De novo design of protein binders to bioactive peptides: from hormones to snake toxins

Susana Vázquez Torres

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
David Baker, Chair
Jesse Zalatan
Neil King

Program authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

De novo design of protein binders to bioactive peptides: from hormones to snake toxins

Susana Vázquez Torres

Chair of the Supervisory Committee:
David Baker

Department of Biochemistry

Recent advancements in deep learning methods for protein design and structure prediction have demonstrated the potential of *de novo* proteins to replace antibodies as affinity reagents, offering cost-effective production, increased stability, and robust platforms for therapeutics and diagnostics.

In this work, we set out to design *de novo* protein binders for two classes of bioactive peptides: hormones which adopt α-helical structures, and lethal snake toxins from the three-finger toxin (3FTX) family. Using deep learning-based methods, we generated high-affinity binders for both targets. These binders enable hormone detection via mass spectrometry and the creation of bioluminescence-based biosensors. Additionally, our designed proteins effectively neutralize 3FTXs *in vitro* and protect mice from lethal neurotoxin challenges.

While further development is necessary to translate these designs into practical diagnostics and therapeutics, this work demonstrates their potential as robust reagents for hormone detection and as safer, widely accessible antivenoms.

Academic Acknowledgments

I would like to thank the University of Washington, the IPD and the BPSD program for providing excellent resources and support during my PhD studies.

I am grateful to my committee members for their guidance and for taking the time to read this thesis:

- David Baker, PhD
- Neil King, PhD
- Jesse G. Zalatan, PhD
- Michael J. MacCoss, PhD
- Frank DiMaio, PhD

I appreciate the assistance from the administrative, lab support, and IT staff at the IPD, as well as the scientists at the IPD Core Labs. Special thanks to Kandise VanWormer, Luki Goldschmidt, Lance Stewart and Lynda Stuart for keeping the Baker lab up and running.

Additionally, I am grateful to Timothy P. Jenkins, Melisa Benard Valle, Stephen P. Mackessy, Stefanie K. Menzies, Nicholas R. Casewell, Shirin Ahmadi, Andreas H. Laustsen, and Thomas J. A. Fryer for their support and collaboration on generating *de novo* proteins to neutralize lethal snake toxins.

Personal acknowledgments

A mis papás, Mateo y Lupita. Gracias por ser unos padres grandiosos, educarme con valores y darme alas para volar. Los amo siempre.

A mis hermanos Violeta, Jorge, y Luis Daniel. Gracias por enseñarme tanto y crecer a mi lado. Son mi mayor ejemplo e inspiración.

A toda la familia Vázquez Torres que se ha extendido y echado nuevas raíces. A mis cuñadas Geraldine y Claudia por ser grandes amigas y aliadas. A mis sobrinos que amo con todo mi corazón: Matías, Sofía, Claudia, y David.

A mi familia extendida. A todos los Vázquez y a todos los Torres. Me siento orgullosa de pertenecer a esta gran familia.

A todos mis compañeros y grandes maestros de vida: Sergio, Ricardo, Bryan, Fernández, Sosa, Geo, Marcia, Alicia, Myrian, Charly, Dinora, Andrea, Arlen, Fátima, Betty, Titi, Alma, Mau, Alberto, Luis, Manuel, Cinthia, Mariana, Nancy y Karla.

A la educación pública y a la Universidad Nacional Autónoma de México que me regaló una educación maravillosa y la oportunidad de explorar el mundo.

A todos los que se fueron y se me olvida nombrar. Gracias por enseñarme tanto y compartir parte de este viaje juntos.

To all my mentors and friends in the Netherlands. Special thanks to the University of Groningen for giving me the opportunity to do my master's studies and opening endless possibilities for me. Special thanks to Ykelien Boersma, Gerrit Poelarends, Mathijs, Amy, Magdalena, Kamil, Marijke, Alejandro, Andrea, Alejandra, Sem, Erika, Minh and all the MPDI crew. Dankjewel.

To Debbie Robinson, Bhawna Chaubey, Bhavna Sharma, Josh, Melanie and all the roomies. Thank you so much for being my family in Seattle and for all the amazing memories.

To Fernando and Melissa Bañales Mejía. Thank you for all your support. I’m so grateful to have you in my life. Los quiero mucho.

To Mohamad, Marc, LJ, Fatima, Arvind, Edin, Yensi, and Miki. Thank you for being a big support on this journey. Your friendship has been an oasis in the Baker lab. I love you all, and I’m sure this friendship will keep flourishing forever.

To Hannah, Gyu Rie, Kathryn, Amir, Indrek, Yakov, Sanaa, Meerit, Harley, Yoann, Jeremiah, Robert, David Kim, and Daniel Silva. Thank you for all the conversations, scientific and personal advice.

To Preetham Venkatesh, Philip Leung, Isaac Lutz and Joseph Watson. Thank you for teaching me so many things inside and outside the lab. It was an honor and a privilege to work with you.

To the Spanish crew Begoña, Reyes, Marc, Joann, Roberta, Diego, and Alfredo. Gracias por acogerme como una más de ustedes y hacerme sentir en casa.

To David Baker. Thank you for giving me an opportunity and changing my life forever. It's been a pleasure to do the most exciting science in your lab.

Table of contents

Chapter 1:

De novo design of high-affinity binders of bioactive helical peptides

This chapter contains additional background for (and reproduction of) content previously published as: Vázquez Torres, S. *et al.* De novo design of high-affinity binders of bioactive helical peptides. *Nature* **626**, 435–442 (2024).

Background

Proteins, composed of linear chains of amino acids, fold into unique three-dimensional structures and are among the most abundant and functionally diverse macromolecules in living systems[1]. Human cells produce approximately 30,000 different proteins, ranging in size from titin, with over 34,000 amino acids, to small peptides like glucagon, consisting of only 29 amino acids[2].

Bioactive peptides, typically shorter than 50 residues[3], play key roles in human biology and serve as well-established biomarkers in clinical care and biomedical research[4]. Among these, peptide hormones, which adopt α-helical structures upon binding to their receptors, aid in the diagnosis and treatment of conditions such as type 2 diabetes, obesity, bone diseases, pain, and cancer[5]. Consequently, there is considerable interest in their sensitive and specific quantification. Currently, this relies on antibodies that require substantial resources to generate, can be difficult to produce with high affinity, and often exhibit less-than-desirable stability and reproducibility[6].

This study addresses the challenge of designing *de novo* protein binders for bioactive helical peptides. Our aim is to develop general methods for designing proteins that bind to peptides in helical conformations and to test their application as detection tools in diagnostic assays.

Design of peptide-binding scaffolds

To fully leverage recent advances in protein design, we explored both parametric and deep learning-based approaches. For parametric generation, we reasoned that helical bundle scaffolds with an open groove for a helical peptide could provide a general solution to the helical peptide-binding problem: the extended interaction surface between the full length of the helical peptide target and the contacting helices on the designed scaffold could enable high-affinity and

specific binding, and the helices flanking the groove could limit self-association of the recessed hydrophobic surfaces. In parallel, we reasoned that deep learning methods, which do not pre-specify scaffold geometries, could permit the exploration of different potential solutions to peptide binding.

We began by exploring parametric methods for generating backbones with overall 'groove' shapes. Using the Crick parameterization of α-helical coiled coils[7], we devised a method to sample scaffolds consisting of a three-helix groove supported by two buttressing helices (Fig. 1b and Methods). We assembled a library sampling a range of supercoiling and helix–helix spacings to accommodate a variety of helical peptide targets (Supplementary Figs. 1–3). We then used this library to design binders to parathyroid hormone (PTH), glucagon (GCG) and neuropeptide Y (NPY), and screened 12 designs for each target using a NanoBiT split-luciferase binding assay (Supplementary Fig. 4). Many of the designs bound their targets (3, 4 and 8 out of 12 to PTH, GCG and NPY, respectively) but with only micromolar affinities (Fig. 1b and Supplementary Fig. 4a–c). These results suggest that groove-shaped scaffolds can be designed to bind helical peptides, but also that design method improvement was necessary to achieve high-affinity binding.

We next explored using RoseTTAFold Inpainting (RF_{joint})[8], a model that can jointly design protein sequences and structures, along with ProteinMPNN[9], an improved sequence design method, to improve the modest affinity of our tightest parametrically designed PTH binder (Fig. 1c, left). We used RF_{joint} Inpainting to extend the binder interfaces and ProteinMPNN to redesign the sequences, reasoning that the combination of these two methods could lead to more favorable interactions with the peptide. Out of 192 designs tested, 44 showed binding against PTH in initial yeast display screening. Following size-exclusion chromatography (SEC), the best binder was found to bind with 6.04 nM affinity to PTH using fluorescence polarization (FP). Binding was specific: very little binding was observed to PTH-related peptide (PTHrp), a related peptide sequence with 34% sequence identity that binds the same receptor as PTH[10] (Fig. 1c, right). Overall, the affinity of the starting PTH binder was improved by approximately three orders of magnitude, and the computational model of the highest-affinity binder had 19% greater surface area contacting the target peptide (the structural extension was critical to the improvement in binding affinity; sequence redesign with ProteinMPNN of the original binding interface did not

measurably increase affinity; Supplementary Fig. 5). We used the same design strategy to generate higher-affinity binders for NPY and GCG. Using weak parametric binders as a starting point, we extended their binding interfaces and redesigned their sequences to generate a 231-nM-affinity binder for GCG and a 3.5-μM binder for NPY after screening 96 designs (Extended Data Fig. 1a,b).

As an alternative to *de novo* parametric design of scaffolds that contain grooves, we explored the threading of helical peptides of interest onto already existing designed scaffolds with interfaces that make extensive interactions with helical peptides[11,12] (Fig. 1d, left, and Methods). We threaded sequences of peptides of interest onto these complexes, and filtered for interfacial hydrophobic interactions between the target sequence and the scaffolds[13,11]. The selected scaffolds were then redesigned in the presence of the threaded target sequence with ProteinMPNN[9] and the complex was predicted with AlphaFold2[14] (AF2; with initial guess[15]) and filtered on AF2 and Rosetta metrics. Initial screening using yeast surface display identified 4/66 binders for secretin (SCT), which were expressed in *E. coli*. After purification, all four of the designs were found to bind with submicromolar affinity using FP, with the highest-affinity design binding with an affinity of 2.7 nM for SCT (Fig. 1d, right); we also made designs with a dissociation constant (K_d) of <100 nM to GCG-like peptide 1 and gastric inhibitory polypeptide (Extended Data Fig. 2a,b). The SCT binder design bound GCG, which has 44% sequence identity to SCT[16], with fourfold weaker affinity than SCT (Fig. 1d, right).

Designing binders using Hallucination

We next explored the use of deep learning Hallucination methods to generate helical peptide binders completely *de novo*, with no pre-specification of the binder or peptide geometry (Fig. 1e, left inset, and Supplementary Fig. 6a). Hallucination or 'activation maximization' approaches start from a network that predicts protein structure from sequence, and carry out an optimization in sequence space for sequences that fold to structures with desired properties. This approach has been used to generate new monomers[17], functional-site scaffolds[8] and cyclic oligomers[18]. Hallucination using AF2 or RoseTTAFold has the advantage that neither the binder nor the peptide structure needs to be specified during the design process, enabling the design of binders to peptides in different conformations (this is useful given the unstructured nature of many

peptides in solution; disordered peptides can bind in different conformations to different binding partners[19]. Hallucination directly optimizes metrics correlated with binding, albeit with the possible hazard of generating adversarial protein sequences[18]. We began by designing binders to the apoptosis-related BH3 domain of Bid (Fig. 1a). The Bid peptide is unstructured in isolation, but adopts an α-helix on binding to Bcl-2 family members[20,21]; it is therefore a model candidate for the design of helix-binding proteins. Starting from only the Bid primary sequence, and a random seed binder sequence (of length 60, 70, 80, 90 or 100 residues), we carried out a Monte Carlo search in sequence space, optimizing for confident binding to the target peptide (AF2 predicted local distance difference test, pLDDT; and predicted alignment error, pAE)[15]. The trajectories typically converged in 5,000 steps (sequence substitutions; Supplementary Fig. 6b), and the output binder structure was subsequently redesigned with ProteinMPNN, as previously described[18]. All designed binders were predicted to bind to Bid in a predominantly helical conformation; the exact conformations differ between designs because only the amino acid sequence of the target is specified in advance. This protocol effectively carries out flexible backbone protein design, which can be a challenge for traditional Rosetta-based design approaches for which deep conformational sampling can be very compute intensive. In line with our prediction that 'groove' scaffolds would offer an ideal topology for helical peptide binding, many of the binders from this approach contain a well-defined 'groove', with the peptide predicted to make extensive interactions with the binder, typically helix–helix interactions (Extended Data Fig. 3a).

We experimentally tested 46 of the Hallucinated designs (Extended Data Fig. 3a) by co-expression of a GFP-tagged Bid peptide and His-tagged binders, with co-elution of GFP and binder used as a readout for binding. Four of these designs were further characterized, and showed soluble, monomeric expression even in the absence of peptide co-expression (Extended Data Fig. 3b), and could be pulled down using Bid BH3 peptide immobilized on beads (Extended Data Fig. 3c). Circular dichroism experiments indicated that the Bid peptide was unstructured in solution, and that helicity increased on interaction with the Hallucinated proteins, in line with the design prediction (Extended Data Fig. 3d). The binders were highly thermostable and, unlike the native Bcl-2 protein Mcl-1, readily refolded after (partial) thermal denaturation at 95 °C (Extended Data Fig. 3e). FP measurements revealed a 7-nM-affinity binder to Bid peptide

(Fig. 1e, right inset), a higher-affinity interaction than that between Bid and the native partner Mcl-1 (Extended Data Fig. 3f,g).

Design refinement with RFdiffusion

We next explored using the RoseTTAFold-based denoising diffusion method RFdiffusion[22]. RFdiffusion directly generates protein structures with diverse topologies, and is much more compute efficient than Hallucination. We first extended RFdiffusion to enable optimization of existing helical peptide-binders.

A long-standing challenge in protein design is to increase the activity of an input native protein or designed protein by exploring the space of plausible closely related conformations for those with predicted higher activity[23]. This is difficult for traditional design methods as extensive full-atom calculations are needed for each sample around a starting structure (using molecular dynamics simulation or Rosetta full-atom relaxation methods), and it is not straightforward to optimize for higher binding affinity without detailed modeling of the binder–target side-chain interactions. We reasoned that, by contrast, RFdiffusion might be able to rapidly generate plausible backbones in the vicinity of a target structure, increasing the extent and quality of interaction with the target guided by the extensive knowledge of protein structure inherent in RoseTTAFold. Typically, during the reverse diffusion (generative) process, RFdiffusion takes random Gaussian noise as input, and iteratively refines this to a new protein structure over many (T) steps (generally 200). Part way through this denoising process, the evolving structure no longer resembles 'pure noise', instead resembling a 'noisy' version of the final structure. We therefore reasoned that ensembles of structure with varying extents of deviation from an input structure could be generated by partially noising initial starting structures to different extents (for example, time step 70), and then denoising to a similar, but not identical, final structure (Fig. 2a; in this case, the input coordinates to RFdiffusion at time step 70 are from a noised starting structure, rather than a partially denoised random distribution).

We implemented this 'partial diffusion' approach (Methods), and sought first to assess the extent to which protein structures could be resampled and refined with partial diffusion. As expected, partial diffusion allowed diversification of a starting protein fold, and the magnitude of this diversity could be tuned by varying how much noise was added to a starting structure (Fig. 2a).

We next explored the ability of partial diffusion to 'regularize' native protein backbones using as a metric AF2 structure prediction from a single sequence. We found that RFdiffusion improves the 'designability' of protein backbones: ProteinMPNN sequence design on partially diffused native backbones (with high similarity to the native fold, Extended Data Fig. 4a,c, middle row) have improved structure recapitulation (self-consistency) by AF2 compared to both the native sequence (Extended Data Fig. 4b, pink versus gray, and Extended Data Fig. 4c, bottom row) and ProteinMPNN sequences generated from the original native backbone (Extended Data Fig. 4b, blue, and Extended Data Fig. 4c, top row). Further, we found in tests on the well-studied colicin–immunity protein system[24] that the small changes in protein backbone that partial diffusion can sample are sufficient to mediate specificity changes within protein families (Supplementary Fig. 7). Thus, partial diffusion enables protein backbone resampling and refinement, the extent of which can be tuned by varying the amount of noise added. Furthermore, partial diffusion can considerably increase the designability of input protein models.

As a first experimental test of partial diffusion, we started from our parametrically designed Inpainted binders to GCG (with 231 nM K_d) and NPY (with 3.5 μM affinity; Extended Data Fig. 1a,b). Following partial noising and denoising, we identified diverse designs (Supplementary Fig. 8) that *in silico* had substantially improved computational metrics compared to the starting design (Supplementary Fig. 9). We used an auxiliary potential during the denoising trajectory[8] that minimized the radius of gyration of the protein–peptide complex to promote additional interaction with the peptide. Initial screening with yeast surface display revealed quite high binding success rates, with 25 out of 96 designs binding GCG, and 20 out of 96 binding NPY at 10 nM peptide concentration. The highest-affinity designs were expressed in *E. coli* and then purified, and their binding affinities were determined using FP to be 5.6 nM to NPY (Fig. 2b, top) and below the limit of detection in the picomolar range to GCG (Fig. 2b, bottom). The designs were quite specific: the GCG binder bound preferentially to GCG over the closely related SCT, and particularly notably, the NPY binder did not show any cross-reactivity to peptide YY (PYY), a member of the NPY and pancreatic polypeptide family[25] with 63% sequence identity to NPY (Fig. 2b).

To gain insight into the structural rearrangements generated by partial diffusion that contribute to the affinity increases, we solved the structures (Extended Data Table 2) of the original Inpainted

GCG binder and the partially diffused higher-affinity GCG binder. Both designs were very close to their design models. Subtle structural changes in the protein backbone between the original Inpainted design model and the partially diffused model are nearly perfectly recapitulated in the corresponding crystal structures (Fig. 2c). Alignment of the two crystal structures (Fig. 2d) on the structurally conserved C-terminal residues (16–29) of GCG (Supplementary Fig. 10) showed that in the partially diffused GCG binder a 2.7-Å shift towards the target in the binder backbone enables an isoleucine to fit into a pocket previously occupied by a phenylalanine side chain at position 13 (Fig. 2e, left inset). Similarly, at position 16, a 3.6-Å shift in the backbone allows a tyrosine residue to pack underneath the peptide and form a hydrogen bond to the peptide backbone where previously a serine could not make any contacts (Fig. 2e, right inset). These backbone movements and accompanying sequence changes increase the interaction shape complementarity (0.62 versus 0.67) and contact molecular surface (431 $Å^2$ versus 522 $Å^2$; computed on the crystal structures). We observed similar improvements in estimated binding energy (Rosetta ddG) and contact molecular surface after running partial diffusion starting from the Inpainted designs for GCG and NPY (Supplementary Fig. 11).

De novo binder design using RFdiffusion

Inspired by this success at optimizing binders with RFdiffusion, we next tested its ability to design binders completely *de novo* through unconditional binder design. We first used the fixed target structure approach of ref.[22], and provided RFdiffusion with the sequence and structures of the two peptides in helical conformations, leaving the topology of the binding protein and the binding mode completely unspecified (Fig. 3a). From this minimal starting information, RFdiffusion generated designs predicted by AF2 to fold and bind to the targets with high *in silico* success rates. A representative design trajectory is shown for PTH in Supplementary Video 1; starting from a random distribution of residues surrounding the PTH peptide in a helical conformation, in sequential denoising steps the residue distribution shifts to surround the peptide and progressively organizes into a folded structure that cradles almost the entire surface of the peptide. We obtained synthetic genes encoding 96 designs for each target. Using yeast surface display, we found that 56 of the 96 designs bound to PTH at 10 nM peptide concentration. The highest-affinity design again bound too tightly for accurate K_d estimation; instead FP data provide an approximate upper bound for the K_d of < 500 pM (Fig. 3b, bottom). Binding was also

highly specific; no binding was observed to the related PTHrp (Fig. 3b, bottom). For Bim, 25/96 of the designs bound by yeast surface display, and FP on the highest-affinity design indicated a K_d of < 500 pM (Fig. 3c, bottom). Circular dichroism temperature melts indicate that both binders are stable at 95 °C (middle panels of Fig. 3b,c). The completely *de novo* diffused binders again had considerable structural similarity to our starting groove binding concept (compare the top panels of Fig. 3b,c to the middle panel of Fig. 1b). We solved the X-ray crystal structure (Extended Data Table 2) of the Bim binder, and found that it closely matched the design model (3.0 Å resolution, 0.57 Å RMSD; Fig. 3d). A kinked helix on the binder adjacent to the interface is well recapitulated in the structure, and a cross-interface hydrogen-bond network designed between Thr73 and Asn77 of the binder and Asn20 of Bim forms in the otherwise hydrophobic interface.

We next sought to generalize RFdiffusion to enable binding to flexible targets from a specification of the target sequence alone (as can be achieved with AF2 Hallucination, detailed above). We fine-tuned RFdiffusion by training on two chain systems from the Protein Data Bank (PDB), noising the structure on one and providing only the sequence on the second. We found that the fine-tuned version could readily design folded structures around a variety of peptides given only sequence information. We used this approach to design binders to PYY (Fig. 3e), which in the cryogenic electron microscopy structure with the NPY Y2 receptor is incompletely resolved and adopts a partially helical structure[26] Starting from only the amino acid sequence of PYY, RFdiffusion generated solutions with the peptide in a range of conformations. A design with the peptide adopting a different conformation from the experimental structure bound PYY with 24.5 nM affinity (Fig. 3e, right). Note that here we explored using shorter binder chain lengths in these calculations, resulting in smaller designs, which probably accounts for the lower affinity than in the fixed structure case above. Lower affinity binders were also obtained for PTH and GCG using this flexible backbone RFdiffusion approach (Extended Data Fig. 5a,b).

Human versus machine problem solving

The deep learning methods largely converged on the overall solution to the helical peptide-binding design problem—groove-shaped scaffolds with helices lining the binding site—that the human designers chose in the initial Rosetta parametric approaches. The increased

affinity of the deep learning designs probably derives at least in part from higher shape complementarity resulting from direct building of the scaffold to match the peptide shape; the average contact molecular surface for the partially diffused GCG binders and NPY increased by 33% and 29% respectively compared to that of the starting models, and the Rosetta ddG improved by 29% and 21% (Supplementary Fig. 11). The ability of RFdiffusion *de novo* design to rapidly 'build to fit' provides a general route to creating high-shape-complementary binders to a wide range of target structures, and as noted above, partial diffusion provides a general route to sampling binders with increased affinity by making small backbone adjustments to enable placement of more space-filling side chains.

Design of protein biosensors

Given our success in generating *de novo* binders to clinically relevant helical peptides, we next sought to test their use as detection tools for use in diagnostic assays. Compared to immunosensors, *de novo* protein-based biosensors can offer a more robust platform with high stability and tunability for diagnostics[27]. To design PTH biosensors, we grafted the 6.1-nM PTH binder into the lucCage system[28] (Fig. 4a), screened eight designs for their luminescence response in the presence of PTH, and identified a sensitive lucCagePTH biosensor (LOD = 10 nM) with ≈21-fold luminescence activation in the presence of PTH (Fig. 4b).

Enrichment for LC–MS/MS detection

We explored the use of our picomolar-affinity RFdiffusion-generated binders to PTH and GCG as capture reagents in immunoaffinity enrichment coupled with liquid chromatography–tandem mass spectrometry (LC–MS/MS), a powerful platform for detecting low-abundance protein biomarkers in human serum[29]. We prepared PTH- and GCG-binder-conjugated beads as described in the Methods. PTH enrichment was quantified on the basis of the analysis of the amino-terminal peptide of a tryptic digestion of PTH in human plasma[30] (Methods and Extended Data Fig. 6a). We found that the designed binder enabled capture of PTH from buffer and human plasma supplemented with PTH (the endogenous levels are too low for reliable detection) with recoveries of 53% and 43%, respectively (Fig. 4d, left). For GCG, enrichment was quantified for recovery of peptide in buffer solution (see Methods and Extended Data Fig. 6b) because recovery was low in extract (further increases in specificity of this design will probably be necessary for

actual applications). The GCG binder beads had comparable peptide capture efficiency to that of monoclonal GCG antibody beads, with 91.1% recovery when normalized to the antibody's 100% recovery rate in a spiked buffer (Fig. 4d, right). In contrast to the antibody-coupled beads, which lost almost all GCG-binding activity after the first use (Fig. 4d, right), the GCG-binder-conjugated beads retained almost full binding activity in a second capture experiment (Fig. 4d, right). This greater robustness to washing and repeated use probably reflects the exceptional stability of the designed binders (middle panels of Fig. 3b,c and Extended Data Fig. 3e), which could substantially lower cost (as they are no longer single use) and extend shelf life compared to using antibodies.

Discussion

Antibodies have served as the industry standard for affinity reagents for many years, but their use is often hampered by variable specificity and stability[31,6]. For binding helical peptides, the computationally designed helical scaffolds described in this paper have a number of structural and biochemical advantages. First, the extensive burial of the full length of an extended helix is difficult to accomplish with antibody loops[32] , but very natural with matching extended α-helices in groove-shaped scaffolds. Second, designed scaffolds are more amenable to incorporation into sensors as illustrated by the lucCagePTH sensor. Third, they are more stable than antibodies, can be produced much less expensively, and can be easily incorporated into affinity matrices for enrichment of peptide hormones from human serum (the striking difference in the robustness of antibody-conjugated versus binder-conjugated beads to repeated use (Fig. 4d, right) highlights the differences in stability of the two modalities). Fourth, computational design avoids the need to immunize animals, which often mount weak responses to highly conserved bioactive molecules[33]. MS-based detection of peptides following enrichment using designed binders could provide a general route forwards for serological detection of a wide range of disease-associated peptide biomarkers.

Our results highlight the emergence of powerful new deep learning methods for protein design. The RF_{joint} and RFdiffusion methods were both able to improve on initial Rosetta designs, and the Hallucination approach generated high-affinity binders without requiring pre-specification of the bound structures. Moreover, the RFdiffusion method rapidly generated very tight (picomolar

K_d values) affinity and specific binders to several helical peptides. RFdiffusion was previously shown to be able to design binders to folded targets[22] ; here we demonstrate further that it can be used to improve starting designs by partial noising and denoising, and can generate binders to peptides starting from no information other than the target sequence. To our knowledge, the Bim- and PTH-binding proteins diffused starting from random noise are the highest-affinity binders to any target (protein, peptide or small molecule) achieved directly by computational design with no experimental optimization. We expect both the RFdiffusion de novo peptide binder design capability and the ability to resample around initial designs (before or after experimental characterization) to be broadly applicable.

Figures

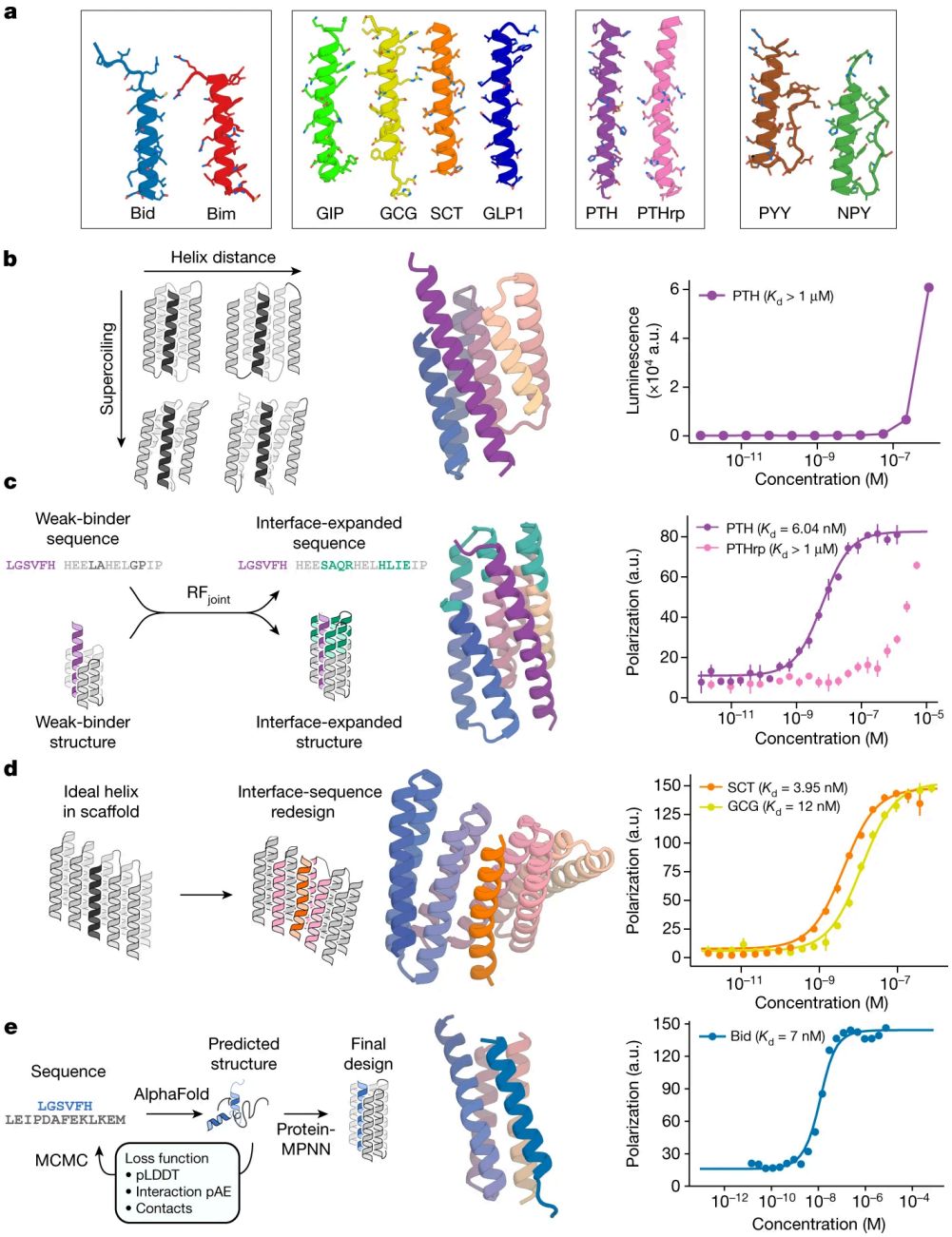


Figure 1: Design strategies for binding helical peptides. (a) Helical peptide targets: apoptosis-related BH3 domains of Bid[34] (PDB ID: 4QVE) and Bim[35] (PDB ID: 3FDL), GCG[36] (PDB ID: 1GCN), gastric inhibitory peptide[37] (GIP; PDB ID: 2QKH), SCT [38] (PDB ID: 6WZG), GCG-like peptide 1[39] (GLP1; PDB ID: 6X18), PTH (DOI: 10.1074/jbc.M001134200) (PDB ID: 1ET1), PTHrp[10] (PDB ID: 7VVJ), PYY[40] (PDB ID: 2DEZ) and NPY[41] (PDB ID: 7X9A). (**b**

Parametric approach. Left: sampling groove scaffolds varying supercoiling and helix distance to fit different targets. Middle: design model (spectrum) and PTH target (purple) of the best parametrically designed PTH binder. Right: split NanoBiT titration of PTH and the binder showed weak binding. a.u., arbitrary units. **(c)** Inpainting binder optimization. Left: redesign of parametrically generated binder designs using RF_{joint} Inpainting to expand the binding interface and ProteinMPNN to redesign the sequences. Middle: AF2 prediction of Inpainted design (spectrum) with extended interface (teal), and PTH target (purple). Right: FP measurements ($n = 4$) indicate 6.04 nM binding to PTH and weak binding to off-target PTHrp. **(d)** Threading approach to peptide binder design. Left: starting with a helix-bound scaffold, a target is threaded onto the bound helix and the interface is redesigned. Middle: AF2 prediction of design (spectrum) and SCT target (orange). Right: FP measurements ($n = 4$) indicate 3.95 nM binding to SCT and 12 nM binding to GCG. **e,** Hallucinating peptide binders. Left: Markov chain Monte Carlo (MCMC) steps are carried out in sequence space. At each step, the peptide sequence is re-predicted, and changes are accepted or rejected on the basis of interfacial contacts and AF2 metrics. The final structure is then redesigned using ProteinMPNN to avoid adversarial sequences. Interaction pAE, predicted alignment error across the interface; pLDDT, predicted local distance difference test. Middle: AF2 prediction of design (spectrum) and Bid target (blue). Right: FP measurements ($n = 4$) indicate 7 nM binding affinity to Bid.

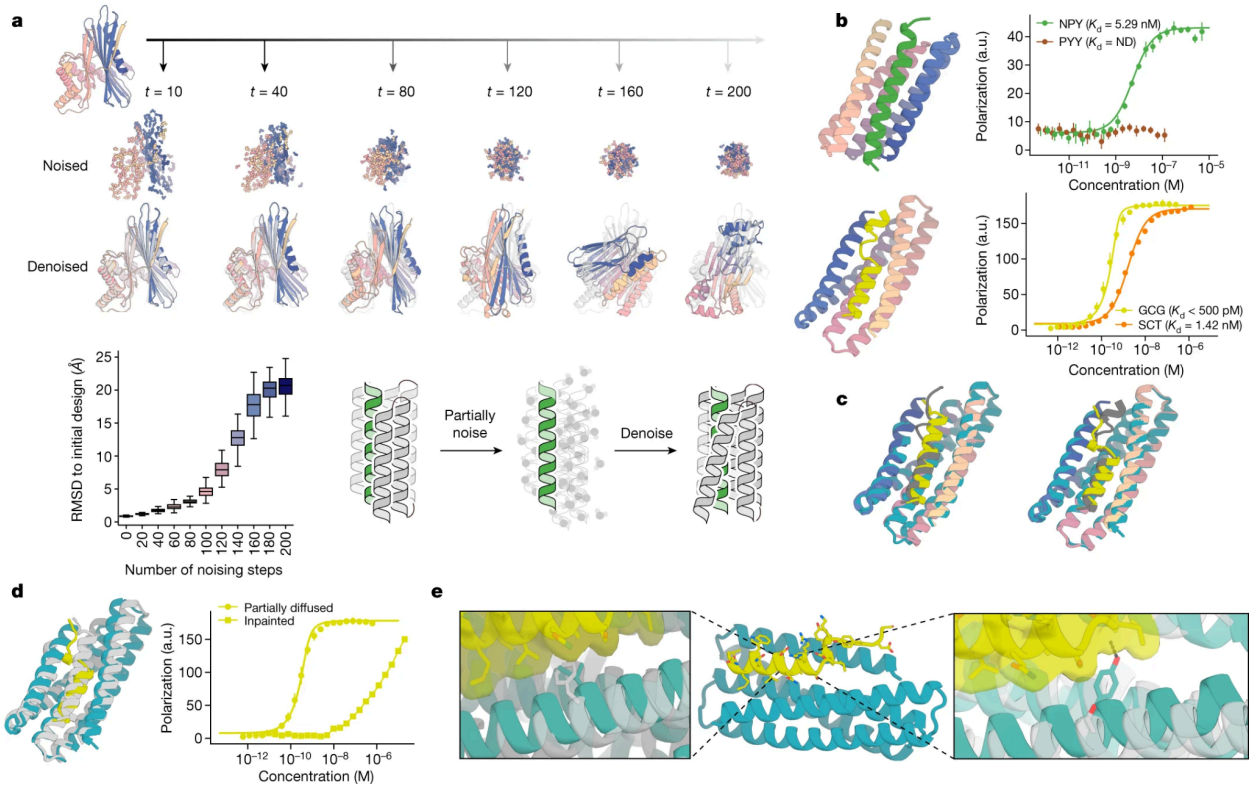


Figure 2: Peptide binder optimization with RFdiffusion. (a) Top: partial diffusion. RFdiffusion is used to denoise a randomly noised starting design (left); varying the extent of initial noising (middle row) enables control over the extent of introduced structural variation (bottom row; colors, new designs; gray, original design). Bottom left: partial diffusion diversifies designs. Note that the greater the amount of noise added, the more dissimilar the outputs are to the starting structure. Bottom right: depiction of the helix binder optimization strategy. **(b)** Top: design model (spectrum) of the partially diffused binder to NPY (green) and FP measurements ($n = 4$) indicating a 5.3 nM binding affinity to NPY target and selectivity over PYY (brown). ND, not detectable. Bottom: design model (spectrum) of the partially diffused binder to GCG (yellow) and FP measurements ($n = 4$) indicating a subnanomolar binding affinity to GCG and selectivity over SCT (orange). **(c)** Left: model (spectrum with GCG in gray) aligns with 0.72 Å RMSD to the 1.95-Å crystal structure (teal and yellow) of the RF_{joint} Inpainted GCG binder. Right: model (spectrum with GCG in gray) aligns with 0.6 Å RMSD to the 1.81-Å crystal structure (teal and yellow) of the partially diffused GCG binder. **(d)** Left: the crystal structures of the Inpainted (gray) and partially diffused (teal and yellow) GCG binders have considerable topological similarity; there are many small readjustments. Right: FP titrations ($n = 4$) with GCG

indicate much tighter binding following partial diffusion. e, Left inset: the crystal structure of the partially diffused backbone (teal) shows how the newly introduced Ile13 increases shape complementarity compared to the phenylalanine in the Inpainted binder (crystal structure in gray; structures aligned on residues 16–29 of GCG). Middle: crystal structure of the partially diffused GCG binder (teal and yellow). Right inset: the backbone shifts in the partially diffused structure (teal) enable Tyr16 to make packing and hydrogen-bonding interactions with the peptide; Ser16 in the original design did not make any peptide contacts (gray).

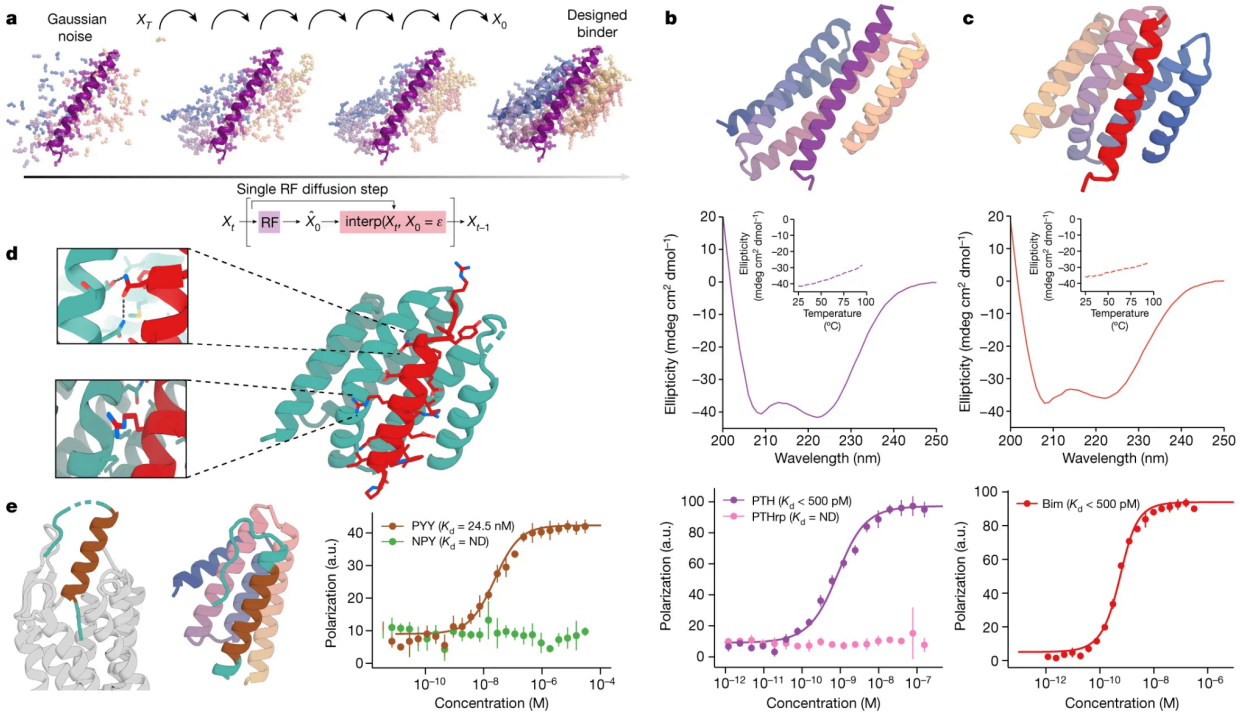


Figure 3: De novo peptide binder design with RFdiffusion. (a) Schematic showing peptide binder design using RFdiffusion. Starting from a random distribution of residues around the target peptide (X_T), successive RFdiffusion denoising steps progressively remove the noise leading at the end of the trajectory to a folded structure, X_0, cradling the peptide. At each step t, RFdiffusion predicts the final structure pX_0 given the current noise sample X_t, and a step that interpolates in this direction is taken to generate the input for the next denoising step X_{t-1}. **(b)** Design of picomolar-affinity PTH binder. Top: design model of PTH binder (spectrum, AF2 metrics in Supplementary Table 9). Middle: circular dichroism data show that the binder has helical secondary structure and is stable at 95 °C (inset). Bottom: FP measurements ($n = 4$) with PTH indicate a subnanomolar binding affinity and no binding to PTHrp indicates high

specificity. **(c)** Design of picomolar-affinity Bim binder. Top: design model of Bim binder (spectrum, AF2 metrics in Supplementary Table 9). Middle: circular dichroism data show that the binder has helical secondary structure and is stable at 95 °C (inset). Bottom: FP measurements ($n = 4$) with Bim indicate a subnanomolar binding affinity. **(d)** Crystal structure of Bim binder (teal and red). Top inset: a cross-interface hydrogen-bond network formed between Asn20 of Bim and Thr73 and Asn77 of the binder. Bottom inset: a kinked helix in the diffused backbone accommodates Arg13 of Bim. **e**, RFdiffusion with PYY sequence input alone. Left: PYY in complex with its native NPY Y2 receptor[26] (PDB ID: 7YON) shows flexibility at its N and C termini (teal). Middle: design model of the binder (spectrum) with PYY target (brown); the peptide is more ordered in both regions (N terminus, teal). Right: FP measurements ($n = 4$) with PYY indicate a 24.5 nM binding affinity

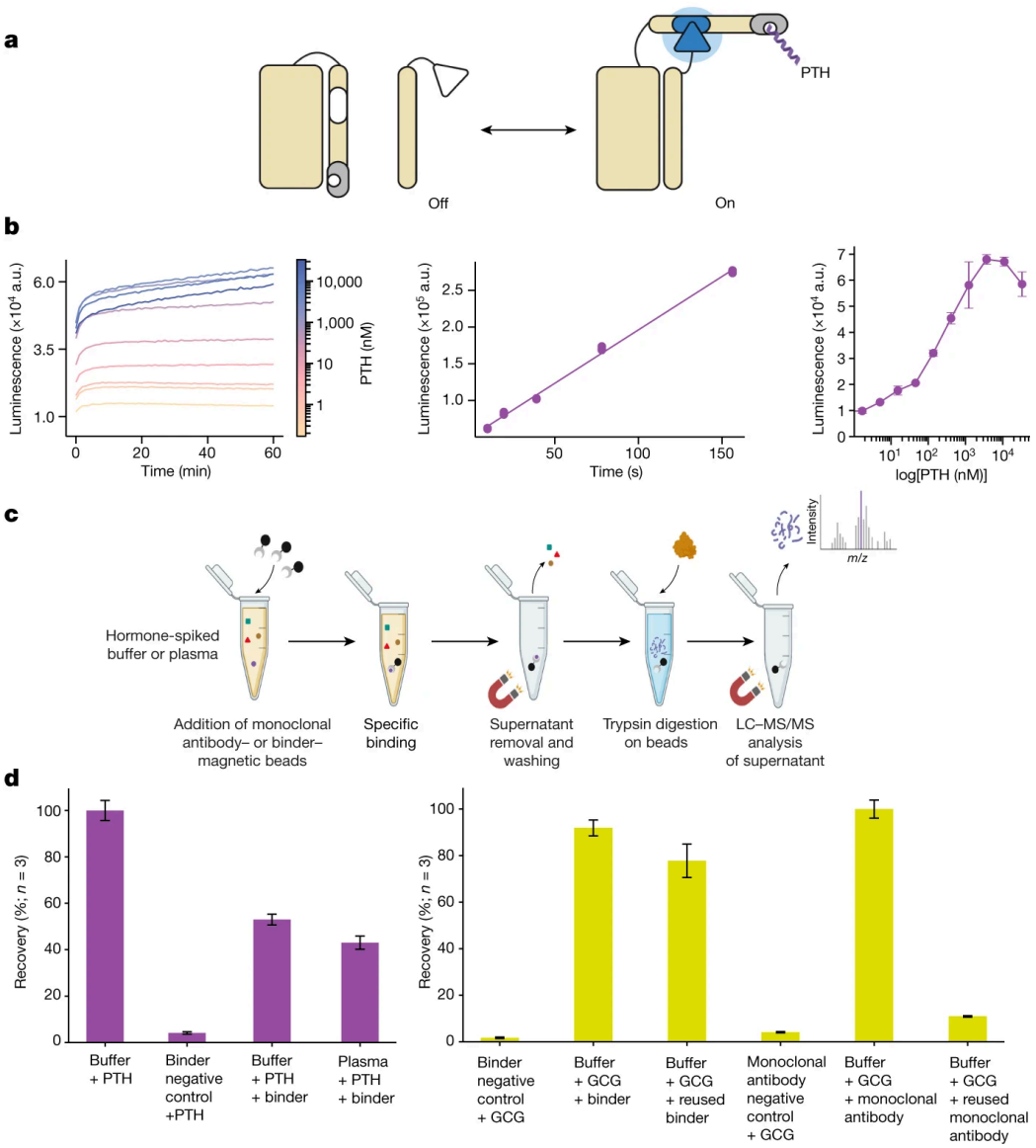


Figure 4: Application of designed binders to sensing and detection. (a) The PTH lucCage biosensor. Cage and latch (left, beige), key (right, beige) and the PTH binder (grey) thermodynamically shift from the off to on state in the presence of PTH peptide target (purple). This conformational change brings two luciferase halves (inactive in white, active in blue) close together leading to luminescence. **(b)** Left: titration of PTH results in luminescence increase ($n = 3$). Middle: response of lucCagePTH biosensor in the linear concentration range, indicating a 10 nM limit of detection (Supplementary Methods). Right: titration curve of 10 nM lucCagePTH + lucKey to various concentrations of PTH ($n = 3$). **(c)** LC–MS/MS enrichment experiment schematic; the trypsin digestion step was skipped for the GCG binder. **(d)** Left: LC–MS/MS recovery percentages for triplicate measurements of an N-terminal tryptic peptide of

PTH. The negative control comprised bovine serum albumin mixed with PTH in a buffer solution. Right: recovery percentage for triplicate measurements of intact GCG peptide normalized to the percentage recovery with a monoclonal antibody ($n = 3$). Following the first binding and elution experiments, beads were extensively washed and resuspended in PBS–CHAPS 0.1%, and then used in a second pulldown experiment. An unrelated binder attached to the magnetic beads mixed with GCG in buffer was used as a negative control. **a**,**c**, Created with BioRender.com.

References

1. Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39–42 (2000).
2. Goodsell, D. S. *The Machinery of Life*. (Copernicus Books, New York, 2009).
3. Duffuler, P., Bhullar, K. S., De Campos Zani, S. C. & Wu, J. Bioactive Peptides: From Basic Research to Clinical Trials and Commercialization. *J. Agric. Food Chem.* **70**, 3585–3595 (2022).
4. Greening, D. W., Kapp, E. A. & Simpson, R. J. The Peptidome Comes of Age: Mass Spectrometry-Based Characterization of the Circulating Cancer Peptidome. in *The Enzymes* vol. 42 27–64 (Elsevier, 2017).
5. Wootten, D. & Miller, L. J. Structural Basis for Allosteric Modulation of Class B G Protein–Coupled Receptors. *Annu. Rev. Pharmacol. Toxicol.* **60**, 89–107 (2020).
6. Bradbury, A. & Plückthun, A. Reproducibility: Standardize antibodies used in research. *Nature* **518**, 27–29 (2015).
7. Grigoryan, G. & DeGrado, W. F. Probing Designability via a Generalized Model of Helical Bundle Geometry. *J. Mol. Biol.* **405**, 1079–1100 (2011).
8. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).
9. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
10. Kobayashi, K. *et al.* Endogenous ligand recognition and structural transition of a human PTH receptor. *Mol. Cell* **82**, 3468-3483.e5 (2022).
11. Yin, H. *et al.* Computational Design of Peptides That Target Transmembrane Helices.

Science **315**, 1817–1822 (2007).

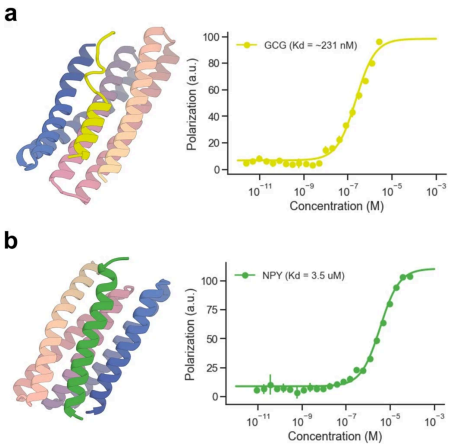
12. Praetorius, F. *et al.* Design of stimulus-responsive two-state hinge proteins. *Science* **381**, 754–760 (2023).
13. Ghirlanda, G., Lear, J. D., Lombardi, A. & DeGrado, W. F. From synthetic coiled coils to functional proteins: automated design of a receptor for the calmodulin-binding domain of calcineurin. *J. Mol. Biol.* **281**, 379–391 (1998).
14. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
15. Bennett, N. R. *et al.* Improving de novo protein binder design with deep learning. *Nat. Commun.* **14**, 2625 (2023).
16. Fukuhara, S. *et al.* Structure of the human secretin receptor coupled to an engineered heterotrimeric G protein. *Biochem. Biophys. Res. Commun.* **533**, 861–866 (2020).
17. Anishchenko, I. *et al.* De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
18. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).
19. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38 (2009).
20. Liu, Q. *et al.* Apoptotic Regulation by MCL-1 through Heterodimerization. *J. Biol. Chem.* **285**, 19615–19624 (2010).
21. Crabtree, M. D., Mendonça, C. A. T. F., Bubb, Q. R. & Clarke, J. Folding and binding pathways of BH3-only proteins are encoded within their intrinsically disordered sequence, not templated by partner proteins. *J. Biol. Chem.* **293**, 9718–9723 (2018).
22. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
23. Khatib, F. *et al.* Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci.* **108**, 18949–18953 (2011).
24. Kühlmann, U. C., Pommer, A. J., Moore, G. R., James, R. & Kleanthous, C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* **301**, 1163–1178 (2000).
25. Larhammar, D. Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul.*

Pept. **62**, 1–11 (1996).

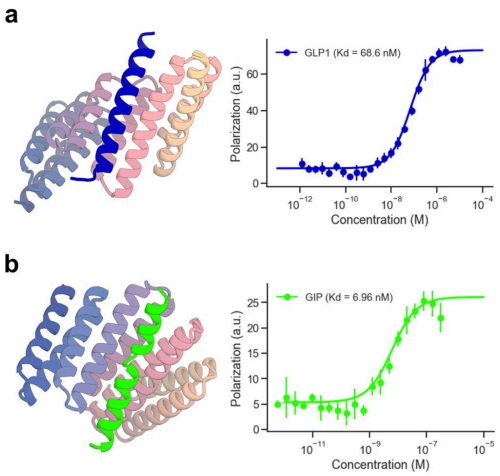
26. Kang, H. *et al.* Structural basis for Y2 receptor-mediated neuropeptide Y and peptide YY signaling. *Structure* **31**, 44-57.e6 (2023).
27. Säll, A. *et al.* Advancing the immunoaffinity platform AFFIRM to targeted measurements of proteins in serum in the pg/ml range. *PLOS ONE* **13**, e0189116 (2018).
28. Quijano-Rubio, A. *et al.* De novo design of modular and tunable protein biosensors. *Nature* **591**, 482–487 (2021).
29. Shi, J. *et al.* A distributable LC-MS/MS method for the measurement of serum thyroglobulin. *J. Mass Spectrom. Adv. Clin. Lab* **26**, 28–33 (2022).
30. Hoofnagle, A. N., Becker, J. O., Wener, M. H. & Heinecke, J. W. Quantification of Thyroglobulin, a Low-Abundance Serum Protein, by Immunoaffinity Peptide Enrichment and Tandem Mass Spectrometry. *Clin. Chem.* **54**, 1796–1804 (2008).
31. Baker, M. Reproducibility crisis: Blame it on the antibodies. *Nature* **521**, 274–276 (2015).
32. Lee, J. H., Yin, R., Ofek, G. & Pierce, B. G. Structural Features of Antibody-Peptide Recognition. *Front. Immunol.* **13**, 910367 (2022).
33. Zhou, H. *et al.* Generation of Monoclonal Antibodies against Highly Conserved Antigens. *PLoS ONE* **4**, e6087 (2009).
34. Rajan, S., Choi, M., Baek, K. & Yoon, H. S. Bh3 induced conformational changes in B cl‐ X $_{l}$ revealed by crystal structure and comparative analysis. *Proteins Struct. Funct. Bioinforma.* **83**, 1262–1272 (2015).
35. Lee, E. F. *et al.* High‐Resolution Structural Characterization of a Helical α/β‐Peptide Foldamer Bound to the Anti‐Apoptotic Protein Bcl‐x $_{L}$. *Angew. Chem. Int. Ed.* **48**, 4318–4322 (2009).
36. Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. & Blundell, T. X-ray analysis of glucagon and its relationship to receptor binding. *Nature* **257**, 751–757 (1975).
37. Parthier, C. *et al.* Crystal structure of the incretin-bound extracellular domain of a G protein-coupled receptor. *Proc. Natl. Acad. Sci.* **104**, 13942–13947 (2007).
38. Dong, M. *et al.* Structure and dynamics of the active Gs-coupled human secretin receptor. *Nat. Commun.* **11**, 4137 (2020).
39. Zhang, X. *et al.* Differential GLP-1R Binding and Activation by Peptide and Non-peptide Agonists. *Mol. Cell* **80**, 485-500.e7 (2020).

40. Nygaard, R., Nielbo, S., Schwartz, T. W. & Poulsen, F. M. The PP-Fold Solution Structure of Human Polypeptide YY and Human PYY3-36 As Determined by NMR ˈ. *Biochemistry* **45**, 8350–8357 (2006).
41. Tang, T. *et al.* Receptor-specific recognition of NPY peptides revealed by structures of NPY receptors. *Sci. Adv.* **8**, eabm1232 (2022).

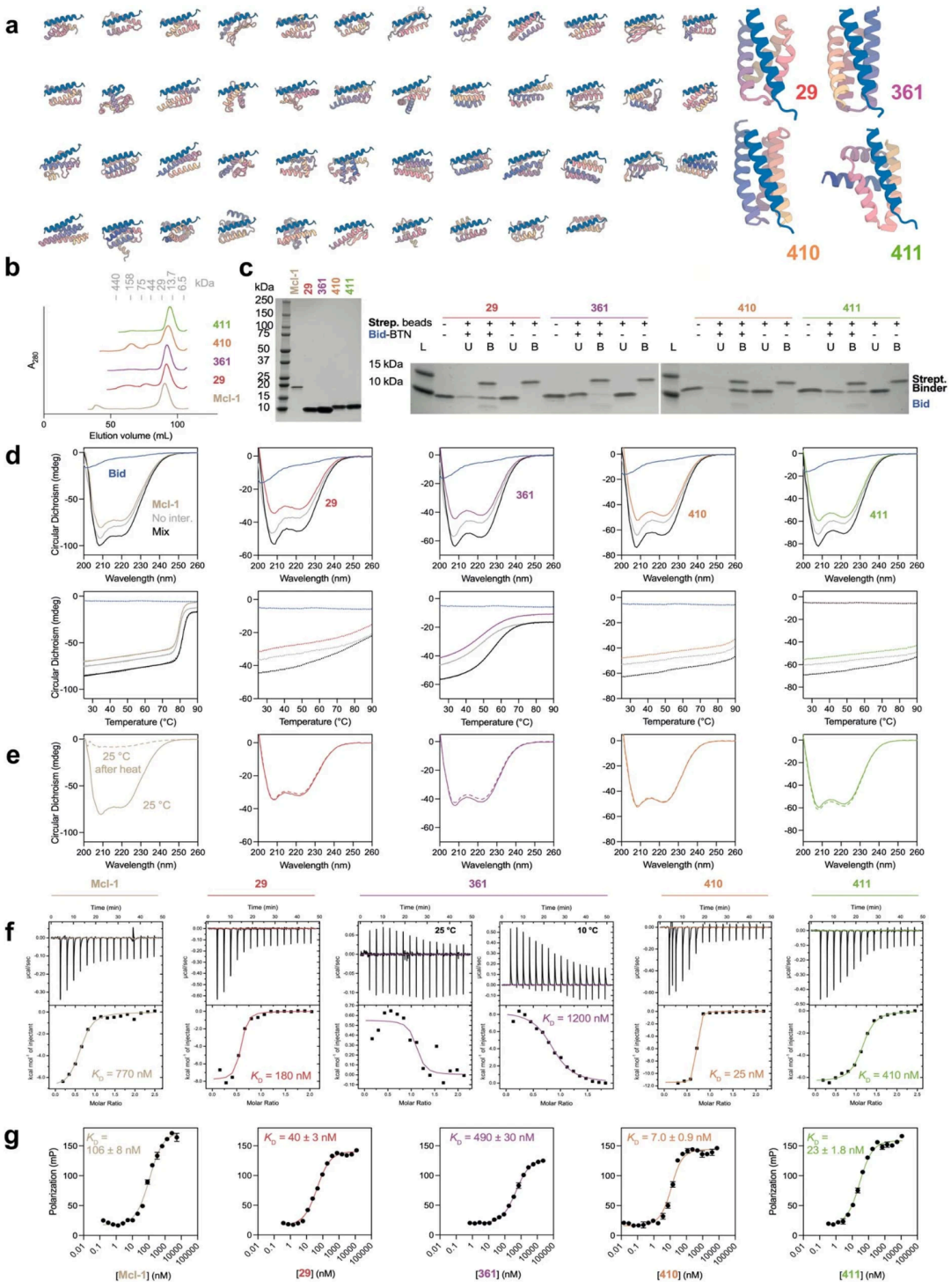
Extended data



Extended Data Fig. 1: Low affinity RF_{joint}-Inpainted binders for NPY and GCG using extended parametric designs. (a) Left: Design model (color spectrum + yellow) of the tightest GCG binder. Right: FP titration (n = 4) for the tightest GCG binder indicates ~ 231 nM binding affinity **(b)** Left: Design model (color spectrum + dark green) of the tightest NPY binder. Right: FP titration (n = 4) for the tightest NPY binder indicates 3.5 μM binding affinity.

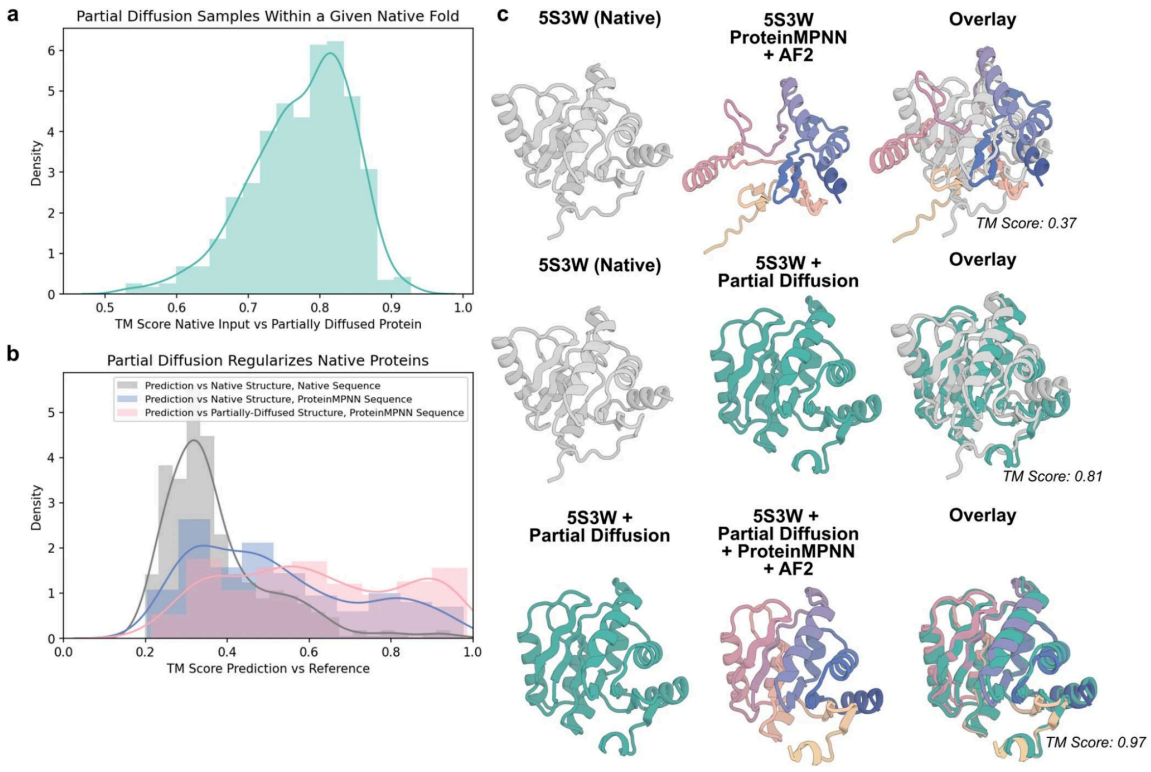


Extended Data Fig. 2: Additional binders made using threading and redesign. (a) Left: Design model (color spectrum + dark blue) of the tightest GLP1 binder. Right: FP titration (n = 4) for the tightest GLP1 binder indicates 68.8 nM binding affinity. **(b)** Left: Design model (colour spectrum + green) of the tightest GIP binder. Right: FP titration (n = 4) for the tightest GIP binder indicates 6.96 nM binding affinity.



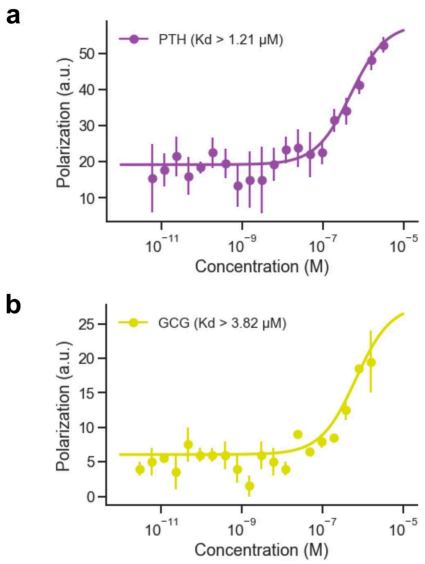
Extended Data Fig. 3: Hallucinated Bid binders are stable and bind Bid peptide with high affinity. (a) 46 Hallucinated designs tested for initial experimental screening. **(b)** 4 designs were chosen for expression without Bid peptide. All expressed as monomeric proteins (assessed by

preparative SEC) and were pure by SDS-PAGE (n = 1). **(c)** All Hallucinations could be pulled-down by biotinylated Bid immobilized on streptavidin magnetic beads. B = bound to bead, U = unbound, in supernatant. L = ladder (n = 1). **(d)** Bid is unstructured in isolation by circular dichroism (CD), whereas all Hallucinations were helical in isolation, as predicted from the Hallucinated structure. A 1:1 molar ratio of binder:Bid (Mix) produced greater helical signal than that predicted by the isolated spectra (No inter.) suggesting binding is inducing helix formation (n = 1). **(e)** Melting with CD showed that Hallucinations were thermostable, and binding to Bid increased thermostability (where measurable) (n = 1). All Hallucinated binders would remain folded, or refold after heating and cooling, in contrast to the natural binder Mcl-1 which precipitated in the process. **(f)** ITC showed that Hallucinations bound to Bid, with μM to nM *K*ds (n = 1). **(g)** FP measurements of designed Bid binders (n = 3).

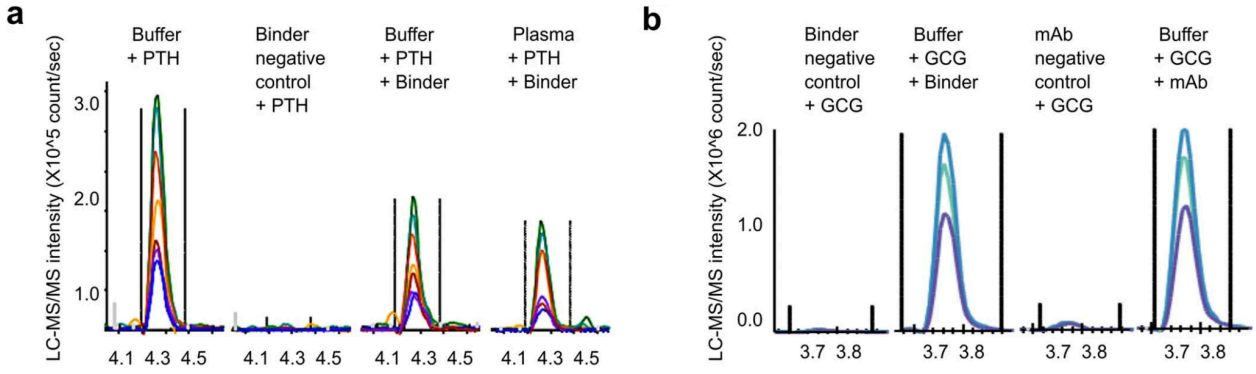


Extended Data Fig. 4: Partial diffusion increases designability of native proteins. 500 native proteins of length 100 to 300 residues were selected from the PDB (< 3.5 Å resolution and no missing residues). Three different methods were applied to these proteins: 1) single sequence AlphaFold2 (AF2), 2) ProteinMPNN combined with AF2, and 3) partial diffusion (60 steps, noise = 1), ProteinMPNN and AF2. **(a)** Partial diffusion generates diverse protein conformations

from the initial fold while maintaining the same overall fold, as indicated by the TM (Template Modeling) score exceeding 0.5. **(b)** The backbones resulting from partial diffusion exhibit higher designability compared to the native backbone, implying that they have been idealized for design purposes. **(c)** Visualization of an example where partial diffusion + ProteinMPNN results in a significantly more designable protein relative to sequence redesign by ProteinMPNN on the native backbone.



Extended Data Fig. 5: PTH and GCG binders designed with RFdiffusion. Representative binding data is shown for PTH **(a)** and GCG **(b)** binders designed by providing sequence input alone. The binding affinities, as measured by FP (n = 4), indicate low micromolar interactions with the respective peptide targets.



Extended Data Fig. 6: LC-MS/MS chromatograms for PTH and GCG binders. (a) LC-MS/MS chromatograms for SVSEIQLMHNLGK, the N-terminal tryptic peptide of PTH;

different peptide fragments detected by the LC-MS/MS assay are in different colors. **(b)** LC-MS/MS chromatograms for the intact GCG peptide HSQGTFTSDYSKYLDSRRAQDFVQWLMNT; different peptide fragments detected by the LC-MS/MS assay are in different colors.

Target	Computational method	Binder amino acid sequence
PTH	Parametric design	SQELIEELIKLAKELAEIKDEEERRKIKRELERLAEELKEAPASSLLRALALLVIALALIQAAESEEERERARELLERLE ELLRELEKQITDERFKEILRELEELAKELKKQL
PTH	Inpainting	SFELLEKLIELSKELYEVAKKYIETGDPELKKKLEEILKKIEEAYKELLESDAHPLEKALAKLILAEAYVVKSFAYISSG KDLEKAQEYLDKAKEILEELLKLLEELKKEETDPEKLEIIEELEKIAEELLKEIEE
GCG	Inpainting	MLDELFSLLNKMFELSDKYRELRKELRKAIESGAPEEELRELLEKMLEIAKKLLELTKELKKLVEDVLKNNPDPVER AKAVLLYAVGVHILYSESSELEVIAERLGFKDIAEKAKEIADKARELKEEVKRKLREIREEVPDPEIRKAAEEAIEML ESNDKRLKEFRKL
NPY	Inpainting	GAAEKLAELYEKFKALREKALEVLKKAVEALENKADKETLLKLIKELKELAEKFEELAEEFERNAGESTTASLNATA AYMARIGLLAALLALAKAAGVPEEELEEIKRRIEETAKRAIEAAERLKALAEARGDTKHVAVGVEAVRMATELYELA QKIIDAF
SCT	Threading	SEELEERLREARERLEEARERLEEAREEGDLREMARALLEEARAVLEIARVAAEAGDDEALREAARRAGEVIRRA GEVGLRAAEEGDTETIREAMLAILEAQRASAVIALHLARDDPEVAEALRVIERLLRTAERALREGQLEVARLATEA VEALADAILRAREIGRPELVREAARLAEEARRLLEAALEALRAGDEEGARERLARARELIREIRERVRRA
GIP	Threading	SPKEKAERLIKEAKEAAEKAKEAAERSGLEEAKKAAEELTKLLEEAAARVAADPEDETKLRALEKIVEAAKEAVKA LEVAIESGDEQLIRAALSLVEAAVHLAKALLAKPESPLVDFGFELLKLAAKTLAAYAEGEDVDKIALKLKAISAMAEA LRLALAGDLERAARAAEEAVRYAIEAGDKELLRLAAEVAAYIARLAEEAGLEEVARRAREAAERAREAAK
GLP1	Threading	SPEEEARRAAREAERAAREAREAARRLGDEESVRVAERLEREARRAERERDLELARRVLRAAEALRLALEGELL AREQGDELGVVVARMITLAARDSALGRGTPELARLLLRVARALLEGDLEEVVRSLAEIAKREIGTERALLAVEAIKL VALESIEEGDFETAELAIEKLREIAEEFEGTEVAEKAREAIEEIEKKKREAE
Bid	Hallucination	TPEDYRRAAELIKEIAREAERYAEGEISAEEALARIRRLRAELEECYEHGLDAVGRSYVDQARPLIDEIERLLQEKL DAE
GCG	RF Diffusion (partial)	SMEKLAEIMQEIIEAYQEVKDAFFKFIKAVHEGAPEEELKKYLEKMKEALEKMKELLERLEKEAKKVIEENKDKKLE LKVLLMLRLAYLLLKVSIELTKIAAEKLGDKELVEELEKESKEVEKKIKELEERIKKLLEEVDDEELKEAYKEVEEME KEAEKFLEKMRKV
NPY	RF Diffusion (partial)	GMEERRKELLEKLKKLKEEVVELFRELAQALRDGASKERLEEIRERAEKLAEEAKKVAEELEKLAEGDAVLQLYLA EAYALEAAALTIEAVAAAELGASKEELEKIKEKIEEALKKAEEAMKKALAEAKARGRERLVRLIEEARKEFEKLSKAI KELLEQV
PTH	RF Diffusion	MREKLEEMLEEFNEVIDELIEITKEDAPELEELRERAEEAVENERLDELEEILDELEVIILEAMFRDLSAAIEMTKAKN DKEKLKELLKQLEELEKRIKELLERAKKRGNKKIIEKLEKLLKEVEKLKKEIEEYLK
Bim	RF Diffusion	EEERKEKREKVRAGLKRAIAELPAEVAARCLALLDDASDEEFIEAVLEVLEAMREALVAMAREGRLDAVRRATSHI NEVLVDAAELALEKGREYFRRLCLIVCDMMIELIRLEPEQTPELRRIRERLEEIRRRLE
PYY	RF Diffusion (sequence only)	GLEEAEKLLEEIFANFEEIVELIKKNIGTERGKKLLKVFVATVDLILARLEQGADLAELAELVKEIAELAKDEEGLEEA EKLVKELTAAR

Extended Data Table 1 Amino acid sequences of peptide binders

	GCG_partdiff (8GJI)	GCG_inpaint (8GJG)	Bim_fulldiff (8T5E)	PTH (8T5F)
Data collection				
Space group	*P* 2_1 2_1 2_1	*P* 2_1	*P* 2_1 2_1 2_1	*P* 4 2_1 2
Cell dimensions				
a, *b*, *c* (Å)	31.26, 50.92, 91.92	37.79, 45.90, 63.60	25.79, 66.67, 74.30	91.32, 91.32, 37.73
α□β□γ□ □□□ □	90, 90, 90	90, 97.31, 90	90, 90, 90	90, 90, 90
Resolution (Å)	91.93 - 1.81 (1.88 - 1.81)	63.09 - 1.95 (2.00 - 1.95)	74.30 - 3.00 (3.18 - 3.00)	40.84 - 1.99 (2.04 - 1.99)
R_{merge}	0.099 (1.581)	0.073 (2.001)	0.064 (0.173)	0.120 (2.069)
I / cI	8.3 (0.90)	10.10 (0.60)	17.8 (8.8)	21.8 (1.3)
Completeness (%)	99.60 (99.40)	98.50 (96.30)	99.9 (100)	98.6 (94.6)
Redundancy	6.2 (6.0)	6.7 (6.8)	7.2 (7.9)	15.1 (15.2)
Refinement				
Resolution (Å)	45.96 - 1.81 (1.88 - 1.81)	63.09 - 1.95 (2.00 - 1.95)	49.62 - 3.00 (49.62 - 3.00)	40.84 - 1.99 (2.19 - 1.99)
No. reflections	13875 (1327)	15425 (2387)	2861 (445)	11302 (738)
R_{work} / R_{free}	0.2080 (0.3752)/ 0.2552 (0.4485)	0.2087 (0.4205)/ 0.2488 (0.4445)	0.2398 (0.2398)/ 0.2617 (0.2617)	0.2201 (0.2506)/ 0.2494 (0.3372)
No. atoms				
Protein	1579	1539	1244	853
Ligand/ion	0	0	0	0
Water	24	26	0	26
B-factors				
Protein	45.14	68.55	77.56	61.14
Ligand/ion	0	0	0	0
Water	47.64	69.57	n/a	62.39
R.m.s. deviations				
Bond lengths (Å)	0.012	0.002	0.003	0.010
Bond angles (°)	1.12	0.440	0.500	1.04

*Single Crystal used for each data collection. *Values in parentheses are for highest-resolution shell.

Extended Data Table 2 Crystallographic data collection and refinement

Experimental Methods

Gene construction of peptide hormone binders

The designed protein sequences were optimized for expression in *E. coli*. Linear DNA fragments (eBlocks, Integrated DNA Technologies) encoding design sequences included overhangs suitable for cloning into pETcon3 vector for yeast display(deposited in Addgene with the number #45121) and Golden Gate cloning into LM627 vector for protein expression (Addgene #191551). For initial testing hallucinated binders to Bid, binders were cloned into a modified LM627 vector. Specifically, Golden Gate cloning was used to generate sfGFP-Bid-STOP-[Binder]-SNAC-HISx6 assemblies.

Yeast display screening

For the yeast transformation, 50-60 ng of digested pETcon3 with the NdeI and XhoI restriction enzymes and 100 ng of insert (eBlocks, Integrated DNA Technologies) were transformed into *S. cerevisiae* EBY100 strain. EBY100 cultures were grown in C-Trp-Ura medium supplemented with 2% (w/v) glucose (CTUG). For induction of expression, yeast cells initially grown in CTUG were transferred to SGCAA medium supplemented with 0.2% (w/v) glucose and induced at 30 °C for 16–24 h. Cells were washed with PBSF (PBS with 1% (w/v) BSA) and labeled for 40 minutes with biotinylated peptide targets at room temperature using without-avidity labeling conditions[7]. After incubation time, cells were washed and resuspended in PBSF for individual sorting of cells harboring each unique design using a 96-well compatible autosampler in the Attune NxT Flow Cytometer (Thermo Fisher Scientific).

NanoBiT screening

Linear gene fragments encoding binder design sequences and target peptide sequences were cloned into *E. coli* expression vectors using Golden Gate assembly; these vectors were pET28b(+) derivatives genetically fusing the smBiT and lgBiT halves of the NanoLuc® Luciferase (Promega) to the binders and peptides respectively. Resulting plasmids were transformed into BL21* (DE3) (Invitrogen) *E. coli* competent cells, then grown in 1mL TBII in 96-deepwell plates at 37ºC and 600 rpm. After 2 hours, expression was induced with IPTG (0.1

mM) and cells were incubated for an additional 4 hours. Cells were harvested by centrifugation (15 min at 4000 x g), then resuspended in 100 µL lysis buffer (10 mM NaP pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mg/mL lysozyme, 10 µg/mL DNAse I, 1 tablet Complete Protease inhibitor / 50 mL). Cells were incubated for 1 hour at room temperature and 600 rpm, then frozen (-80C for 30min) and thawed (37ºC at 600 rpm for 30min) twice. Lysate was cleared by centrifugation (20 min at 4000 x g), and the soluble fraction was then transferred to a 96-well plate for use as stock protein/peptide for conducting the nanoBiT screen. Screens were assembled in 96-well Half Area Black Flat Bottom Polystyrene NBS Microplates (Corning 3686). Binder design smBiT lysate was diluted 12 µL into 1400 µL assay buffer (10 mM NaP pH 7.4, 150 mM NaCl), while target peptide lgBiT lysate was diluted 6 µL into 1400 µL assay buffer. Stock rows in the assay plate were prepared by mixing 40 µL substrate (499.2 µL assay buffer, 20.8 µL Nano-Glo® Luciferase Assay Substrate (Promega)) with 40 µL diluted binder design smBiT lysate, while experimental rows were prepared by adding 50 µL diluted target peptide lgBiT lysate. At read time, 50 µL of the stock row was added to the 50 µL experimental row and mixed quickly and carefully, then luminescence was read immediately for 5 min using a plate reader (Biotek Synergy Neo2).

Identification of weak binder hits from parametric designs

The first helical peptide binder hits were identified in experiments screening for binding using the nanoBiT split luciferase assay. These kinetic binding experiments were performed in cell lysate with no control over protein concentration, so candidate binders were selected qualitatively for showing some increase in luminescence signal over time above background noise, indicating likely binding activity. Additional binding curve experiments (Supplementary Fig. S4) indicated that this binding activity was at very weak affinities, likely >100 nM (indistinguishable from the background signal of weak luciferase binding for the assay). Therefore, these initial candidates were not further characterized, but rather selected for additional design to yield higher affinity binders.

Bicistronic protein expression

Hallucinated binders to Bid were screened by bicistronic expression with the Bid peptide. Plasmids encoding sfGFP-Bid-STOP-[Binder]-SNAC-HISx6 were cloned into *E. coli,* and 2 mL cultures of each of the 47 designs were grown overnight in LB. Cultures were diluted into TB medium, and grown to approximately OD280 0.6, before induction with 1 mM IPTG for 4 hours at 37ºC. Bacteria were lysed for 15 minutes in 300 B-PER (Thermo) + 1 mM PMSF, 0.1 mg/mL Lysozyme (Sigma), 0.01 mg/mL DNAse I. Lysates were clarified by centrifugation at 4000 *g* for 10 minutes, before purification on Ni-NTA resin (wash buffer: 20 mM Tris pH 8.0, 150 mM NaCl, 20 mM Imidazole; elution buffer: 20 mM Tris pH 8.0, 150 mM NaCl, 250 mM Imidazole). Eluates were assessed for GFP fluorescence on a fluorescence plate reader.

Peptide synthesis and purification

The PTH-TAMRA peptide was synthesized in-house on a CEM Liberty Blue microwave synthesizer. All L- and D-amino acids were purchased from P3 Biosystems. Oxyma Pure was purchased from CEM, DIC was purchased from Oakwood Chemical, diisopropyl ethylamine (DIEA) and piperidine were purchased from Sigma- Aldrich. Dimethylformamide (DMF) was purchased from Fisher Scientific and treated with an Aldraamine trapping pack prior to use. Synthesis was done on a 0.1 mmol scale on CEM Cl-TCP(Cl) resin. Five equivalents of each amino acid were activated using 0.1 M Oxyma with 2% (v/v) DIEA in DMF, 15.4% (v/v) DIC, and coupled on resin for 4 min with double coupling if needed. This was followed by deprotection using 5 mL of 20% piperidine in DMF for 2 min at 95 °C. Global deprotection was accomplished TFA/Water/TIPS (95:2.5:2.5) for 3 hours. This deprotection mixture was precipitated in 30 mL of ice-cold ethyl ether, centrifuged and decanted, then washed twice more with fresh ether and dried under nitrogen to yield crude peptide for high pressure liquid chromatography (HPLC) purification.

The TAMRA-Bid peptide was synthesized using a PurePep Chorus peptide synthesizer (Gyros Protein Technologies). Fmoc-protected amino acids were purchased from Ambeed. The peptide synthesis was performed at a 0.05 mmol scale on Rink amide AM resin (Matrix Innovation). Following reagents were prepared in dimethylformamide (DMF, VWR) for amino acid coupling: Fmoc-protected amino acids (0.2M),

O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU, 0.5 M, Ambeed), and diisopropyl ethylamine (DIEA, 1.0 M, Thermo Scientific). For peptide elongation, the resin was washed with DMF (3x 3 mL) before fmoc-deprotection was carried out using 3 mL 20 % (v/v) piperidine (Sigma-Aldrich) in DMF (2x, 2 min, room temperature, agitating at 300 rpm). After a DMF washing step (3x 3 mL), amino acid/HCTU/DIEA (1:1.25:2.5) was added to the resin in 5-fold excess and agitated (2x 10 min, 50 °C). After a final fmoc-deprotection, N-terminal coupling of 5-(and-6)-carboxytetramethylrhodamine (TAMRA, Anaspec Inc.) was accomplished using following conditions: TAMRA:benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP, Novabiochem):DIEA (1.5:1.5:3) in N-Methyl-2-pyrrolidone (NMP, Sigma-Aldrich) for 16 h at room temperature and 300 rpm. Tin foil was used to cover the reaction vessel to avoid photobleaching of the fluorophore. For global deprotection and cleavage from the resin, the peptide was incubated with TFA/Water/triisopropyl silane (TIPS, Sigma-Aldrich)/ 1,2-Bis(2-mercaptoethoxy)ethane, 3,6-Dioxa-1,8-octane-dithiol (DODT, Sigma-Aldrich) (92.5:2.5:2.5:2.5) for 2 h at room temperature while agitating. 30 mL of ice-cold diethyl ether (VWR) was added for peptide precipitation and the precipitated peptide was centrifuged at 2500 x g for 10 min at 4 °C. The supernatant was discarded, and the peptide was washed with 20 mL of diethyl ether. The peptide pellet was dissolved in water/acetonitrile/TFA (50:50:0.1) and lyophilized.

The crude peptide was dried and dissolved in a mixture of ACN and water where the entire crude is soluble. This solution was purified on a C18 column in an Agilent HPLC instrument. A linear gradient of increasing ACN with 0.1% TFA was used to purify the samples. UV signal was monitored at 214 nm and all peaks were collected. Peaks were checked using ESI mass spectroscopy for the correct peptide mass. The purified peptide was then lyophilized for further use.

Protein expression and purification in *E. coli* for peptide hormone binders

Protein expression was performed using 50 mL of the Studier autoinduction media supplemented with kanamycin, and grown overnight at 37°C. The cells were harvested by spinning at 4,000 x g for 10 min and then resuspended in lysis buffer (100 mM Tris-HCl, 200

mM NaCl, 50 mM imidazole) supplemented with protease inhibitor tablets (Pierce™ Protease Inhibitor Tablets, EDTA-free). Then, the cells were lysed by sonication in a Qsonica, Q500 with a: 4-pronged horn for 2:30 min ON total, with an amplitude of 80%. Soluble fractions were clarified by centrifugation at 14,000 x g for 40 minutes, and were subsequently purified by affinity chromatography using bed Ni-NTA resin (Qiagen or Thermo Fisher) on a vacuum manifold. A series of washes using Low-salt buffer (20 mM Tris-HCl, 200 mM NaCl, 50 mM imidazole) and High- salt buffer (20 mM Tris-HCl, 1000 mM NaCl, 50 mM imidazole) were performed prior to elution with Elution buffer (20 mM Tris-HCl, 200 mM NaCl, 500 mM imidazole). After elution, protein samples were filtered and injected into an autosampler-equipped Akta pure system on a Superdex S75 Increase 10/300 GL column at room temperature. The SEC running buffer was 20mM Tris-HCl, 100mM NaCl pH 8. We pooled the largest abundance monodisperse peak fractions and concentrated using Spin filters (3 kDa molecular weight cutoff, Amicon, Millipore Sigma) and stored at 4 °C before downstream characterizations. Protein concentrations were determined by absorbance at 280 nm using a NanoDrop spectrophotometer (Thermo Scientific) using their extinction coefficients and molecular weights obtained from their amino acid sequences using the ProtParam tool. We additionally verified the monodispersity of the pooled fractions by reinjecting them on the same column for the majority of the binders after 24h (Supplementary Fig. S18)

Fluorescence polarization

Fluorescence polarization binding assays were carried out in 96-well plates (Corning 3686), with two-fold serial dilution of designed peptide binders in the presence of 0.5 nM fluorescently labeled peptide targets. Protein and peptide were diluted from their stock concentration into 20mM Tris-HCl pH 8, 100mM NaCl, 0.1% v/v Tween 20, and the protein was titrated in 2-fold serial dilutions onto constant peptide. After incubating the peptide and binder for one hour at room temperature, the fluorescence polarization was measured at the excitation and emission wavelengths of the FAM dye (485/530 nm) or the TAMRA dye (530/590 nm), in a Synergy Neo2 multi-mode plate reader. Titrations were conducted in replicate, and the Kd was fitted with SciPy59. Specifically, curves were fit to observations of an observed signal, $Signal_i$, at titrated concentrations $[A_{tot}]_i$ according to the following equation:

$$Signal_i = Baseline + Amplitude \frac{AB_{conc}([A_{tot}]_i, [B_{tot}], K_d)}{[B_{tot}]},$$

Where $[B_{tot}]$ is the known total concentration of the binder, Baseline and Amplitude are free parameters, and the concentration of the bound state [AB] is computed as:

$$AB_{con}([A_{tot}]_i, ([B_{tot}], K_d) = ((([A_{tot}] + [B_{tot}] + K_d) \pm \sqrt{(([A_{tot}] + [B_{tot}] + K_d)^2 - 4[A_{tot}][B_{tot}])}/2$$

The unknown parameters (K_D, Baseline and Amplitude) were fit using scipy.optimize.curve_fit, $[B_{tot}]$ was additionally fit in the optimization, but only allowed to within 0.5 nM $\pm$ 0.1%.

Fluorescence polarization measurements for TAMRA-Bid were performed in 96-well, flat-bottom, half-area microplates (Corning 3881) using a CLARIOstar Plus plate reader (BMG Labtech) set to excitation and emission wavelengths of 540 and 590 nm, respectively. TAMRA-Bid peptide and hallucinated proteins were diluted into a buffer containing 50 mM sodium phosphate pH 7.0 and 0.05 % v/v Tween-20. The concentration of TAMRA-Bid peptide was kept constant at 10 nM. For the proteins, a 2-fold serial dilution was performed and added to the peptide. After one to four hours of incubation at room temperature, FP measurements were conducted. Titrations were carried out in triplicates.

Peptides used for the assay are shown in Supplementary Table 1

Supplementary Table 1. Fluorophore-labeled peptides used in Fluorescence polarization assays

Peptide name	Sequence	Supplier	Cat #	Fluorophore
PTH-TAMRA	SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNF	In-house	NA	5-TAMRA

PTHrp-FAM	AVSEHQLLHDKGKSIQDLRRRFFLHHLIAEIHTAEIA	Phoenix Pharmaceuticals, Inc.	FG-056- 08A	FAM
SCT-FAM	HSDGTFTSELSRLREGARLQRLLQGLV	Phoenix Pharmaceuticals, Inc.	FG-067- 03A	FAM
GCG-FAM	HSQGTFTSDYSKYLDSRRAQDFVQWLMNT	Addex Bio	ABBFO2033	FAM
NPY-FAM	SKPDNPGEDAPAEDMARYYSALRHYINLITRQR	Phoenix Pharmaceuticals, Inc.	FG-049- 04A	FAM
PPY-FAM	IKPEAAGEDASPEELNRYYASLRHYLNLVTRQRY	Phoenix Pharmaceuticals, Inc.	FG-059- 02A	FAM
TAMRA-Bid	TAMRA-(b-A)-QEDIIRNIARHLAQVGDSMDRSIPPG	In-house	NA	5-TAMRA

Cloning, expression and purification of Bid-binding hallucinations, Avi-tagged Bid peptide and MCL-1

Bid-binding Hallucinations were cloned into a pET28 vector, containing an N-terminal His_{10} and a PreScission cleavage site, using TEDA cloning and transformed into XL-1-Blue chemically competent cells, single clones isolated and amplified and sequences confirmed by Sanger sequencing. Plasmids transformed into chemically competent BL21 DE3 *E. coli*, and plated onto LB agar plates supplemented with 100 µg/mL kanamycin. Single colonies were used to make starter cultures of LB with 100 µg/mL kanamycin and incubated overnight at 37 °C. 1:100 volume starter culture was added to autoinduction media Overnight Express Instant TB Medium (Novagen) in Ultra-Yield flasks (Thomson), with 100 µg/mL kanamycin, incubated at 37 °C for 5 hours, then 18 °C for 18 hrs. Cells harvested by centrifugation 6,000 rpm, 20 mins, 4 °C, and pellets were frozen at -80°C.

Defrosted cell pellets were resuspended in approx. 10 mL/g Lysis Buffer (50 mM potassium phosphate pH 7.0, 300 mM NaCl, 5 mM imidazole, 2 mM b-mercaptoethanol, 10% glycerol), supplemented with 60 µg/mL lysozyme, 1.4 µg/mL DNaseI, 0.05 mM PMSF. Cells were lysed by passing through the French press twice, 18 kpsi. Lysate was clarified by centrifugation 18,000 x g, 45 mins, 4 °C, and loaded onto HIS-Select Nickel affinity resin (Sigma) by gravity, resin washed with Wash Buffer (50 mM potassium phosphate pH 7.0, 100 mM NaCl, 5 mM imidazole, 2 mM b-mercaptoethanol, 10% glycerol) and eluted with Wash Buffer containing 350 mM imidazole. Protein containing fractions (assessed by A_{280}) were combined, and further purified by size exclusion chromatography (SEC) using HiLoad 16/600 200 pg Superdex column (Cytiva) using ÅKTA FPLC system (Cytiva) equilibrated in 50 mM sodium phosphate pH 7.0, 1 mM DTT. Fractions were concentrated, concentration measured using A_{280} and predicted extinction coefficients, then flash frozen $N_{2(l)}$ for storage at -80 °C.

DNA corresponding to BH3 motif of human Bid Q79-G144 (Uniprot: P55957) was assembled by complementary oligos (IDT) and primer extension using Klenow fragment (NEB), and cloned using TEDA into pET28 with an N-terminal His_{10}, SUMO and C-terminal Avi. Expression and purification was carried out as for the Hallucinations, except for co-transformation with a chloramphenicol-resistant BirA expressing plasmid, the addition of

chloramphenicol 25 μg/mL in all cultures, with the addition of 40 μM BTN to the media before temperature was reduced to 18°C. After SEC, His_{10}-SUMO was cleaved using ULP-1 protease, and His_{10}-SUMO removed using Ni resin, Bid-Avi peptide concentration was measured using A_{280}, and stored at -80°C. To express human Mcl-1 P166-G327 (Uniprot: Q07820) a pEQ80L vector with N-terminal His_6 and Avi-tag, for co-expression with BirA. Expression and purification was carried out as for the Bid-binding hallucinations, with the addition of 40 μM BTN to the media before temperature was reduced to 18 °C.

ITC

Isothermal titration calorimetry was carried out with an ITC200 (Micocal). Bid peptide was in the syringe, at ~300 μM, and binder (Hallucination of Mcl-1) was kept in the cell (~25 μM), with both peptide and binder in matched buffer (sodium phosphate pH 7.0, 1 mM DTT). Temperature was held at 25 °C or 10 °C, as indicated. Fitting of titrations was carried out using 1-site binding, using manufacturers software (OriginLab).

Circular dichroism

Spectra were recorded for Bid peptide alone, Bid in complex with binders (Hallucination or Mcl-1) and binders alone. All concentrations were 10 μM, in a 2 cm pathlength quartz cuvette. Spectra recorded on J-1500 Circular Dichroism Spectrophotometer, with temperature held at 25°C, or ramped at 1 °C/min.

Pull-down

10 μL bead slurry Dynabeads M-280 Streptavidin (Thermo Fisher Scientific) were washed with Pull-Down Buffer (sodium phosphate pH 7.0, 1 mM DTT, 0.05% Tween20), incubated with saturating amounts of (Avi-tagged) Bid peptide 15 mins, 4 °C with rotation, beads were then incubated with free biotin 25 μM, and washed three times with ice cold Pull-Down Buffer. 10 μL of 2 μM binder (hallucination or Mcl-1) was incubated with pelleted beads for 30 mins, 4 °C, with rotation. Supernatant was recovered and the beads washed three times before resuspension in 10 μL Pull-Down Buffer. Both supernatant and washed beads were loaded onto denaturing SDS-PAGE, with protein detection by InstantBlue Coomassie staining.

Crystallization and Structure Determination

All crystallization experiments were conducted using the sitting drop vapour diffusion method. Crystallisation trials were set up in 200 nL drops using the 96-well plate format at 20 ˚C. Crystallisation plates were set up using a Mosquito LCP from SPT Labtech, then imaged using UVEX microscopes and UVEX PS-256 from JAN Scientific. Diffraction quality crystals formed in 0.2 M Ammonium chloride 0.1 M Tris pH 8 20% (w/v) PEG 6000 for GCG_partdiff; in 0.9 M Halogens, 0.1 M Tris- Bicine pH 8.5 Buffer, and 37.5% of 25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350 mixture for GCG_inpaint. In 0.1 M Citric acid pH 2.5, 20% (w/v) PEG 6000. PHT 0.2 M Sodium chloride for Bim_fulldiff, and 0.1 M Sodium acetate pH 4.5, and 1.26 M Ammonium sulfate for PTH peptide only.

Diffraction data was collected at the Advanced Photon Source (APS) beamline 24-ID-C. X-ray intensities and data reduction were evaluated and integrated using XDS and merged/scaled using Pointless/Aimless in the CCP4 program suite. Starting phases were obtained by molecular replacement using Phaser using the designed model for the structures. Following molecular replacement, the models were improved using phenix.autobuild; efforts were made to reduce model bias by setting rebuild-in-place to false, and using simulated annealing and prime-and-switch phasing. Structures were refined in Phenix. Model building was performed using COOT. The final model was evaluated using MolProbity. Data collection and refinement statistics are recorded in Supplementary Table 9. Data deposition, atomic coordinates, and structure factors reported in this paper have been deposited in the Protein Data Bank (PDB), http://www.rcsb.org/ with accession code 8GJI (GCG_partdiff), 8GJG (GCG_inpaint), 8T5E (Bim_fulldiff), and 8T5F (PTH peptide only).

Design and characterization of lucCagePTH biosensor for parathyroid hormone detection

The detailed design protocol for the lucCage and lucKey sensor system was described previously. In brief, the amino acid sequence (FELLDKLIELLRELIELTREYI) at the N-terminal end of the 6.1 nM PTH binder was grafted onto the latch region (residues 323 to 353) of lucCage. The Rosetta models were visually inspected and eight of them were selected for experimental validation. We produced, purified, and screened for the luminescence signal emitted from each biosensor in the presence of 5 μM PTH. From this process, we identified

several hits showing increased luminescence upon adding PTH, of which we assigned the best one with a 21-fold activation as lucCagePTH. We then set up assays to evaluate the response of lucCagePTH with a range of PTH concentrations. 10 µl of 10 nM lucCagePTH, 10 µl of 10 nM lucKey, 10 µl of serial diluted PTH, and 40 µl of buffer (50% HBS-EP/50% Nano-Glo luciferase assay buffer) were pre-mixed and 30 µl of 100× diluted furimazine was injected immediately before luminescence kinetic acquisition. The luminescence measurements were taken every 1 min (0.1 s integration and 10 s shaking during intervals) for a total of 60 mins by Neo2 microplate reader. The linear region of luminescence responses to the corresponding PTH concentrations was fitted to a linear regression curve and the LOD was calculated as 3 × standard deviation of the response / the slope of the calibration curve.

Affinity enrichment of PTH analyzed by LC-MS/MS

PTH-minibinder Bead preparation

PTH binder-conjugated beads were prepared by binding 1-10 mg of cysteine-containing protein binder at the C-terminus to SulfoLink™ Coupling Resin (ThermoFisher). Following the binding step, the excess active groups on the beads were blocked by adding 1 mL of 50 mM L-Cysteine•HCl and incubating for 30 minutes.

Sample description

Recombinant human PTH protein was purchased from Sigma (#SAE 0192_100 µg, MA, USA) and reconstituted at 100 µg/mL in a 10 % acetonitrile, 0.1 % formic acid,1 mg/mL bovine serum albumin solution and stored in 40 µL aliquots at -20 °C. Dilutions at 1000 ng/mL and 62.5 ng/mL were prepared freshly as needed by dilution in the same acetonitrile, formic acid, albumin solution.

The plasma samples used were de-identified clinical samples obtained from the clinical laboratories at the University of Washington Medical Center. The use of de-identified leftover clinical samples was reviewed by the University of Washington Human Subjects Division (STUDY00013706).

The evaluation of PTH immunoaffinity enrichment in buffer and plasma was performed in three process replicates using 8 different types of samples:

- Series A: Reconstitution buffer (10 % acetonitrile, 0.1 % formic acid,1 mg/mL bovine serum albumin in water) served as the blank.
- Series B: Reconstitution buffer spiked with PTH at 7.2 ng/mL was directly digested without the addition of beads and served as the Control sample (representing 100% recovery of PTH).
- Series C: Reconstitution buffer spiked with PTH at 7.2 ng/mL was incubated with beads blocked by bovine serum albumin before washing and digestion, which served as the negative control, to quantify non-specific binding in buffer.
- Series D: Reconstitution buffer spiked with PTH at 7.2 ng/mL was incubated with designed binder-conjugated beads before washing and digestion, which was used to quantify the affinity precipitation of PTH from buffer.
- Series E: Plasma was incubated with beads blocked by bovine serum albumin before washing and digestion, which was used to quantify non-specific binding in unspiked plasma.
- Series F: Plasma was incubated with designed binder-conjugated beads before washing and digestion, which was used to quantify affinity precipitation of PTH in plasma.
- Series G: Plasma spiked with PTH at 7.2 ng/mL was incubated with beads blocked by bovine serum albumin before washing and digestion, which was used to quantify non-specific binding in spiked plasma.
- Series H: Plasma spiked with PTH at 7.2 ng/mL was incubated with designed binder-conjugated beads before washing and digestion, which was used to quantify the affinity precipitation of PTH in spiked plasma.

Sample preparation and LC-MS/MS conditions

Affinity enrichment was performed in buffer or plasma at the protein level. Designed binders were conjugated to tosyl-activated Dynabeads M-280 according to the manufacturer's instructions and subsequently blocked using bovine serum albumin and Tris. The amino terminal peptide was analyzed after tryptic digestion of either pure protein in buffer, or after trypsin

digestion of PTH that had been affinity precipitated by the designed binder-conjugated beads (or by the control/blocked magnetic beads). Briefly, PTH proteins in buffer/plasma were purified using PTH mini-binder conjugated-paramagnetic beads at room temperature, for 1 h. The beads were then washed 4 times with phosphate-buffered saline supplemented with CHAPS (0.1% 3-((3cholomidopropyl) dimethylammonio)-1-propanesulfate to reduce nonspecific interactions). The proteins that were affinity precipitated by the designed binder-conjugated-paramagnetic beads were suspended in 10 µL of a solution containing 10 % acetonitrile, 0.1 % formic acid, 1 mg/mL bovine serum albumin. The washed beads were then suspended with 30 µL of 30% isopropanol,100 mM ammonium bicarbonate, and digested at 37°C for 30 min after adding 100 µL of 0.01 mg/mL trypsin in 10 mM hydrochloric acid. The liberated peptides were then removed from the beads using a magnet and analysed using LC-MS/MS.

Peptides were analyzed by liquid chromatography-tandem mass spectrometry in the multiple reaction monitoring acquisition mode using an UHPLC I-Class Chromatography system coupled to a Xevo TQ-S triple quadrupole tandem mass spectrometer (Waters, MA, USA). Peptides were eluted from an Acquity UPLC HSS T3 1.8µm (C18, 2.1x50 mm, pore size 100 Å) analytical column (Waters) at 45 °C using 0.1 % formic acid, 2 % dimethylsulfoxide in LC-MS grade water as mobile phase A and 0.1 % formic acid, 2 % dimethylsulfoxide in LC-MS grade methanol as mobile phase B.

The liquid chromatography and mass spectrometry conditions are detailed in Supplementary Table 2, 3 and 4.

Supplementary Table 2. PTH Liquid chromatography conditions	
Mobile phase	Phase A: 0.1 % formic acid, 2 % dimethylsulfoxide in water 0.1 % formic acid, 2 % dimethylsulfoxide in methanol
Column	Acquity UPLC HSS T3 1.8µm (C18, 2.1x50 mm, pore size 100 Å)
Temperature	45±5 °C
Flow rate	0.3 mL/min
Injection volume	20 µL

Gradient	0-0.5 min: 2% B at 0.3 mL/min 7.5: 98% B at 0.3 mL/min 7.6: 98% B at 0.6 mL/min 8.6: 2% B at 0.6 mL/min 9.9: 2% at 0.3 mL/min

Supplementary Table 3. PTH Mass spectrometry conditions

Source polarity	ESI+
Capillary voltage	3.25 kV
Source Offset voltage	50 V
Desolvation Temp	600 °C
Desolvation Gas Flow	1000 L/h
Cone Gas Flow	150 L/h

Supplementary Table 4. PTH Multiple reaction monitoring conditions

Peptide sequences	Q1 (m/z)	Q3 (m/z)	Cone (V)	Collision (V)	Ion type
HLNSMER.2	443.7136	218.1047	35	15	y3
HLNSMER.2	443.7136	261.6207	35	15	y4
HLNSMER.2	443.7136	318.6421	35	15	y5
HLNSMER.3	296.1448	218.1047	35	9	y3
HLNSMER.3	296.1448	261.6207	35	9	y4
HLNSMER.3	296.1448	318.6421	35	9	y5

HLNSMER.3	296.1448	435.202	35	9	y3
HLNSMER.3	296.1448	522.2341	35	9	y4
HLNSMER.3	296.1448	636.277	35	9	y5
HLNSM(+15.994915)ER.2	451.7111	226.1021	35	16	y3
HLNSM(+15.994915)ER.2	451.7111	269.6181	35	16	y4
HLNSM(+15.994915)ER.2	451.7111	326.6396	35	16	y5
HLNSM(+15.994915)ER.2	451.7111	451.1969	35	16	y3
HLNSM(+15.994915)ER.2	451.7111	538.229	35	16	y4
HLNSM(+15.994915)ER.2	451.7111	652.2719	35	16	y5
HLNSM(+15.994915)ER.3	301.4765	226.1021	35	10	y3

HLNSM(+15.994915)ER.3	301.4765	269.6181	35	10	y4
HLNSM(+15.994915)ER.3	301.4765	326.6396	35	10	y5
HLNSM(+15.994915)ER.3	301.4765	451.1969	35	10	y3
HLNSM(+15.994915)ER.3	301.4765	538.229	35	10	y4
ADVNVLTK.2	430.2478	574.3559	35	15	y5
ADVNVLTK.2	430.2478	673.4243	35	15	y6
ADVNVLTK.3	287.1676	181.1259	35	9	y3
ADVNVLTK.3	287.1676	230.6601	35	9	y4

ADVNVLTK.3	287.1676	361.2445	35	9	y3
ADVNVLTK.3	287.1676	460.313	35	9	y4
SLGEADK.2	360.1821	167.0921	35	12	y3
SLGEADK.2	360.1821	231.6134	35	12	y4
SLGEADK.2	360.1821	260.1241	35	12	y5
SLGEADK.2	360.1821	333.1769	35	12	y3
SLGEADK.2	360.1821	462.2195	35	12	y4
SLGEADK.2	360.1821	519.2409	35	12	y5

SLGEADK.3	240.4572	167.0921	35	7	y3
SLGEADK.3	240.4572	260.1241	35	7	y5
SLGEADK.3	240.4572	333.1769	35	7	y3
SLGEADK.3	240.4572	462.2195	35	7	y4
VEWLR.2	351.7003	229.1183	35	12	b2
VEWLR.2	351.7003	474.2823	35	12	y3
EDNVLVESHEK.2	649.8148	629.2889	35	23	y5
EDNVLVESHEK.2	649.8148	728.3573	35	23	y6

EDNVLVESHEK.2	649.8148	841.4414	35	23	y7
EDNVLVESHEK.3	433.5456	315.1481	35	14	y5
EDNVLVESHEK.3	433.5456	364.6823	35	14	y6
EDNVLVESHEK.3	433.5456	421.2243	35	14	y7
EDNVLVESHEK.3	433.5456	629.2889	35	14	y5
DAGSQRPR.2	443.7281	322.1854	35	15	y5
DAGSQRPR.2	443.7281	643.3634	35	15	y5
DAGSQRPR.2	443.7281	700.3849	35	15	y6

DAGSQRPR.3	296.1545	214.6401	35	9	y3
DAGSQRPR.3	296.1545	278.6693	35	9	y4
DAGSQRPR.3	296.1545	322.1854	35	9	y5
DAGSQRPR.3	296.1545	350.6961	35	9	y6
DAGSQRPR.3	296.1545	428.2728	35	9	y3
DAGSQRPR.3	296.1545	556.3314	35	9	y4
DAGSQRPR.3	296.1545	643.3634	35	9	y5
DAGSQRPR.3	296.1545	700.3849	35	9	y6

SVSEIQLMHNLGK.2	728.3849	527.2973	35	26	y9
SVSEIQLMHNLGK.2	728.3849	568.3202	35	26	y5
SVSEIQLMHNLGK.2	728.3849	635.3346	35	26	y11
SVSEIQLMHNLGK.2	728.3849	699.3607	35	26	y6
SVSEIQLMHNLGK.2	728.3849	812.4447	35	26	y7
SVSEIQLMHNLGK.2	728.3849	940.5033	35	26	y8
SVSEIQLMHNLGK.2	728.3849	1053.587	35	26	y9
SVSEIQLMHNLGK.2	728.3849	1269.662	35	26	y11

SVSEIQLMHNLGK.3	485.9257	159.1128	35	16	y3
SVSEIQLMHNLGK.3	485.9257	431.2613	35	16	y4
SVSEIQLMHNLGK.3	485.9257	470.7553	35	16	y8
SVSEIQLMHNLGK.3	485.9257	527.2973	35	16	y9
SVSEIQLMHNLGK.3	485.9257	568.3202	35	16	y5
SVSEIQLMHNLGK.3	485.9257	591.8186	35	16	y10
SVSEIQLMHNLGK.3	485.9257	635.3346	35	16	y11
SVSEIQLMHNLGK.3	485.9257	699.3607	35	16	y6

SVSEIQLMHNLGK.3	485.9257	812.4447	35	16	y7
SVSEIQLMHNLGK.3	485.9257	940.5033	35	16	y8
SVSEIQLM(+15.994915)HNLGK.2	736.3823	159.1128	35	26	y3
SVSEIQLM(+15.994915)HNLGK.2	736.3823	317.2183	35	26	y3
SVSEIQLM(+15.994915)HNLGK.2	736.3823	431.2613	35	26	y4
SVSEIQLM(+15.994915)HNLGK.2	736.3823	535.2948	35	26	y9
SVSEIQLM(+15.994915)HNLGK.2	736.3823	568.3202	35	26	y5
SVSEIQLM(+15.994915)HNLGK.2	736.3823	643.3321	35	26	y11

SVSEIQLM(+15.994915)HNLGK.2	736.3823	715.3556	35	26	y6
SVSEIQLM(+15.994915)HNLGK.2	736.3823	828.4396	35	26	y7
SVSEIQLM(+15.994915)HNLGK.2	736.3823	956.4982	35	26	y8
SVSEIQLM(+15.994915)HNLGK.2	736.3823	1069.582	35	26	y9
SVSEIQLM(+15.994915)HNLGK.3	491.2573	159.1128	35	16	y3
SVSEIQLM(+15.994915)HNLGK.3	491.2573	643.3321	35	16	y11

Data treatment

Data processing was performed with Skyline Daily version 21.1.1.223. The peak area for each peptide was determined as the sum of the peak areas of all selected transitions. The recovery over blocked-beads (RE) in spiked buffer and in spiked plasma was estimated using Equations 1, and 2, respectively.

Affinity Enrichment of Glucagon and Analysis by LC-MS/MS Bead Preparation

Minibinder Beads

GCG binder-conjugated beads were prepared by binding 1-10 mg of cysteine-containing protein binder at the C-terminus to 30 mg of BcMag™ long-arm iodoacetyl-activated Magnetic Beads (Bioclone). Following the binding step, the excess active groups on the beads were blocked by adding 1 mL of 50 mM L-Cysteine•HCl and incubating for 60 minutes. As a control for binder specificity, an unrelated *de novo* protein binder was used in place of the GCG binder for the binder-negative control beads.

Monoclonal Antibody Beads

Anti-glucagon monoclonal antibody bound beads were prepared by binding 5 µg of in-house monoclonal antibody per prepared sample to 7.5 µL of tosyl M-280 beads (Thermofisher) Post-binding, beads are blocked with a Tris 0.1% BSA solution. Two in-house monoclonal antibodies are bound separately and are pooled together prior to use. For monoclonal antibody negative control beads bovine gamma globulin was used in place of monoclonal antibody.

Immuno-Affinity Enrichment

Enrichment of glucagon peptide was assayed in triplicate for four different bead types: anti-glucagon monoclonal antibody tosyl beads, negative control tosyl beads, minibinder beads and negative control minibinder beads. PBS-CHAPS 0.1% was spiked with 62.5 pM exogenous glucagon, HSQGTFTSDYSKYLDSRRAQDFVQWLMNT (Anaspec) and 500 µL was added to 10 µL prepared monoclonal antibody beads and 50 µL of prepared minibinder beads. After 45 minutes of agitation at 1400 rpm at room temperature, the paramagnetic beads were washed 4 times 200 µL of PBS-CHAPS 0.1%. Samples were eluted for 8 minutes at 1400 rpm with 50 µL elution solvent (20% acetic acid 10% dimethylsulfoxide 10% acetonitrile 0.01% BSA in water) that was spiked with stable isotope labeled internal standard (New England Peptide). Elution solvent was also spiked with exogenous glucagon to mimic the peptide levels obtained if 100% of the peptide was captured by the antibody or minibinder ligands.

To prepare the beads for reuse after the elution step beads were resuspended in PBS-CHAPS 0.1% and pooled back together. On a subsequent day the beads were washed three times with elution solvent. Then they were washed three times in PBS-CHAPS 0.1%. Minibinder beads

fresh, minibinder beads reused, mAb beads fresh and mAb beads reused were each tested in triplicate according to the same protocol listed above for the initial comparison.

LC-MS/MS Analysis

Samples were transferred to glass total recovery vials (Waters) and run on a Waters Acquity UPLC I-Class system with a flow through needle and coupled with a Waters TQ-S tandem mass spectrometer running in multiple reaction monitoring (MRM) mode. Liquid chromatography, mass spectrometer and MRM methods are summarized in Supplementary Table 5, 6 and 7, respectively. Enriched samples were injected in singlicate and the 100% recovery sample was injected in quintuplicate.

Supplementary Table 5. Glucagon Liquid Chromatography Conditions		
Mobile Phase	A: 0.1% formic acid and 2% dimethylsulfoxide in water	
	B: 0.1% formic acid and 2% dimethylsulfoxide in methanol	
Column	Acquity UPLC HSS T3 1.8µm (C18, 2.1x50 mm, pore size 100 Å)	
Temperature	45±5 °C	
Injection Volume	35µL	
Gradient	0-0.5min	30% B at 0.3 mL/min
	7.0 min	5% B at 0.3mL/min
	7.1 min	95% B at 0.6 mL/min
	8.1min	30% B at 0.6 mL/min
	9.4 min	30% B at 0.3 mL/min

Supplementary Table 6. Glucagon Mass Spectrometry Conditions	
Source Polarity	ESI+
Capillary Voltage	1.4 kVA

Cone Offset	60 V
Desolvation Temperature	600 ℃
Desolvation Gas Flow	1000 L / hour
Cone Gas Flow	150 L / hour
Cone	35 V
Collision	24 V

Supplementary Table 7. Glucagon Multiple Reaction Monitoring Conditions

Peptide Sequence	Q1 (m/z)	Q2 (m/z)	Ion Type
HSQGTFTSDYSKYLDSRRAQDFVQWLMNT	871.70	1002.74	b25
		1040.46	b26
		1084.19	b27
HSQGTF^TSDYSKYLDSRRAQDF^VQWLMNT F^ = 13C(+9)	876.17	1008.69	b25 (IS)
		1046.41	b26 (IS)
		1090.14	b27 (IS)

Chromatograms were analyzed in Skyline Daily version 23.0.9.239. Integrated peak areas were exported to Excel for further analysis. Results were calculated per these three equations. The results summary is shown in figure 4d (right) and representative chromatograms found in supplemental figure S17b.

$$Peak\ Area\ Ratio = \frac{sum\ Peak\ Area\ (b25 + b26 + b27)}{sum\ Internal\ Standard\ Peak\ area\ (b25(IS) + b26\ (IS) + b27\ (IS))}$$

$$Percent\ Recovery\ = \frac{Peak\ Area\ Ratio\ of\ Sample}{Average\ Peak\ Area\ Ratio\ of\ 100\%\ Recovery\ sample\ (n = 5)}$$

$$Normalized\ Percent\ Recovery\ = \frac{Percent\ Recovery\ of\ Sample}{Average\ Percent\ Recovery\ of\ mAb\ sample\ (n = 3)}$$

Computational methods

Parametric design of groove-shaped scaffold library and use for binder design

The parametric groove-shaped scaffold library was sampled using a random sampling approach, for which key parameters were selected randomly from specific distributions. An even distribution of bundle 'lengths' was sampled, for which each parametric helix was 15–19 residues long. A supercoiling value was randomly selected from a biased distribution favoring more supercoiled scaffolds, given that these scaffolds were more likely to fail in the subsequent looping step (Fig. 1b and Supplementary Fig. 1). This biased sampling strategy was chosen to achieve a more uniform distribution of supercoiling within the final scaffold library, with sufficient numbers of highly supercoiled bundles. An average helix neighbor distance value was randomly selected from a Gaussian distribution informed by native protein helical bundle geometries (Fig. 1b and Supplementary Fig. 1). The distance of each helix from its neighbors was independently randomly selected from a much tighter Gaussian distribution centered at the preselected average helix neighbor distance value, to provide some noise within a given scaffold to helix distances and allow for heterogeneous amino acid selections (Supplementary Fig. 2). Values for helix phase and *z* displacement were randomly sampled for each helix. The 'groove' consisting of three helices was first sampled as a helical bundle using the Crick parameterization of α-helical coiled coils, around an imaginary central helix where the target was to later be docked. Next, the two buttressing helices were sampled with the same parameterization, but moved radially outwards with randomly sampled helix neighbor distances as well as an additional randomly sampled tilt. This process was used to sample a set of 200,000 arrangements

of 5 helices. Next, the Rosetta ConnectChainsMover[51] was used to loop this set into approximately 135,000 successful scaffold backbones. These backbones were designed and filtered using Rosetta (including flexible backbone design) to yield a final library of 18,000 scaffolds. Backbones were filtered on metrics including buried nonpolar surface area per residue, Rosetta score per residue, percentage alanine, exposed hydrophobics per residue, and Rosetta 'holes'.

RF_{joint} Inpainting

To sample around an initial putative binder, and to extend the binding interface to make additional contacts with the bound peptide, the RF_{joint} Inpainting network was used[23], in conjunction with ProteinMPNN. Rosetta-designed binders to PTH, GCG and NPY were used as input to RF_{joint}. RF_{joint} is deterministic, and hence, to generate diversity, additional length was added (randomly and independently sampled) at the loop junctions between the binder helices. Additionally, one whole helix was completely rebuilt by RF_{joint}, to further permit diversification. RF_{joint} designs were subsequently sequence redesigned with ProteinMPNN, validated and filtered in silico by AF2 with initial guess, and subsequently tested experimentally.

Sequence threading to generate peptide binders

We started from a library of several thousand all-helical scaffolds bound to designed single helices. We then threaded sequences of peptides of interest onto the bound single helix and filtered to obtain threaded conformations that maximized the number of target sequence positions that formed hydrophobic interactions at the interface to the binder scaffold. The resulting binders were then redesigned in the presence of the threaded target sequence with ProteinMPNN (forbidding cysteine) and the complex was predicted with AF2 with initial guess. Another round of ProteinMPNN and AF2 + initial guess was carried out on the AF2 models that passed gate filters. Both rounds had gate filters of interface pAE < 10, mean pLDDT > 92, predicted template modelling (pTM) score > 0.8 and RMSD to input backbone <1.75 Å. AF2 models from both rounds that passed gate filters were further filtered on AF2 metrics and filtered on Rosetta metrics to select sequences to order. Sequences were filtered against membrane insertion potential, contact_molecular_surface, ddG[7], interface pAE and monomer pAE.

AF2 Hallucination for flexible peptide binder design

Code for running Hallucination with AF2 was modified with custom losses developed to promote binding of the Hallucinated protein to the input peptide sequence. AF2 model_4_ptm was used for all experiments.

Initial sequence sampling

The initial binder sequence was sampled randomly, with amino acid probabilities corresponding to background amino acid frequencies in BLOSUM62. The target sequence (but no template structure) is also provided, separated by a chain break (+32 residue positional index offset). Residues were then mutated, with probabilities related to their background frequency in BLOSUM62. The number of amino acid changes proposed at each step is decayed throughout the trajectory (1,250 ×3 steps, 2,500 ×2 steps, 1,250 ×1 step). Multiple simultaneous amino acid changes initially helps speed up Hallucination, and a lower rate of changes later on allows more gradual refinement. To further speed up convergence, alterations were selectively made to residues with the lowest 50% of AF2 pLDDTs.

Losses used for Hallucination

Losses used for Hallucination were as follows.

- pLDDT of the bound state: average pLDDT of the binder–peptide complex.
- pTM of the bound state: the pTM score of the binder–peptide complex.
- Radius of gyration: the radius of gyration was calculated as the mean squared distance of residues from the centre of mass of the protein. To approximately standardize the scaling with length of the protein, this was empirically normalized by dividing the radius of gyration by the radius of a sphere of volume related to the length of the Hallucinated protein.
- Contact probability: calculated as total probability that a residue in the target is in contact (closer than 8 Å) with the target peptide (the summed probability over the sub-8-Å bins of the distogram output from AF2). This was averaged across all target residues.

- Interface pAE: the mean pAE between the binder and peptide chains.

For all examples shown in this work, the losses were weighted with relative weights of 1:1:0.1:3:5.

Simulated annealing

To optimize the designed binder, Monte Carlo simulated annealing was carried out, with a starting temperature of 0.01, and the half-life of the exponential decay set to 500 steps. Alterations were accepted or rejected using the Metropolis criterion. A total of 5,000 steps were carried out during design.

ProteinMPNN

Previous work has demonstrated that AF2 Hallucination yields adversarial sequences that do not work experimentally. However, designs can be rescued with ProteinMPNN redesign of the sequences. Sixty-four sequences were designed per backbone, and were subsequently filtered on the basis of AF2 pLDDT, pTM, RMSD to the design model, RMSD of the monomer to the binder model (without the peptide), and Rosetta ddG. The precise values used for filtering were chosen to reduce the set down to 46 designs.

Partial diffusion to optimize binders

RFdiffusion was modified to allow the input structure to be noised only up to a user-specified time step instead of completing the full noising schedule. The starting point of the denoising trajectory is therefore not a random distribution. Rather, it contains information about the input distribution resulting in denoised structures that are structurally similar to the input (Fig. 2a). The AF2 models of the highest-affinity designs from Inpainting for GCG and NPY were used as inputs to partial diffusion. The models were subjected to 40 noising time steps out of a total of 200 time steps in the noising schedule, and subsequently denoised. An auxiliary potential minimizing the radius of gyration of the binder–peptide complex was used (described below). Approximately 2,000 partially diffused designs were generated for each target. The backbones in the resulting library were sequence designed using ProteinMPNN (and ProteinMPNN after Rosetta FastRelax), followed by AF2 + initial guess. The resulting libraries were filtered on AF2 pAE, pLDDT, RMSD to the design model, RMSD of the monomer to the binder model (without

the peptide) and Rosetta ddG. The precise values used for filtering were chosen to reduce the set down to 96 designs for each target.

De novo peptide binder design using RFdiffusion

The AF2 model of the PTH peptide in the highest-affinity binder from Inpainting was used as input to RFdiffusion. For Bim, there was no previously designed binder and therefore the crystal structure of Bim (PDB: 6X8O) was used as input. An auxiliary potential minimizing the radius of gyration of the binder–peptide complex was used during denoising (described below). Approximately 2,000 diffused designs were generated for each target. The backbones in the resulting library were sequence designed using ProteinMPNN (and ProteinMPNN following FastRelax), followed by AF2 + initial guess. The resulting libraries were filtered on AF2 PAE, pLDDT, RMSD to the design model, RMSD of the monomer to the binder model (without the peptide) and Rosetta ddG. The precise values used for filtering were chosen to reduce the set down to 96 designs for each target.

Radius of gyration potential

RFdiffusion enables the use of external guiding potentials during inference, which help in the design of proteins with a certain desired property. In this work, we take advantage of guiding potentials to minimize the radius of gyration of the binder–peptide complex. The radius of gyration is calculated as the root mean square of the distance of all the Cα atoms from the centroid. It is more important to apply the potential at the initial denoising steps, and less so towards the end when the quaternary structure is largely fixed. Therefore, the scaling factor with which the gradients are multiplied has a cubic decay over the course of the denoising trajectory.

Training RFdiffusion for designing binders to targets from sequence alone

A modified version of RFdiffusion was trained to permit the design of protein binders to targets, for which only the sequence of the target was specified. The training strategy largely followed the training strategy used for the original RFdiffusion model, with some modifications. A summary is provided below.

Overview of 'base' RFdiffusion training

RFdiffusion is a denoising diffusion probabilistic model fine-tuned from a pretrained structure prediction model; RoseTTAFold. RFdiffusion is trained with a forward noising process that iteratively, over 200 time steps, noises residue translations and orientations to distributions that are indistinguishable from random distributions (three-dimensional Gaussian distribution and a uniform distribution on SO(3), respectively). RFdiffusion is then trained to reverse this corruption process, predicting the ground truth (X_0) at each time step of prediction. Mean-squared-error losses are used to minimize the error between the forward and reverse processes. Full training details are extensively described in ref.

Modifications to RFdiffusion for binder design to sequence inputs alone

RFdiffusion was trained on both monomers (<384 amino acids) and heterocomplexes (one chain, denoted the 'binder chain', <250 amino acids) from the PDB. Coordinates were scaled by a factor of four, in line with the original RFdiffusion model. In 20% of cases, no sequence or structure was provided to the model (for unconditional generation). In the other 80% of cases, 20–100% of the protein was noised. In contrast to RFdiffusion, however, the structure of up to 50% of the protein (monomer or 'target chain') was noised (diffused), while providing the sequence of those residues. Thus, RFdiffusion learns to condition its predictions on the sequence (without structure) of part of a protein (the monomer) or of a target to bind to. This version of RFdiffusion was trained for seven epochs.

Computational filtering

Precise metrics cutoffs changed for each design campaign to get to an orderable set, but largely focused on pAE (<10), pLDDT (>80) and Rosetta ddG (<−40).

Supplementary Figures

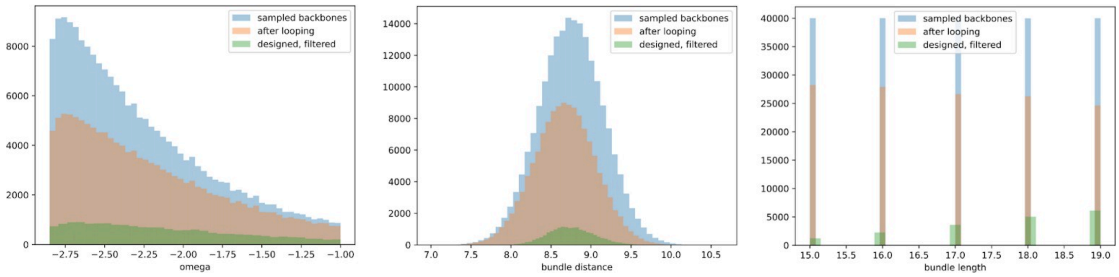


Figure S1: Sampled parameter distributions for parametric groove scaffolds. Left: Distributions of supercoiling (omega). Middle: Average distance between neighboring helices (bundle distance). Middle: Length of helices in scaffold bundle (bundle length) for sampled parametric groove scaffolds. Distributions are depicted for each step of the design process: initial parametrically sampled helix arrangements, scaffold backbones after loop closure, and final designed scaffolds after filtering.

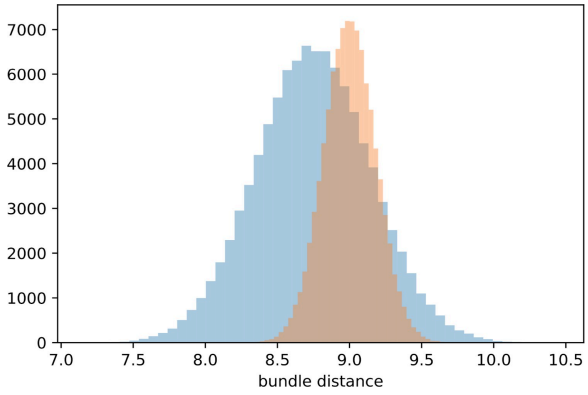


Figure S2: Sampling strategy for average helical neighbor distance. First, a global average helix neighbor distance (bundle distance) was randomly selected for each sampled scaffold from a Gaussian distribution informed by native protein helical bundle geometries (blue). Then, the distance of each helix from its neighboring helices for a single scaffold was independently randomly sampled from a tighter normal distribution (orange) centered around the selected global average helix neighbor distance (for this demonstrative example 9.0).

Figure S3: Parametric groove scaffold library: 45 scaffolds from the library of 18 thousand parametric groove scaffolds, demonstrating a range of supercoiling and helix distances to accommodate a range of helical peptide targets.

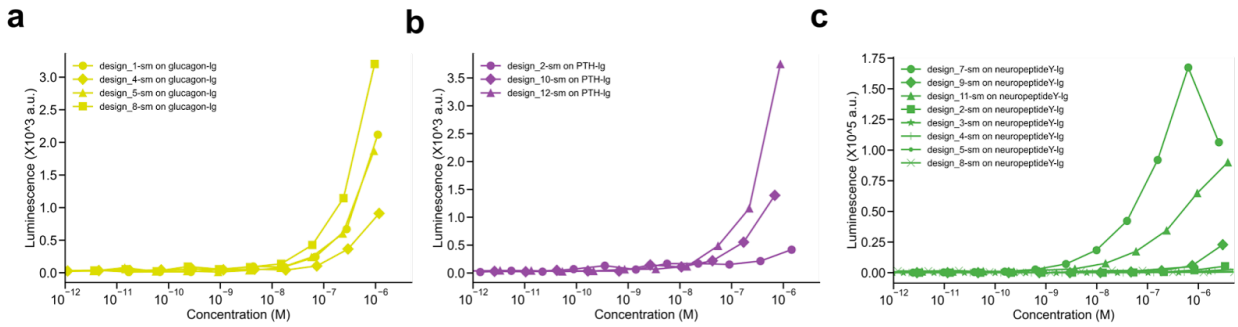


Figure S4: NanoBiT binding curves for initial parametric binder hits show weak affinities. **(a)** Glucagon (yellow), **(b)** PTH (purple), and **(c)** NPY (green) titration binding curves as measured through a nanoBiT split luciferase assay show weak binding, in most cases indistinguishable from the background signal of the two split luciferase halves binding (190 μM).

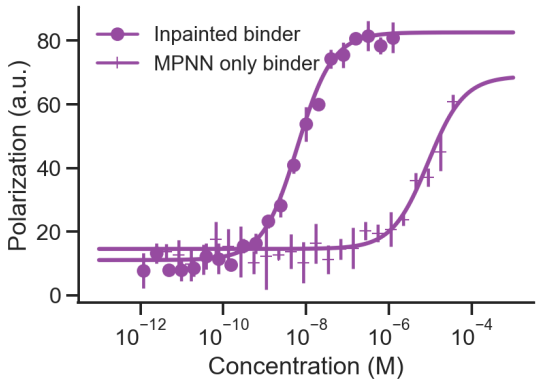


Figure S5. Comparison of binding affinities between the PTH Inpainted binder and the parametrically redesigned binder using ProteinMPNN only. Structural extension with RF_{joint} Inpainting and ProteinMNN sequence redesign resulted in a significant increase in the binding affinity of the parametrically designed PTH binder. Conversely, ProteinMNN redesign performed solely on the original parametrically designed binder did not lead to an increase in affinity.

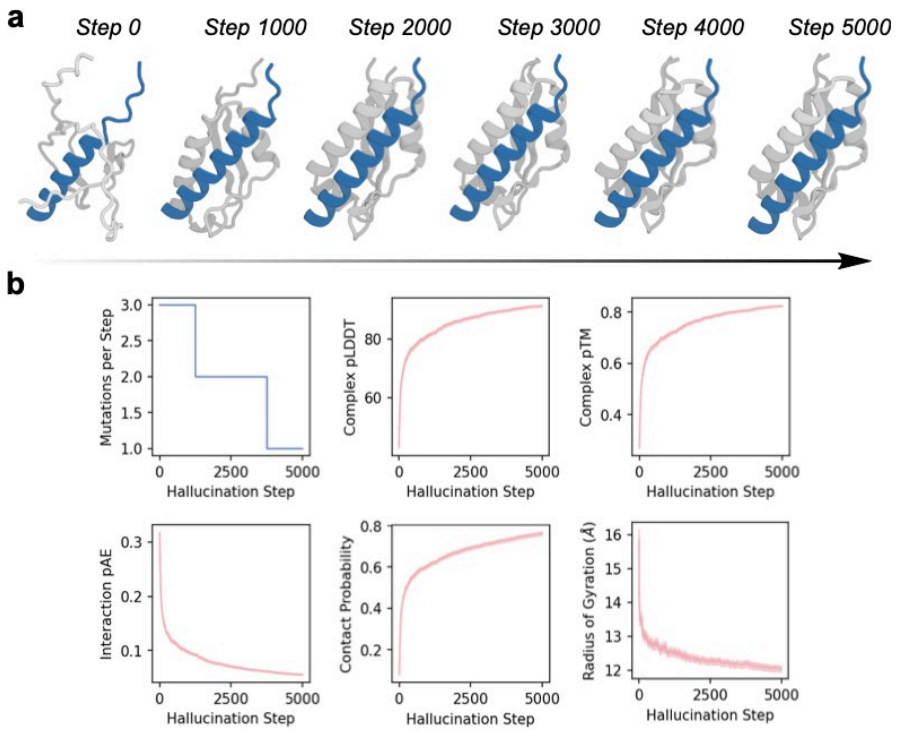


Figure S6: Hallucinating Bid Binders with AlphaFold2: (a) Example hallucination trajectory generating 70 amino acid binders to the peptide Bid (blue). Initially, AlphaFold2 predicts an unstructured "binder", but over 5000 steps, a binder is built up around the peptide. Crucially, no template structure is provided for the Bid peptide, allowing AF2 to predict its structure throughout. Note the predicted elongation of the helical structure in the peptide (blue, top) over the hallucination trajectory. **(b)** Hallucination trajectories approximately converge after 5000 steps. Left to right, top to bottom: The mutation rate at each step is decayed throughout the trajectory (1250 x 3 steps, 2500 x 2 steps, 1250 x 1 step). More mutations initially helps speed up hallucination, while a lower rate later on allows more gradual refinement. The AF2 confidence (pLDDT, pTM) in the bound structure increases throughout trajectories, while the interaction pAE decreases. The contact probability also trends to convergence over the trajectories, while the proteins typically become more compact (radius of gyration). N=96 trajectories.

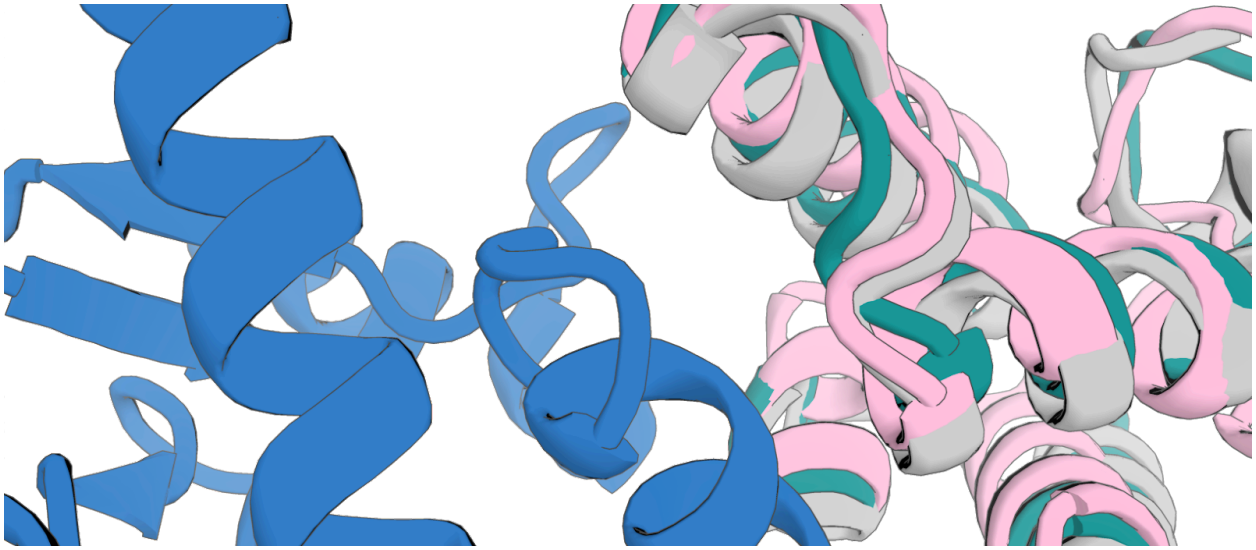


Figure S7. Specificity redesign of a native immunity protein. Colicin-E2 immunity protein (teal, PDB ID: 3U43) was superimposed on the PDB structure (PDB ID: 1EMV) of the native Immunity protein E9 (gray) - colicin E9 structure (blue). The superimposed structure serves as the starting point for conducting 100 designs using partial diffusion (60 steps, noise=1) and ProteinMPNN (T=0.1) vs ProteinMPNN alone (T=0.1). The latter resulted in only one structure being predicted as a complex using AlphaFold (interaction pAE < 20) while the former had 7 structures predicted to form a complex. The AlphaFold model of the top-ranked design is shown in pink. Notably, the model reveals a subtle movement in the partially diffused structure of the "α1–α2 motif", a major specificity determinant, which accurately recapitulates the native interaction observed in the colicin E9 complex. The native complex was not part of the training set used to train the diffusion model.

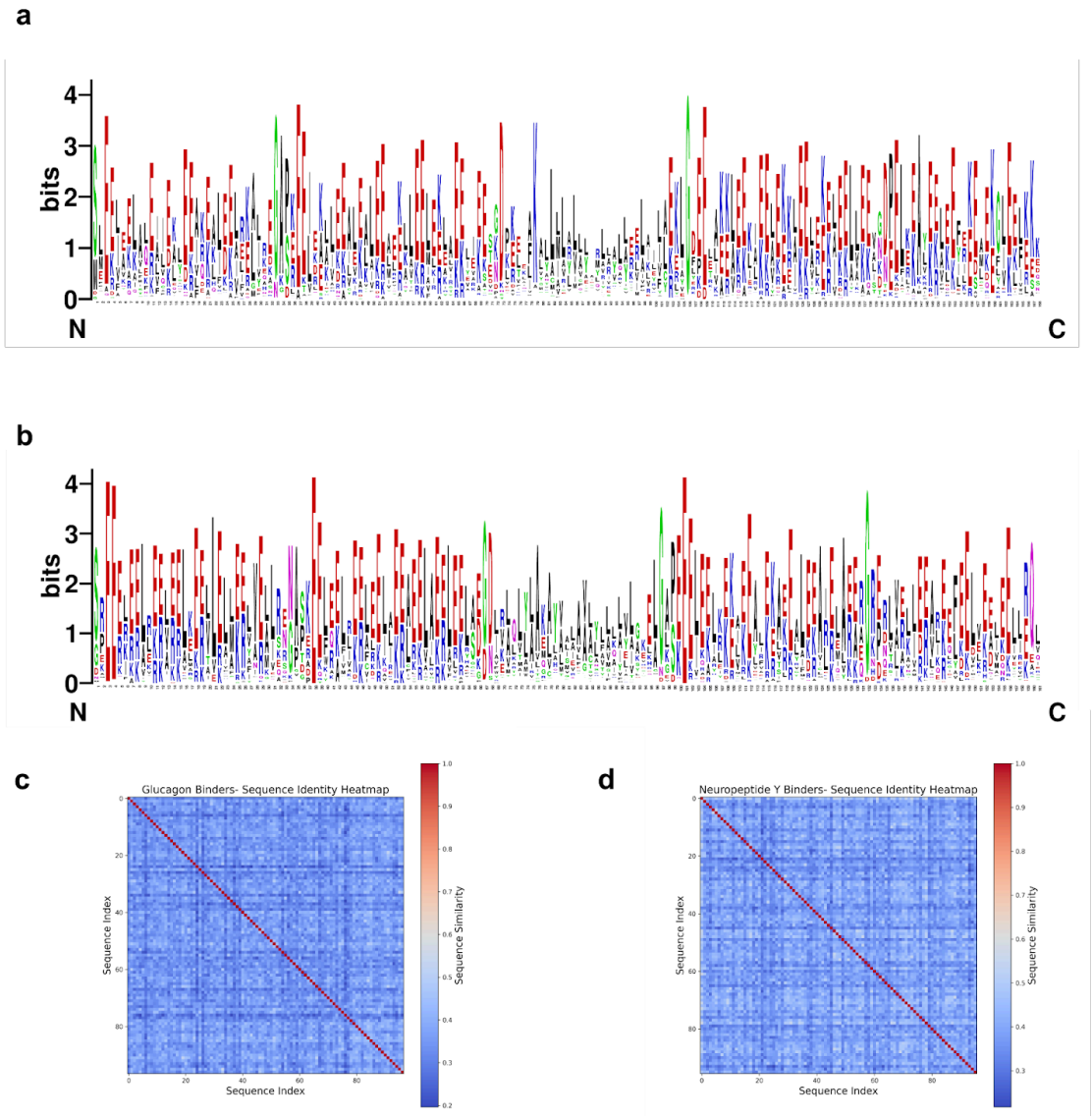


Figure S8. Sequences designed after partial diffusion are highly diverse. (a) Sequence logo for the designed binders generated through partial diffusion for GCG and **(b)** NPY. In both cases, the computationally designed binders show rich diversity in sequence space as visualised in their sequence logos. **(c)** All-by-all heatmap of sequence identity for the designs tested against GCG and **(d)** NPY further demonstrate sequence diversity of the design proteins.

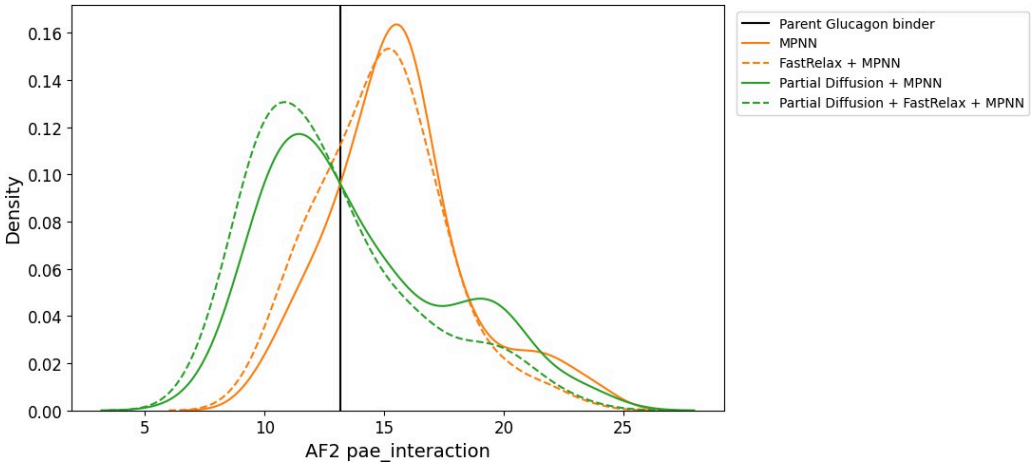


Figure S9. Partial Diffusion is necessary for significant improvement in binding metrics. An AlphaFold2 predicted model of a weak binder to GCG (K_d ~ 200 nM) was used as a starting structure to improve in-silico binding metrics through different methods. Using this parent design, we ran 1) ProteinMPNN (T=0.1) alone, 2) ProteinMPNN (T=0.1) followed by FastRelax and a second round of ProteinMPNN (T=0.1), 3) Partial diffusion (40 steps) followed by ProteinMPNN (T=0.1), and 4) Partial diffusion (40 steps), followed by ProteinMPNN (T=0.1), FastRelax and a second round of ProteinMPNN (T=0.1). 200 sequences were generated using each method. Binding metrics based on AlphaFold2 clearly show that while ProteinMPNN and FastRelax are occasionally able to improve the parent design, the distribution is significantly improved when partial diffusion is used to generate the backbone ensemble first. FastRelax offers a further marginal improvement.

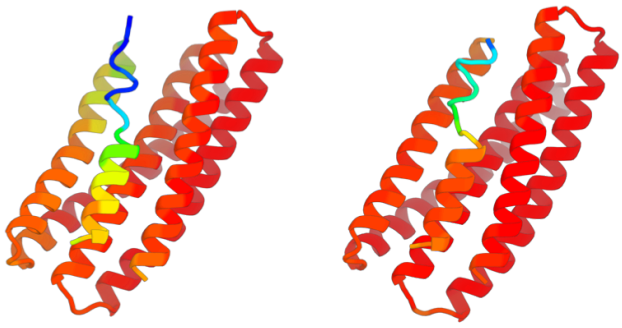


Figure S10. AF2 predictions of the Inpainted and partially diffused GCG binders. The Inpainted (left) and partially diffused (right) binders to GCG are coloured (spectrum blue to red, low to high) by pLDDT.

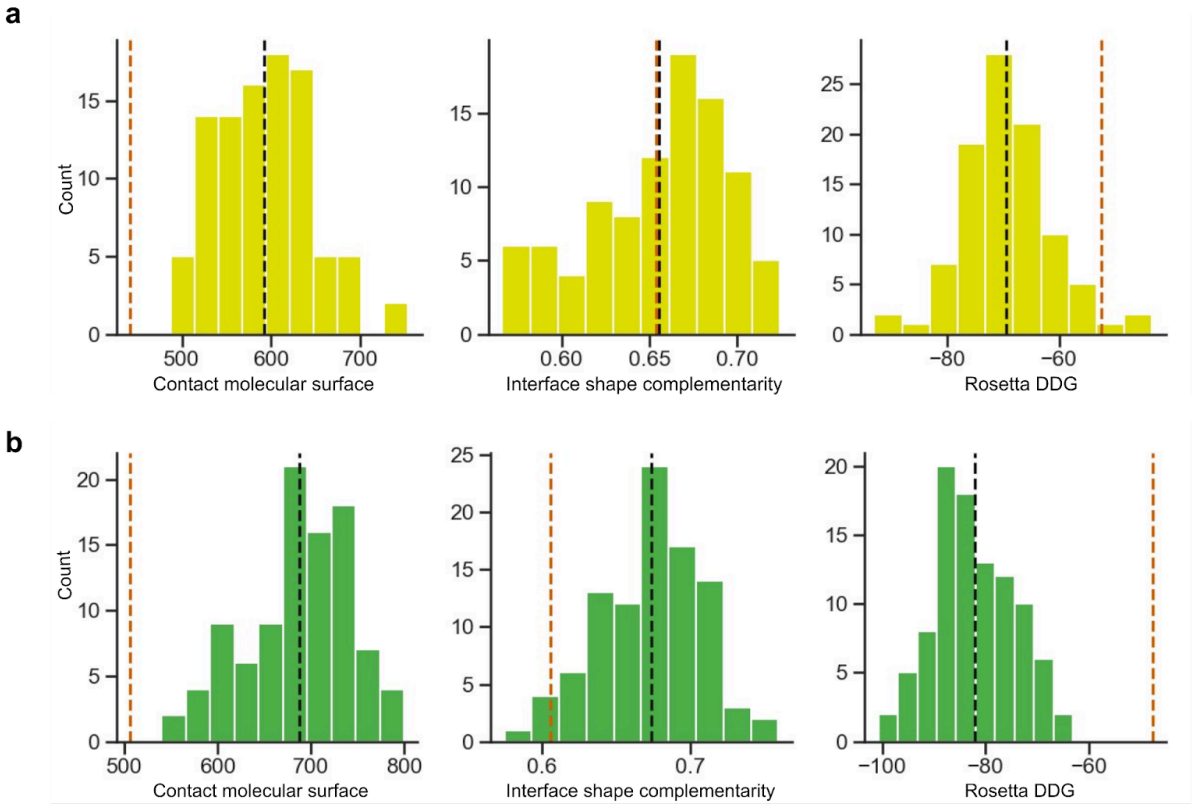


Figure S11. Binding metrics for partially diffused binders. (a) Computational metrics for 96 tested partially diffused glucagon binders showed significant improvement in contact molecular surface (a measure of interface size and quality) and Rosetta ddG (a measure of interface predicted energy) over the starting design (vertical red lines). Distribution means are shown in black. **(b)** Computational metrics for 96 ordered partially diffused NPY binders showed significant improvement in contact molecular surface, Rosetta ddG, and interface shape complementarity (a measure of interface quality) over the starting design (vertical red lines). Means are shown in black.

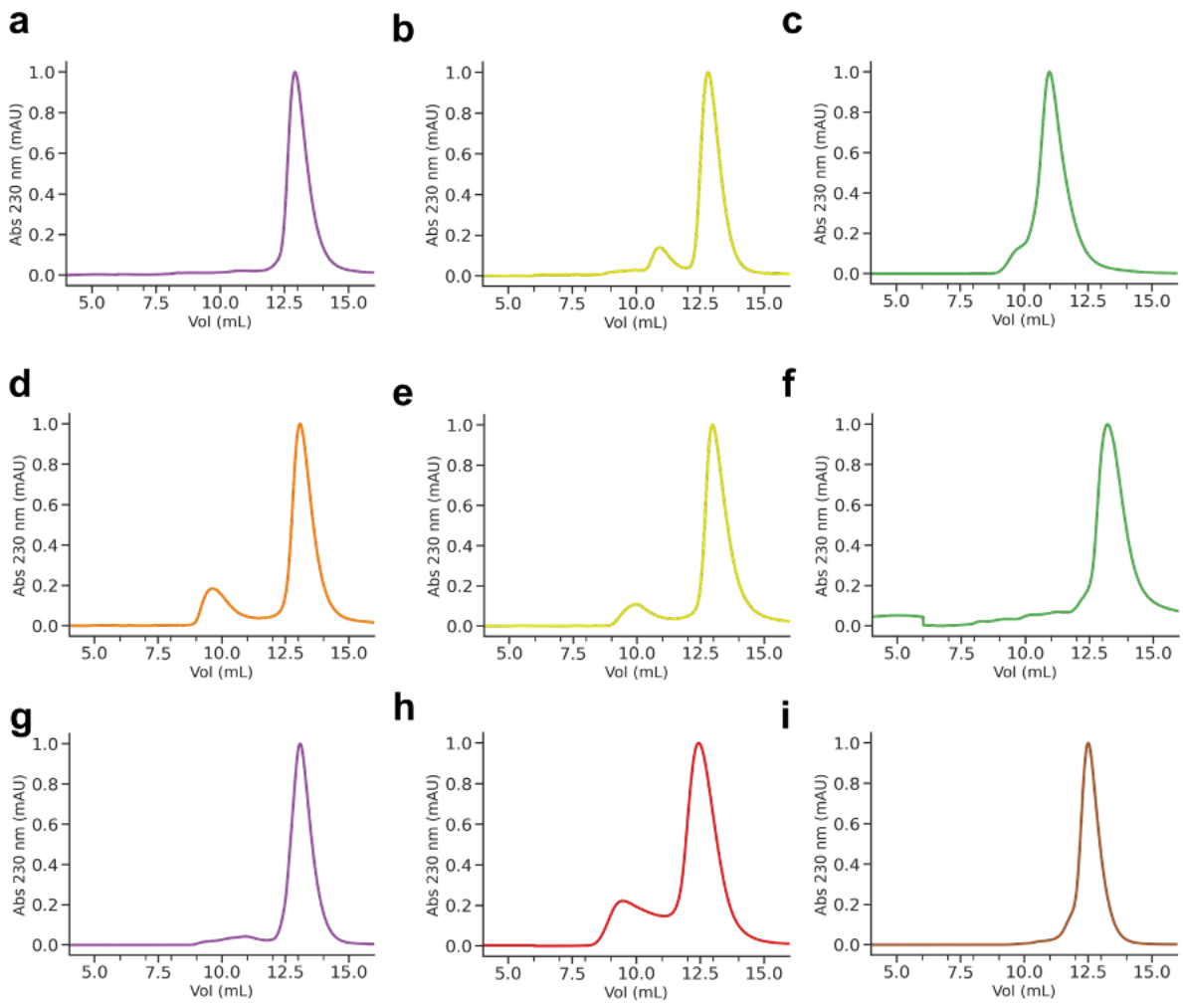
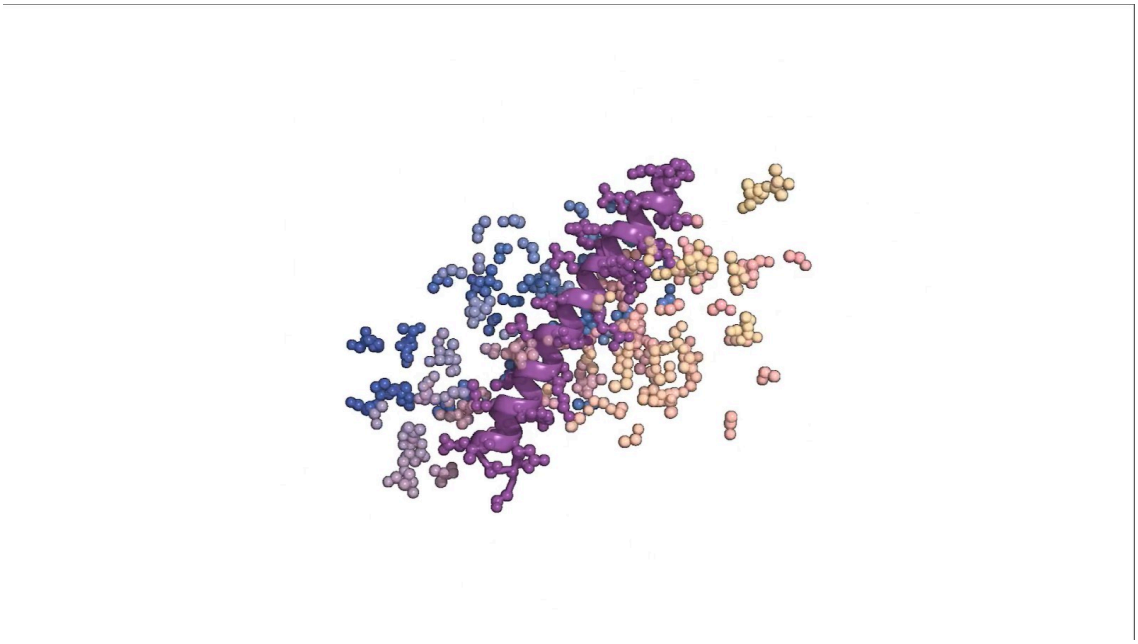


Figure S12. SEC traces of peptide binders. (a) SEC reinjection of the Inpainted PTH binder. **(b)** SEC reinjection of the Inpainted GCG binder. **(c)** SEC reinjection of the Inpainted NPY binder. **(d)** SEC reinjection of the threaded SCT binder. **(e)** SEC reinjection of the partially diffused GCG binder. **(f)** SEC reinjection of the partially diffused NPY binder. **(g)** SEC reinjection of the fully diffused PTH binder. **(h)** Initial SEC of the fully diffused Bim binder. **(i)** SEC reinjection of the flexibly diffused PYY binder. All reinjections were of the most abundant monodisperse peak and were performed 24 hours after initial SEC.



Supplementary Video 1. A video of the diffusion trajectory for the fully diffused PTH binder can be seen at:
https://www.bakerlab.org/wp-content/uploads/2022/11/diffusion_animation_PTHbinder_v6.mp4

Computational method	Targeted peptides
Parametric design	GCG, PTH and NPY
Hallucination	Bid
RF_{joint} Inpainting	Parametrically designed binders for GCG, PTH and NPY
Threading design	PTH, NPY, GCG, SCT, GIP, GLP1, GLP2
Partial diffusion	Inpainted GCG and NPY binders
Unconditional diffusion	PTH and Bim
RFdiffusion	PYY, NPY, GCG, Puma and PTH

Supplementary Table 8. Targeted peptides by each of the presented computational design approaches

	GCG Binder	NPY Binder	PTH Binder	Bim Binder
RMSD AF2 vs Design	0.62 Å	0.61 Å	0.78 Å	0.80 Å
AF2 interaction pAE	9.25	8.29	4.40	4.50
AF2 pLDDT for binder	95.52	93.41	94.3	96.6

Supplementary Table 9. AlphaFold2 metrics for partially and fully diffused binders.

Chapter 2:

De novo designed proteins neutralize lethal snake venom toxins

This chapter contains additional background for (and reproduction of) content previously published as: Vázquez Torres, S. *et al.* De novo designed proteins neutralize lethal snake venom toxins. *Res. Sq.* rs.3.rs-4402792 (2024) doi:10.21203/rs.3.rs-4402792/v1.

Background

Snakebite envenoming remains a devastating and neglected tropical disease, claiming over 100,000 lives annually and causing severe complications and long-lasting disabilities for many more[1,2]. Three-finger toxins (3FTXs) are highly toxic components of elapid snake venoms that can cause severe tissue damage[3] and life-threatening neurotoxicity by inhibiting nicotinic acetylcholine receptors (nAChRs)[4]. Currently, the only available treatments for snakebite consist of polyclonal antibodies derived from the plasma of immunized animals, which have high cost and limited efficacy against 3FTXs[5,6,7]

This study addresses the challenge of designing *de novo* protein binders for 3FTXs. Our aim is to use the deep learning-based RFdiffusion method[8] to design antivenoms for short- and long-chain α-neurotoxins and cytotoxins from the 3FTX snake venom toxin family. Such potent, stable, and readily manufacturable toxin-neutralizing proteins could form the basis for safer, cost-effective, and widely accessible next-generation antivenom therapeutics.

Design of α-neurotoxin binding proteins

α-neurotoxins, a prominent subclass of 3FTxs, adopt a multi-stranded β-structure with three extended loops protruding from a hydrophobic compact core stabilized by highly-conserved disulfide bridges[9,10] (Figure 1a). Short-chain and long-chain α-neurotoxins differ in length and number of disulfide bonds. Despite sequence homology, α-neurotoxins display distinct pharmacological profiles across nAChR subtypes: short- and long-chain α-neurotoxins inhibit muscle-type nAChRs, but only long-chain α-neurotoxins strongly bind to neuronal α7 nAChRs[11] (Figure 1c). As many elapid snake species possess venoms that derive their lethal effect from

these toxins, it is crucial to neutralize both types of α-neurotoxins to achieve therapeutic efficacy and prevent venom-induced lethality in victims envenomed by these snakes.

We chose to target our design efforts against the neurotoxin edge β-strands, unlike previously discovered monoclonal antibodies that mimic the nAChR binding site[12,13,14]. *De novo* design of binders performs best when targeting rigid regions, rather than flexible loops. Thus, while the loops presented the functionally most relevant epitopes, we decided to focus on the proximal beta-sheets. By targeting the neurotoxin edge β-strands, we aimed for high-affinity and specific binding while blocking neurotoxin binding to nAChRs through steric hindrance. Secondary structure and block adjacency tensors were provided to the RFdiffusion model to specify desired β-strand interactions between the designed binder and target α-neurotoxins (see Methods). For each secondary structure tensor, interactions between one binder β-strand and a target neurotoxin β-strand were encoded in a block adjacency tensor, preconditioning RFdiffusion towards β-strand pairing complexes. Following backbone generation through RFdiffusion denoising trajectories, sequence design was carried out using ProteinMPNN, the resulting designs were filtered based on AF2 *initial guess*[15] and Rosetta metrics, and the most promising candidates selected for experimental characterization. Top candidates for *in vitro* validation were selected based on their ability to interfere with neurotoxin binding to the nicotinic acetylcholine receptor (nAChR). This selection was guided by structural alignments comparing the neurotoxin bound to the receptor, with and without the designed binder.

We targeted short-chain α-neurotoxins using a previously designed consensus toxin derived from elapid snakes (ScNtx)[16] as a representative template (Supplementary Figure S8). Synthetic genes encoding 44 designs targeting ScNtx were screened via yeast surface display (YSD), and one candidate was identified to bind ScNtx with a dissociation constant (K_d) of 842 nM, as confirmed by bio-layer interferometry (BLI) (Supplementary Figure S1). Following partial diffusion optimization[17], 11 out of 78 designs had higher affinity than the initial hit (Supplementary Figure S9), with the best (SHRT) having a binding affinity of 0.9 nM, as determined by surface plasmon resonance (SPR) (Figure 2c, top row; a very similar value of 0.7 nM was obtained by BLI (Supplementary Figure S2)). SHRT displayed a single monomeric peak on size exclusion chromatography (SEC), characteristic αβ-protein circular dichroism (CD) spectra, and thermal stability with a melting temperature (T_m) of 78 °C (Figure 2d, top row). Using X-ray

crystallography, we determined the structure of the SHRT design in the apo state, which closely matched the computational design model (2.58 Å resolution; 1.04 Å RMSD) (Figure 3a). The binder interacts with loop III of the neurotoxin, which is key for the toxin to bind to the muscle-type α7 nAChRs[18] (Fig. 3a). A β-strand in SHRT makes extensive backbone hydrogen bonds with an edge β-strand in the toxin (Fig. 3a, right inset), and tyrosine 45, which is located on an alpha helix in SHRT, forms a backbone hydrogen bond with cysteine 41 on ScNtx (Fig. 3a, left inset).

As a representative native long-chain α-neurotoxin, we chose α‑cobratoxin (P01391) from *Naja kaouthia*, one of the most extensively characterized toxins from the 3FTx family[19] (Figure 1). From 42 RFdiffusion designs against α-cobratoxin, one candidate had a binding affinity of 1.3 μM using BLI (Supplementary Figure S3). Partial diffusion optimization of the binding interface generated 38 designs, of which 6 had higher affinity than the initial hit (Supplementary Figure S10). The highest affinity binder, LNG, had a K_d of 1.9 nM as measured by SPR (Figure 2c, middle row; BLI yielded a value of 6.7 nM (Supplementary Figure S4)). CD melting experiments revealed very high thermal stability (T_m > 95°C; Figure 2d, middle row).

Using X-ray crystallography, we determined the structure of the LNG α‑cobratoxin binder in complex with the target, which closely matched the computational design model (2.68 Å resolution; 0.42 Å RMSD over design, 0.61 Å over toxin; there is a slight deviation in the positioning of the toxin relative to the binder). As in the design model, the binder interacts with the central loop II of the neurotoxin, which is crucial for interaction of the toxin with muscle-type and neuronal α7 nAChRs[18,20]. This interaction is primarily mediated by backbone hydrogen bonding between a β-strand in LNG and a β-strand in the toxin (Fig. 3b). Arginine 33, located at the tip of loop II of α-cobratoxin, forms extensive interactions with LNG (Fig. 3b, left inset); this toxin residue also makes extensive interactions with the AChBP[21].

Design of cytotoxin binding proteins

Cytotoxins, a prominent functional group within the 3FTx family found in cobra venoms, exert cytotoxic effects and induce local tissue damage by destabilizing phospholipid membranes[22] (Figure 1b). Neutralizing these toxins is crucial to prevent severe sequelae, such as limb deformity, amputation, and lasting disabilities in snakebite victims[23].

For targeting cytotoxins, we hypothesized that relying solely on β-strand pairing interactions might not adequately prevent cytotoxin insertion into membranes due to the critical role of their three-finger loops in membrane interaction and disruption[24,25,26,27] (Figure 1b). Instead, we focused on binding directly to the cytotoxin three-finger loops by generating RFdiffusion-based protein backbones with hotspot residues defined within these regions (Figure 2a, bottom row). To increase the breadth of neutralization, we targeted a consensus sequence derived from 86 different snake cytotoxins (Type IA cytotoxin sub-subfamily; see Methods). Following ProteinMPNN and AF2 screening, partial diffusion was used to further optimize designs with the best metrics. A total of 55 protein designs were recombinantly expressed using *Escherichia coli*, and following SEC purification, the 18 designs with monomeric populations were tested in a luminescent cell viability assay. Of these, one protein binder (CYTX) had high solubility, with a single monomeric peak in SEC and had high neutralization activity against *Naja pallida* and *Naja nigricollis* whole venoms, known for their high cytotoxin content[28] (Supplementary Figure S5). The K_d for the cytotoxin from *Naja pallida* was determined to be 271 nM via SPR (Figure 2c, bottom row). CYTX exhibited characteristic αβ-protein CD spectrum and was thermostable, with a T_m of 61°C (Figure 2d, bottom row).

Few designed binders have targeted loops, and hence we sought to solve the crystal structure of CYTX in complex with *Naja pallida* cytotoxin (Fig 3c). To reduce flexibility to favor crystallization, a disulfide bond was introduced within a flexible loop connecting the β-sheet segment to the two α-helices of CYTX, yielding a candidate (CYTX_B10) with improved thermal stability (T_m= 70.3°C) and monomeric profile during SEC, but a slightly weaker K_d of 740 nM for *Naja pallida* cytotoxin (Supplementary Fig.S6). The structure of CYTX_B10 in complex with the target closely matched the computational design model (resolution: 2.0 Å; RMSD: 1.32 Å over design, 0.58 Å over toxin), revealing extensive electrostatic interactions involving side chain–main chain hydrogen bonds between cytotoxin loops II and III and the CYTX_B10 binder (Figure 3c, left inset). The unusual open fold of CYTX_B10 highlights the power of RFdiffusion to custom generate scaffolds shape-matched with protein targets, and the power of proteinMPNN to stabilize structures which violate common rules of protein structure (in this case lacking a central hydrophobic core).

In vitro neutralization

We assessed the ability of the designs to functionally neutralize α-neurotoxins in patch-clamp experiments using a human-derived rhabdomyosarcoma cell line expressing muscle-type nAChRs. When preincubated with ScNtx, the SHRT design achieved complete neutralization at a 1:1 molar ratio (toxin:binder), better than a previously characterized ScNtx nanobody (TPL1163_02_A01)[29] (Figure 4a; a control nanobody targeting phospholipase A_2 (PLA_2) had no effect). Similarly, the LNG design had better neutralizing efficacy than a previously characterized α-cobratoxin nanobody (TPL1158_01_C09)[29], achieving full protection at a 1:1 molar ratio (toxin:binder) (Figure 4b).

We used a cytotoxicity assay to evaluate the cross-reactivity of the CYTX design against various cobra whole venoms. Immortalized human keratinocytes (N/TERTs) were exposed to venoms from seven different *Naja* species, which prior proteomic analyses suggest consist primarily (~70%) of cytotoxins[30]. Pre-incubating CYTX with venoms (2 IC_{50}s) at a 1:5 molar ratio (toxin:binder) provided 70-90% protection against venom-induced cytotoxicity (Figure 4c). Similarly, pre-incubation of the cytotoxin binder with isolated cytotoxin from *Naja pallida* (2 IC_{50}s) at a 1:5 molar ratio (toxin:binder) gave 85% protection against cytotoxicity (Figure 4c). However, preliminary studies indicated that the CYTX design, in 1:1, 1:2.5, and 1:5 molar ratios (toxin:binder), did not significantly decrease the size of the dermonecrotic lesions induced by intradermal *Naja nigricollis* venom administration in a previously published murine model[31] (Supplementary Fig.S7); the affinity of CYTX likely needs to be further optimized for full *in vivo* neutralization of cytotoxins.

In vivo protection

Given the encouraging *in vitro* neutralization for our anti-neurotoxin designs, we proceeded to *in vivo* studies. We determined the mean lethal dose (LD_{50}) values for α-neurotoxins in male non-Swiss albino (NSA) mice via intraperitoneal (IP) administration; α-cobratoxin had an LD_{50} of 0.098 μg/g and ScNtx had an LD_{50} of 0.087 μg/g, in agreement with prior intravenous LD_{50} doses of these toxins (0.1 μg/g)[32]. To evaluate the *in vivo* neutralizing efficacy of our neurotoxin-targeting protein binders, we monitored survival for 24 hours post-lethal neurotoxin challenge following administration of purified toxins (3 LD_{50}s) (Figure 4d). The SHRT binder

provided complete protection (100%) to mice when pre-incubated and administered intraperitoneally with the corresponding short-chain neurotoxin at a 1:10 molar ratio of toxin to binder, but as expected did not neutralize the non-target α-cobratoxin. The LNG binder exhibited comparable efficacy, completely neutralizing α-cobratoxin but not the non-target ScNtx (Figure 4d, left). In rescue assays better mimicking a real-life snakebite scenario, complete protection (100%) was achieved when short-chain or long-chain α-neurotoxin binders were administered intraperitoneally at a 1:10 molar ratio (toxin:binder) 15 minutes after a lethal α-neurotoxin challenge (3 LD_{50}s) (Figure 4d, middle). Administering the SHRT binder 30 minutes post-toxin injection also provided 100% protection against ScNtx, while the LNG binder conferred 60% protection against α-cobratoxin (Figure 4d, right). All surviving mice showed no evidence of limb or respiratory paralysis. At a 1:5 molar ratio (toxin:binder), IP administration of the SHRT design 15 minutes after toxin injection (3 LD_{50}s) resulted in 100% survival, while the LNG binder provided 80% protection. Mice injected with the binder alone showed no negative effects at 24 and 48 hours post-injection, nor up to two weeks post-injection.

Discussion

Antivenoms based on animal-derived polyclonal antibodies have long been the cornerstone of snakebite envenoming therapy, but their application is hampered by their limited efficacy against toxins with low immunogenicity, their propensity to cause severe adverse reactions, and the inherent batch-to-batch variations and high production costs associated with their manufacture[33]. Thus, there has been a search for alternatives, with recombinant human monoclonal antibodies and nanobodies presenting a solution that can help overcome some of these limitations[34]. Our designed neurotoxin binders demonstrate comparable potency to the best immunoglobulin G antibodies and nanobodies reported in literature[34], are highly stable and readily producible in microbial systems, and their small size (~100 amino acids) may possibly enable them to penetrate rapidly into deep tissue[35]. More generally, our *in silico* design approach avoids animal immunization and/or construction and multiple rounds of selection and/or screening of large libraries, providing a low-cost methodology for rapid development of toxin binders to the many components of snake venom when structural or sequence data exists for these targets. *De novo* designed proteins have high stability and are amenable to low cost manufacturing, which is key to effectively addressing snakebite envenoming as a neglected tropical disease. From the design

perspective, the crystal structure of our cytotoxin binder highlights the ability of RFdiffusion to custom design scaffolds to match almost any target shape, and to generate binders to loop regions of proteins, with the inhibitory activity of the anti-cytotoxin designs directly supporting a role for the loops in membrane disruption.

Advancing the field to provide effective solutions for snakebite victims requires a collaborative effort involving the scientific community, the pharmaceutical industry, public health systems, and governments[2]. While traditional antivenoms will likely remain a cornerstone in snakebite treatment for the immediate future, our *de novo* designed binders offer potential as fortifying agents to enhance antivenom efficacy. This approach could be especially beneficial in treating elapid envenomations, where low-molecular-weight toxins with limited immunogenicity, but high medical importance, dominate the toxic effects of the venoms (and therefore must be neutralized)[36]. Beyond fortification, generative binder design could be used to generate neutralizing proteins against other medically relevant toxins, thereby expediting the discovery of antivenoms with broader species coverage. More generally, as *in silico* protein design is less resource intensive than traditional antibody development, our approach could aid in the democratization of drug design and discovery, enabling researchers residing in low and middle-income countries to better contribute to the development of effective treatments for snakebite envenoming and other neglected tropical diseases.

Figures

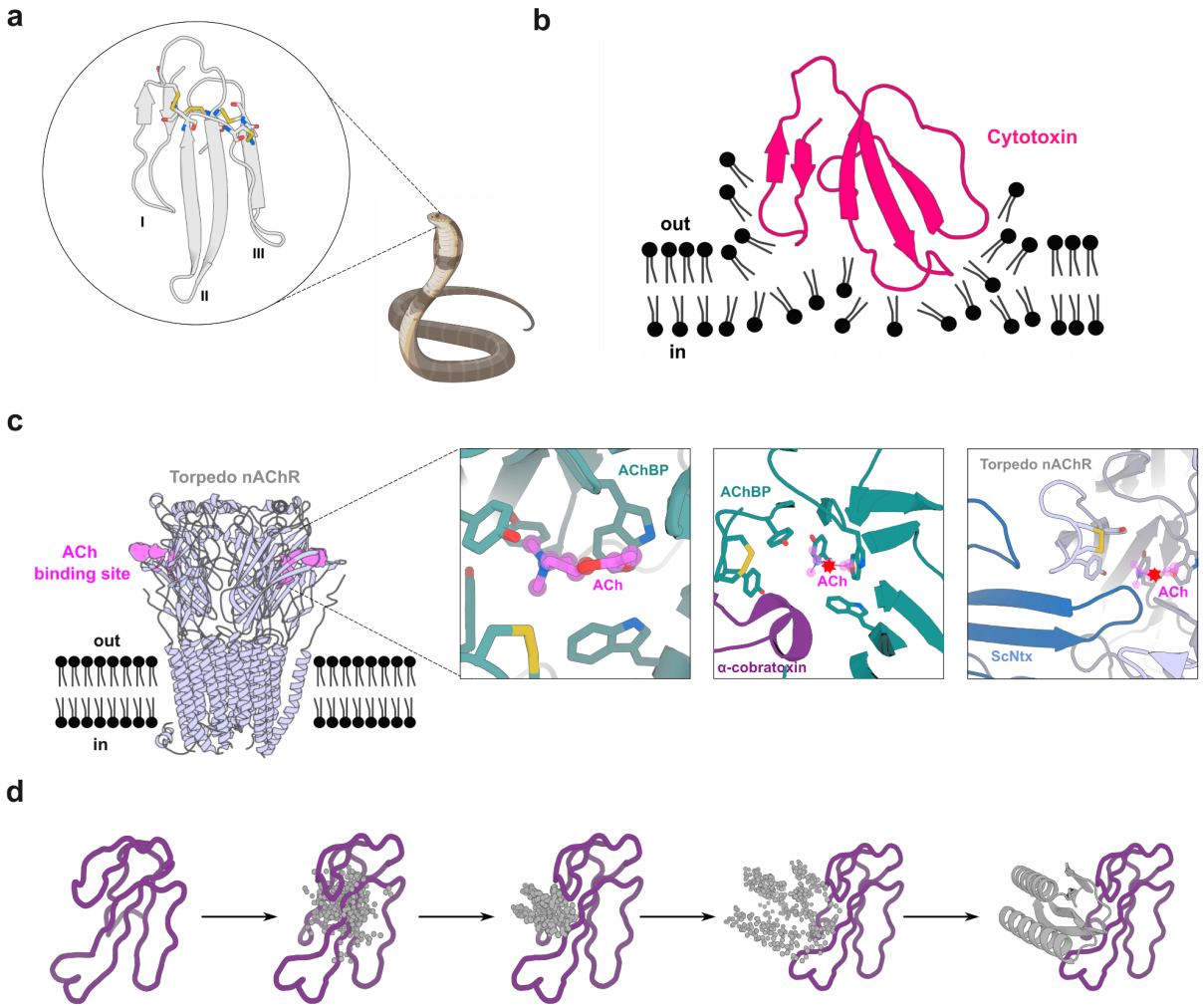


Figure 1. Targets of 3 finger snake toxins (3FTxs). **(a)** Structure of 3FTxs[37] (PDB ID: 1QKD). Highly conserved cysteine residues are highlighted in sticks and each of the three fingers indicated (I-III). **(b)** Representation of a type IA cytotoxin[38] (dark pink) (PDB ID: 5NQ4) interacting with a lipid bilayer. **(c)** Muscle acetylcholine (Torpedo) receptor (light blue) (PDB ID: 7Z14)[11]. Acetylcholine (ACh) binding site is depicted in violet. Left inset: Close-up of the acetylcholine binding protein (AChBP) (teal) (PDB ID: 3WIP) bound to ACh[39] (violet). A set of aromatic residues form a cage around the neurotransmitter. Middle: Close-up of α-cobratoxin (dark purple) blocking access to the ACh binding site in AChBP (teal) (PDB ID: 1YI5)[21]. A molecule of ACh is depicted to illustrate its binding site. Right: Close-up of ScNtx (dark blue) blocking access to the ACh binding site in the Torpedo receptor (light blue) (PDB ID: 7Z14)[11]. A

molecule of ACh is depicted to illustrate its binding site. **(d)** Schematic showing α-cobratoxin binder design using RFdiffusion. Starting from a random distribution of residues around the specified β-strands in the target toxin (dark purple), successive RFdiffusion denoising steps progressively remove the noise leading at the end of the trajectory to a folded structure interacting with α-cobratoxin β-strands. Panels **(a)**, **(b)**, and **(c)** were created with BioRender.com

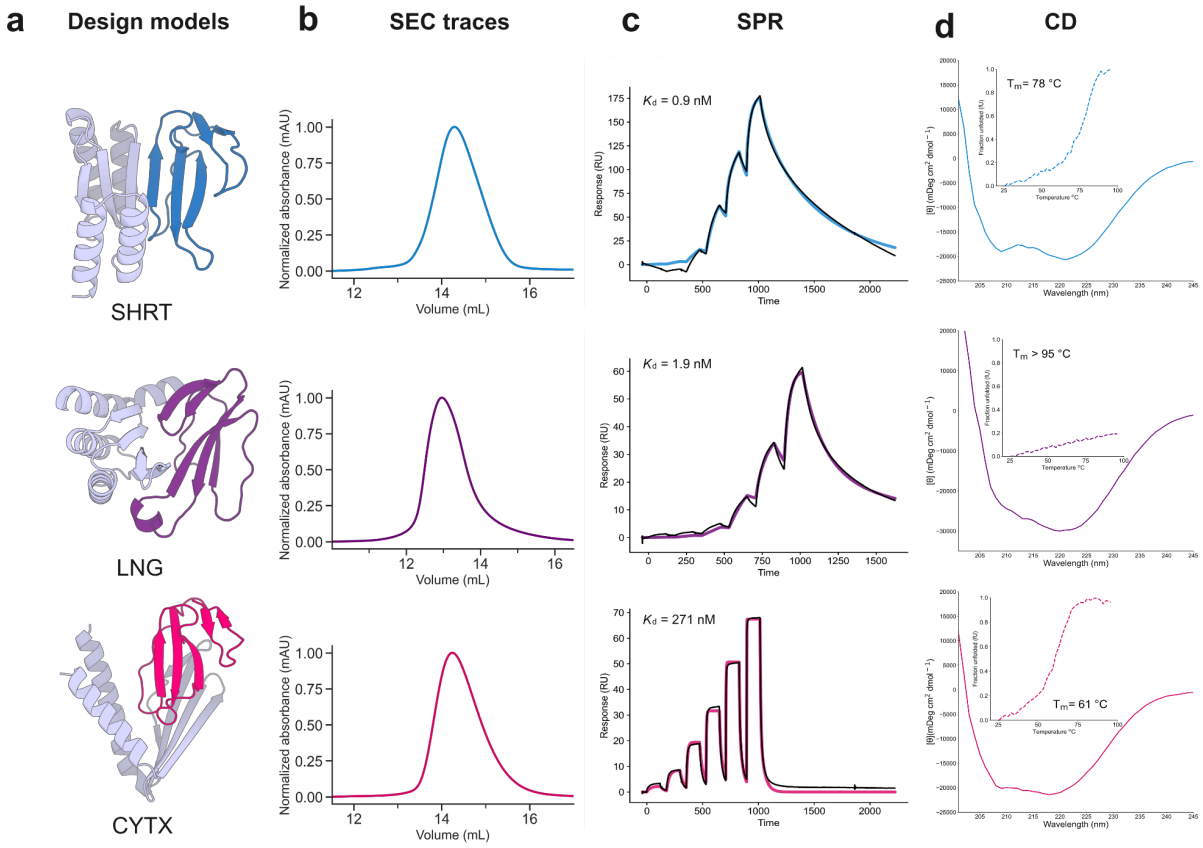


Figure 2. Experimental characterization of 3FTx binding proteins. (a) Design models of protein binders (gray) bound to their 3FTx targets (dark blue: ScNtx, dark purple: α-cobratoxin, dark pink: consensus cytotoxin). **(b)** SEC traces of purified proteins. **(c)** SPR binding affinity measurements. Colored solid lines represent fits using the heterogeneous ligand model, with dissociation constant (K_d) values derived from these fits**. (d)** CD data confirms the presence of αβ-secondary structure in the 3FTx binding proteins and their thermal stability (inset).

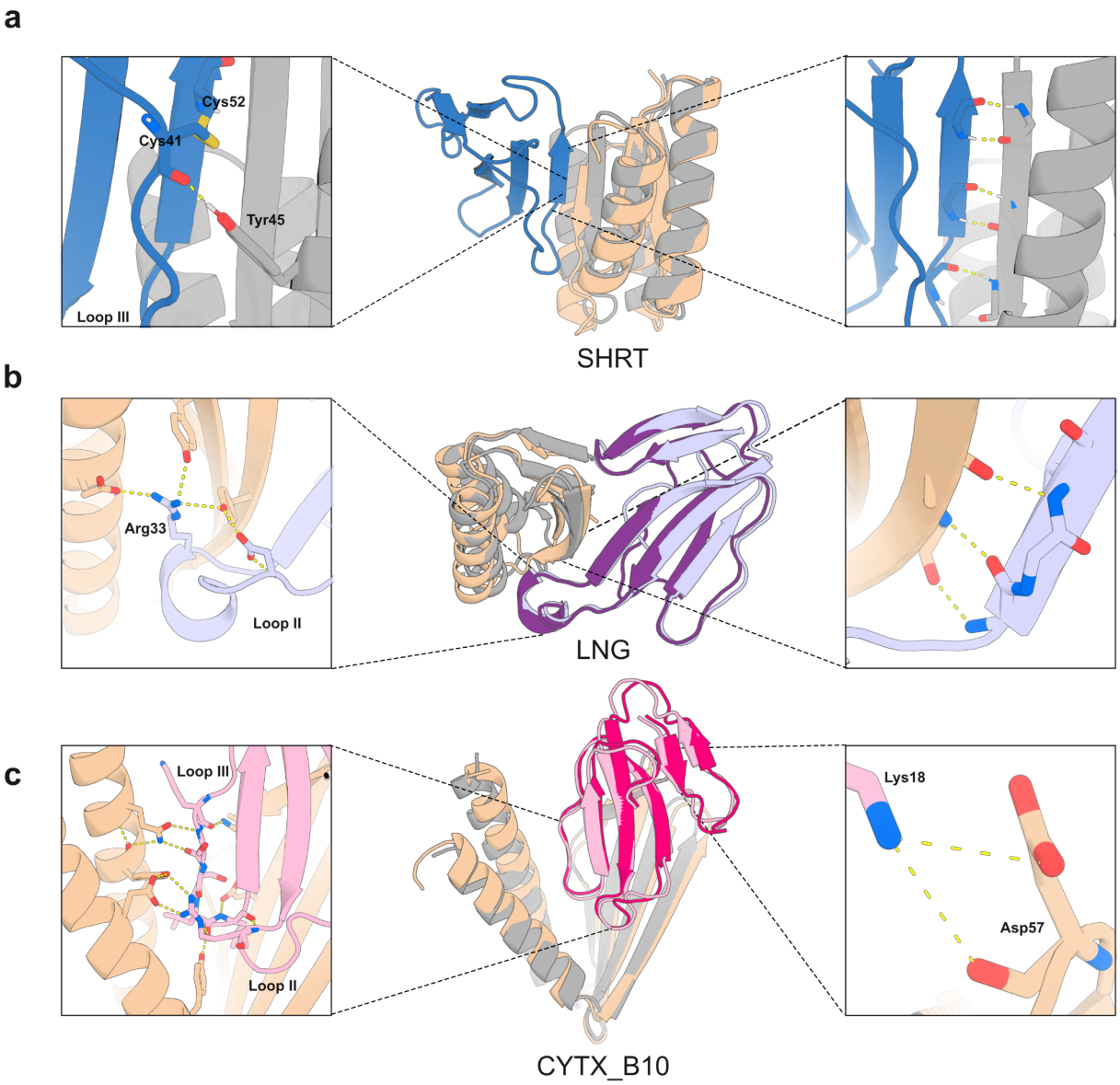


Figure 3. Crystal structures of 3FTx binding proteins closely match design models. (a) Apo-state crystal structure of SHRT design. Left: Hydrogen bonding between the carbonyl oxygen of Cys41 in ScNtx (dark blue) and the side chain of Tyr45 in the SHRT design model (gray). Middle: Overlay of SHRT design model (gray) with crystal structure (wheat). Right: Backbone hydrogen bonding between the SHRT design model (gray) and ScNtx (dark blue) β strands. **(b)** Crystal structure of LNG design in complex with α-cobratoxin. Left: Cross-interface hydrogen-bond network involving Arg33 at loop II in α-cobratoxin (light purple) and Glu69, Tyr40, and Tyr49 in LNG crystal structure (wheat). Middle: Overlay of LNG design model (gray) bound to α-cobratoxin (dark purple) with crystal structure of binder (wheat) bound to

toxin (light purple). Right: Backbone hydrogen bonding between crystal structure of designed binder (wheat) and α-cobratoxin (light purple) β strands. **(c)** Crystal structure of CYTX_B10 design in complex with *Naja pallida* cytotoxin. Left: Cross-interface electrostatic interaction network between loops III and II of *Naja pallida* cytotoxin (light pink) and binder crystal structure (wheat). Middle: Overlay of CYTX_B10 design model (gray) bound to toxin (dark pink) with crystal structure of binder (wheat) bound to *Naja pallida* cytotoxin (light pink). Right: Salt bridge between positively charged Lys18 in cytotoxin (light pink) and Asp57 in the binder crystal structure (wheat).

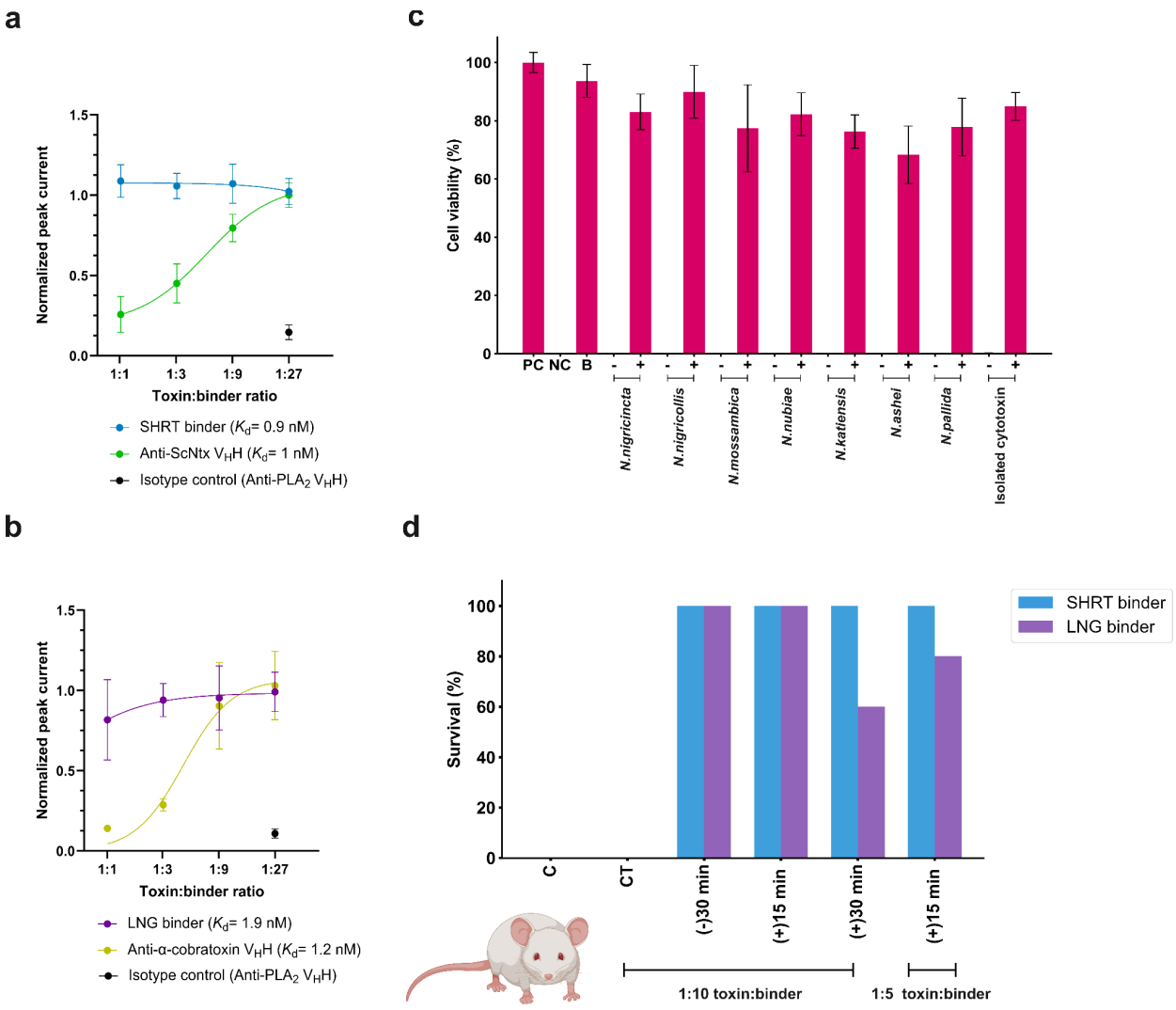


Figure 4. ***In vitro*** **and** ***in vivo*** **efficacy of designed proteins against snake venom toxins. (a)** Concentration-response curves comparing SHRT binder and anti-ScNtx V_HH efficacy in preventing nAChR blocking by one IC_{80} of ScNtx. Data represent the toxin's inhibition of ACh response, normalized to full ACh response, averaged within each group (*n*=16). **(b)** Concentration-response curves comparing the efficacy of LNG binder and anti-α-cobratoxin V_HH in preventing nAChR blocking by one IC_{80} of α-cobratoxin. **(c)** Neutralization of the cytolytic effects of whole venoms from seven different *Naja* species and isolated cytotoxin by the CYTX binder. 2 IC_{50} of the whole venoms or toxin were pre-incubated with CYTX at a 1:5 molar ratio (toxin:binder). This ratio was estimated assuming 70% of the whole venom consists of cytotoxins based on prior proteomic analyses[30]. Keratinocyte media was used as a positive control (PC). Triton X-100 was used as a negative control (NC). CYTX binder (B) was used as a

positive control. (-) denotes 2 IC_{50} of the whole venoms without binder, and (+) denotes venoms incubated with binder. Experiments were performed in triplicates, and results are expressed as mean ± SD. **(d)** Mice survival following lethal neurotoxin challenge (*n*=5). 3 LD_{50}s of ScNtx or α-cobratoxin were preincubated for 30 minutes (-30 min) with the corresponding protein binders at 1:10 ratios and then administered IP into groups of five mice. Toxins administered IP following IP administration of binders at 1:10 or 1:5 molar ratios (toxin:binder) either after 15 (+15 min) or 30 minutes (+30 min) post-toxin injection. Controls included mice receiving toxins alone (C). Specificity was assessed via cross-treatment (CT) experiments, where non-target binders were preincubated with 3 LD_{50}s of ScNtx or α-cobratoxin and administered IP. Signs of toxicity were observed, and deaths were recorded for a period of 24 hours. **(d)** was created with BioRender.com.

References

1. GBD 2019 Snakebite Envenomation Collaborators *et al.* Global mortality of snakebite envenoming between 1990 and 2019. *Nat. Commun.* **13**, 6160 (2022).
2. Gutiérrez, J. M. *et al.* Snakebite envenoming. *Nat. Rev. Dis. Primer* **3**, 17063 (2017).
3. Royte, L. & Sawarkar, A. Snake bite - cytotoxic effects of snake venom: a rare clinical image. *Pan Afr. Med. J.* **44**, (2023).
4. Barber, C. M., Isbister, G. K. & Hodgson, W. C. Alpha neurotoxins. *Toxicon* **66**, 47–58 (2013).
5. Deka, A., Gogoi, A., Das, D., Purkayastha, J. & Doley, R. Proteomics of Naja kaouthia venom from North East India and assessment of Indian polyvalent antivenom by third generation antivenomics. *J. Proteomics* **207**, 103463 (2019).
6. Laustsen, A. H. *et al.* Snake venomics of monocled cobra (Naja kaouthia) and investigation of human IgG response against venom toxins. *Toxicon* **99**, 23–35 (2015).
7. Tan, K. Y., Tan, C. H., Fung, S. Y. & Tan, N. H. Venomics, lethality and neutralization of Naja kaouthia (monocled cobra) venoms from three different geographical regions of Southeast Asia. *J. Proteomics* **120**, 105–125 (2015).
8. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
9. Utkin, Y. N. Last decade update for three-finger toxins: Newly emerging structures and

biological activities. *World J. Biol. Chem.* **10**, 17–27 (2019).

10. Kalogeropoulos, K. *et al.* A comparative study of protein structure prediction tools for challenging targets: Snake venom toxins. *Toxicon* **238**, 107559 (2024).
11. Nys, M. *et al.* The molecular mechanism of snake short-chain α-neurotoxin binding to muscle-type nicotinic acetylcholine receptors. *Nat. Commun.* **13**, 4543 (2022).
12. Khalek, I. S. *et al.* Synthetic development of a broadly neutralizing antibody against snake venom long-chain α-neurotoxins. *Sci. Transl. Med.* **16**, eadk1867 (2024).
13. Ledsgaard, L. *et al.* Discovery and optimization of a broadly-neutralizing human monoclonal antibody against long-chain α-neurotoxins from snakes. *Nat. Commun.* **14**, 682 (2023).
14. Glanville, J. *et al.* Venom protection by broadly neutralizing antibody from a snakebite subject. Preprint at https://doi.org/10.1101/2022.09.26.507364 (2022).
15. Bennett, N. R. *et al.* Improving de novo protein binder design with deep learning. *Nat. Commun.* **14**, 2625 (2023).
16. De La Rosa, G., Corrales-García, L. L., Rodriguez-Ruiz, X., López-Vera, E. & Corzo, G. Short-chain consensus alpha-neurotoxin: a synthetic 60-mer peptide with generic traits and enhanced immunogenic properties. *Amino Acids* **50**, 885–895 (2018).
17. Vázquez Torres, S. *et al.* De novo design of high-affinity binders of bioactive helical peptides. *Nature* **626**, 435–442 (2024).
18. Bekbossynova, A., Zharylgap, A. & Filchakova, O. Venom-Derived Neurotoxins Targeting Nicotinic Acetylcholine Receptors. *Molecules* **26**, 3373 (2021).
19. Miersch, S. *et al.* Synthetic antibodies block receptor binding and current‐inhibiting effects of α‐cobratoxin from *Naja kaouthia*. *Protein Sci.* **31**, e4296 (2022).
20. Antil-Delbeke, S. *et al.* Molecular Determinants by Which a Long Chain Toxin from Snake Venom Interacts with the Neuronal α7-Nicotinic Acetylcholine Receptor. *J. Biol. Chem.* **275**, 29594–29601 (2000).
21. Bourne, Y., Talley, T. T., Hansen, S. B., Taylor, P. & Marchot, P. Crystal structure of a Cbtx–AChBP complex reveals essential interactions between snake α-neurotoxins and nicotinic receptors. *EMBO J.* **24**, 1512–1522 (2005).
22. Hiu, J. J. & Yap, M. K. K. The myth of cobra venom cytotoxin: More than just direct cytolytic actions. *Toxicon X* **14**, 100123 (2022).
23. Chong, H. P., Tan, K. Y., Liu, B.-S., Sung, W.-C. & Tan, C. H. Cytotoxicity of Venoms and

Cytotoxins from Asiatic Cobras (Naja kaouthia, Naja sumatrana, Naja atra) and Neutralization by Antivenoms from Thailand, Vietnam, and Taiwan. *Toxins* **14**, 334 (2022).

24. Hiu, J. J., Fung, J. K. Y., Tan, H. S. & Yap, M. K. K. Unveiling the functional epitopes of cobra venom cytotoxin by immunoinformatics and epitope-omic analyses. *Sci. Rep.* **13**, 12271 (2023).
25. Dubovskii, P. V., Lesovoy, D. M., Dubinnyi, M. A., Utkin, Y. N. & Arseniev, A. S. Interaction of the P‑type cardiotoxin with phospholipid membranes. *Eur. J. Biochem.* **270**, 2038–2046 (2003).
26. Li, F., Shrivastava, I. H., Hanlon, P., Dagda, R. K. & Gasanoff, E. S. Molecular Mechanism by Which Cobra Venom Cardiotoxins Interact with the Outer Mitochondrial Membrane. *Toxins* **12**, 425 (2020).
27. G. Konshina, A., V. Dubovskii, P. & G. Efremov, R. Structure and Dynamics of Cardiotoxins. *Curr. Protein Pept. Sci.* **13**, 570–584 (2012).
28. Kazandjian, T. D. *et al.* Convergent evolution of pain-inducing defensive venom components in spitting cobras. *Science* **371**, 386–390 (2021).
29. Jensen, Anna. *Design of Consensus Toxins and Their Use for the Discovery of Broadly Neutralizing Antibodies*.
30. Petras, D. *et al.* Snake Venomics of African Spitting Cobras: Toxin Composition and Assessment of Congeneric Cross-Reactivity of the Pan-African EchiTAb-Plus-ICP Antivenom by Antivenomics and Neutralization Approaches. *J. Proteome Res.* **10**, 1266–1280 (2011).
31. Bartlett, K. E. *et al.* Dermonecrosis caused by a spitting cobra snakebite results from toxin potentiation and is prevented by the repurposed drug varespladib. *Proc. Natl. Acad. Sci.* **121**, e2315597121 (2024).
32. Modahl, C. M., Mukherjee, A. K. & Mackessy, S. P. An analysis of venom ontogeny and prey-specific toxicity in the Monocled Cobra (Naja kaouthia). *Toxicon* **119**, 8–20 (2016).
33. Jenkins, T. P. & Laustsen, A. H. Cost of Manufacturing for Recombinant Snakebite Antivenoms. *Front. Bioeng. Biotechnol.* **8**, 703 (2020).
34. Laustsen, A. H. Recombinant snake antivenoms get closer to the clinic. *Trends Immunol.* **45**, 225–227 (2024).
35. Nessler, I. *et al.* Increased Tumor Penetration of Single-Domain Antibody–Drug Conjugates

Improves *In Vivo* Efficacy in Prostate Cancer Models. *Cancer Res.* **80**, 1268–1278 (2020).

36. Laustsen, A. H., Lohse, B., Lomonte, B., Engmark, M. & Gutiérrez, J. M. Selecting key toxins for focused development of elapid snake antivenoms and inhibitors guided by a Toxicity Score. *Toxicon* **104**, 43–45 (2015).
37. Nastopoulos, V., Kanellopoulos, P. N. & Tsernoglou, D. Structure of Dimeric and Monomeric Erabutoxin a Refined at 1.5 Å Resolution. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 964–974 (1998).
38. Dubovskii, P. V. *et al.* Impact of membrane partitioning on the spatial structure of an S-type cobra cytotoxin. *J. Biomol. Struct. Dyn.* **36**, 3463–3478 (2018).
39. Olsen, J. A., Balle, T., Gajhede, M., Ahring, P. K. & Kastrup, J. S. Molecular Recognition of the Neurotransmitter Acetylcholine by an Acetylcholine Binding Protein Reveals Determinants of Binding to Nicotinic Acetylcholine Receptors. *PLoS ONE* **9**, e91232 (2014).

Methods

Cytotoxin consensus sequence design

Amino acid sequences for cytotoxins were collected from the UniProt website using family: "snake three-finger toxin family Short-chain subfamily Type IA cytotoxin sub-subfamily" as a query. The resultant 86 unique CTX sequences thereafter underwent multiple sequence alignment (MSA) in Clustal Omega. Using these alignments a consensus sequence was designed to represent the most common amino acids at each position across the aligned sequences. In this process, each column of the sequence alignment was analyzed to select the most frequent amino acid. In scenarios where no single amino acid dominated, a consensus symbol was used to represent a group of similar amino acids based on properties like charge or hydrophobicity. This approach allowed for the representation of conserved biochemical properties rather than specific amino acid identities at positions with high variability.

Secondary structure and block adjacency tensors

In order to generate desired binder-target β-strand pairing interactions using RFdiffusion, fold-conditioning tensors describing single binder β-strands interacting with target β-strands in a matrix format were supplied to RFdiffusion at inference. This information is supplied via two tensors: a [L,4] secondary one-hot tensor (0=α-helix, 1=β-strand, 2=loop, 3=masked secondary structure identity) to indicate the secondary structure classification of each residue in the binder-target complex, and an [L,L,3] adjacency one-hot tensor (0=non-adjacent, 1=adjacent, 2=masked adjacency) to indicate interacting partner residues for each residue in the binder-target complex. For the design of the binders described here, the secondary structure tensor indicated an entirely masked binder structure with the exception of binder residues set to β-strand identities, while the adjacency tensor indicated a masked adjacency between binder-target residues with the exception of the pre-defined strand residues being adjacent to the defined target strand residues.

***De novo* 3FTX binder design using RFdiffusion**

The crystal structures of ScNtx (PDB ID: 7Z14) and α-cobratoxin (PDB ID: 1YI5) served as input for RFdiffusion. In the case of the consensus cytotoxin, its AF2 model was utilized. A total of approximately two thousand diffused designs were generated for each target, employing secondary structure and block adjacency tensors in the RFdiffusion model. The resulting backbone libraries underwent sequence design using ProteinMPNN, followed by FastRelax and AF2 + initial guess. The resulting libraries were filtered based on AF2 PAE <10, pLDDT > 80, and Rosetta ddg < -40.

Partial diffusion to optimize binders

The AF2 models of the highest-affinity designs for each toxin target were used as inputs to partial diffusion. The models were subjected to 10 and 20 noising timesteps out of a total of 50 timesteps in the noising schedule, and subsequently denoised ("diffuser.partial_T" input values of 10 and 20). Approximately two thousand partially diffused designs were generated for each target. The resulting library of backbones were sequence designed using ProteinMPNN after Rosetta FastRelax, followed by AF2+initial guess. The resulting libraries were filtered based on AF2 PAE <10, pLDDT > 80, and Rosetta ddg < -40.

Recombinant expression of ScNtx

ScNtx was recombinantly expressed from the methylotrophic yeast *Komagataella phaffii* (formerly known as *Pichia pastoris*). The ScNtx sequence was codon-optimized for expression in yeast and included a N-terminal His_6 tag, followed by a Biotin Acceptor Peptide, and a Tobacco Etch Virus (TEV) proteolytic site. The culture media was dialyzed overnight against a wash buffer (50 mM Sodium phosphate buffer, pH 8.0, 20 mM imidazole). Purification was carried out using an NGC™ chromatography system (Bio-Rad) with a 5 mL IMAC Nickel column (Bio-Rad). After loading, the column was washed with 5 column volumes of wash buffer to remove non-specifically bound proteins. The protein was then eluted using a gradient of 250 mM imidazole over 10 column volumes. Fractions with high absorbance at 280 nm were pooled,

dialyzed against 50 mM Sodium phosphate buffer, pH 8.0. Purity was assessed on SDS-PAGE to confirm the size. The protein solution was aliquoted, and stored at -20°C for further use.

Toxins

α-cobratoxin (L8114) was obtained from Latoxan (Portes lés Valence, France). Cytotoxin from *Naja pallida* was obtained from Sigma-Aldrich (217503).

Venoms

Whole venoms for initial neutralization screening from *Naja nigricollis* (CV01089563VEN) and *Naja pallida* (CV01089566VEN) were obtained in lyophilized form from Amerigo Scientific. Catalog numbers are provided in parentheses.

For *in vitro* neutralization experiments in human keratinocytes, whole venoms from *Naja nigricollis* (L1327), *Naja nigricincta* (L1368), *Naja mossambica* (L1376), *Naja nubiae* (L1342), *Naja katiensis* (L1317), *Naja ashei* (L1375), and *Naja pallida* (L1321) were purchased in lyophilized form from Latoxan (Portes lés Valence, France). Catalog numbers are provided in parentheses.

For the anti-cytotoxin *in vivo s*tudy, *Naja nigricollis* venom was sourced from wild-caught Tanzanian specimens housed in the herpetarium of the Liverpool School of Tropical Medicine.

Gene construction of 3FTX binders

The designed protein sequences were optimized for expression in *E. coli.* Linear DNA fragments (eBlocks, Integrated DNA Technologies) encoding the design sequences contained overhangs suitable to cloning into the pETcon3 vector for yeast display (deposited in Addgene as #45121) and LM627 vector for protein expression (Addgene #191551) via Golden Gate cloning.

Yeast display screening

For the yeast transformation, 50-60 ng of pETcon3, digested with NdeI and XhoI restriction enzymes, and 100 ng of the insert (eBlocks, Integrated DNA Technologies) were transformed into *S. cerevisiae* EBY100. EBY100 cultures were cultivated in C-Trp-Ura medium with 2%

(w/v) glucose (CTUG). To induce expression, yeast cells initially grown in CTUG were transferred to SGCAA medium with 0.2% (w/v) glucose and induced at 30 °C for 16–24 h. After induction, cells were washed with PBSF (PBS with 1% (w/v) BSA) and labeled for 40 minutes with biotinylated toxin targets at room temperature using the without-avidity labeling condition. Subsequently, cells were washed, resuspended in PBSF, and individually sorted based on each unique design using a 96-well compatible autosampler in the Attune NxT Flow Cytometer (Thermo Fisher Scientific).

Protein expression and purification in *E. coli* for 3FTX binders

Protein expression was conducted in 50 mL of Studier autoinduction media supplemented with kanamycin, and cultures were grown overnight at 37°C. Cells were harvested by centrifugation at 4,000 x g for 10 min and resuspended in lysis buffer (100 mM Tris-HCl, 200 mM NaCl, 50 mM imidazole) supplemented with Pierce™ Protease Inhibitor Tablets (EDTA-free). Cell lysis was achieved by sonication using a Qsonica Q500 instrument with a 4-pronged horn for 2:30 min ON total, at an amplitude of 80%. Soluble fractions were clarified by centrifugation at 14,000 x g for 40 minutes and subsequently purified by affinity chromatography using Ni-NTA resin (Qiagen) on a vacuum manifold. Washes were performed using Low-salt buffer (20 mM Tris-HCl, 200 mM NaCl, 50 mM imidazole) and High-salt buffer (20 mM Tris-HCl, 1000 mM NaCl, 50 mM imidazole) before elution with Elution buffer (20 mM Tris-HCl, 200 mM NaCl, 500 mM imidazole). Eluted protein samples were filtered and injected into an autosampler-equipped Akta pure system on a Superdex S75 Increase 10/300 GL column at room temperature, using SEC running buffer (20 mM Tris-HCl, 100 mM NaCl, pH 8). Monodisperse peak fractions were pooled, concentrated using Spin filters (3 kDa molecular weight cutoff, Amicon, Millipore Sigma), and stored at 4°C before downstream characterizations. Protein concentrations were determined by absorbance at 280 nm using a NanoDrop spectrophotometer (Thermo Scientific) using the molecular weights and extinction coefficients obtained from their amino acid sequences using the ProtParam tool.

Bio-layer Inferometry (BLI) Binding Experiments

BLI experiments were performed on an Octet Red96 (ForteBio) instrument, with streptavidin coated tips (Sartorius Item no. 18-5019). Buffer comprised 1X HBS-EP+ buffer (Cytiva

BR100669) supplemented with 0.1% w/v bovine serum albumin. Tips were pre-incubated in the buffer for at least 10 minutes before use. Tips were then sequentially incubated in biotinylated toxin target, buffer, designed binder, and buffer.

Affinity measurements by surface plasmon resonance (SPR)

SPR experiments were conducted using a Biacore™ 8K instrument (Cytiva) and analyzed with the accompanying evaluation software. Biotinylated α-cobratoxin was immobilized on a streptavidin sensor chip (Cytiva). For ScNtx and *Naja pallida's* cytotoxin, immobilization involved the activation of carboxymethyl groups on a dextran-coated chip through reaction with N-hydroxysuccinimide. The ligands were then covalently bonded to the chip surface via amide linkages, and excess activated carboxyls were blocked with ethanolamine (doi: 10.1007/978-1-59745-523-7_20). Increasing concentrations of protein binders were flown over the chip in 1X HBS-EP+ buffer (Cytiva BR100669).

Circular dichroism (CD)

Secondary structure content was evaluated by CD in a Jasco J-1500 CD spectrometer coupled to a Peltier system (EXOS) for temperature control. The experiments were performed on quartz cells with an optical path of 0.1 cm, covering a wavelength range from 200-260 nm. CD signal is reported as molar ellipticity [θ]. The thermal unfolding experiments were followed by a change in the ellipticity signal at 222 nm as a function of temperature. Proteins were denatured by heating the proteins at 1°C/min from 20 to 95°C.

Crystallization and Structure Determination

Crystallization experiments for the binder complex were conducted using the sitting drop vapor diffusion method. Crystallization trials were setup in 200 nL drops using 96-well format by Mosquito LCP from SPT Labtech. Crystals drops were imaged using the UVEX crystal plate hotel system by JANSi. Diffraction quality crystals for LNG binder-complex appeared in 1.5 M Ammonium Sulfate and 25% (v/v) glycerol in 2 weeks. Diffraction quality crystals for SHRT binder appeared in 0.08 M Sodium acetate trihydrate pH 4.6, 1.6 M Ammonium sulfate and 20% (v/v) glycerol. For CYTX_B10-complex diffraction quality crystals appeared in 0.1 M MES pH

6, 0.01 M Zinc chloride, 20% (w/v) PEG 6000 and 10 % (v/v) Ethylene glycol. Crystals were flash-cooled in liquid nitrogen before shipping to the synchrotron for diffraction experiment.

Diffraction data were collected at the NSLS2 beamline AMX (17-ID-1). X-ray intensities and data reduction were evaluated and integrated by XDS and merged/scaled by Pointless/Aimless in the CCP4i2 program suite. Structure determined by molecular replacement using a designed model using Phaser. Following molecular replacement model was improved and refined by Phenix. Model building was performed by COOT in between refinement cycles. Final model was evaluated by MolProbity. Data collection and refinement statistics were reported in Table 1. Final atomic coordinates, mmCIF and structure factors were deposited in the Protein Data Bank (PDB) with accession codes 9BK5, 9BK6 and 9BK7.

	LNG binder (holo) (PDB ID: 9BK5)	**B10_CYTX binder (holo) (PDB ID: 9BK6)**	**SHRT_binder (apo) (PDB ID: 9BK7)**
Resolution range	34.06 - 2.68 (2.85 - 2.68)	33.17 - 2.00 (2.05 - 2.00)	32.17 - 2.58 (2.84 - 2.58)
Space group	I 4_1 2 2	P 2_1 2_1 2_1	I 4_1 2 2
Unit cell	77.79, 77.79, 173.52; 90, 90, 90	34.56, 63.66, 77.72; 90, 90, 90	75.33, 75.33, 108.34; 90, 90, 90
Unique reflections	9109 (1483)	12102 (864)	5179 (1262)
Multiplicity	24.3 (25.8)	6.4 (6.2)	24.6 (25.2)
Completeness (%)	99.86 (99.53)	99.7 (99.7)	99.83 (99.92)
Mean I/sigma(I)	17.12 (1.08)	10.4 (2.6)	13.30 (4.13)
Wilson B-factor	82.46	33.64	60.21

R-merge	0.105 (3.172)	0.088 (0.599)	0.246 (1.069)
R-pim	0.022 (0.631)	0.041 (0.281)	0.051 (0.215)
$CC_{1/2}$	0.999 (0.522)	0.994 (0.929)	0.999 (0.982)
Reflections used in refinement	7836 (1264)	12047 (2928)	5179 (1262)
R-work	0.2387 (0.3049)	0.2496 (0.3301)	0.1970 (0.2954)
R-free	0.2681 (0.3349)	0.2850 (0.4167)	0.2235 (0.3560)
Number of non-hydrogen atoms	1118	1298	739
macromolecules	1118	1241	739
solvent	0	57	0
Protein residues	146	161	102
RMS(bonds)	0.003	0.002	0.004
RMS(angles)	0.55	0.45	0.62
Ramachandran favored (%)	92.25	96.82	97.00
Ramachandran allowed (%)	7.75	2.55	3.00
Ramachandran outliers (%)	0.0	0.64	0.00

Average B-factor	101.38	48.62	67.50
macromolecules	101.38	48.61	67.50
solvent	n/a	48.81	n/a

Table 1. Data collection and refinement statistics.

The highest-resolution shell are shown in parentheses.

***In vitro* neutralization using electrophysiology**

Human-derived Rhabdomyosarcoma RD cells (American Type Culture Collection, ATCC), endogenously expressing the muscle-type nAChR were used for electrophysiology experiments. Planar whole-cell patch-clamp recordings were conducted on a QUBE automated electrophysiology platform (Sophion Bioscience), with 384-channel patch chips (patch hole resistance 2.00 ± 0.02 MΩ), following the protocol detailed in. Protein binders were preincubated with approximately 1 IC_{80} of α-cobratoxin or ScNtx at various toxin to binder molar ratios (1:1, 1:3, 1:9, and 1:27) and then added to the cells. The toxin's ability to inhibit an acetylcholine (ACh, 70 μM) response, in the presence or absence of binders was normalized to the full ACh response and averaged within each group (n=16), and represented in a non-cumulative concentration-response plot. Data analysis was performed using Sophion Analyzer v6.6.70 (Sophion Bioscience) and GraphPad Prism v10.1.1 (GraphPad Software).Data analysis was performed using Sophion Analyzer v6.6.70 (Sophion Bioscience) and GraphPad Prism v10.1.1 (GraphPad Software).

Initial neutralization screening of whole venoms using cell viability assay

HEK293T cells were cultured in DMEM (Dulbecco's Modified Eagle Medium, Gibco) medium with 10% fetal bovine serum at 37 °C, and 5% CO_2. Cells were subjected to commercial whole venoms from *Naja pallida* (34 μg/mL) and *Naja nigricollis* (42 μg/mL) either in the absence or presence of 1:1 or 5:1 molar ratio of toxin:binder. Buffer and binder-only controls were run in

parallel and all samples were pre-incubated for 30 minutes at room temperature before addition to the HEK293T cells. To determine the percentage of viable cells, the RealTime-Glo™ MT Cell Viability Assay (Promega, Madison, WI, USA) was performed according to the manufacturer's protocol. Experiments were performed in triplicates, and results were expressed as mean ± SD.

***In vitro* neutralization of whole venoms using cell viability assay**

N/TERT immortalized keratinocytes were cultured as previously described. After determining the IC_{50} for seven venoms of Afronaja snakes, N/TERT cells were subjected to 2X the IC_{50} of each venom either in the absence or presence of a 1:5 molar ratio of venom:binder. Buffer and binder-only controls were run in parallel and all samples were pre-incubated (30 min at 37 °C) before addition to the N/TERT cells. To determine the percentage of viable cells, the CellTiter-Glo luminescent cell viability assay (Promega, Madison, WI, USA) was performed according to the manufacturer's protocol. Experiments were performed in triplicates, and results were expressed as mean ± SD.

LD_{50} determinations for α-neurotoxins

All assays used male NSA mice (20-30 g) and all doses were mass adjusted. Toxins assayed were α-cobratoxin (7820 Da, from *Naja kaouthia* venom obtained from Latoxan SAS, France) and short-chain neurotoxin ScNtx (8944 Da, recombinantly expressed). Toxins were solubilized in PBS at 1.0 mg/mL and then diluted in PBS as needed. For toxin LD_{50} determinations, five doses with 3 mice/dose were used, and a 100 µL bolus was injected IP in the right lower abdominal region; controls received only PBS. Injected mice were observed for the first 2 hours and then again at 24 hours; LD_{50} values were calculated using the "Quest Graph™ LD_{50} Calculator" (AAT Bioquest, Inc.; https://www.aatbio.com/tools/ld50-calculator).

***In vivo* neurotoxicity protein binder protection assays**

In pre-incubation experiments, three LD_{50}s of the toxins (α-cobratoxin: 0.294 μg/g mouse; ScNtx: 0.261 μg/g mouse) were mixed with a 10-fold molar excess of their respective protein binders in PBS and incubated at room temperature for 30 minutes before intraperitoneal (i.p.) administration. Groups of five mice were injected with the binder:toxin mixture and observed at 2 and 24 hours. In rescue experiments, toxins (3 LD_{50}s) were administered i.p. 15 or 30 minutes before the corresponding binder, given i.p. at either 10- or 5-fold molar excess to groups of five mice. Protection against lethality was measured as percent mortality at 24 hours.

***In vivo* dermonecrosis protein binder protection assays**

Animal experiments were conducted under protocols approved by the Animal Welfare and Ethical Review Boards of the Liverpool School of Tropical Medicine and the University of Liverpool, and under project licence P58464F90 approved by the UK Home Office in accordance with the UK Animal (Scientific Procedures) Act 1986.

CD1 male mice (Charles River, 18-20 g) were acclimated for one week before experimentation in specific pathogen-free conditions. Holding room conditions were 23°C with 45-65% humidity and 12/12 hour light cycles (350 lux). Mice were housed in Techniplast GM500 cages (floor area 501 cm2) containing 120 g Lignocell wood fibre bedding (JRS, Germany), Z-nest biodegradable paper-based material for nesting and environmental enrichment (red house, clear polycarbonate tunnel and loft). Mice had *ad lib* access to irradiated PicoLab food (Lab Diet, USA) and reverse osmosis water in an automatic water system. Animals were split into cages (experimental units) upon arrival and no further randomisation was performed.

All mice were pre-treated with 5 mg/kg morphine (injected subcutaneously) before receiving intradermal injections in a 100 μL volume into the ventral abdominal region (rear side flank region). A venom-only control group of 5 mice received 63 μg of *Naja nigricollis* (Tanzania) venom (dissolved in PBS). For protection assays, crude venom was pre-incubated (30 min at 37 °C) with varying cytotoxin:binder ratios of 1:1, 1:2.5 and 1:5 prior to injection (n=3) (ratios estimated from the proportion of cytotoxin in the venom). Prior to this, a control group (N=3) received injections of cytotoxin binder alone (278 μM, equivalent to the 1:5 cytotoxin:binder dose) to check tolerance of the cytotoxin binder. For sample size, N of 3 was used for groups receiving cytotoxin binder as this was a pilot experiment. N of 5 was used for the venom-only

control group due to variation in lesion size, and this is the size recommended by WHO. In total 17 mice were used. No inclusion or exclusion criteria were used during the experiment, and all data points were used in the analysis. No strategy was used to control for confounders. All experimenters were aware of the group allocation during the experiment and analysis.

After 72 hours, mice were euthanized by rising concentrations of CO_2 and the lesions were excised. The outcome measured was the lesion size. Photographs of lesions were taken using a digital camera immediately after excision and the severity and size of dermonecrotic lesions was determined using VIDAL.

Supplementary Materials

Target	**Binder ID**	**Binder amino acid sequence**
ScNtx	SHRT	PKTVVVRLSPSMNEEQAA EIGREAGKAALAAGDRLV FVGPADQSYAAMKAAME AGLPEVTMYALDFSDAES ALKAAEVAEDEGDEEVAE VAREIAEEIKA
α-cobratoxin	LNG	SPAEREAGRIVVRGDVAIA EAVVRKVGEVAGKEVILLI SYRKNGEWITYQRNLEAT PEDVERTIAVIREIYEESGG DFILAIFSDPEVGAAGRAV AAAAA
Consensus cytotoxin	CYTX	SAEEKLKELEKKIEEFYKS AESVTSSISESNGTLTVTT TLKIGDATVTATETYKIIN

		GSPYKDGEVTYSASAEGK ALADELAKLDKEYREAQ EKVEEEKK
Consensus cytotoxin	CYTX_B10	TPEERLAQLEKEIQALYDA ADEVVDEVEEKDGKMTV TRTLTIGDGTVTLVETLKI VDGAPVKDGEIEVICNPE CEELGKRLKALAKEYEKA QEEVEKAKA

Supplementary Table 1. Amino acid sequences of 3FTx binding proteins.

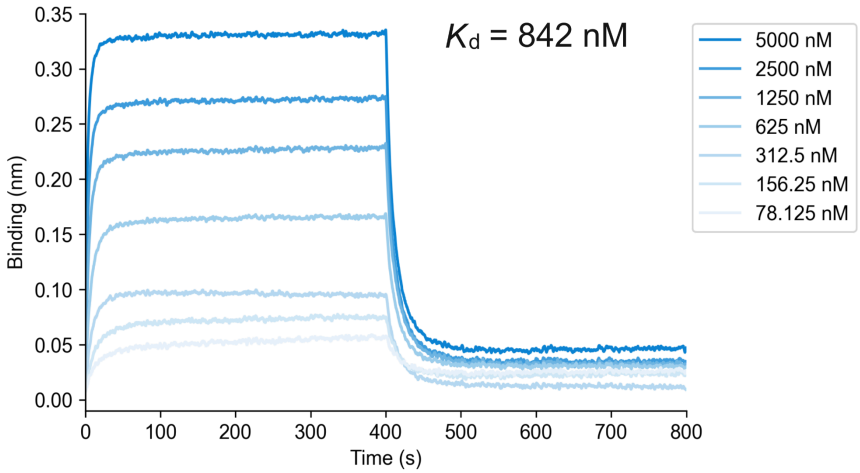


Figure S1. BLI traces of initial hit for ScNtx. Measurements indicate 842 nM binding affinity to ScNTx.

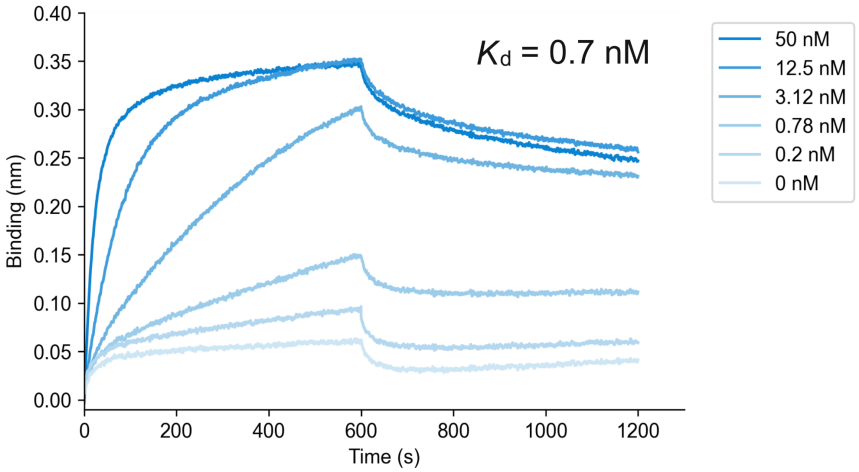


Figure S2. BLI traces of SHRT binder. Measurements indicate 0.7 nM binding affinity to ScNTx.

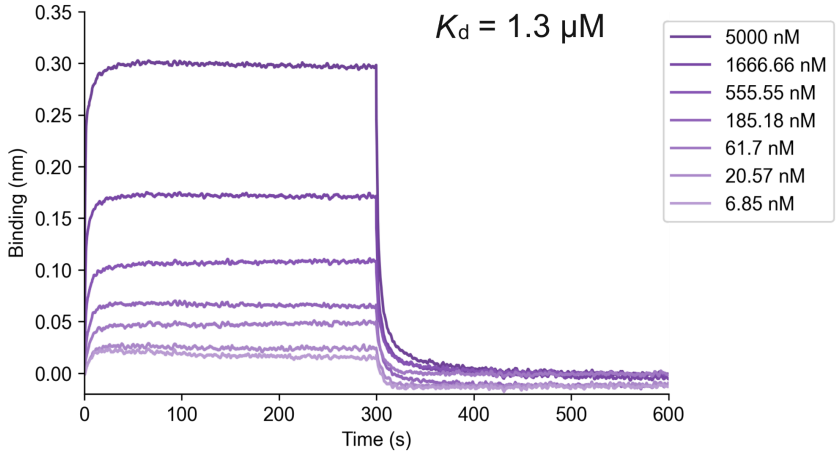


Figure S3. BLI traces of initial hit for α-cobratoxin. Measurements indicate 1.3 µM binding affinity to α-cobratoxin.

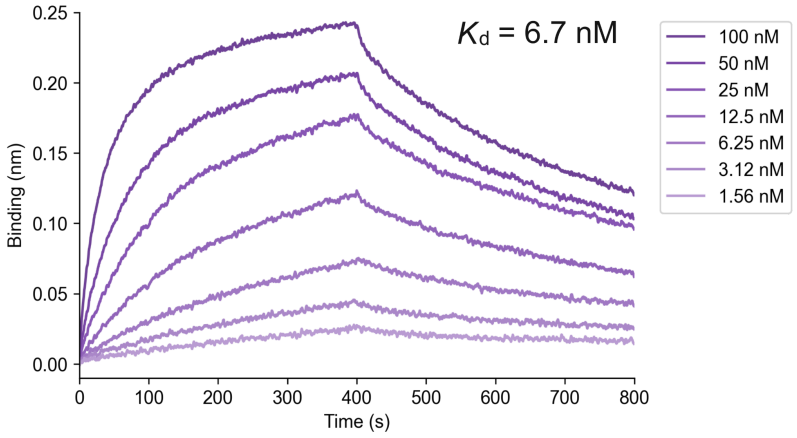


Figure S4. BLI traces of LNG binder. Measurements indicate 6.7 nM binding affinity to α-cobratoxin.

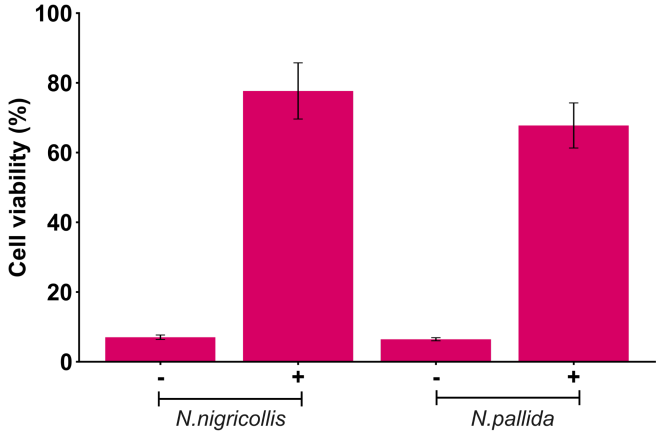


Figure S5. Initial screening of whole venoms for CYTX binder in HEK293T cells. Neutralization of the cytolytic effects of whole venoms from two different *Naja* species by the CYTX binder. Two IC_{50}s of the whole venoms were pre-incubated with the binder at a 1:5 molar ratio (toxin:binder).This ratio was estimated assuming 70% of the whole venom consists of cytotoxins based on prior proteomic analyses. Experiments were performed in triplicates, and results are expressed as mean ± SD.

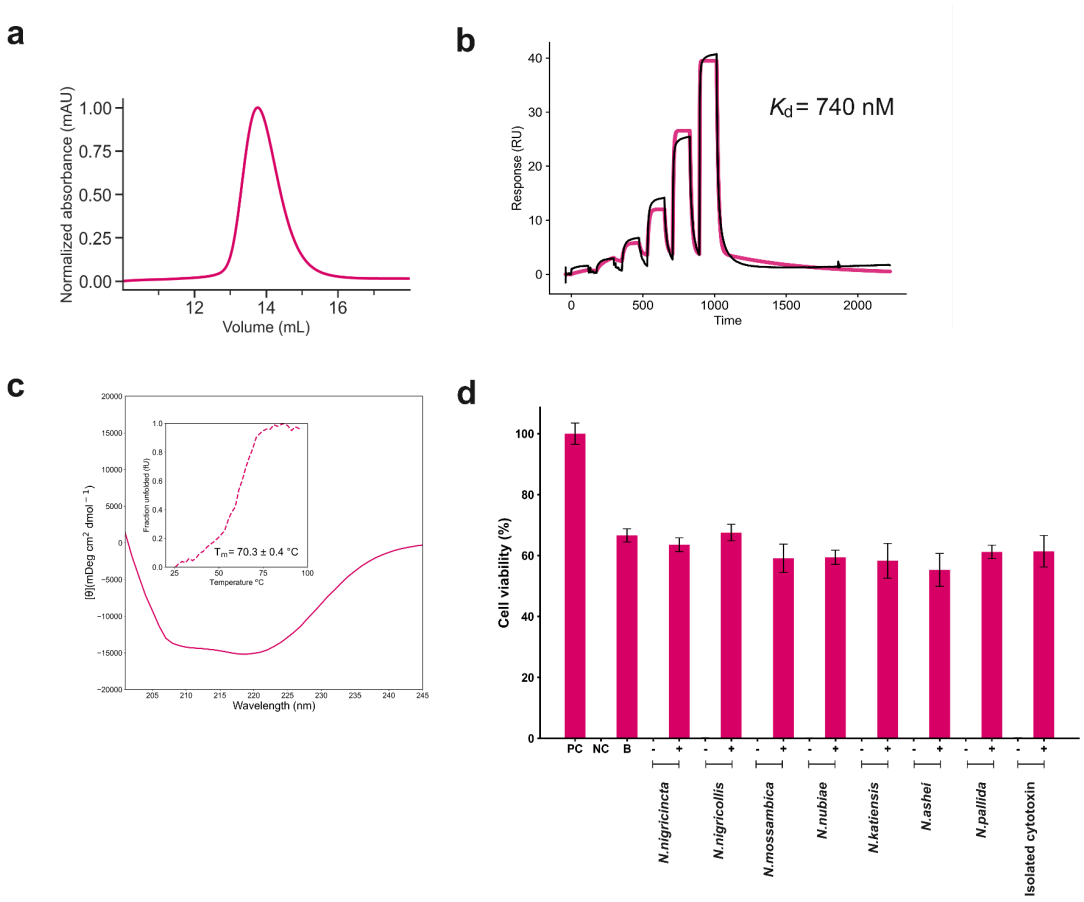


Figure S6. Experimental characterization of CYTX_B10 binder. (a) SEC trace of purified protein. **(b)** SPR binding affinity measurement. Colored solid line represent the fit using the heterogeneous ligand model, with dissociation constant (K_d) value derived from this fit**. (c)** CD data confirms the presence of αβ-secondary structure in CYTX_B10 and its thermal stability (inset). **(d)** Neutralization of the cytolytic effects of whole venoms from seven different *Naja* species and isolated cytotoxin by the CYTX_B10 binder. Two IC_{50}s of the whole venoms or toxin were pre-incubated with CYTX_B10 at a 1:5 molar ratio (toxin:binder). This ratio was estimated assuming 70% of the whole venom consists of cytotoxins based on prior proteomic analyses. Keratinocyte media was used as a positive control (PC). Triton X-100 was used as a negative control (NC). CYTX_B10 binder (B) was used as a positive control. (-) denotes two IC_{50}s of the whole venoms without binder, and (+) denotes venoms incubated with binder. Experiments were performed in triplicates, and results are expressed as mean ± SD.

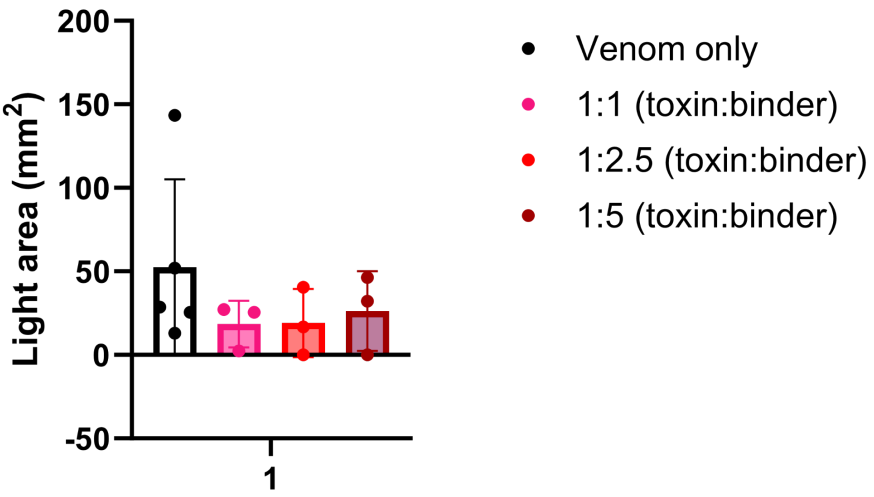


Figure S7. Histopathological analysis of light lesions induced by West African *N. nigricollis* whole venom by VIDAL. Lesions induced by the whole venom, preincubated with varying concentrations of CYTX binder, do not exhibit a significant reduction in size. An outlier within the venom-only group is observed, likely attributable to biological variation.

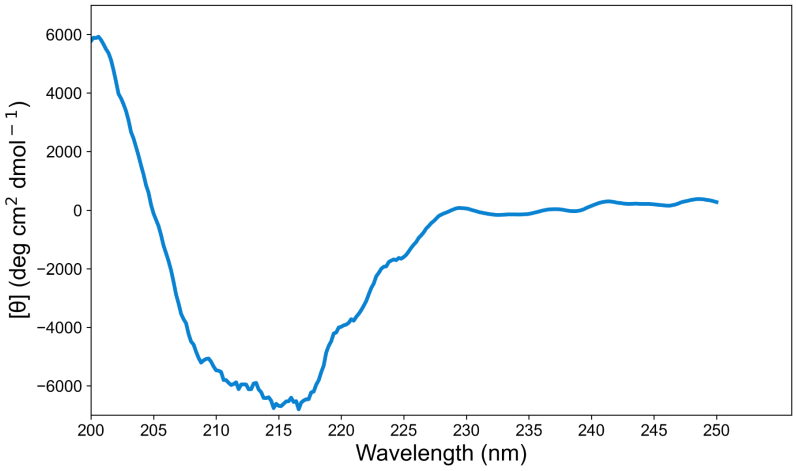


Figure S8. CD spectra of ScNtx. CD data confirms the presence of β-secondary structure of the recombinantly expressed toxin.

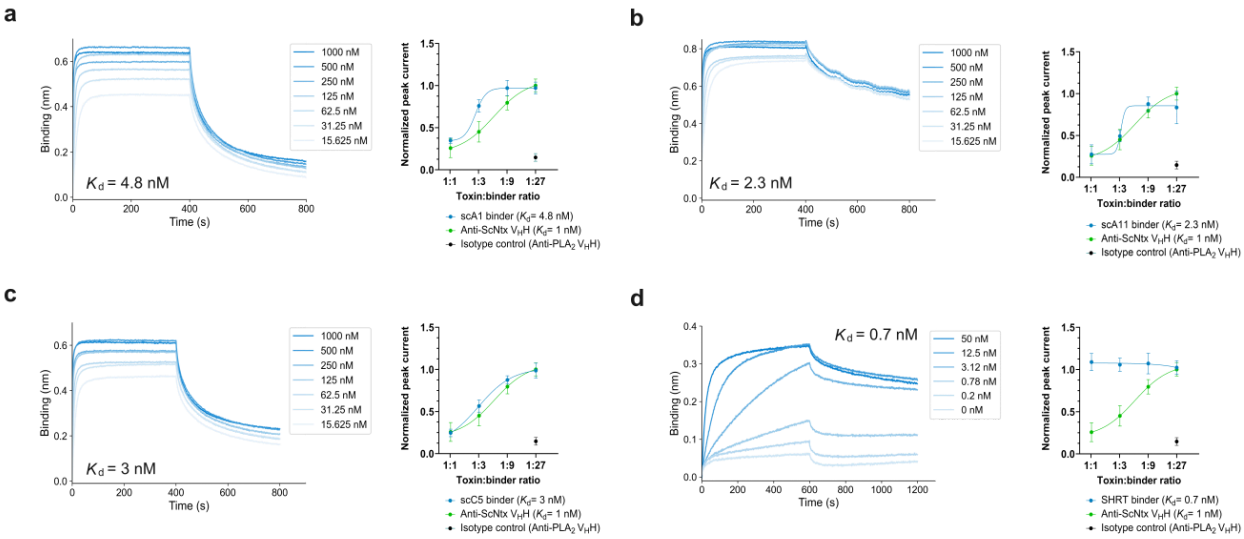


Figure S9. BLI traces and patch-clamp functional neutralization of top ScNtx binding proteins. (a) scA1 binder. Left: BLI measurements indicate 4.8 nM binding affinity to ScNTx. Right: concentration-response curves comparing the efficacy of scA1 binder and anti-ScNtx V_HH in preventing nAChR blocking by one IC_{80} of ScNtx. **(b)** scA11 binder. Left: BLI measurements indicate 2.3 nM binding affinity to ScNTx. Right: concentration-response curves comparing the efficacy of scA11 binder and anti-ScNtx V_HH in preventing nAChR blocking by one IC_{80} of ScNtx. **(c)** scC5 binder. Left: BLI measurements indicate 3 nM binding affinity to ScNTx. Right: concentration-response curves comparing the efficacy of scC5 binder and anti-ScNtx V_HH in preventing nAChR blocking by one IC_{80} of ScNtx. **(d)** SHRT binder. Left: BLI measurements indicate 0.7 nM binding affinity to ScNTx. Right: concentration-response curves comparing the efficacy of SHRT binder and anti-ScNtx V_HH in preventing nAChR blocking by one IC_{80} of ScNtx.

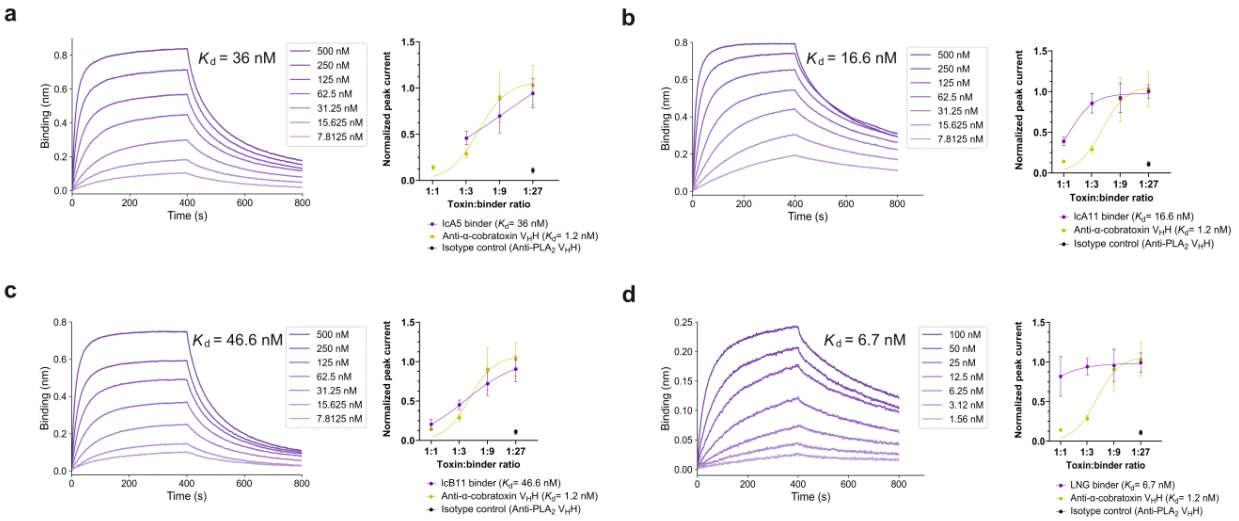


Figure S10. BLI traces and patch-clamp functional neutralization of top α-cobratoxin binding proteins. (a) lcA5 binder. Left: BLI measurements indicate 36 nM binding affinity to α-cobratoxin. Right:concentration-response curves comparing the efficacy of lcA5 binder and anti-α-cobratoxin V_HH in preventing nAChR blocking by one IC_{80} of α-cobratoxin. **(b)** lcA11 binder. Left: BLI measurements indicate 16.6 nM binding affinity to α-cobratoxin. Right: concentration-response curves comparing the efficacy of lcA11 binder and anti-α-cobratoxin V_HH in preventing nAChR blocking by one IC_{80} of α-cobratoxin. **(c)** lcB11 binder. Left: BLI measurements indicate 46.6 nM binding affinity to α-cobratoxin. Right: concentration-response curves comparing the efficacy of lcB11 binder and anti-α-cobratoxin V_HH in preventing nAChR blocking by one IC_{80} of α-cobratoxin. **(d)** LNG binder. Left: BLI measurements indicate 6.7 nM binding affinity to α-cobratoxin. Right: concentration-response curves comparing the efficacy of LNG binder and anti-α-cobratoxin V_HH in preventing nAChR blocking by one IC_{80} of α-cobratoxin.

Conclusions

The computational methods described in Chapter 1 demonstrate the potential of new deep learning techniques in designing *de novo* protein binders for bioactive peptides. Both traditional Rosetta and deep learning methods yielded groove-shaped scaffolds with high shape complementarity to peptide targets. The RF_{joint} and RFdiffusion approaches significantly improved initial Rosetta designs, while Hallucination yielded high-affinity binders without pre-specified structures. Notably, RFdiffusion achieved picomolar affinity through purely computational design, without experimental optimization. The designed proteins offer advantages over traditional antibodies, including greater stability, reduced production costs, and easier integration into sensors and affinity matrices, making them promising for serological detection and advanced diagnostic tools for disease-associated peptide biomarkers.

Chapter 2 extends RFdiffusion to design protein binders for α-neurotoxins and cytotoxins from the 3FTx snake venom toxin family. This chapter illustrates how computational methods can revolutionize snakebite treatment by rapidly identifying high-affinity binders with therapeutic potential. The designed proteins can be cost-effectively manufactured through microbial fermentation and exhibit improved thermostability. This *in silico* approach avoids the need for animal immunization and extensive library screening, providing a low-cost and rapid method for developing toxin neutralizing proteins when structural or sequence data is available.

Overall, advancements in protein design through deep learning methods hold significant promise for creating accessible diagnostic tools and democratizing drug design. This approach could enable researchers, especially in low and middle-income countries, to improve early disease diagnosis and address neglected tropical diseases such as snakebite envenoming.