Designability / refolding accuracy	
`filter_rmsd`	Root mean square deviation used for filtering, either backbone RMSD (`from_inverse_folded=True`) or all-atom RMSD (lower is better).
`filter_rmsd_design`	RMSD of the designed structure only, used for filtering (lower is better).
`designfolding-filter_rmsd`	RMSD when refolding the design in isolation (without target), ensuring design stability (lower is better).
2 Interaction metrics	
2.1 Binding interface analysis	
`plip_hbonds_refolded`	Number of hydrogen bonds between design and target in refolded structure (higher is better).
`plip_saltbridge_refolded`	Number of salt bridge interactions between design and target in refolded structure (higher is better).
2.2 Binding site adherence	
`bindsite_under_[cutoff]rmsd`	Fraction of binding site residues within [cutoff] Å of designed residues, where [cutoff] is 3, 4, 5, 6, 7, 8, or 9 (higher is better).
3 Solvent accessibility and hydrophobicity	
3.1 Surface area analysis	
`delta_sasa_refolded`	Change in solvent accessible surface area when binder is present vs absent, computed on refolded structure (higher indicates better burial).
`delta_sasa_original`	Change in solvent accessible surface area when binder is present vs absent, computed on original structure (higher indicates better burial).
3.2 Hydrophobicity metrics	
`design_chain_hydrophobicity`	Hydrophobicity score of the entire designed chain sequence.
`design_hydrophobicity`	Hydrophobicity score of only the designed residues.
`neg_design_hydrophobicity`	Negative of `design_hydrophobicity` for ranking purposes .
`design_largest_hydrophobic_patch_refolded`	Area of the largest hydrophobic patch in the refolded design structure (lower is better for solubility).
4 Sequence composition and structure	
4.1 Amino acid composition	
`num_design`	Number of designed residues in the sequence.
`[amino_acid]_fraction`	Fraction of specific amino acid residues in the designed sequence.
`UNK_fraction`	Fraction of unknown (X) residues in the designed sequence (0 preferred to avoid unknown residues).
4.2 Secondary structure	
`loop`	Fraction of residues in loop conformation (0–1 scale).
`helix`	Fraction of residues in helical conformation (0–1 scale).
`sheet`	Fraction of residues in β-sheet conformation (0–1 scale).

Continued on next page

Metric Name	Description
5 Liability analysis	
5.1 Overall liability scores	
liability_score	Overall developability score combining all liability assessments (lower is better).
liability_num_violations	Total number of liability violations detected in the sequence (lower is better).
liability_high_severity_violations	Number of high-severity liability violations (lower is better).
liability_medium_severity_violations	Number of medium-severity liability violations (lower is better).
liability_low_severity_violations	Number of low-severity liability violations (lower is better).
liability_violations_summary	Human-readable summary of all liability violations detected.
liability_details	Consolidated details string combining all motif-specific liability information.
5.2 Specific liability motifs	
liability_[motif]_count	Number of instances of the specific liability motif found (lower is better). Examples: HydroPatch detects hydrophobic patches like "FILVWY", DPP4 detects cleavage sites like "AP", MetOx detects methionine oxidation sites.
liability_[motif]_position	Position of the first occurrence of the motif in the sequence (residue index).
liability_[motif]_length	Length of the liability motif in residues.
liability_[motif]_severity	Severity score for this motif instance (lower is better). Examples: HydroPatch severity increases with patch size (3+ consecutive hydrophobic residues = high severity), MetOx has moderate severity, DPP4 cleavage sites have high severity.
liability_[motif]_details	Specific details about the motif violation.
liability_[motif]_positions	All positions where this motif occurs (comma-separated).
liability_[motif]_num_positions	Total number of positions where this motif occurs.
liability_[motif]_global_details	Global context details for this motif.
liability_[motif]_avg_severity	Average severity across all instances of this motif.
7 Affinity prediction (small molecule binders)	
affinity_probability_binary1	Probability of binary binding classification (higher is better).
8 Filtering and ranking metrics	
8.1 Aggregated ranking and filtering	
final_rank	Final ranking position after quality and diversity optimization (1 = best).
num_filters_passed	Number of filter criteria that the design passed.
pass_filters	Binary flag indicating whether design passed all filters (1 = pass, 0 = fail).

B Additional Computational Results and Details

B.1 Calibrating Filtering Algorithm for Protein-Protein Complexes

The relative importance of the Boltz-2 confidence metrics and interaction metrics used to rank designs is calibrated on a benchmark of 11,000 validated binders across 11 target proteins, based on data

from Cao et al. [2022]. These weights serve as default values and are manually adjusted for wetlab design experiments based on domain expert feedback.

Binder designs selection benchmark. For each of the 11 targets (InsulinR, FGFR2, EGFR, H3, IL7Ra, PDGFR, SARS-CoV-2 RBD, TGFb, Tie2, TrkA, and VirB8), we sample up to 100 positive examples (i.e., 4μM binders) and fill the remainder up to 1,000 designs with negative examples (i.e., non-binders), resulting in a balanced subset of 11,000 designs.

When exploring different methods to prioritize designs, we optimize the mean enrichment factor at top 25 and top 50 designs. Following Zambaldi et al. [2024], due to the high variance of metric values across targets, we do not directly optimize the mean enrichment score. Instead, we optimize the mean rank across targets, where each target's rank is based on its individual enrichment value. When computing enrichment factors, we normalize by the original binder-to-non-binder ratio from the full dataset, rather than our 11,000-design subset.

Binder designs selection method. In our initial experiments, we trained a decision tree to predict binary binder labels based on the given metric values. However, we find that learning decision thresholds for individual metrics is not optimal, as the best threshold values can be specific to the target protein (for example, optimal interaction count metric thresholds can vary depending on protein size). Instead of using absolute thresholds, we develop a design selection scheme that prioritizes candidates with the best worst-case ranks across all metrics (Algorithm 2). Each metric is assigned a single learnable weight from the set {0, 1, 1.2, 1.5, 2}, where 0 means that the metric is not used. We calibrate these weights by maximizing the enrichment factor on our benchmark of 11,000 experimentally validated binder designs. We do not run this optimization over all metrics but only over a representative subset of non-correlated ones (Figure 22). We include both `design_iiptm` and `neg_min_design_to_target_pae` despite their high correlation, as variations of these two metrics have been shown to be complementary [Zambaldi et al., 2024].

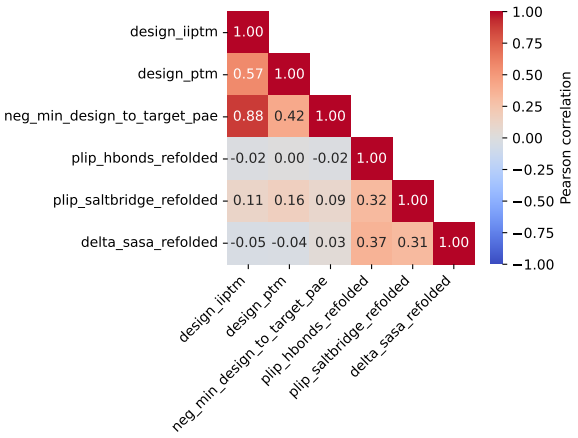


Figure 22: **Final Metrics Used For Binder Design Filtering And Their Pairwise Correlations.** Correlations were calculated on 11,000 BoltzGen-designed binders for 11 targets from our benchmark derived from Cao et al. [2022].

Our enrichment factor optimization, combined with wetlab experimental feedback, yields the following final combination of metrics and weights for Algorithm 2: `design_iiptm`: 1, `design_ptm`: 2, `neg_min_design_to_target_pae`: 1, `plip_hbonds_refolded`: 2, `plip_saltbridge_refolded`: 2, and `delta_sasa_refolded`: 2. When designing small-molecule binders, we slightly modify the weights of Boltz-2 metrics: `design_iiptm`: 1.1, `design_ptm`: 1.1, `neg_min_design_to_target_pae`: 1.1, `plip_hbonds_refolded`: 2, `plip_saltbridge_refolded`: 2, and `delta_sasa_refolded`: 2.

B.2 Baseline methods

RFdiffusion. We employ RFdiffusion [Watson et al., 2023] as a baseline for binder generation, using the official implementation (`https://github.com/RosettaCommons/RFdiffusion`). We apply the standard settings (`diffuser.T=100`) and reduce the inference noise to improve design quality (`denoiser.noise_scale_ca=0`, `denoiser.noise_scale_frame=0`), following the configuration used in the official binder design example (`https://github.com/RosettaCommons/RFdiffusion/blob/main/examples/design_ppi.sh`).

We use ProteinMPNN [Dauparas et al., 2022] for inverse folding, as implemented in the official

LigandMPNN [Dauparas et al., 2025] repository (https://github.com/dauparas/LigandMPNN). We use the standard checkpoint `proteinmpnn_v_48_020.pt`. Side-chain packing is performed with the following settings: `pack_side_chains=1`, `number_of_packs_per_design=1`, and `pack_with_ligand_context=1`.

RFdiffusionAA. We use RFdiffusion All-Atom (RFdiffusionAA) [Krishna et al., 2024] as a baseline for generating protein binders against small molecules. We employ the official implementation (https://github.com/baker-laboratory/rf_diffusion_all_atom) with a standard configuration (`diffuser.T=150`, `inference.ckpt_path=RFDiffusionAA_paper_weights.pt`). Inverse folding is performed using LigandMPNN [Dauparas et al., 2025], as described above for RFdiffusion.

BoltzGen only requires a SMILES representation of the input small molecule and performs cofolding during the design process. In contrast, RFdiffusionAA uses a fixed ligand structure to generate a protein binder. To ensure a fair comparison between the two methods, we generate ligand conformers for RFdiffusionAA using RDKit [Bento et al., 2020]. Specifically, inspired by DiffDock [Corso et al., 2022], we employ the `AllChem.ETKDGv3` algorithm and, if it fails, fall back to initializing random coordinates followed by optimization with `AllChem.MMFFOptimizeMolecule`. We generate a single conformer per input ligand, as we observe that increasing the number of conformers to diversify designs has no significant impact on RFdiffusionAA performance.

B.3 BoltzIF Inverse Folding Model

We verify whether BoltzIF behaves similarly to Protein MPNN (PMPNN) and Soluble (SMPNN) on a set of 64 monomer targets from the PDB. We evaluate each model's ability to inverse fold both native and designed structures. For native ability, we inverse fold the targets themselves and evaluate 50 sequences for each one. For designed ability, we generate 50 binders for each target with BoltzGen and evaluate 1 inverse-folded sequence per design.

Method	Native Backbones		Designed Backbones	
	RMSD < 2.5	Hydrophobic	RMSD < 2.5	Hydrophobic
ProteinMPNN	0.55	1343.30	0.31	1448.65
SolubleMPNN	0.55	1025.72	0.33	1143.31
BoltzIF	0.55	1185.98	0.32	1400.73

Table 13: **Inverse Folding Model Comparison.** "RMSD<2.5" denotes the success rate with which designed sequences refold into the inverse folded structure (using Boltz-2). "Hydrophobic" indicates the surface area of the inverse folded protein's largest hydrophobic patch.

Table 13 reports both the backbone designability of the refolded sequences to the original structures as well as the size of the largest hydrophobic patch, which is relevant for protein expressibility. We see that BoltzIF attains the same designabilities as ProteinMPNN and SolubleMPNN and its hydrophobicity scores fall between the two models.

Fig. 23 shows the amino acid distribution over all sequences from each of the models. BoltzIF's residue frequencies mostly fall between that of ProteinMPNN and SolubleMPNN. Likely, the reason for the higher hydrophobicity scores than SolubleMPNN is that it has been trained on crops (hydrophobic cores can be viewed as solvent-facing when it is the surface of a crop).

B.4 Motif Scaffolding Benchmark Results

We ran the following Motif-scaffolding performance benchmark performed in [Geffner et al., 2025]:

- For each motif scaffolding task, we generate 1000 backbones.

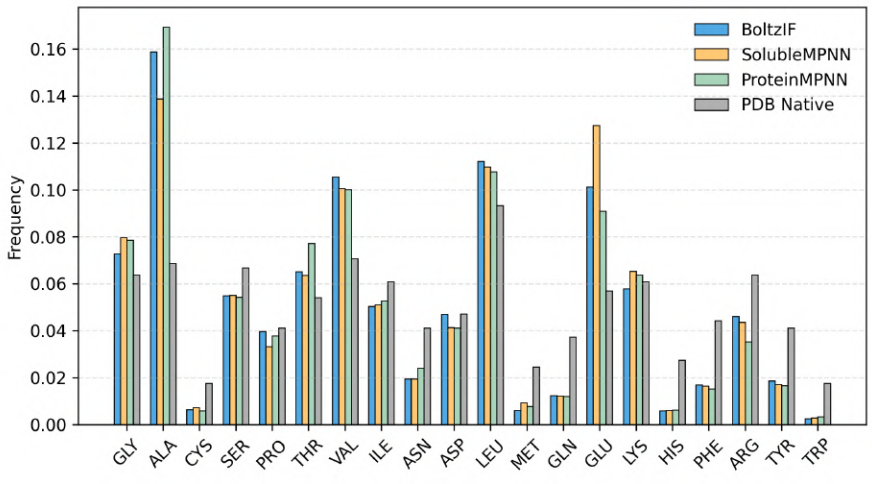


Figure 23: Amino acid distributions when inverse folding BoltzGen's designed binders against 64 monomers in PDB. "PDB Native" denotes the amino acid distribution in our PDB training data.

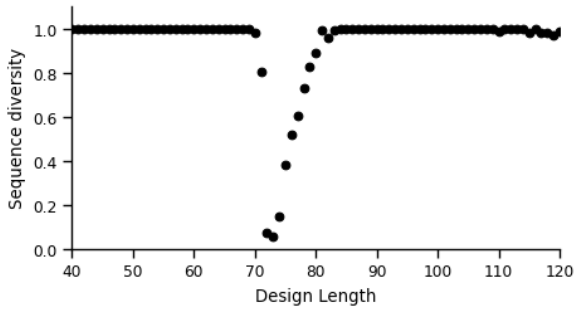


Figure 24: **The Ubiquitin Memorization Issue.** Shown is the sequence diversity (number of unique sequences divided by total number of designs) for designs with varying lengths against 9 targets (33,600 designs). The gap in the 73-76 region stems from BoltzGen's bias toward Ubiquitin in that length range. The bias likely stems from Ubiquitin's overrepresentation in the training data (>1000 entries in the PDB).

- For each backbone, 8 sequences are generated by ProteinMPNN with fixed sequences in the motif region.
- All 8 sequences are refolded via ESMfold and the C_α-RMSD and the motifRMSD are computed between the ground truth and the prediction.
- A backbone is categorized as a success when one of the ProteinMPNN sequences satisfy C_α-RMSD ≤ 2Å, motifRMSD ≤ 1Å, pLDDT ≥ 70, and pAE ≤ 5.
- Hierarchical clustering with single linkage and TM-score threshold 0.6 is performed on all successful backbones to get a clustering to get the final unique successes.

BoltzGen has the highest number of sole best method in 8 tasks (compared to Proteína that wins in 6 tasks).

B.5 Memorization of Ubiquitin

In a few design campaigns, we observed diminished sequence diversity and low filter pass rates for BoltzGen minibinder designs in the 73-76 amino acid length range. This is visualized in an analysis of

Task Name	BoltzGen	Proteina	Genie2	RFDiffusion	FrameFlow
6E6R_short	**83**	56	26	23	25
5TRV_med	**25**	22	23	10	21
5YUI	**36**	5	3	1	1
6EXZ_short	**11**	3	2	1	3
5TRV_short	**6**	1	3	1	1
4JHW	**3**	0	0	0	0
5IUS	**3**	1	1	1	0
1PRW	**2**	1	1	1	1
6E6R_long	289	**713**	415	381	110
6EXZ_long	59	290	326	167	**403**
6E6R_medium	164	**417**	272	151	99
1YCR	102	**249**	134	7	149
5TRV_long	155	**179**	97	23	77
4ZYP	1	**11**	3	6	4
6EXZ_med	32	43	54	25	**110**
7MRX_128	64	51	27	**66**	35
7MRX_85	22	**31**	23	13	22
3IXT	10	8	**14**	3	8
5TPN	3	4	**8**	5	6
7MRX_60	4	2	**5**	1	1
1QJG	0	3	5	1	**18**
1BCF	1	1	1	1	1
5WN9	2	2	1	0	**3**
2KL8	1	1	1	1	1

Table 14: Number of unique successes on the RFDiffusion benchmark for BoltzGen and 4 other methods, for 1000 backbones.

33,600 designs against 5 protein and 4 small molecule targets in Figure 24. Inspection of the sequences revealed that the pipeline (backbone design followed by inverse folding) is frequently recapitulating the sequence for ubiquitin in this length range. For example, in this analysis all designs of length 73 (n=156) had >97% sequence identity to "MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP-PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRL".

Likely, this arises since this sequence is present, often in complex, in >1000 entries in the PDB. In future versions of the model, we plan to down-sample this interaction during training.

C Learnings from BoltzGenv0

Template bug. A previous version of BoltzGen, which we term BoltzGenv0, had a serious flaw resulting in close-to-random ranking and filtering. We can judge the impact of the bug based on the nanobody design results in Section 2.7, where, with the current version of BoltzGen, we obtain a 1/7 hit rate for the Penguinpox target and a 7/7 hit rate against hemagglutinin. BoltzGenv0's hit rates were 0 for both targets. The same failure case occurred for an attempt to design helicons against a pMHC complex.

The nature of this bug was in our handling of which residues are considered to be part of the target and which part of the design. In cases where the designed binder contains fixed residues, such as when designing helicons or nanobodies, BoltzGenv0 considered the fixed residues of the designs as being part of the target. The implications of this are that their relative position with respect to the target is provided in the refolding step via the templates that we employ for the targets. Thus, the resulting structure prediction is bound to recapitulate the generated structure, without providing any filtering power that enriches for binders. Furthermore, metrics such as the minimum interaction pAE and the ipTM would be influenced by the fixed residues that are in the design. For instance, in the overwhelming

majority of cases, the minimum interaction pAE would not correspond to an interaction between the design and the target, but rather an interaction between a designed residue and a fixed residue in the designed binder. Hence, these scores also do not provide any filtering power for BoltzGenv0.

Designfolding. Another improvement in BoltzGen over BoltzGenv0 is its "designfolding" step. In this step, we additionally refold the design in the absence of the target and compute the RMSD to the design in the generated structure. This serves as a proxy to assess whether the binder could attain the designed structure by itself, which we use to filter out designs that are likely to require a significant conformational change upon binding or do not express since they do not fold into a stable structure by themselves.

We introduced this improvement after a protein-protein binder design attempt where all 12 of BoltzGenv0's designs failed to express. These designs would often partially or completely "envelop" the target and require a large conformational change to bind.

D Wetlab Experimental Method Details

D.1 Target Selection Process for 10 Hard Adaptyv Bio Targets

For assembling a panel of hard protein targets for Adaptyv Bio to test experimentally, we use the following criterion:

1. **PDB Monomer** — The biological assembly has to be a monomer with exactly one protein polymer instance (`oligomeric_state = Monomer`, `polymer_entity_instance_count_protein = 1`).
2. **Monomer-only sequence cluster** — Each chain is either a singleton or a member of a sequence cluster (30% identity threshold, `mmseqs easy-cluster` with `-min-seq-id 0.30` and `-c 0.0`) in which *every* member is a PDB monomer (satisfies Condition 1).
3. **Catalog availability** — Must be available in the *Sino Biological* catalog (via mapping to a Swiss-Prot accession listed there).

With this selection method, we aim to ensure that each protein we keep is a **monomeric PDB entry** that has no close sequence homolog (MMseqs2 sequence identity $\geq 30\%$) anywhere in the PDB that appears in a multimeric or ligand-bound assembly. This makes the targets genuinely "hard" for our binder design: the model has not seen a closely related protein in a bound context during training.

We verify our target's sequence identities $<30\%$ to any non-monomeric protein in PDB as follows. For each target sequence, we ran MMseqs2 easy-search against non-monomer PDBs and keep the top identity hit that passes our coverage filter:

```
mmseqs easy-search queries.fa nonmonomer_db out.m8 tmp \
  --threads 32 \
  --min-seq-id 0.0 \
  --alignment-mode 3 \
  -e 1e5 -s 9.5 \
  --prefilter-mode 2 \
  --cov-mode 2 -c 0.9 \
```

- **Coverage policy.** We require *near-global coverage on the target* via `-cov-mode 2 -c 0.9`, i.e.

$$\text{tcov} = \frac{\text{aligned residues}}{\text{target length}} \geq 0.9.$$

  This enforces that the match spans (almost) the entire *target* chain. We initially used `-c 1.0`, but very long targets often fail to align end-to-end stably; we therefore relaxed to 0.9.

- **Zeros in the table.** Some entries appear as 0.0% (e.g., **1JQD**, **2A1X**). This indicates that no hit met the near-global coverage threshold, not that a full-length alignment with 0% identity was found. These two sequences are also the longest among the ten (292 and 308 aa), making full-length alignment to known non-monomeric chains particularly hard under tcov ≥ 0.9.

Table 15: **Maximum sequence identity** (against non-monomeric PDB chains) for the selected targets.

Target	Max sequence identity (%)
1G13	25.4
1JQD	0.0
1NB0	24.6
2A1X	0.0
2PNY	23.0
3APU	19.8
3CH4	22.7
3QKG	27.9
7AAH	24.3

D.2 BLI and SPR Details for Sections "Designing Nanobodies and Proteins against 9 Novel Targets" and "Designing Nanobodies and Proteins against 5 Benchmark Targets"

The binding affinity assays were carried out by the contract research organization Adaptyv Bio.

Biolayer Interferometry (BLI) affinity characterization Ligand constructs were designed by reverse-translating target protein sequences and optimizing codon usage for expression in a prokaryotic cell-free system. A C-terminal assay tag was included to facilitate capture on biosensors. Gene fragments (Twist Bioscience) were assembled using NEBuilder HiFi DNA Assembly (NEB) in 2 μL reactions. Assembly products were validated via capillary electrophoresis (Agilent ZAG DNA Analyzer) and quantified using the Qubit dsDNA assay (Invitrogen).

Proteins were expressed in 8 μL reactions using an optimized in vitro transcription/translation system supplemented with 4 nM DNA template. Reactions were incubated at 37 °C for 8 hours. After expression, total protein levels were measured using an affinity-based detection assay and normalized across samples prior to binding analysis.

BLI experiments were carried out using a Gator Bio instrument with Strep-Tactin XT biosensors. Twin-Strep-tagged ligands were captured on the biosensors using the following protocol:

- **Baseline 1:** 120 s in running buffer
- **Ligand loading:** 120 s (target shift 0.5–1.0 nm)
- **Baseline 2:** 200 s in running buffer

The running buffer consisted of 50 mM HEPES, 100 mM NaCl, and 0.5% Triton X-100 at pH 7.4. All steps were performed at 25 °C with data collected at 5 Hz. Binding was assessed using a multi-cycle format with four antigen concentrations (30–1000 nM, half-log dilution series). Each kinetic cycle consisted of a 220 s association phase in antigen solution followed by a 240 s dissociation phase in running buffer. After each cycle, biosensors were regenerated with 10 mM glycine-HCl (pH 1.5) applied five times for 10 s each, followed by a wash step in running buffer to restore baseline. Buffer-only and non-binding ligand controls were used for reference subtraction and signal drift correction.

Surface Plasmon Resonance (SPR) affinity characterization SPR measurements were performed on a Carterra LSA XT system. DNA constructs encoding ligands with C-terminal Twin-Strep tags were synthesized (Twist Bioscience), assembled using NEBuilder HiFi DNA Assembly, and validated using capillary electrophoresis and Qubit fluorometry. Proteins were expressed in a prokaryotic in vitro translation system and normalized post-expression using an affinity-based quantification assay. Sensor chip surfaces were functionalized by covalently attaching Strep-Tactin XT to a carboxymethylated surface using EDC/NHS coupling. The chip preparation procedure included:

- **Conditioning:** 50 mM NaOH
- **Activation:** EDC/NHS solution
- **Capture:** Strep-Tactin XT (50 μg/mL in 10 mM sodium acetate, pH 4.5)
- **Quenching:** 1 M ethanolamine hydrochloride, pH 8.5
- **Wash:** 0.1 M sodium borate, 1 M NaCl, pH 9.0

Twin-Strep-tagged ligands were captured on the chip using a 96-channel printhead under bidirectional flow for 750 s, followed by a 600 s baseline step in running buffer. Antigens were diluted in running buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% Tween-20, pH 7.4) and injected at seven concentrations (1–1000 nM, half-log dilution series) in a single-cycle kinetic format. Each cycle consisted of a 60 s baseline step in running buffer, a 300 s antigen association phase, and a 600 s dissociation phase in running buffer. After the completion of each injection series, the chip was regenerated with 10 mM glycine-HCl (pH 1.5) for 5 minutes, followed by a 20-minute wash in running buffer.

Data analysis and binder classification Sensorgrams from both BLI and SPR assays were analyzed using Adaptyv Fitting software. Preprocessing included trimming to relevant kinetic phases (association and dissociation), correcting for signal jumps at buffer transitions, aligning phases, and subtracting signals from both baseline and reference channels.

Data were fit to a 1:1 Langmuir binding model using a global fitting approach across all antigen concentrations. If global fits were not feasible, alternative fitting strategies—such as dissociation-only or slope-based methods—were employed. In cases where individual curve fits could not be achieved, group-level models (e.g., equilibrium, flat, or linear) were used to approximate binding behavior.

Final kinetic parameters (k_{on}, k_{off}, and K_D) were determined based on the best available fit. Under global fitting, k_{off} and K_D were estimated directly, with k_{on} calculated as k_{off}/K_D. Ligands were labeled as **binders (True)** or **non-binders (False)** based on the presence of quantifiable sensorgrams and successful kinetic model fitting. In cases where ligands generated a large signal during the association phase (at least 300% greater than the negative control) but could not be reliably fit, binding classification was assigned based on the observed magnitude of the shift.

D.3 Designing Proteins to Bind Bioactive Peptides with Diverse Structures

Experiments by A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, and William F. DeGrado

Visual Inspection The top 100 computationally ranked designs were manually examined in PyMOL. Visual inspection criteria included: (i) extent of peptide burial within the binding pocket, (ii) number and geometry of hydrogen bonds between binder and target peptide, (iii) overall packing density and complementarity at the binding interface, and (iv) internal packing quality of the apo binder in the absence of the target peptide. The top 6 designs exhibiting consistent burial, multiple well-oriented hydrogen bonds, and tightly packed interfaces were prioritized for experimental characterization.

Solid phase peptide synthesis and purification Protegrin-1 was purchased from MedChemExpress (catalog #HY-P1633)/ Melittin and Indolicidin were synthesized following the procedure below.

Melittin and Indolicidin were synthesized on a Biotage Initiator Alstra microwave synthesizer using standard Fmoc solid-phase peptide synthesis on a TentaGel S Ram resin. Resin (417 mg, 0.24 mmol) was swollen in DMF for 10-15 mins prior to synthesis. The general synthetic steps included: (a) Fmoc deprotection with 20% (v/v) piperidine in DMF, (b) resin washing (3x, DMF), (c) amino acid (0.50 mmol) coupling with DIPEA (0.50 mmol) and HCTU (0.50 mmol) for 5 min at 75 °C, (d) resin washing (3x, DMF), and (e) repeat deprotection/coupling until sequence completion.

The peptides were globally deprotected in a 10 mL solution of TFA:H2O:TIPS (95:2.5:2.5) for 3 h. The solution was then filtered, with filtrate concentrated under the flow of nitrogen. The concentrate was precipitated in cold diethyl ether (40 mL), centrifuged, and the pellet dissolved in 5 mL H2O:ACN (1:1, 0.1% TFA).

Crude peptides were purified by reverse-phase HPLC on a C18 column using H2O/ACN (0.1% TFA) at a 10 mL/min gradient of 5-100% ACN (0.1% TFA) for 50 min. Pure fractions were identified by analytical HPLC and MALDI-TOF, pooled, and lyophilized. The final products were white powders with $\geq$95% purity.

Protein expression and purification Codon-optimized genes encoding the designed candidate proteins with an N-terminal 6×His tag and a TEV protease cleavage site (HHHHHHENLYFQS) were synthesized and obtained from Twist Bioscience. To facilitate Gibson assembly into the pET-28a(+) vector, short sequences were added at the 5' end (CTCTAGAAATAATTTTGTTTAACTTTAA-GAAGGAGATATACC) and 3' end (GATCCGGCTGCTAACAAAGCCCGAAAG) of each gene. The recombinant plasmids were transformed into Escherichia coli strain E. cloni BL21(DE3). A single colony was picked from an LB agar plate and inoculated into LB medium supplemented with kanamycin (50 μg/mL) for overnight growth. The culture was then transferred into 200 mL of TB medium containing kanamycin (50 μg/mL) and incubated at 37 °C until reaching an OD600 of 0.6 - 0.8. Protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cultures were incubated overnight at 30 °C. Cells were harvested by centrifugation and resuspended in 25 mL PBS buffer (10 mM Na2HPO4, 1.8 mM KH2PO4, 2.7 mM KCl, 137 mM NaCl, pH 7.4) supplemented with 20 mM imidazole. Cells were lysed by ultrasonication (Sonic Dismembrator Model 500, Fisher Scientific), and the lysate was clarified by centrifugation (35,000g, 30 min). The supernatant was loaded onto a gravity column containing Ni-NTA agarose resin (HisPur, Thermo Fisher, 1.0 mL or 3.0 mL). The resin was washed with three column volumes (CVs) of PBS buffer containing 20 mM imidazole, and bound proteins were eluted with 7 mL PBS buffer containing 250 mM imidazole. The eluted proteins were concentrated and subjected to three rounds of buffer exchange with PBS buffer using a 15 mL, 10 kDa cutoff centrifugal filter unit (EMD Millipore).

Circular dichroism Protein samples were prepared at 10 μM in sterile filtered 10 mM sodium phosphate with 50 mM NaCl, pH 7.4. A280 was measured via UV-Vis spectroscopy in a 0.1 mm Quartz cuvette, and protein concentration was calculated from A280 via Beer's law using the extinction coefficient calculated from protein sequence via ExPasy ProtParam. Circular dichroism measurements were performed on a Jasco J-810 spectropolarimeter. Spectra were collected from 200-250nm in continuous scanning mode at 50 nm/min and 1nm band width with six accumulations per sample. CD spectra were converted from millidegrees to molar ellipticity using the equation $m \cdot M/(10 \cdot L \cdot C)$ where C is concentration in g/L (derived from A280 signal in UV-Vis experiments), M is the average molecular weight (g/mol) and L is the path length of the cell.

Analytical size exclusion chromatography The oligomeric state of binder samples was assessed via analytical size exclusion chromatography using a Superdex 75 5/150 analytical gel filtration column (Cytiva) on an AKTA FPLC. Samples (50 μL) were prepared at 100 μM in sterile filtered 1X Phosphate Buffered Saline (PBS), pH 7.4 and centrifuged in a microfuge at 21,000g for 15 minutes before loading onto the FPLC. Chromatography runs were conducted at 0.2 mL/min for 1.2 column volumes and absorbance was measured at 220 and 280 nm. For measurement of protein:peptide complexes, protein and peptide were mixed at a 1:1 ratio (50 μM each) and incubated overnight at 4 °C. After equilibration to room temperature, samples were centrifuged at 21,000g for 15 minutes before loading onto the FPLC.

Change in intrinsic tryptophan fluorescence for *in vitro* assessment of peptide binding *In vitro* binding was assessed via tryptophan quenching in which either peptide or protein was held constant with the other binding partner varied depending on which species contained tryptophan residues. All peptides were solubilized in DMSO at > 1mg/mL prior to dilution into sterile filtered 1X PBS, pH 7.4 for binding experiments. All binding assays were conducted in PBS in non-binding 96-well half-area black plates (Corning 3686) and tryptophan quenching was measured as endpoint fluorescence intensity measurements ($\lambda_{ex} = 295$ nm, $\lambda_{em} = 330$ nm) in a BioTek Synergy Neo-2 multi-mode plate reader. For melittin and indolicidin, which both contain tryptophan residues, assays were conducted with constant [peptide], and binder sequences were designed to exclude tryptophan. [Melittin] was fixed at 10 μM and [melittin binder] was varied from 0 to 40 μM. [Indolicidin] was initially fixed at 10 μM, with [binder] varying from 0-20 μM. For subsequent global fitting experiments, [Indolicidin] was held constant at either 5, 7.5 or 10 μM and [binder] was varied from 0-25μM. Samples were prepared in triplicate and incubated overnight at 4 °C. After equilibrating to room temperature for 30 minutes, tryptophan fluorescence was measured as described. As a control, a gradient of [binder] without peptide was included for background subtraction. For all samples, the A295 and A330 of the binder was below 0.1; thus, protein fluorescence was subtracted as background instead of being treated with inner filter effect correction. For protegrin, which contains no tryptophan residues, [binder] was held constant while [peptide] was varied. All protegrin binder designs contain at least one tryptophan. For initial experiments, [protegrin] was held constant at 5 μM and [binder] was varied from 0-30μM. Samples were prepared in triplicate and incubated at room temperature for 3h before tryptophan fluorescence was measured. Here, a gradient of [protegrin] without binder was included for background subtraction.

Binding was fit with the following quadratic binding equation in Prism (GraphPad): $Y = M + ((Q - M)/(2 \cdot P)) \cdot ((1/K) + X + P - \sqrt{((1/K) + X + P)^2 - 4 * P * X})$ where M = fully unbound signal (baseline signal), Q = fully bound (saturation) signal, P = concentration of fixed species, X = ligand concentration, Y = observed fluorescence intensity, and K = association constant. M, Q, and K were not constrained, and P was fixed. For global fitting, K was shared for all datasets.

paragraphSurface plasmon resonance for validation of indo4-indolicidin binding The binding of the highest affinity peptide/binder pair was further validated by surface plasmon resonance (SPR). SPR was performed on a Bruker SPR-24 Pro Instrument using an NTA derivatied SPR chip (SPR sensor prism NiHC1000M; Xantec bioanalytics). The surface was preconditioned with 350mM ethylenediaminetetraacetic acid (EDTA) and running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 50 µM EDTA, 0.05% Tween-20) prior to loading with 5mM Ni2+. Hexahistidine tagged indo4 binder was immobilized on the surface prior to exposure to analyte (indolicidin). Indolicidin was solubilized in water to $\tilde{4}$mg/mL to afford a stock solution. From the stock solution, a concentration gradient of 0-20μM indolidicin was prepared in running buffer. The analyte solutions were flowed over the immobilized protein surface for 80 seconds at 25 μL/min flow rate from low to high concentration and 120 second dissociation time. Blank (running buffer only) injections interspersed between the analyte injections to confirm that analyte was dissociating between injections. Following the cycle of injections, binding affinity was calculated by plotting the pre-injection stop point signal (RU) versus protein concentration. High concentration samples were omitted because of bulk shift from buffer mismatch from preparation of the samples from stock solution solubilized in water. Affinity was calculated via a Langmuir fit of the response units (RU) at the pre-injection stop point.

Bacterial growth assays to measure neutralization of antimicrobial activity Minimum inhibitory concentrations (MICs) of melittin, indolicidin, and protegrin-1 were determined against Bacillus subtilis (ATCC 23857). Peptides were prepared at 100 μg/mL and serially diluted 2-fold in Mueller Hinton Broth (MHB). A glycerol stock of B. subtilis was inoculated into 10 mL MHB and grown overnight at 37°C. The following day, 100 µL of starter culture was added to 10 mL MHB and grown to an OD600 of 0.6–0.8, then diluted to OD600 = 0.001 and added to the peptide dilutions in a 96-well non-treated cell culture plate (Gen-Clone 25-104). Absorbance at 600 nm was measured using a BioTek Synergy Neo-2 plate reader at 37°C with cycles of 7 min shaking and 3 min rest for 15 hours. The MIC was defined as the lowest peptide concentration that completely inhibited growth, and was 1.1 μM, 1.70 μM, and 1.16 μM for melittin, indolicidin, and protegrin-1, respectively.

For neutralization assays, peptides were held constant at their respective MICs, and protein binders were serially diluted 2-fold starting from 40X its target peptide MIC. Peptide was added to the protein

binders, transferred to a 96-well plate, and then the diluted B. subtilis culture prepared as described above was added. Samples were prepared in triplicates. Absorbance at 600 nm was measured as previously described, with the exception that the indolicidin assays were conducted for 7.5 hours instead of 15 hours to account for peptide degradation. For the protegrin binders, protein and peptide dilutions were incubated overnight at 4°C before measuring absorbance. % neutralization was calculated with the following equation: % neutralization = $((A_{\text{obs}} - A_{\text{min}})/(A_{\text{max}} - A_{\text{min}}))cdot100$ where A_{obs} is the observed endpoint A600 for the varying protein concentrations, A_{min} is the observed endpoint A600 of the peptide and bacteria only control, and A_{max} is the observed endpoint A600 of the bacteria only control.

Hemolysis assays Sheep red blood cells (25 mL) were transferred into a 50 mL conical tube and centrifuged at 500 x g for 5 min. The plasma layer was aspirated, leaving the pellet. The cells were resuspended in 150 mM NaCl solution to 25 mL, with gentle inversion and centrifugation (500 x g, 5 min). The supernatant was aspirated, and washed once more with 150 mM NaCl solution. The pellet was then resuspended in 1X PBS (pH 7.4), centrifuged (500 x g, 5 min), and aspirated. The pellet after the PBS wash, was then resuspended in 1X PBS to 25 mL and stored at 4 °C. For hemolysis assessment, 190 μL of RBC (1:100 in 1X PBS) was added per well of a 96-well plate, followed by the addition of 10 μL of either 1X PBS, 20% Triton X-100, or melittin. Cells were treated with serial dilutions of melittin (0.08 to 10 μM final). The plate was incubated at 37 °C with gently shaking for 1 h, followed by centrifugation (500 x g, 5 min). From each well, 100 μL of supernatant was transferred to a fresh plate ensuring pellets were undisturbed. Absorbance was measured at 400 nm using a microplate reader (SpectraMax M5). Values were normalized to PBS and Triton X-100 controls (N= 4, performed in duplicate). To assess melittin's hemolytic activity in the presence of its binders, 10 μL of melittin (1.2 μM final) was added to wells of a 96-well plate, followed by serial dilution of protein (0.05 μM to 6 μM final). This was incubated for 1h at room temperature. Then, 180 μL of red blood cells (1:100 in 1X PBS) was added to each well and incubated at 37 °C for 1 h. After centrifugation, 100 μL of supernatants were transferred to a fresh plate and absorbance measured at 400 nm. Absorbance values were normalized to PBS and melittin only (1.2 μM) controls (N= 4, performed in duplicate).

D.4 Designing Peptides to Bind the Disordered Region of NPM1.

Experiments by Yaotian Zhang, and Denes Hnisz

pRK5-msfGFP-NPM1binder plasmids were constructed by amplifying msfGFP from pRK5_msfGFP-HMGB1-Shuffled 1 (Addgene #237650) (PMID: 40468084). The NPM1binder sequences were inserted at the C-terminus of msfGFP through primer sequences. The amplicons were assembled into AgeI + XbaI-digested pRK5 backbone (Addgene #194548) (PMID: 36755093) using the NEBuilder HiFi DNA assembly master mix.

Cell culture Cells were cultured under standard conditions (37 ℃ and 5% CO_2) in sterile, TC-treated, non-pyrogenic, polystyrene tissue culture dishes (Corning). U2-OS (ATCC, HTB-96) cells were cultured in DMEM GlutaMAX (Gibco, 31966047). The culture medium included 10% FBS (Gibco, 10438-026) and 100 U ml^{-1} penicillin–streptomycin (Gibco, 15140148). All cell lines tested negative for mycoplasma using the LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich, MP0035) or the PCR Mycoplasma Test Kit II (Applichem, A8994). Mycoplasma testing was performed on 0.2–1 ml of culture medium taken from tissue culture dishes containing confluent monolayers of cells, on a routine basis at least twice a year.

Live-cell imaging All live-cell imaging experiments were performed using the LSM880 Airyscan microscope equipped with a Plan-Apochromat 63×/1.40 oil differential interference contrast objective, while incubating cells at 37 ℃ and 5% CO_2. Cells were seeded onto eight-well chamber slides (Ibidi, 80826-90) at 40,000 cells per well, transfected 24 h later, and imaged 24 h after transfection. U2OS cells were transfected using FuGENE HD according to the manufacturer's instructions. Hoechst 33342 (0.2 µg ml^{-1}, Thermo Fisher Scientific, 62249) was added to the cell culture medium for nuclear staining.

Live-cell imaging For immunofluorescence experiments, U2OS cells were seeded on eight-well chamber slides (Ibidi, 80826-90) at 40,000 cells per well, transfected 24 h later, and fixed 24 h after transfection with 4% PFA in PBS for 10 min. Cells were permeabilized with 0.5% Triton X-100 (Thermo Fisher Scientific, 85111) in PBS for 30 min, incubated in blocking buffer containing 1% BSA (BSA Fraction V, Gibco, 15260037) and 0.1% Triton X-100 in PBS for 1 h, and stained with primary antibodies at room temperature for 1 h with gentle rotation. Slides were washed five times with blocking buffer, incubated with secondary antibodies (AlexaFluor 594 donkey anti-mouse antibody, Jackson ImmunoResearch, 715-585-150; and AlexaFluor 594 donkey anti-mouse antibody, Jackson ImmunoResearch, 711-605-152; 1:1,000) in blocking buffer for 1 h at room temperature, washed twice with blocking buffer, stained with 0.5 µg ml^{-1} DAPI in PBS (Invitrogen, D1306), and washed three times with PBS. The following primary antibodies were used: NPM1 (B23) (Santa Cruz, sc-271737, 1:100) and SURF6 (Abcam, ab221990, 1:1000). Imaging was performed using the LSM880 Airyscan microscope equipped with a Plan-Apochromat 63×/1.40 oil differential interference contrast objective.

D.5 Designing Peptides to Bind a Specific Site of RagC and the RagA:RagC Dimer

Experiments by Shamayeeta Ray, Jonathan T. Goldstein, and David M. Sabatini.

Expression and purification of the Rag A: Rag C GTPase heterodimer E.coli LOBSTR [Andersen et al., 2013] cells carrying a pETDuet-1 vector encoding codon optimized, C-terminally His-tagged RagA with a mutation (T21N) that favors the GDP loaded state[Shen et al., 2017, Kim et al., 2008, Yang et al., 2020] and tagless wildtype (WT) RagC in its state (Addgene: 99664), were grown in Terrific broth (TB) and protein expression induced with an overnight IPTG treatment at 18°C b. The purification follows a protocol described previously[Shen et al., 2017]. The complex was purified using Ni-NTA affinity chromatography followed by ion-exchange chromatography (IEX) using a Capto HisRes Q anion-exchange column and then size-exclusion chromatography (SEC) using a Superdex200 column. The protein corresponding to the heterodimer was concentrated in a final buffer containing 50 mM HEPES (pH7.5), 100 mM NaCl and 2 mM MgCl2 and used for SPR studies.

Binding studies of peptides to the Rag GTPase heterodimer using a high-throughput SPR instrument Surface plasmon resonance (SPR) experiments were performed on a Cytiva Biacore 8K instrument. 0.2 μM of the Rag GTPase heterodimer with the His-tag on RagA was immobilized on a Biacore NTA sensor chip using a Ni-NTA-Histag immobilization technique. All the peptides, at a concentration range from 0-100 μM were flown over the protein-bound NTA sensor chip as 'analyte' and their binding responses were recorded. The Rag GTPase heterodimer was first loaded with 1 μM GDP after immobilization prior to each peptide run. The protein, GDP, and the peptide samples were prepared in a buffer containing 50 mM HEPES (pH7.5), 100 mM NaCl and 2 mM MgCl2. Rag GTPase heterodimer was immobilized on the NTA sensor chip for 200 sec and all the runs were performed at 25°C. For each peptide run at each concentration, the association and dissociation times were 120 and 300 sec, respectively. The sensor chip was regenerated after each run using 0.35 M EDTA and subsequently reused throughout the entire run. Association and dissociation kinetics along with binding affinities were analyzed using the Biacore™ Insight Software. Each sensogram corresponding to a single peptide concentration fit best using a two-state binding model that indicates an initial weak binding state followed by a conformational change to obtain a strong binding state. For each peptide that showed a detectable binding response, a log-plot of concentration (x-axis) vs relative response (y-axis) based on the individual fits was generated and a single dissociation constant (KD) was computed using the Biacore™ Insight Software based on the two-state model (Table 4).

D.6 Designing Nanobodies that Bind Penguinpox and Hemagglutinin

Experiments by Jacob A. Hambalek, Anshika Gupta, Diego Taquiri Diaz, and Chang C. Liu.

Each nanobody design was cloned into plasmid pMAA28 Hendel et al. [2025], which is a CEN/ARS plasmid that encodes the nanobody as an N-terminal fusion to Aga2 (i.e., N-nanobody-HA-tag-Aga2-C) for display under the control of a pER promoter Yang et al. [2023]. Each plasmid was transformed

into yeast strain yAP174 Wong et al. [2024] and then plated on synthetic complete media lacking histidine, uracil, and leucine (SC-HLU). Colonies for each design were picked into SC-HLU media and grown separately for 18–20 hours at 30°C with 200 rpm. Expression of the surface protein Aga1 was induced with 200 nM β-estradiol, eliciting surface display of the constitutively expressed nanobody–HA tag–Aga2 fusion.

For designs targeting cGAMP PDE, the nanobody-expressing cells were incubated with the labeled cGAMP PDE and an Alexa Fluor 488 conjugated antibody (R&D Systems, catalog #IC6875G), which targets the nanobody's HA tag, in the incubation buffer HBSBM (20 mM HEPES, pH 7.5; 100 mM NaCl; 1 g/L BSA; 1.8 g/L maltose) for 1 hour at 4°C. The cGAMP PDE protein (a gift from Philip J. Kranzusch and Samuel J. Hobbs, Harvard Medical School) was labeled with a reporter dye Alexa Fluor 647 using an NHS-AlexaFluor 647 labeling kit (Thermo Fisher).

For the FhaB-targeting designs, the nanobody-expressing cells were incubated with the FhaB protein for 1 hour at 4°C, followed by incubation with fluorescently labeled Anti-His Alexa Fluor 647 antibody (R&D Systems; catalog #IC0501R) and Anti-HA Alexa Fluor 488 for 30 minutes at 4°C. The FhaB protein (a gift from Celia W. Goulding and Christine D. Hardy, UC Irvine) contains a His-tag for protein purification and detection. After antigen incubation and reporter incubation, 2.5 µg propidium iodide (Sigma-Aldrich; catalog #81845) was added to stain dead cells. The cells were then washed with two volumes of HBSBM and resuspended in 125 µL HBSBM. Each cell population was interrogated for fluorescence using the Attune NxT Flow Cytometer (Thermo Fisher).

D.7 Designing Proteins that Bind to Small Molecules

Experiments by A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, and William F. DeGrado

Rational inspection The top 100 computationally ranked designs were examined based on the number and geometry of potential hydrogen bonds formed between rucaparib and each designed binder. rucaparib was conceptually fragmented into three hydrogen-bonding functional groups: carboxamide, indole NH, and secondary amine. Hydrogen bonds were defined by the distance between oxygen or nitrogen atoms of these rucaparib fragments and those of the binder residues within 3.2 Å. The presence of hydrogen bonds involving the carboxamide group was given the highest priority during selection. Six candidate designs were subsequently chosen by visual inspection, considering both burial within the binding pocket and diversity of the protein scaffolds.

Protein expression and purification Codon-optimized genes encoding the designed candidate proteins with an N-terminal 6×His tag and a TEV protease cleavage site (HHHHHHENLYFQS) were synthesized and obtained from Twist Bioscience. To facilitate Gibson assembly into the pET-28a(+) vector, short sequences were added at the 5' end (CTCTAGAAATAATTTTGTTTAACTTTAA-GAAGGAGATATACC) and 3' end (GATCCGGCTGCTAACAAAGCCCGAAAG) of each gene. The recombinant plasmids were transformed into Escherichia coli strain E. cloni BL21(DE3). A single colony was picked from an LB agar plate and inoculated into LB medium supplemented with kanamycin (50 μg/mL) for overnight growth. The culture was then transferred into 200 mL of TB medium containing kanamycin (50 μg/mL) and incubated at 37 °C until reaching an OD600 of 0.6 - 0.8. Protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cultures were incubated overnight at 30 °C. Cells were harvested by centrifugation and resuspended in 25 mL PBS buffer (10 mM Na2HPO4, 1.8 mM KH2PO4, 2.7 mM KCl, 137 mM NaCl, pH 7.4) supplemented with 20 mM imidazole. Cells were lysed by ultrasonication (Sonic Dismembrator Model 500, Fisher Scientific), and the lysate was clarified by centrifugation (35,000 g, 30 min). The supernatant was loaded onto a gravity column containing Ni-NTA agarose resin (HisPur, Thermo Fisher, 1.0 mL or 3.0 mL). The resin was washed with three column volumes (CVs) of PBS buffer containing 20 mM imidazole, and bound proteins were eluted with 7 mL PBS buffer containing 250 mM imidazole. The eluted proteins were concentrated and subjected to three rounds of buffer exchange with PBS buffer using a 15 mL, 10 kDa cutoff centrifugal filter unit (EMD Millipore).

Fluorescence emission and fluorescence polarization assays: Fluorescence emission and fluorescence polarization spectra for assessment of rucarparib binding: To assess rucaparib binding, rucaparib

dissolved in DMSO was mixed with proteins in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4) to a final DMSO concentration below 2%, and incubated for 5 min prior to measurement. Fluorescence emission spectra were recorded in black, flat-bottom 96-well plates using a BioTek Synergy Neo-2 plate reader with an excitation wavelength of 355 nm. Protein aliquots from 10 or 100 μM stocks in PBS were combined to make 200 μL samples containing 10 μM rucaparib. Each condition was measured in triplicate. Fluorescence polarization (FP) assays were performed using the same samples on a BioTek Synergy 2 plate readerequipped with excitation and emission filters of 405 nm and 516 nm, respectively. FP values were recorded in polarization (P) units. The polarization values were plotted against protein concentration, and the data were fitted to a one-site binding model using nonlinear regression in GraphPad Prism 10 to determine the dissociation constant (K_d).

D.8 Designing Antimicrobial Peptides that Inhibit the GyrA to GyrA Interaction

Experiments by Andrew Savinov, and Gene-Wei Li

A library of DNA templates encoding designed variants, mutated variants with 3 alanine substitutions at the binding interface, alongside a library encoding protein fragments tiling GyrA and eGFP, was generated (Twist Biosciences), and massively parallel relative growth measurements in E. coli were performed as previously [Savinov et al., 2022, 2025].

Specifically, the library of coding sequences was cloned into the pET-9a expression vector (Novagen) exactly as previously [Savinov et al., 2022, 2025]. The plasmid library encoding designed binders and protein fragments was then transformed into electrocompetent *E. coli* BL21 (DE3) (Sigma-Aldritch) at $\geq$110-fold coverage of the library size, and following 1-hr recovery from transformation, cells were immediately diluted into LB media (Gibco) containing kananycin (selecting for presence of the library) and 10 uM IPTG (inducer for library expression), beginning the massively parallel inhibition measurements. Cells were then grown to an OD_{600nm} of 1.5, at which point they were harvested. These experiments were performed in triplicate (3 biological replicates). Plasmids were extracted from each sample (Qiagen miniprep kit), and DNA from each output sample as well as the plasmid library input was prepared for high-throughput sequencing as previously [Savinov et al., 2022]. Paired-end sequencing was performed on a Singular G4 platform. From these measurements we determined designed peptide and protein fragment frequencies in the population (f) at the initial and final growth assay timepoints, allowing calculation of the enrichment E = log2(f_initial/f_final). The inhibition score for each peptide was then calculated as Inhibition = -E, such that larger positive values correspond to stronger inhibitory effects. Results across biological replicates of these measurements were highly reproducible as in prior work [Savinov et al., 2022, 2025].

The specificity of designed binders for the designed binding mode to GyrA was calculated as Δ(Inhibition) = Inhibition(designed binder) - Inhibition(mutated binder). Positive Δ(Inhibition) values therefore correspond to binders which are more inhibitory than their corresponding variants where 3 interface residues are mutated to alanine. Designed binders and protein fragments that substantially inhibit cell growth were defined as those with inhibition scores $E \leq$ -2, corresponding to a $\geq$4-fold reduction in relative growth. Results were similar with a less stringent threshold of $E \leq$ -1, and so the more conservative threshold was employed. This threshold picks out previously identified inhibitory peaks from protein fragments tiling across GyrA [Savinov et al., 2022, 2025].

References

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 2024.

Kasper R Andersen, Nina C Leksa, and Thomas U Schwartz. Optimized e. coli expression strain lobstr eliminates common contaminants from his-tag purification. *Proteins: Structure, Function, and Bioinformatics*, 81(11):1857–1861, 2013.

Ernesto Anoz-Carbonell, Maribel Rivero, Victor Polo, Adrián Velázquez-Campoy, and Milagros Medina. Human riboflavin kinase: Species-specific traits in the biosynthesis of the fmn cofactor. *The FASEB Journal*, 34(8):10871–10886, 2020.

Benjamin D. Bax, Garib Murshudov, Anthony Maxwell, and Thomas Germe. Dna topoisomerase inhibitors: Trapping a dna-cleaving machine in motion. *Journal of Molecular Biology*, 431(18): 3427–3449, 2019. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2019.07.008. URL https://www.sciencedirect.com/science/article/pii/S0022283619304322. The molecular basis of antibiotic action and resistance.

Els Beirnaert, Aline Desmyter, Silvia Spinelli, Marc Lauwereys, Lucien Aarden, Torsten Dreier, Remy Loris, Karen Silence, Caroline Pollet, Christian Cambillau, et al. Bivalent llama single-domain antibody fragments against tumor necrosis factor have picomolar potencies due to intramolecular interactions. *Frontiers in immunology*, 8:867, 2017.

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, et al. Atomically accurate de novo design of antibodies with rfdiffusion. *bioRxiv*, pages 2024–03, 2025. doi: 10.1101/2024.03.14.585103.

Antonio P. Bento, Anne Hersey, Eddy Félix, et al. An open source chemical structure curation pipeline using rdkit. *Journal of Cheminformatics*, 12(1):51, 2020. doi: 10.1186/s13321-020-00456-1. URL https://doi.org/10.1186/s13321-020-00456-1.

Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28, 2000.

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, May 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04654-9. URL https://doi.org/10.1038/s41586-022-04654-9.

Chai-Discovery, Jacques Boitreaud, Jack Dent, Danny Geisz, Matthew McPartlon, Joshua Meier, Zhuoran Qiao, Alex Rogozhnikov, Nathan Rollins, Paul Wollenhaupt, et al. Zero-shot antibody design in a 24-well plate. *bioRxiv*, pages 2025–07, 2025.

Venkatesh Chanukuppa, Debasish Paul, Khushman Taunk, Tathagata Chatterjee, Sanjeevan Sharma, Amey Shirolkar, Sehbanul Islam, Manas K Santra, and Srikanth Rapole. Proteomics and functional study reveal marginal zone b and b1 cell specific protein as a candidate marker of multiple myeloma. *International Journal of Oncology*, 57(1):325–337, 2020.

Zhiqiang Chen, Xinyi Zhou, Xiaojun Zhou, Yi Tang, Mingzhu Lu, Jianhong Zhao, Chenhui Tian, Mingzhi Wu, Yanliang Liu, Edward V Prochownik, et al. Phosphomevalonate kinase controls β-catenin signaling via the metabolite 5-diphosphomevalonate. *Advanced science*, 10(12):2204909, 2023.

E Conzelmann and K Sandhoff. Ab variant of infantile gm2 gangliosidosis: deficiency of a factor necessary for stimulation of hexosaminidase a-catalyzed degradation of ganglioside gm2 and glycolipid ga2. *Proceedings of the National Academy of Sciences*, 75(8):3979–3983, 1978.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Michael S. Costello, Bryan C. Neumann, Bonnie J. Cuthbert, Jana Holubová, Mia W. Raimondi, Fernando Garza-Sánchez, Abdul Samad, Ladislav Bumba, Jacob A. Torres, Nickolas Holznecht, Jessica Mendoza, Ondrej Stanek, Sasiprapa Prombhul, Thomas Weimbs, Meghan A. Morrissey, Diego Acosta-Alvear, David A. Low, Peter Šebo, Celia W. Goulding, Shane Gonen, and Christopher S. Hayes. Bacteria deliver a microtubule-binding protein into mammalian cells to promote colonization. *bioRxiv*, 2025. doi: 10.1101/2025.06.17.660209.

Tristan I Croll, Brian J Smith, Mai B Margetts, Jonathan Whittaker, Michael A Weiss, Colin W Ward, and Michael C Lawrence. Higher-resolution structure of the human insulin receptor ectodomain: multi-modal inclusion of the insert domain. *Structure*, 24(3):469–476, 2016.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

Justas Dauparas, Grace R. Lee, Ross Pecoraro, et al. Atomic context-conditioned protein sequence design using ligandmpnn. *Nature Methods*, 22:717–723, 2025. doi: 10.1038/s41592-025-02626-1. URL https://doi.org/10.1038/s41592-025-02626-1.

WF DeGrado, GF Musso, M Lieber, ET Kaiser, and FJ Kezdy. Kinetics and mechanism of hemolysis induced by melittin and by a synthetic melittin analogue. *Biophysical journal*, 37(1):329–338, 1982.

Christopher E Dempsey. The actions of melittin on membranes. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes*, 1031(2):143–161, 1990.

Lena Erlandsson, Aurélien Ducat, Johann Castille, Isac Zia, Grigorios Kalapotharakos, Erik Hedström, Jean-Luc Vilotte, Daniel Vaiman, and Stefan R Hansson. Alpha-1 microglobulin as a potential therapeutic candidate for treatment of hypertension and oxidative stress in the stox1 preeclampsia mouse model. *Scientific reports*, 9(1):8561, 2019.

Timothy J Falla, D Nedra Karunaratne, and Robert EW Hancock. Mode of action of the antimicrobial peptide indolicidin. *Journal of Biological Chemistry*, 271(32):19298–19303, 1996.

Moran Farhi, Elena Marhevka, Tania Masci, Evgeniya Marcos, Yoram Eyal, Mariana Ovadis, Hagai Abeliovich, and Alexander Vainstein. Harnessing yeast subcellular compartments for the production of plant terpenoids. *Metabolic engineering*, 13(5):474–481, 2011.

Dan Friedman and Adji Bousso Dieng. The vendi score: A diversity evaluation metric for machine learning. *arXiv preprint arXiv:2210.02410*, 2022. doi: 10.48550/arXiv.2210.02410.

Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, et al. Proteina: Scaling flow-based protein structure generative models. *arXiv preprint arXiv:2503.00710*, 2025.

David Gidalevitz, Yuji Ishitsuka, Adrian S Muresan, Oleg Konovalov, Alan J Waring, Robert I Lehrer, and Ka Yee C Lee. Interaction of antimicrobial peptide protegrin with biomembranes. *Proceedings of the National Academy of Sciences*, 100(11):6302–6307, 2003.

Casper A Goverde, Martin Pacesa, Nicolas Goldbach, Lars J Dornfeld, Petra EM Balbi, Sandrine Georgeon, Stéphane Rosset, Srajan Kapoor, Jagrity Choudhury, Justas Dauparas, et al. Computational design of soluble and functional membrane protein analogues. *Nature*, 631(8020):449–458, 2024.

Shantanu Guha, Ryan P Ferrie, Jenisha Ghimire, Cristina R Ventura, Eric Wu, Leisheng Sun, Sarah Y Kim, Gregory R Wiedman, Kalina Hristova, and Wimley C Wimley. Applications and evolution of melittin, the quintessential membrane active peptide. *Biochemical pharmacology*, 193:114769, 2021.

Nathan Hendel, Hannes Stark, Philip J. Kranzusch, and Jennifer A. Doudna. Yeast surface display of nanobodies enables rapid characterization of protein binders. *bioRxiv*, 2025. doi: 10.1101/2025.10.04.680454. URL https://www.biorxiv.org/content/10.1101/2025.10.04.680454v1. Preprint.

Julie A Hixon, Caroline Andrews, Lila Kashi, Casey L Kohnhorst, Emilee Senkevitch, Kelli Czarra, Joao T Barata, Wenqing Li, Joel P Schneider, Scott TR Walsh, et al. New anti-il-7rα monoclonal antibodies show efficacy against t cell acute lymphoblastic leukemia in pre-clinical models. *Leukemia*, 34(1):35–49, 2020.

Samuel J. Hobbs, Jason Nomburg, Jennifer A. Doudna, and Philip J. Kranzusch. Animal and bacterial viruses share conserved mechanisms of immune evasion. *Cell*, 187(20):5530–5539.e8, Oct 2024. ISSN 0092-8674. doi: 10.1016/j.cell.2024.07.057. URL https://doi.org/10.1016/j.cell.2024.07.057.

Kalina Hristova, Christopher E Dempsey, and Stephen H White. Structure, location, and lipid perturbations of melittin at the membrane interface. *Biophysical journal*, 80(2):801–811, 2001.

Yi-Chen Hsieh, Joseph Negri, Amy He, Richard V Pearse, Lei Liu, Duc M Duong, Lori B Chibnik, David A Bennett, Nicholas T Seyfried, and Tracy L Young-Pearse. Elevated ganglioside gm2 activator (gm2a) in human brain tissue reduces neurite integrity and spontaneous neuronal activity. *Molecular neurodegeneration*, 17(1):61, 2022.

Arttu Jolma, Yimeng Yin, Kazuhiro R Nitta, Kashyap Dave, Alexander Popov, Minna Taipale, Martin Enge, Teemu Kivioja, Ekaterina Morgunova, and Jussi Taipale. Dna-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, 527, 2015.

Ivana Jovčevska and Serge Muyldermans. The therapeutic potential of nanobodies. *BioDrugs*, 34(1): 11–26, 2020.

Ioanna Kalvari, Eric P Nawrocki, Nancy Ontiveros-Palacios, Joanna Argasinska, Kevin Lamkiewicz, Manja Marz, Sam Griffiths-Jones, Claire Toffano-Nioche, Daniel Gautheret, Zasha Weinberg, et al. Rfam 14: expanded coverage of metagenomic, viral and microrna families. *Nucleic acids research*, 49, 2021.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35, 2022.

Tabassum Khan, Kaksha Sankhe, Vasanti Suvarna, Atul Sherje, Kavitkumar Patel, and Bhushan Dravyakar. Dna gyrase inhibitors: Progress and synthesis of potent compounds as antibacterial agents. *Biomedicine and Pharmacotherapy*, 103:923–938, 2018. ISSN 0753-3322. doi: https://doi.org/10.1016/j.biopha.2018.04.021. URL https://www.sciencedirect.com/science/article/pii/S0753332217365794.

Eunjung Kim, Pankuri Goraksha-Hicks, Li Li, Thomas P Neufeld, and Kun-Liang Guan. Regulation of torc1 by rag gtpases in nutrient response. *Nature cell biology*, 10(8):935–945, 2008.

Ki-Myo Kim, Kang-Gu Lee, Saseong Lee, Bong-Ki Hong, Heejae Yun, Yune-Jung Park, Seung-Ah Yoo, and Wan-Uk Kim. The acute phase reactant orosomucoid-2 directly promotes rheumatoid inflammation. *Experimental & Molecular Medicine*, 56(4):890–903, 2024.

Rohith Krishna et al. Generalized biomolecular modeling and design with rosettafold all-atom. *Science*, 384:eadl2528, 2024. doi: 10.1126/science.adl2528.

Alexey S Ladokhin, Michael E Selsted, and Stephen H White. Bilayer interactions of indolicidin, a small antimicrobial peptide rich in tryptophan, proline, and basic amino acids. *Biophysical Journal*, 72(2):794–805, 1997.

Alexey S Ladokhin, Michael E Selsted, and Stephen H White. Cd spectra of indolicidin antimicrobial peptides suggest turns, not polyproline helix. *Biochemistry*, 38(38):12313–12319, 1999.

Qiang Li, Yaju Wang, Xiangshu Meng, Wenjing Wang, Feifan Duan, Shuya Chen, Yukun Zhang, Zhiyong Sheng, Yu Gao, and Lei Zhou. Mettl16 inhibits papillary thyroid cancer tumorigenicity through m6a/ythdc2/scd1-regulated lipid metabolism. *Cellular and Molecular Life Sciences*, 81(1): 81, 2024.

Christopher A. Lipinski, Franco Lombardo, Bruce W. Dominy, and Paul J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 46(1-3):3–26, March 2001. doi: 10.1016/S0169-409X(00)00129-0.

Lei Lu, Xuxu Gou, Sophia K. Tan, Samuel I. Mann, Hyunjun Yang, Xiaofang Zhong, Dimitrios Gazgalis, Jesús Valdiviezo, Hyunil Jo, Yibing Wu, Morgan E. Diolaiti, Alan Ashworth, Nicholas F. Polizzi, and William F. DeGrado. De novo design of drug-binding proteins with predictable binding energy and specificity. *Science*, 384(6691):106–112, 2024. doi: 10.1126/science.adl5364. URL https://www.science.org/doi/abs/10.1126/science.adl5364.

Ujjini H. Manjunatha, Anthony Maxwell, and Valakunja Nagaraja. A monoclonal antibody that inhibits mycobacterial dna gyrase by a novel mechanism. *Nucleic Acids Research*, 33(10):3085–3094, 01 2005. ISSN 0305-1048. doi: 10.1093/nar/gki622. URL https://doi.org/10.1093/nar/gki622.

Luis S. Mille-Fragoso, John N. Wang, Claudia L. Driscoll, Haoyu Dai, Talal Widatalla, Xiaowei Zhang, Brian L. Hie, and Xiaojing J. Gao. Efficient generation of epitope-targeted de novo antibodies with germinal. *bioRxiv*, 2025. doi: 10.1101/2025.09.19.677421. URL https://www.biorxiv.org/content/early/2025/09/25/2025.09.19.677421.

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, et al. Bindcraft: one-shot design of functional protein binders. *bioRxiv*, 2024.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707. URL https://www.biorxiv.org/content/early/2025/06/18/2025.06.14.659707.

Ya-Nan Qi, Zhu Liu, Lian-Lian Hong, Pei Li, and Zhi-Qiang Ling. Methyltransferase-like proteins in cancer biology and potential therapeutic targeting. *Journal of hematology & oncology*, 16(1):89, 2023.

Deborah Renaud and Michael Brodsky. Gm2-gangliosidosis, ab variant: clinical, ophthalmological, mri, and molecular findings. In *JIMD Reports, Volume 25*, pages 83–86. Springer, 2015.

Andrew Savinov, Andres Fernandez, and Stanley Fields. Mapping functional regions of essential bacterial proteins with dominant-negative protein fragments. *Proceedings of the National Academy of Sciences*, 119(26):e2200124119, 2022. doi: 10.1073/pnas.2200124119. URL https://www.pnas.org/doi/abs/10.1073/pnas.2200124119.

Andrew Savinov, Sebastian Swanson, Amy E. Keating, and Gene-Wei Li. High-throughput discovery of inhibitory protein fragments with alphafold. *Proceedings of the National Academy of Sciences*, 122(6):e2322412122, 2025. doi: 10.1073/pnas.2322412122. URL https://www.pnas.org/doi/abs/10.1073/pnas.2322412122.

Kuang Shen, Abigail Choe, and David M Sabatini. Intersubunit crosstalk in the rag gtpase heterodimer enables mtorc1 to respond rapidly to amino acid availability. *Molecular cell*, 68(3):552–565, 2017.

Deborah A Steinberg, Malinda A Hurst, Craig A Fujii, AH Kung, JF Ho, FC Cheng, David J Loury, and John C Fiddes. Protegrin-1: a broad-spectrum, rapidly microbicidal peptide with in vivo activity. *Antimicrobial agents and chemotherapy*, 41(8):1738–1742, 1997.

Martin Steinegger and Johannes Söding. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, 35, 2017.

Aleksandra Szczepankiewicz, Anna Bręborowicz, Paulina Sobkowiak, and Anna Popiel. Polymorphisms of two histamine-metabolizing enzymes genes and childhood allergic asthma: a case control study. *Clinical and Molecular Allergy*, 8(1):14, 2010.

Karel HM van Wely, Jelto Swaving, Roland Freudl, and Arnold JM Driessen. Translocation of proteins across the cell envelope of gram-positive bacteria. *FEMS microbiology reviews*, 25(4):437–454, 2001.

Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, et al. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50, 2022.

Horst Vogel and Fritz Jähnig. The structure of melittin in membranes. *Biophysical journal*, 50(4): 573–582, 1986.

Teresa R Wagner and Ulrich Rothbauer. Nanobodies right in the middle: intrabodies as toolbox to visualize and modulate antigens in the living cell. *Biomolecules*, 10(12):1701, 2020.

Xiao-Kun Wang, Ya-Wei Zhang, Chun-Ming Wang, Bo Li, Tian-Zhi Zhang, Wen-Jie Zhou, Lyu-jia Cheng, Ming-Yu Huo, Chang-Hua Zhang, and Yu-Long He. Mettl16 promotes cell proliferation by up-regulating cyclin d1 expression in gastric cancer. *Journal of cellular and molecular medicine*, 25 (14):6602–6617, 2021.

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, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620, 2023.

Wei Wei, Zhong-Yuan Zhang, Bin Shi, Yike Cai, Hou-Shun Zhang, Chun-Lei Sun, Yun-Fei Fei, Wen Zhong, Shuang Zhang, Chen Wang, et al. Mettl16 promotes glycolytic metabolism reprogramming and colorectal cancer progression. *Journal of Experimental & Clinical Cancer Research*, 42(1):151, 2023.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, and Regina Barzilay. Boltz-1 Democratizing Biomolecular Interaction Modeling, 2025.

Jia Yin Wong, Christine D. Hardy, Celia W. Goulding, and Chang C. Liu. A modular yeast surface display platform for rapid screening of nanobody libraries. *ACS Synthetic Biology*, 13(8):1750–1764, 2024. doi: 10.1021/acssynbio.4c00370. URL https://pubs.acs.org/doi/full/10.1021/acssynbio.4c00370.

Shu Yang, Yingbiao Zhang, Chun-Yuan Ting, Lucia Bettedi, Kuikwon Kim, Elena Ghaniam, and Mary A Lilly. The rag gtpase regulates the dynamic behavior of tsc downstream of both amino acid and growth factor restriction. *Developmental cell*, 55(3):272–288, 2020.

Yuqi Yang, Chong Na, and et al. Engineering yeast for modular surface display of proteins under a beta-estradiol–inducible promoter. *ACS Synthetic Biology*, 12(4):843–854, 2023. doi: 10.1021/acssynbio.2c00423. URL https://pubs.acs.org/doi/full/10.1021/acssynbio.2c00423.

Yimeng Yin, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K Das, Teemu Kivioja, Kashyap Dave, Fan Zhong, et al. Impact of cytosine methylation on dna binding specificities of human transcription factors. *Science*, 356, 2017.

Takeo Yoshikawa, Tadaho Nakamura, and Kazuhiko Yanai. Histamine n-methyltransferase in the brain. *International journal of molecular sciences*, 20(3):737, 2019.

Hao Yu, Juntao Zhuang, Zijian Zhou, Qiang Song, Jiancheng Lv, Xiao Yang, Haiwei Yang, and Qiang Lu. Mettl16 suppressed the proliferation and cisplatin-chemoresistance of bladder cancer by degrading pmepa1 mrna in a m6a manner through autophagy pathway. *International Journal of Biological Sciences*, 20(4):1471, 2024.

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, 2024. URL https://arxiv.org/abs/2409.08022.

Fei Zhang, Hudie Wei, Xiaoxiao Wang, Yu Bai, Pilin Wang, Jiawei Wu, Xiaoyong Jiang, Yugang Wang, Haiyan Cai, Ting Xu, et al. Structural basis of a novel pd-l1 nanobody for immune checkpoint blockade. *Cell discovery*, 3(1):1–12, 2017.

Qiu Zhengqi, Guo Zezhi, Jiang Lei, Qiu He, Pan Jinyao, and Ao Ying. Prognostic role of phyh for overall survival (os) in clear cell renal cell carcinoma (ccrcc). *European Journal of Medical Research*, 26(1):9, 2021.