# BindCraft: one-shot design of functional protein binders

Martin Pacesa[1]‡, Lennart Nickel[1]‡, Joseph Schmidt[1], Ekaterina Pyatova[1], Christian Schellhaas[1], Lucas Kissling[2], Ana Alcaraz-Serna[3], Yehlin Cho[4], Kourosh H. Ghamary[5], Laura Vinué[5], Brahm J. Yachnin[5], Andrew M. Wollacott[5], Stephen Buckley[1], Sandrine Georgeon[1], Casper A. Goverde[1], Georgios N. Hatzopoulos[6], Pierre Gönczy[6], Yannick D. Muller[3], Gerald Schwank[2], Sergey Ovchinnikov[4]*, Bruno E. Correia[1]*

[1] Laboratory of Protein Design and Immunoengineering, École Polytechnique Fédérale de Lausanne and Swiss Institute of Bioinformatics; Lausanne, Switzerland

[2] Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

[3] Division of Immunology and Allergy, Lausanne University Hospital and University of Lausanne; Lausanne, Switzerland

[4] Massachusetts Institute of Technology; Cambridge, MA, USA

[5] Visterra Inc., Waltham, MA, USA

[6] Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland.

‡ These authors contributed equally

* Corresponding authors. Email: bruno.correia@epfl.ch, so3@mit.edu

**Abstract:**

Protein–protein interactions (PPIs) are at the core of all key biological processes. However, the complexity of the structural features that determine PPIs makes their design challenging. We present BindCraft, an open-source and automated pipeline for *de novo* protein binder design with experimental success rates of 10-100%. BindCraft leverages the trained deep learning weights of AlphaFold2[1] to generate nanomolar binders without the need for high-throughput screening or experimental optimization, even in the absence of known binding sites. We successfully designed binders against a diverse set of challenging targets, including cell-surface receptors, common allergens, *de novo* designed proteins, and multi-domain nucleases, such as CRISPR-Cas9. We showcase their functional and therapeutic potential by demonstrating that designed binders can reduce IgE binding to birch allergen in patient-derived samples. This work represents a significant advancement towards a “one design-one binder” approach in computational design, with immense potential in therapeutics, diagnostics, and biotechnology.

## Introduction

Proteins are versatile biomolecules capable of mediating a diverse range of biological functions, including catalysis, molecular recognition, structural support and others. However, proteins rarely perform their biological functions in isolation but rather rely on protein–protein interactions (PPIs) to execute complex biological processes, such as signal transduction, antibody-mediated immunity, cellular communication, etc.

Designing protein binders that can specifically target and regulate PPIs holds immense therapeutic and biotechnological potential. Such binders can be utilized to modulate protein interaction networks and signaling pathways, design therapeutic antibodies, inhibit pathogenic agents, or create biotechnological tools for research and industry. Traditional methods for generating protein binders, such as immunization, antibody library screening, or directed evolution of binding scaffolds are, however, often laborious, time-consuming, and provide limited control over the target epitope.

Computational protein design provides a powerful alternative, where binders can be designed and tailored to a specific protein target and epitope, enabling the exploration of a much broader sequence and structure space. For example, *de novo* designed protein binders have been previously used to block viral entry[2], modulate immune and inflammatory response[3,4], prevent amyloid assembly[5], or control cell differentiation pathways[6].

Physics-based methods like Rosetta have been instrumental in early binder designs through scaffolding and sidechain optimization[7–9]. However, such methods suffer from very low experimental success rates (typically less than 0.1%) and require the generation and sampling of a vast number of designs, ranging from hundreds of thousands to millions[7,9–11]. Moreover, because such methods typically require the docking of predefined scaffolds onto a fixed target structure, incompatibilities between the target and binder surfaces can result in suboptimal binding interactions or even preclude the targeting of certain epitopes.

Recent breakthroughs in deep learning have revolutionized the field of biomolecular modelling, particularly the prediction of protein structure. Models like AlphaFold2 (AF2)[1] and RoseTTAFold2 (RF2)[12], trained on large protein structure and sequence datasets, have demonstrated remarkable capabilities in accurately predicting protein structures and modeling complex PPIs. Indeed, AF2 filtering has been shown to significantly increase the success rates of binder design by evaluating the plausibility of predicted complexes[10,11]. Deep learning has also been successfully applied for *de novo* design of proteins and binders. The current state-of-the-art methods involve the use of RFdiffusion[10] for plausible backbone generation coupled with ProteinMPNN sequence generation[13,14]. When applied to binder design, this approach has successfully designed binders against a variety of protein targets, with improved success rates compared to previous methods[10]. However, this requires the generation of thousands to tens of thousands designs *in silico* and extensive experimental *in vitro* screening to identify suitable binders, which remains a significant limitation for most research groups.

Given the utility of AF2 in improving binder filtering success, we hypothesized that we could harness its trained weights and learned patterns of protein structures directly for the design of protein binders. We present BindCraft, a user-friendly pipeline for d*e novo* design of protein binders that requires minimal user intervention and computational expertise. BindCraft leverages backpropagation through the AF2 network to efficiently hallucinate novel binders and interfaces without the need for extensive sampling (**Fig. 1a**). We demonstrate the efficiency of our pipeline on ten diverse, challenging, and therapeutically relevant protein targets (**Fig. 1b**) and identify several high affinity binders without the need for high throughput screening of hundreds to thousands of designs experimentally. This marks an important advancement in the design of protein binders on demand, a long standing problem in protein design; furthermore, this makes it accessible to research groups without means for high throughput screening facilities or large computational resources.

## Results

### Deep learning-based design of *de novo* binders

Our goal was to create an accessible, efficient, and automated pipeline that leverages the capabilities of AF2 models for accurate binder design, reducing the need for large-scale sampling. To this end, the user inputs a structure of the target protein, the size range of binders, and the final number of desired designs. Target hotspots can be specified, or the AF2 network can automatically select an optimal binding site by identifying the one that satisfies the loss function most effectively. We utilize the ColabDesign implementation of AF2 to backpropagate hallucinated binder sequences through the trained AF2 weights and calculate an error gradient. This error gradient is used to update and optimize the binder sequence to fit specified loss functions and design criteria. These include the AF2 confidence scores, intra- and intermolecular contacts, helical content, and radius of gyration. By iterating over the network, we can generate high quality designs, essentially enabling the generation of binder structure, sequence and interface concurrently (**Fig. 1a**). Unlike other design methods, such as RFdiffusion[10] or RIFdock[7,11], which keep the target backbone fixed during design, BindCraft co-folds both the target and binder at every iteration. This allows for defined levels of flexibility on the side-chain and backbone level for both binder and target, resulting in backbones and interfaces that are molded to the target binding site.

We utilize AF2 multimer[15] for designing initial binders, as this version of AF2 was trained on protein complexes and would likely be able to more accurately model PPIs. We utilize all 5 trained model weights of AF2 multimer to avoid overfitting of sequences to a single model. However, we and others[14,16] have previously demonstrated that AF2-hallucinated proteins can exhibit low levels of expression when tested experimentally. We therefore subsequently optimize the sequence of the binder core and surface using $MPNN_{sol}$[13,16] while keeping the interface intact (**Fig. 1a**). The optimized binder sequences are repredicted using the AF2 monomer model[1]. This model was exclusively trained on

monomeric proteins, which minimizes prediction bias of PPIs and enables robust filtering for high quality interfaces. Lastly, we filter the predicted designs based on AF2 confidence metrics, as well as Rosetta physics-based scoring metrics, as deep learning models have been shown to sporadically produce physically improbable results[1,15].

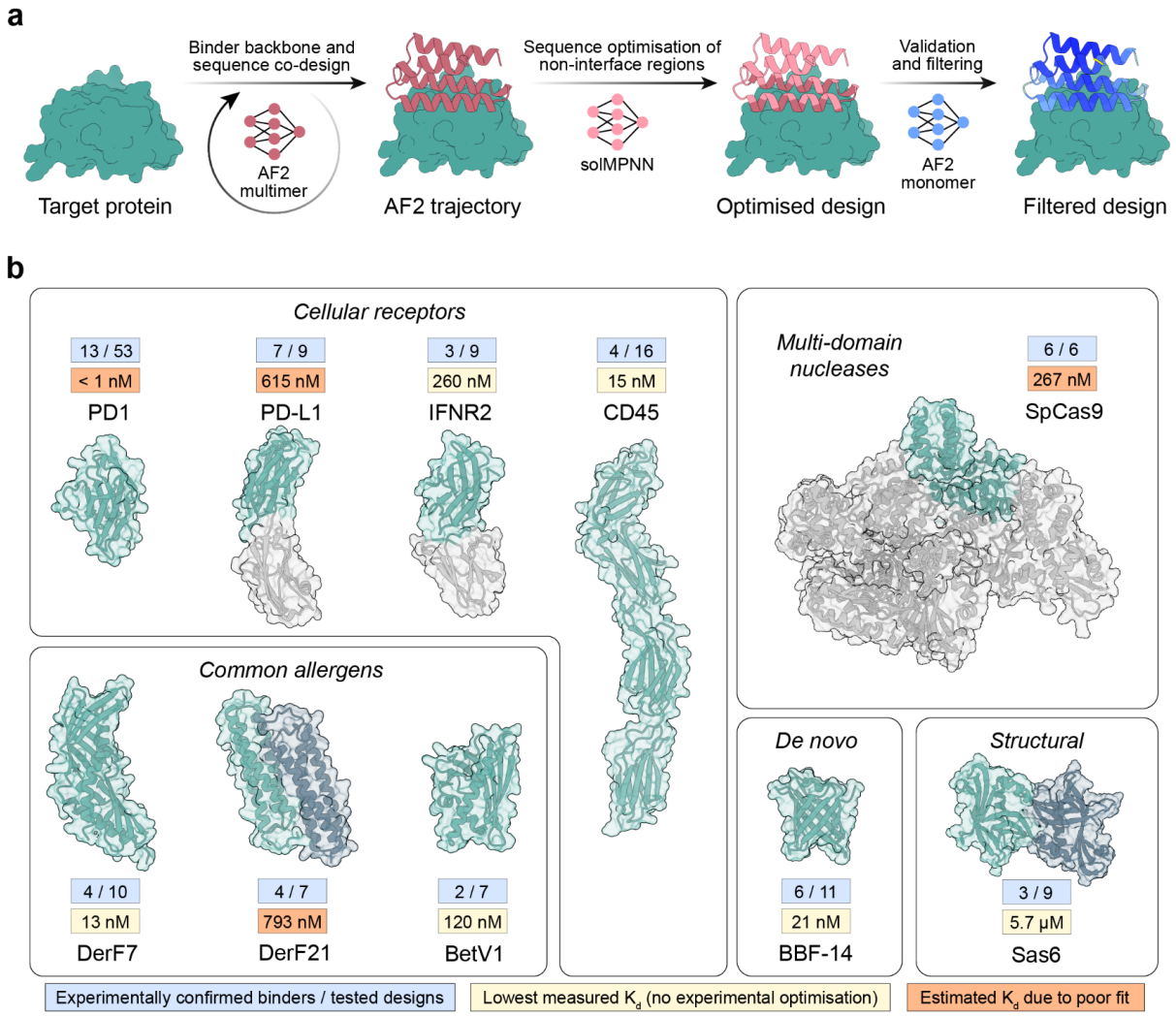


**Figure 1 | *De novo* binder design using BindCraft. a,** Schematic representation of the BindCraft binder design pipeline. Given a target protein structure, a binder backbone and sequence is generated using AF2 multimer, then the surface and core are optimized using $MPNN_{sol}$ while keeping the interface intact, and finally designs are filtered using the AF2 monomer model. **b,** Overview of protein targets for binder design. Parts of the model colored in green were considered during design, gray areas were excluded. Values in the blue box indicate the number of successful designs, where binding was observed on SPR measurement versus the total number of designs tested. Values in the yellow box indicate the measured $K_d$ of the highest affinity binder without experimental sequence optimization, while values in orange boxes indicate estimated $K_d$ values due to poor fit and will be re-measured.

All of the described steps are automated into a single workflow, where designs are filtered automatically, statistics are stored in a user-friendly format, and settings have been optimized to ensure the design procedure is generalizable across different targets.

We hope that by minimizing human intervention needed to generate and sort high quality binder designs, BindCraft will help to democratize protein binder design and make it accessible to a broader scientific community for direct application.

**Targeting cell surface receptors**

To test the performance of our pipeline, we designed binders against therapeutically-relevant cell surface receptors and tested them for binding activity *in vitro.* Receptors are ideal targets for binder design due to the presence of well-characterized binding sites exposed on the extracellular domain, either through interactions with endogenous binding partners or therapeutic antibodies. We first generated designs that could bind the human PD-1 protein, a key immune checkpoint receptor expressed primarily on the surface of T cells[17]. We initially purified and screened 53 designs for binding using bio-layer interferometry (BLI) in a bivalent Fc-fusion format. We observed a binding signal for 13 binders, with the best binder displaying an apparent dissociation constant ($K_d$) lower than 1 nM (**Fig. 2a**), although the exact $K_d$ could not be determined due to the extremely slow dissociation rate and avidity effect from the Fc-fusion. To confirm the binding site, we performed a competition assay with the well characterized anti-PD-1 monoclonal antibody, pembrolizumab, which should engage the same binding site. Indeed, our binder could not outcompete the antibody binding ($K_d$ = 27 pM), indicating it is targeting overlapping epitopes (**Fig. 2a**).

Encouraged by these results we reduced the number of designs tested experimentally for all subsequent targets to test whether we can minimize the need for experimental screening. We next designed binders against PD-L1[17] and the interferon 2 receptor (IFNAR2)[18], both important modulators of immune signaling, where specific binders could allow us to design novel tumor or antiviral therapies. We tested 9 designs against PD-L1 out of which 7 showed binding signal, while for IFNAR2 we could detect binding for 3 out of 9 designs (**Fig. 1b**). The top performing binder4 against PD-L1 displayed a $K_d$ of 615 nM and an expected alpha-helical signature as measured by circular dichroism (CD) (**Fig. 2b**). The best performing binder5 against IFNAR2 displayed an affinity of 260 nM determined by surface plasmon resonance (SPR) and similarly a typical alpha-helical signature by CD, validating the fold integrity and the thermal stability of our designs (**Fig. 2c**). To validate their binding mode, we performed competition assays using SPR. We probed the binding of our PD-L1 binder4 using a previously characterized *de novo* binder[8] and could confirm they compete for the intended target binding site (**Fig. 2d**). Similarly, we probed the binding of IFNAR2 binder5 against the native cytokine interferon alpha 2 (IFNA2)[18]. We observe competition for the native IFNA2 binding site, validating our designed binding mode (**Fig. 2e**). These results demonstrate we are able to efficiently design binders, straight from the computational design pipeline, against known binding sites, without the need for extensive screening to identify hits with nanomolar affinity.

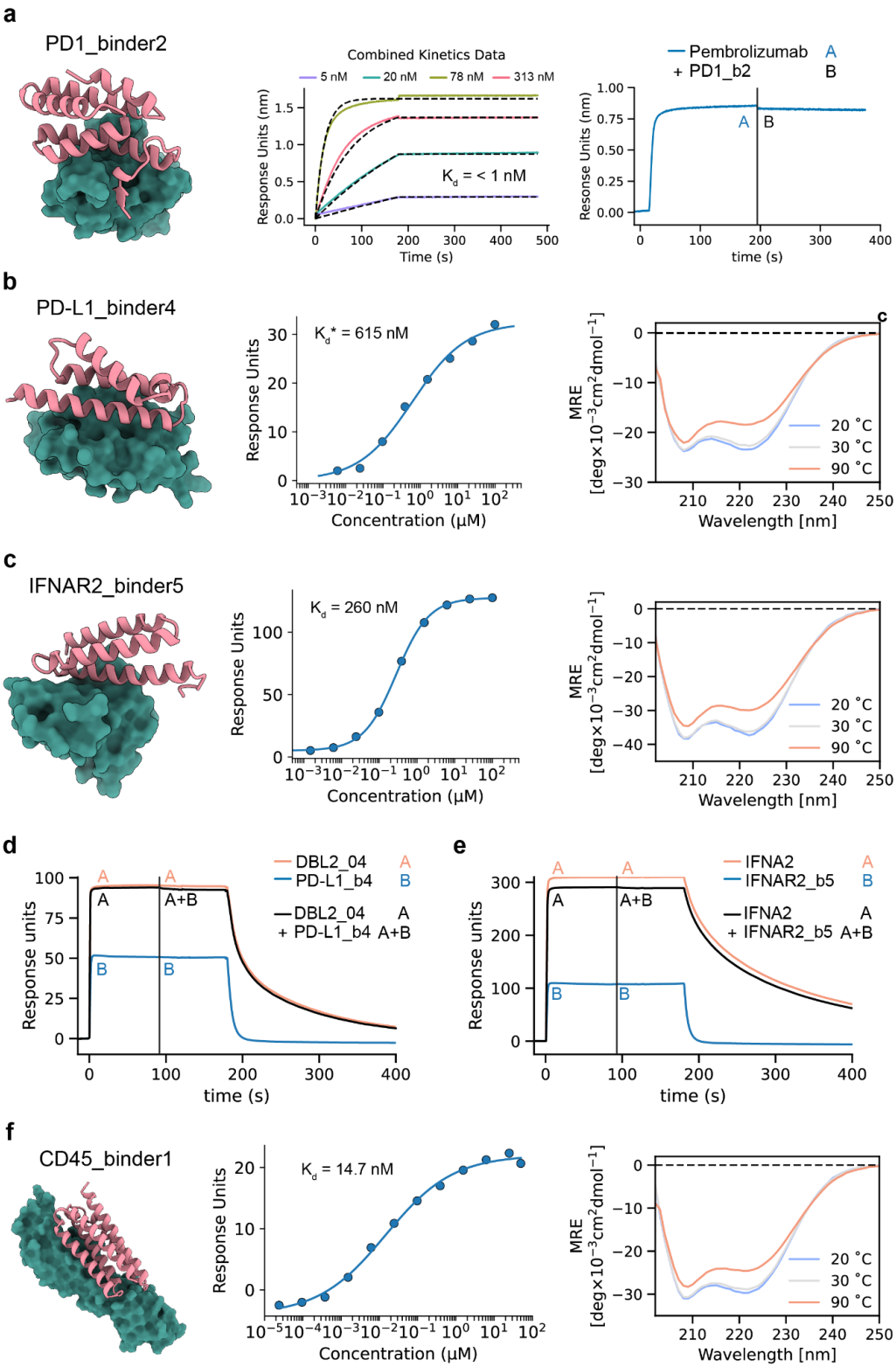


**Figure 2 | Binder design targeting cell surface receptors. a,** Left: Design model of binder2 binding to PD-1. Center: Representative BLI sensorgram displaying binding kinetics of binder2 (bivalent Fc-fusion) to PD-1. Right: Competition assay displaying that the anti-PD-1 antibody, pembrolizumab, occupies the same binding site on PD1 as binder2. **b,** Left: Design model of binder4 binding to PD-L1. Center: SPR binding affinity fit for binder4. Right: CD spectra of binder4 at different temperatures. **c,** Left: Design model of

binder5 binding to IFNAR2. Center: SPR binding affinity fit for binder5. Right: CD spectra of binder5 at different temperatures. **d,** Competition assay displaying previously published *de novo* binder DBL2_04 occupying the same binding site as PD-L1 binder4. **e,** Competition assay displaying the natural IFNA2 binding partner occupies an overlapping binding site with IFNAR2 binder5. **f,** Left: Design model of binder1 binding to CD45. Center: SPR binding affinity fit for binder1. Right: CD spectra of binder1 at different temperatures. $K_d$ values with asterisk indicate estimated $K_d$ due to suboptimal fits.

Next, we sought to determine whether our pipeline could design binders against extracellular receptors lacking well-characterized binding sites. We selected CD45 as a target, due to the structural complexity of its extracellular domain (ECD), comprising four β-sandwich domains d1-d4 with heavy N-glycosylation in the smallest isoform[19]. CD45 is a transmembrane tyrosine phosphatase involved in critical pathways of T cell function. Designing binders against the ECD of CD45 could allow us to modulate its signaling activity and fine tune T-cell activation thresholds for anti-tumor therapies. As the ECD is large, we designed binders both against individual domains and pairs of adjacent domains, while removing any binders that overlapped with known glycosylation sites. We tested 16 binders experimentally and could observe satisfactory binding signals for 4 binders on SPR (**Fig. 1b**). The best performing binder1 displayed a binding affinity of 14.7 nM and is designed to bind at the interface of domains d3 and d4 (**Fig. 2f**). We also observed the expected alpha-helical signal in CD, validating the correct folding of our design. These results indicate we can effectively design binders even against novel or previously uncharacterized binding sites.

**Targeting proteins in non-native binding sites**

To further assess the generalizability of our pipeline for targeting proteins or surfaces lacking known binding sites. First, we designed binders against proteins with no known sequence homologs in the PDB. For this we chose the completely *de novo* designed beta barrel fold BBF-14[16] as a target, additionally because beta barrels are not commonly regarded as PPI partners. We purified 11 top scoring designs and observed binding signal for 6 binders (**Fig. 1b**). The best binding binder4 (**Fig. 3a**) is composed of a mixed alpha-beta topology, where the interface is formed by both the split beta-sheets and a helix motif. Interestingly, the beta-sheet interface is not mediated by backbone hydrogen bonding but rather by sidechain interactions. The binder4 exhibited a 20.9 nM affinity to BBF-14 as determined by SPR (**Fig. 3b**). To assess the fidelity of our design procedure, we solved a 3.1 Å structure of BBF-14 bound to binder4 (**Fig. 3c**). When aligned on the BBF-14 target, the binder4 exhibits a backbone $RMSD_{C\alpha}$ of 1.7 Å, confirming both the accuracy of the fold as well as of the designed binding mode. This also underscores our ability to generate binders purely based on structural information without relying on existing binding sites or any influence of co-evolutionary data.

We additionally chose the highly conserved structural protein SAS-6 as a design target. SAS-6 assembles into higher-order oligomers and is essential for centriole biogenesis across the eukaryotic tree of life[20]. A major challenge in the study of the centriole architecture has been the lack of tools that allow precise modulation of the

assembly process. We attempted binder design against *Chlamydomonas reinhardtii* SAS-6 previously using published computational methods[8] but were unable to obtain satisfactory binders. Using BindCraft, we were able to generate several designs passing computational filters, and experimentally tested 9 top scoring designs. We identified binder4 (**Fig. 3d**) that bound with 5.7 μM affinity to the monomeric form of CrSAS-6 (**Fig. 3e**) and 4.2 μM affinity to the dimeric form (**Fig. 3f**), indicating compatibility with its oligomeric form. The binder4 targets an overlapping epitope with the previously reported monobody $MB_{CRS6}$-15, which binds to the N-terminal head domain of CrSAS-6 and causes a shift in its assembly mechanism, transforming the ring-like structure into a helical assembly[21]. We speculate that we can now design binders on-demand against challenging targets to probe their biological function, even in the context of higher order assemblies.

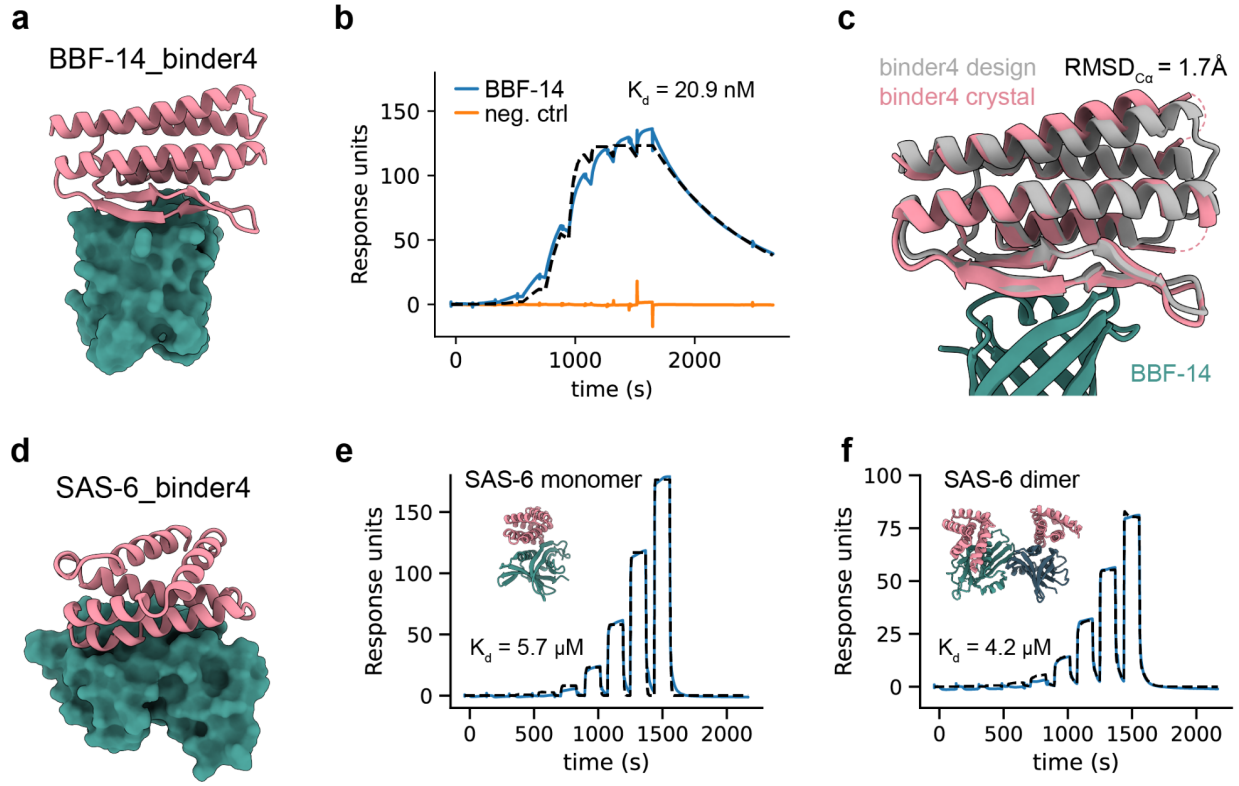


**Figure 3 | Targeting *de novo* and structural proteins in non-native sites. a,** Design model of binder4 binding to *de novo* designed beta barrel BBF-14. **b,** Single cycle kinetic analysis with SPR of binder4 binding to BBF-14. **c,** Crystal structure (colored) of the BBF-14_binder4 complex overlaid with the design model (gray). **d,** Design model of binder4 binding to the challenging structural protein target SAS-6. **e,** SPR binding traces of binder4 to CrSAS-6 monomeric form. **f,** SPR binding traces of binder4 to CrSAS-6 dimeric form.

## Blocking immunogenic epitopes of common allergens

The prevalence of allergic rhinitis has been steadily on the rise and seasonal allergies have been estimated to affect up to 50% of the population in some countries[22]. Current treatments primarily focus on reducing global inflammation with

immunosuppressants and monoclonal antibodies. However, neutralizing allergic reactions could potentially offer a more effective strategy for managing allergies. Allergens comprise a diverse group of proteins with different folds, biological functions, and highly charged surfaces[23]. Generally, hydrophobic epitopes are considered more tractable for computational binder design[7], making allergens more challenging targets.

To test the capabilities of BindCraft at targeting allergens, we designed binders against the dust mite allergens Der f7 and Der f21, and the major birch allergen Bet v1 responsible for up to 95% of birch-related allergies[24]. We examined 10 designs against Der f7 experimentally and identified 4 binders (**Fig. 1b**), with binder2 exhibiting the highest binding affinity with a $K_d$ of 12.8 nM (**Fig. 4a**). To confirm the binding mode of binder2, we solved a 2.2 Å crystal structure of it bound to Der f7 (**Fig. 4b**). When aligned on the allergen, the backbone $RMSD_{C\alpha}$ of binder2 crystal structure compared to the design model is 1.7 Å, validating the structural accuracy of the design method. Interestingly, binder2 exhibits a helical topology that wraps around connecting loops of the mixed beta sheet and helical tip of the protein (**Fig. 4b**). While no structures exist, mouse monoclonal antibodies raised against Der f7 have been shown to bind to the same epitope through mutational studies[25], indicating we are able to design binders against known immunogenic epitopes of allergens.

Similarly, we evaluated 7 binders against Der f21 and could detect binding for 4 designs on SPR (**Fig. 1b**). The best performing binder10 displayed an apparent affinity of 793 nM (**Fig. 4c**) and we attempted to solve the crystal structure to validate the design. The 2.6 Å resolution crystal structure allowed us to validate the mode of binding of binder10 against a highly charged helical epitope of Der f21, with a backbone $RMSD_{C\alpha}$ of 3.1 Å caused by an alternative rotamer conformation of an interface tyrosine (**Fig. 4d**). Similarly to Der f7, no bound structures of Der f21 are available, however mutational analysis in patient sera indicated that our binders target epitopes distinct from the IgE sera of allergic individuals[26].

Lastly, we tested 7 binders against the birch allergen Bet v1 and could identify 2 successful binders (**Fig. 1b**). Binder2 exhibited a 120 nM binding affinity on SPR (**Fig. 4e**) and we could validate the binding using size exclusion chromatography with multi-angle static light scattering (SEC-MALS) where the complex shows the expected mass of 27.8 kDa (**Fig. 4f**). The binder2 exhibits a warped helical topology, where its C-terminal helix inserts itself deep into the ligand binding pocket of Bet v1[27]. Previously, an antibody cocktail mix of three antibodies that bind three different immunogenic epitopes of Bet v1 was developed to prevent allergic response[28].

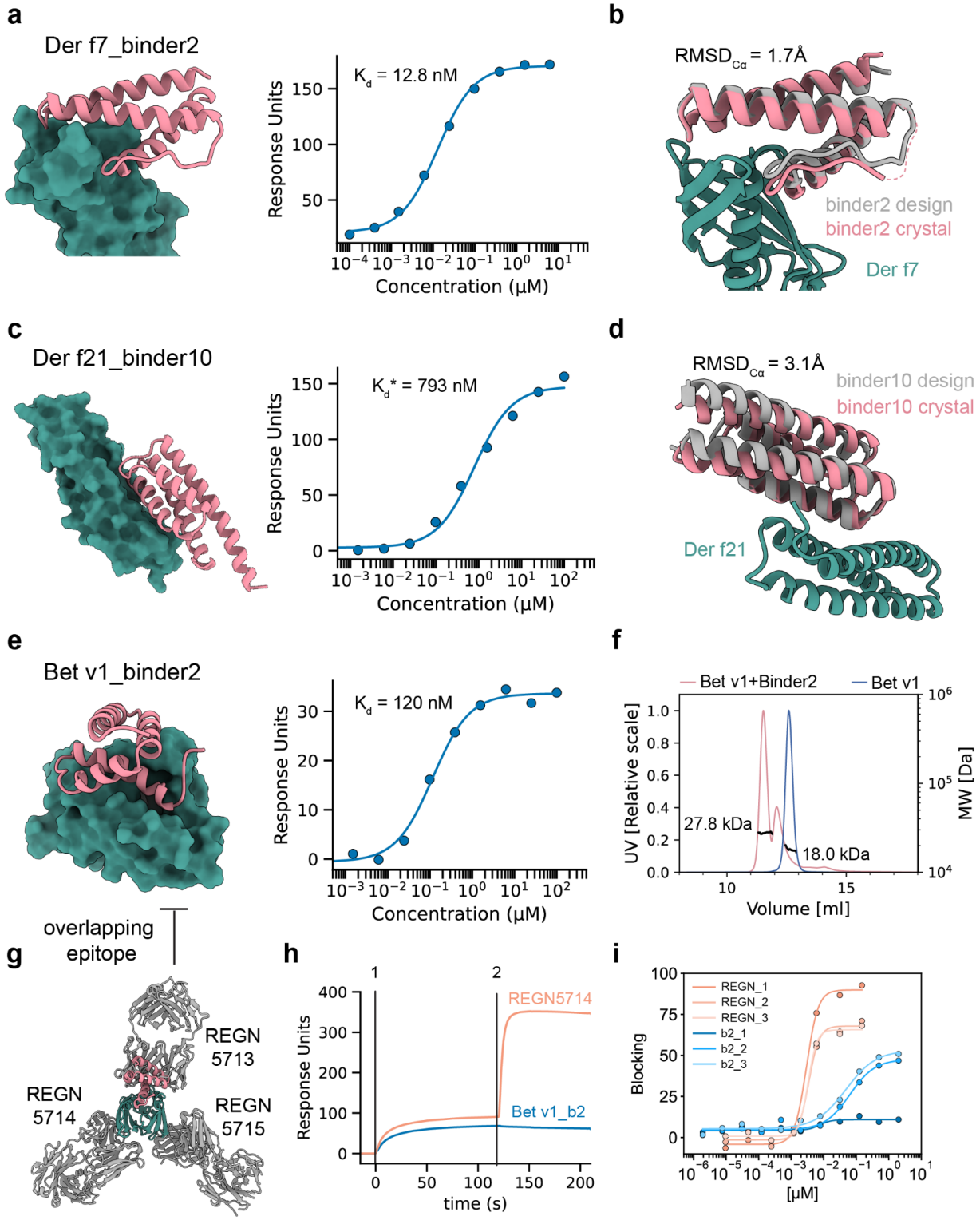


**Figure 4 | Designs occluding epitopes of common allergens. a,** Left: Design model of binder2 against dust mite allergen Der f7. Right: SPR binding affinity fit for binder2. **b,** Crystal structure (colored) of the Der f7_binder2 complex overlaid with the design model (gray). **c,** Left: Design model of binder10 against dust mite allergen Der f21. Right: SPR binding affinity fit for binder10. **d,** Crystal structure (colored) of the DerF21_binder10 complex overlaid with the design model (gray). **e,** Left: Design model of binder2 against birch allergen Bet v1. Right: SPR binding affinity fit for binder2. **f,** SEC-MALS analysis of Bet v1 allergen

(blue) and Bet v1 mixed with binder2 (orange). **g,** CryoEM structure (PDB: 7MXL) of Bet v1 bound to commercial anti-Bet v1 REGN antibody mix. **h,** Competition assay on immobilized REGN5713-Bet v1 complex binding of the REGN5714 antibody but not Bet v1 binder2, confirming binding at the designed site. **i,** Blocking ELISA showing the capacity of the REGN antibody mix (orange) or binder2 (blue) to prevent the binding of Bet v1 to IgE from three birch-allergic patient sera. Number suffix represents an individual serum from a patient.

The published cryoEM structure indicates that our binder targets a known epitope targeted by the REGN5713 antibody, albeit with a different binding mode (**Fig. 4g**). To probe the binding mode, we immobilized REGN5713 on SPR and loaded the Bet v1 allergen on it. We observe a binding signal when REGN5714 is injected, but not when we inject binder2, confirming that it targets an overlapping epitope with REGN5713 (**Fig. 4h**).

We further hypothesized that our binders can compete with Bet v1 specific IgE present in serum samples from birch allergic patients, similarly to the REGN antibody mix[28]. To test the neutralization activity of our anti-Bet v1 binder2, we performed a blocking ELISA using the serum of three birch allergic patients with high titers of anti-Bet v1 IgE. In this assay, biotinylated Bet v1 was preincubated with either the REGN antibody cocktail or our designed binder2 (**Fig. 4i**). While the REGN three antibody mix was able to block up to 90% of the binding of Bet v1 to IgE at low concentrations, our single binder exhibited blocking rates of up to 50% in 2 out of 3 donors. This is on par with blocking rates of single antibodies[28], indicating that there is therapeutic potential for *de novo* designed binders in neutralizing allergic response when targeting multiple epitopes.

**Modulating the function of large multi-domain nucleases**

Nucleic acid interaction interfaces in proteins have long been considered to be undruggable[29]. This is due to their highly charged, convex and large interfaces, which are difficult to target with small molecules[29]. Protein binders offer a promising alternative to modulate protein-nucleic acid interactions in biotechnological and therapeutic applications. We decided to test the applicability of our pipeline to such interfaces on the large multi-domain CRISPR-Cas9 nuclease from *Streptococcus pyogenes* (SpCas9). SpCas9 has been adapted for gene editing applications due to its easy programmability and has since revolutionized synthetic biology and medicine[30]. However, CRISPR-Cas9 is originally a prokaryotic immune system that protects bacteria against invading genetic elements[31]. To counter this, phages have evolved small proteins, termed anti-CRISPRs (Acrs), that can block CRISPR-Cas nuclease activity by directly occluding nucleic acid binding sites[32]. We wondered whether we could design artificial Acrs that could emulate a similar function.

We designed binders against the bipartite REC1 domain of SpCas9, which contains a highly charged pocket for the binding of the guide RNA[33] (**Fig. 5a**). We tested 6 binders experimentally and strikingly all 6 binders exhibited binding activity against the full length apo SpCas9 enzyme (**Fig. 1b**). The top performing binder 3 (**Fig. 5b**) and 10 (**Fig. 5c**) exhibited apparent binding affinities in the range of 300 nM on the SPR, although no plateau was reached and the real affinity might differ. To validate their binding mode,

we solved cryoEM structures of binder3 and binder10 bound to the full length SpCas9 apo enzyme (**Fig. 5d-e**). We observe clear density in the binding pocket of the REC1 domain and can confidently dock both binders, validating the designed binding mode.

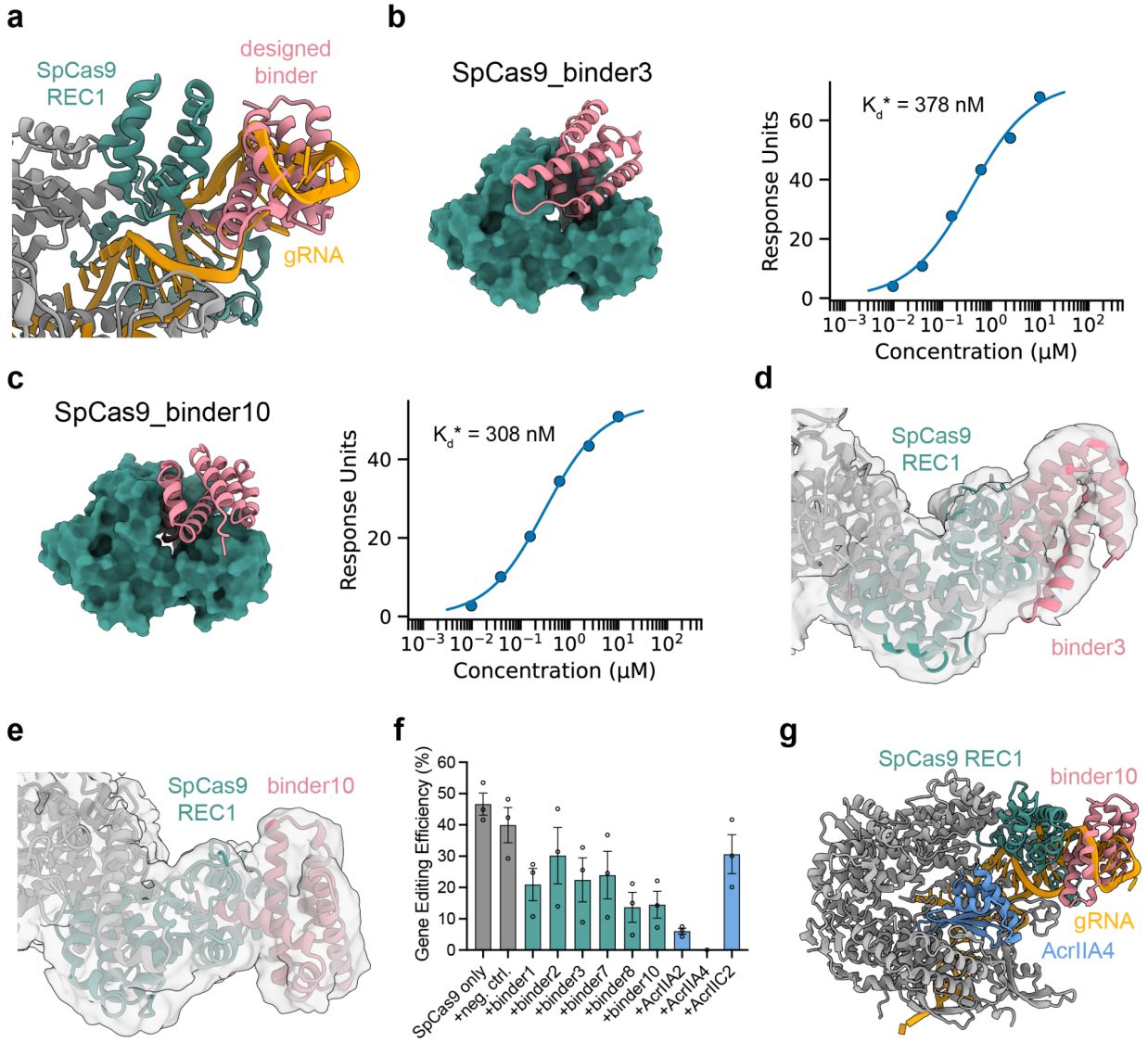


**Figure 5 | Targeting nucleic acid interactions with *de novo* binders against CRISPR-Cas9. a,** Zoom in on the SpCas9 REC1 domain with bound guide RNA (PDB: 4ZT0). A designed binder is overlaid in the binding pocket. **b,** Left: Design model of binder3 against the REC1 domain of SpCas9. Right: SPR binding affinity fit for binder3. Asterisk indicates apparent $K_d$ due to poor fit. **c,** Left: Design model of binder10 against the REC1 domain of SpCas9. Right: SPR binding affinity fit for binder10. Asterisk indicates apparent $K_d$ due to poor fit. **d,** CryoEM structure of binder3 bound to the apo form of SpCas9. REC1 domain is highlighted in green, rest of Cas9 in grey. CryoEM density overlaid in grey. **e,** CryoEM structure of binder10 bound to the apo form of SpCas9. REC1 domain is highlighted in green, rest of Cas9 colored in grey. CryoEM density overlaid in grey. **f,** SpCas9-based editing of HEK293T cells (gray bars) in the presence of designed binders (green bars) or natural Acrs (blue bars). **g,** CryoEM structure of AcrIIA4-bound Cas9 binary complex (PDB: 5VW1) with docked binder10.

To evaluate the functional consequence of this binding, we co-transfected HEK293T cells with CRISPR-SpCas9 and either our designed binders or natural anti-CRISPR proteins[34–36]. Strikingly, we observe a significant reduction of SpCas9 gene editing activity in the presence of our designed binders (**Fig. 5f**). They outperform the natural AcrIIC2, which has also been shown to be an inhibitor of guide RNA loading, albeit using a different targeting mechanism[35]. AcrIIA2 and AcrIIA4, which inhibit target DNA binding (**Fig. 5g**), nearly eliminate gene editing activity, underscoring the differences in the effectiveness of various inhibition strategies. These results suggest that we can design protein binders even against challenging nucleic acid binding sites, potentially opening paths towards novel types of protein-based therapeutics.

## Discussion

The design of *de novo* PPIs by computational means has been a cornerstone problem in protein design. This is primarily due to our lack of detailed understanding of the determinants of molecular recognition that drives PPIs and protein-ligand interactions. Recent advances in deep learning, particularly the development of accurate structure prediction networks such as AF2, have revolutionized the field and enabled more accurate filtering of *de novo* designs. Here we introduce a robust pipeline for binder design based on backpropagation through the AF2 network capable of hallucinating protein binders. Unlike the majority of previously described approaches, BindCraft allows for flexibility on the side of the target protein, which given the intrinsic flexibility of protein structures could be critical for capturing binding-induced changes essential for effective molecular recognition.

Our results demonstrate the effectiveness of BindCraft in designing binders against a diverse set of 10 challenging targets. The binder affinities lie predominantly in the nanomolar range, with one at the μM level, and one binder displaying an apparent $K_d$ even in the subnanomolar range (with avidity). The success rates range from 24.5 to 100%, with an average success rate of 49.5%, which is remarkable for designs resulting from a purely computational approach. These rates allow for the screening of far fewer designs experimentally to identify functioning binders, when compared to the current state of the art RFdiffusion[10] and the recently described closed-source AlphaProteo binder design pipeline[37]. Notably, a binder design from our pipeline recently ranked first in an international binder design competition, displaying nanomolar affinity against the challenging EGFR target. However, we do expect success rates to vary based on the target protein and binding site.

One of the main challenges of PPI design is the choice of a favorable target site[7,11,38]. Prototypical binding sites are often composed of hydrophobic patches with mostly flat surfaces. Here we targeted a wide range of structural sites, some with previously described binding interfaces, such as in the case of cell surface receptors, as well as unexplored surface regions in *de novo* proteins, allergens, and CRISPR-Cas nucleases. In the absence of a defined target epitope, BindCraft is able to sample optimal

binding sites by making use of the trained AF2 multimer model[15], suggesting that the network has likely learned which sites have a high propensity for forming PPIs. This is the case even for challenging epitopes, such as protein-nucleic acid interfaces, which will potentially unlock new avenues for the design of transcription factor modulators, whose aberrant activity is the underlying cause of many oncogenic diseases[29].

The structural accuracy of our method, validated through both crystallography and cryoEM, not only allows us to create proteins that bind to defined surfaces but also enables their potential for biotechnological and therapeutic applications. We demonstrate this by utilizing our designed binders to reduce the binding of birch allergen Bet v1 to specific IgE from patient-derived serum samples. While a single binder displayed limited blocking activity compared to an antibody cocktail, we anticipate that covering a larger part of the antigen surface could produce comparable results. *De novo* binders would therefore offer a promising alternative to antibodies for such treatments, due to their high stability. However, due to the synthetic nature of our binders and their relatively large size (60–240 amino acids), concerns about immunogenicity and effective delivery persist, though these issues are gradually being addressed in preclinical models[3].

Despite the design successes outlined here, there are limitations to the BindCraft design approach. Backpropagation through the AF2 network requires the use of a GPU with large amounts of memory. For instance, a target-binder complex 500 amino acids in size allocates about 30 Gb of GPU memory. This sometimes requires the trimming or splitting of large proteins and complexes during design. Additionally, since we utilize AF2 monomer in single sequence mode for filtering, it is possible that we filter out prospective high affinity binders at the cost of a robust binding predictor. AF2 has also been shown to be insensitive to the predicted effects of point mutations[39], which could be detrimental at PPI interfaces, where a single mutation can abrogate or significantly enhance binding. The addition of an orthogonal physics-based scoring method, such as Rosetta, have been shown to add more discriminatory power to binder identification[40]. Lastly, a potential limitation is the use of the AF2 i_pTM metric for the ranking of designs, which has been shown to be a powerful binary predictor of binding activity, but does not correlate with the interaction affinity[41]. However, accurate prediction of affinity remains highly challenging and alternative ranking metrics may similarly struggle to address this complexity.

Looking forward, we aim to further improve our design procedure by testing the limits on even more challenging and diverse targets. We also aim to diversify the structure of our design towards more natural and complex folds. As with most other confidence-based deep learning design approaches[10,11], most of our top ranking designs were alpha-helical. To mitigate this, we recently incorporated a “negative helicity loss”, that allows the generation of purely beta sheeted proteins, although with reduced *in silico* design success rates. We hope to improve upon this concept to potentially generate nanobody-based binders and other more relevant molecular formats for clinical translation. Through iterative refinement of our pipeline, we aim to eventually reach a ‘one design, one binder’

stage, enabling the rapid generation of binders for applications in research, biotechnology, and therapeutics.

**Code availability:**

The full BindCraft code along with installation instructions and binder design protocols are available on GitHub under MIT license (https://github.com/martinpacesa/BindCraft). A Google Colab notebook for running BindCraft is available at https://github.com/martinpacesa/BindCraft/blob/main/notebooks/BindCraft.ipynb.

**Data availability:**

Atomic coordinates and structure factors of the reported X-ray structures and cryoEM densities will be deposited in the Protein Data Bank and the Electron Microscopy Data Bank, respectively.

**Acknowledgments:**

We thank SCITAS at EPFL for support in running design trajectories. We thank Anthony Marchand, Petra E.M. Balbi, Ahmed Sadek, and Simon Mauro for support with protein purification. We thank Florence Pojer, Kelvin Lau, and Amedé Larabi (Protein Production and Structure Characterization Core Facility, EPFL, Switzerland) for help with crystallization, biochemical characterization, and providing SpCas9 protein. We thank Didier Nurizzo and Max Nanao (European Synchrotron Radiation Facility, MASSIF-3 and ID30B beamline, Grenoble, France) for assistance with crystallographic data collection. We thank Alexander Myasnikov, Bertrand Beckert, and Sergey Nazarov (Dubochet Center for Imaging, EPFL-UNIL-UNIGE, Switzerland) for assistance with cryo-EM data collection. We thank the group of Ricardo Fernandes for generously providing purified CD45 ECD protein.

**Author contributions:**

M.P., L.N., and B.E.C. conceived the study and designed experiments. M.P., L.N., Y.C., C.A.G., and S.O. developed the code base. M.P. and L.N. generated protein designs. M.P., L.N., J.S., and S.G. purified proteins. L.N. and K.H.G. performed protein binding assays. E.P. and S.B. performed CD45 binder characterisation. C.S. performed SAS-6 binder characterisation. K.H.G. and L.V. developed and performed PD-1 binder characterisation assays. G.N.H. purified SAS-6. L.K. performed gene editing assays. A.A-S. performed blocking assays for birch allergen. M.P. and L.N. solved crystal and cryoEM structures. B.J.Y., A.M.W., P.G., Y.D.M, G.S., S.O. and B.E.C. supervised the work and acquired core funding. M.P., L.N., and B.E.C. wrote the initial manuscript. All authors read and contributed to the manuscript. M.P. and L.N. agree to rearrange the order of their respective names according to their individual interests.

**Funding:**

M.P. was supported by the Peter und Traudl Engelhorn Stiftung. B.E.C. and G.N.H. were supported by the Swiss National Science Foundation, the NCCR in Chemical Biology, the NCCR in Molecular Systems Engineering. S.O. and Y.C. were supported by NIH DP5OD026389, NSF MCB2032259 and Amgen. Y.D.M. was funded by the Gabriella Giorgi-Cavaglieri Foundation. A.A-S. was funded by Fondation Machaon. G.S. was supported by the Swiss National Science Foundation grant no. 214936. L.K. was funded by the University of Zurich Research Priority Program ITINERARE.

**Competing interests:**

K.H.G., L.V., B.J.Y., and A.M.W. are employees of Visterra Inc., USA. Rest of the authors declare no competing interests.

## Materials and Methods

*BindCraft design protocol*

The input and design settings for running the BindCraft pipeline are organized into user-friendly JSON files. To initiate design trajectories, a target PDB format structure needs to be specified, along with the desired minimum and maximum length of the binders, and the desired number of final filtered designs. A target hotspot can be specified as either individual residues or entire chains, or can be omitted completely in which case a binding site is selected according to the combined design loss.

The binder hallucination process is performed using the ColabDesign implementation of AF2. The design process is initialized with a random sequence for the binder, which is predicted in single sequence mode, and a structural input template for the target. This is passed through the AF2 network to obtain a structure prediction and calculate the design loss. The design loss function is composed of multiple terms, with default weight values indicated in parentheses:

a) Binder confidence pLDDT (weight 0.1)
b) Interface confidence i_pTM (weight 0.05)
c) Normalized predicted alignment error (pAE) within the binder (weight 0.4)
d) Normalized predicted alignment error (pAE) between binder and target (weight 0.1)
e) Residue contact loss within binder (weight 1.0)
f) Residue contact loss between the target and binder - if hotspots are specified, the rest of the target is masked from this loss (weight 1.0)
g) Radius of gyration of binder (weight 0.3)
h) “Helicity loss” - penalize or promote backbone contacts every in a 3 residue offset to promote the hallucination of helical or non-helical designs (weight -0.3)

The loss function is used to calculate position specific errors, which are then backpropagated through the AF2 network to produce a L x 20 error gradient, where L is the sequence length. Using multiple iterations and stochastic gradient descent optimization, this error gradient is recomputed and used to optimize the input binder sequence for the next iteration to minimize the resulting loss. We backpropagate through the AF2 multimer model weights[15] and swap randomly between the 5 trained models at each iteration to ensure robust sequence generation and reduce the risk of overfitting to a single model.

Since our goal is to arrive at a real discrete sequence for the binding interface, the sequence optimization is performed in four stages. The first sequence optimization stage is performed in a continuous sequence space using logit inputs. At each step, the sequence representation is based on linear combination of $(1-\lambda) * \text{logits} + \lambda * \text{softmax}(\text{logits}/\text{temp})$, where $\lambda = (\text{step}+1)/\text{iterations}$ and temp = 1.0. Here, multiple amino acids are considered per each binder position, which allows the exploration of a larger and less constrained sequence-structure space. After 50 iterations, we terminate trajectories exhibiting poor AF2 confidence scores, as we found that such trajectories rarely converge to high confidence designs. Additionally, if a beta-sheeted trajectory is

detected, we increase the number of recycles during design from 1 to 3 to ensure accurate prediction. The continuous sequence space optimization is then continued for additional 25 iterations. During the second optimization stage, the sequence logits are normalized to sequence probabilities using the softmax function for 45 iterations to funnel the design space towards a more realistic sequence representation defined as softmax(logits/temp) At each step, the temperature is lowered, where temp = (1e-2 + (1 - 1e-2) * (1 - (step + 1) / iterations)**2). The temperature is also used to scale the learning rate for rate decay. For the third stage, we implement the straight-through estimator, allowing the model to see the one-hot representation, but backprop through the softmax representation. This procedure is performed for 5 iterations. For the final fourth stage the sequence inputs are converted to a one-hot discrete encoding. At each step, X random mutations are independently sampled and tested from the probability distribution of the softmax representation from the previous stage, and mutations with best loss are fixed. X is defined based on the length of the binder sequence (0.05 * binder length). This procedure is performed for 15 iterations. At the end, trajectories with pLDDT below 0.7, less than 7 interface contacts, or significant backbone clashes are rejected.

Successful binder design trajectories are subjected to $MPNN_{sol}$ sequence optimization to improve stability and solubility[16]. To this end, we preserve binder residues in a 4 Å radius around the target interface, and design 20 new sequences for the remaining binder core and surface residues using the soluble weights of ProteinMPNN[13], with a temperature of 0.1 and 0.0 backbone noise. These optimized sequences are then re-predicted using the AF2 monomer model, with 3 recycles and 2-template based models[42] in single sequence mode, to ensure robust and unbiased complex assessment. Each of the two resulting models is then energy minimized using Rosetta's FastRelax protocol[43] with 200 iterations, and interface scores are computed using the InterfaceAnalyzer mover[44] with sidechain and backbone movement enabled.

Designs are finally filtered using a set of predefined filters to ensure the selection of high quality designs for experimental testing. Filters were initially defined based on experimental observations from previous binder design studies[7,8,10,11] and refined over the course of this work. These include:

a) AF2 confidence pLDDT score of the predicted complex (> 0.8)
b) AF2 interface predicted confidence score (i_pTM) (> 0.5)
c) AF2 interface predicted alignment error (i_pAE) (> 0.35)
d) Rosetta interface shape complementarity (> 0.55)
e) Number of unsaturated hydrogen bonds at the interface (< 3)
f) Hydrophobicity of binder surface (< 35%)
g) RMSD of binder predicted in bound and unbound form (< 3.5 Å)

We allow only 2 $MPNN_{sol}$ generated sequences per individual AF2 trajectory to pass filters to promote interface diversity amongst selected binders. This design procedure is set up to loop until a defined number of final desired designs is reached. For optimal results, we recommend running the design pipeline until at least 100 designs pass

computational filters. We then usually pick 10 designs from the top 20 (ranked by i_pTM) for experimental testing.

*Design settings for individual target proteins*

To generate designs against targets described in the results section, we utilized the following input structures, binder specifications, and hotspot designations. For AF2 predictions, we used input sequences from Uniprot. In all cases, the amino acid cysteine was excluded from sequence design.

| *Target* | *Input PDB* | *Hotspots* | *Binder length* |
|---|---|---|---|
| PD-1 | AF2 prediction, trimmed to 32-146 | 64, 126, 129, 133 | 80-150 |
| PD-L1 | AF2 prediction, trimmed to 18-132 | 54, 56, 66, 115 | 65-155 |
| IFNAR2 | 2LAG[45], trimmed to 8-110 | 52, 80, 82, 84, 96, 98 | 60-175 |
| CD45 | AF2 prediction of 5FMV[46], split by extracellular fibronectin domains | d1, d2, d3, d4, d1-d2, d2-d3, d3-d4 | 80-200 |
| BBF-14 | Crystal structure (in preparation) | none | 70-250 |
| CrSAS-6 | AF2 prediction, trimmed to 15-160 | none | 90-200 |
| Der f7 | AF2 prediction with 3UV1 template[47], trimmed to 18-213 | 132 or 178 or 201 | 70-185 |
| Der f21 | AF2 dimer prediction with 5YNY template[26], trimmed to 25-136 | 34 or 52 or 87 or 105 or 116 | 70-185 |
| Bet v1 | AF2 prediction | 24 | 70-185 |
| SpCas9 | AF2 prediction with 4ZT0[33] template, trimmed to 96-174 + 306-446 | 360 | 70-150 |

*Protein expression, purification, and characterization*

DNA sequences of designed proteins, as well as BBF-14, Der f7, Der f21, and Bet v1 targets were ordered from Twist Biosciences with Gibson cloning adapters for cloning into bacterial expression vectors pET21b or pET11. Proteins were expressed in *Escherichia coli* BL21 Codon Plus (DE3) cells (Novagen) by inducing with 0.5 mM IPTG for 6 hours at 18 °C. Pellets were resuspended and lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol, 1 mg/ml lysozyme, 1 mg/ml PMSF and 1 µg/ml DNAse) using sonication. Cell lysates were clarified using ultracentrifugation, loaded on a 1 ml Ni-NTA Superflow column (QIAGEN) and washed with 7 column volumes of 50

mM Tris-HCl pH 7.5, 500 mM NaCl, 10 mM imidazole. Proteins were eluted with 10 column volumes of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 500 mM imidazole.

Fc-fused PD-L1 target[8], IFNAR2 target, the IFNA2 cytokine, and antibodies were expressed using a mammalian Expi293 secreted expression system (Thermo Fisher Scientific, A14635). Six days post transfection the supernatants are collected, cleared and purified either using a 1 ml Ni-NTA Superflow column (QIAGEN) or protein A affinity column (QIAGEN). SAS-6[21] and SpCas9[48] have been purified as described previously.

Both bacterial and mammalian expressed proteins were then concentrated and injected onto a Superdex 75 16/600 gel filtration column (GE Healthcare) in 50 mM Tris-HCl pH 7.5, 250 mM KCl. Proteins after size exclusion were concentrated, frozen in liquid nitrogen, and stored at -80 °C. Molar mass, sample homogeneity, and multimeric state were confirmed using SEC-MALS (miniDAWN TREOS, Wyatt) by injecting 100 µg of protein in PBS. Folding, secondary structure content, and melting temperatures were assessed using circular dichroism in a Chirascan V100 instrument from Applied Photophysics in PBS at a concentration of 0.1-0.3 mg/ml.

*Expression and purification of PD-1 target and binders*

DNA sequences were synthesized in the pcDNA3.4 vector with an osteonectin secretion signal at the N-terminus (Twist Biosciences). *De novo* designs were fused to the N-terminus of human IgG1 Fc. The extracellular domain (25-167) of human PD-1 (UniProtKB: Q15116) was fused to a C-terminal AviTag™ and His tag. Plasmid DNA was prepared from glycerol stocks (Twist Biosciences) using Cowin Biosciences GoldVac EndoFree plasmid maxi kit. Plasmids were transfected into 3 mL or 50 mL cultures of Expi293F™ (Gibco) cells per the manufacturer's recommendations. Cells incubated at 37 ºC for 4-5 days prior to harvest. Following protein expression, the cell culture supernatant was filtered through a 0.22 µM filter and purified using MabSelect protein A affinity chromatography resin (Cytiva). The column was washed with PBS and the protein was eluted in Tris glycine buffer pH 2.5. Following elution, proteins were dialyzed into PBS using a 10 kDa MWCO dialysis cassette. For production of biotinylated PD-1 protein, the PD-1 plasmid was co-transfected with BirA plasmid (2:1 ratio). The BirA plasmid contains the BirA sequence (UniProtKB: P06709) with a C-terminal Flag tag in the pcDNA3.4 vector.

*Binding characterization of PD-1*

Designs were initially screened for binding to biotinylated human PD-1 or a random protein using biolayer interferometry (Sartorius OctetRED384). Biotinylated human PD-1 protein and biotinylated lysozyme (GeneTex) were prepared at 500 nM in PBS containing 0.1% BSA (PBSA). The designs were diluted to 5 µM in PBSA. Streptavidin-labeled biosensors were saturated with either biotinylated human PD-1 or biotinylated chicken lysozyme. The designs were then allowed to associate with the immobilized ligand for 60 seconds, followed by a dissociation step in PBSA. The baseline subtracted signal (nm)

was calculated and used to prioritize human PD-1 specific binders for further characterization.

To determine the affinity of selected designs, 100 nM biotinylated human PD-1 prepared in PBSA was immobilized onto a streptavidin labeled biosensor for 15 seconds. Serial dilutions of the designs (from 2.5 μM to 5 nM) were then allowed to associate with the immobilized ligand for 180 seconds, followed by a dissociation step in PBSA for 300 seconds. Following background subtraction of the BLI binding curves using the buffer only (PBSA) curve, the $K_d$ was determined using the 1:1 model in the Data Analysis HT 11.1 curve fitting module.

To determine if the designed protein competed with pembrolizumab for binding to PD-1, 100 nM biotinylated human PD-1 in PBSA was immobilized onto streptavidin coated biosensors for 15 seconds. An initial association with 200 nM pembrolizumab prepared in PBSA was performed for 180 seconds, followed by a second association with 200 nM design prepared in PBSA for 180 seconds.

*Surface Plasmon Resonance (SPR) binding and competition assays*

SPR measurements were performed using the Biacore 8K system (Cytiva) in HBS-EP+ buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P20 GE Healthcare). Target proteins were immobilized on a CM5 chip (GE Healthcare) through amide coupling in 10 mM NaOAc pH 4.5 for 250s at a flow rate of 10 μl/min until 100 relative response units were immobilized. Designed binders or control proteins were injected as analytes in either a single 10 μM concentration during binder pre-screening or in serial dilutions to assess binding kinetics. These were injected at a flow rate of 30 μl/min for a varying contact time, followed by dissociation. If necessary, the chip surface was regenerated after each injection using 10 mM Glycine-HCl pH 2.5 for 30s at a flow rate of 30 μl/min. Binding curves were fitted with a 1:1 Langmuir binding model in the Biacore 8K analysis software. Steady-state response units were plotted against analyte concentration and a sigmoid function was fitted to the experimental data in Python 3.9 to derive the $K_D$.

Competition assays were performed as follows. For PD-L1 and IFNAR2: Target receptors were immobilized, and binders and competitors were injected as analytes. Two subsequent injections were performed either with only competitor (A,1 μM), only design (B,1 μM) or first competitor (1 μM, A) and then design+competitor (both1 μM, A+B). For BetV1: REGN5713 (Antibody format) was immobilized on the SPR chip and in a first injection (1) loaded with BetV1 allergen (1 μM), before either REGN5714 (Fab format) or Birch_binder2 were injected (both 1 μM) (2).

*Protein crystallization and structure determination*

The BBF-14_binder4 complex was crystallized at a concentration of 5 mg/ml using sitting drop vapor diffusion at 16 °C in 0.1 M MES pH 6.0, 0.2 M Na acetate trihydrate, 20% w/v PEG 8000 buffer (SG1-Eco Screen, Molecular Dimensions). The DerF7_binder2

complex was crystallized at a concentration of 15 mg/ml using sitting drop vapor diffusion at 16 °C in 0.1 M MES pH 6.5, 0.2 M KSCN, 25% w/v PEG 2000 MME buffer (Clear Strategy Screen I, Molecular Dimensions). The DerF21_binder10 complex was crystallized at a concentration of 30 mg/ml using sitting drop vapor diffusion at 16 °C in 0.1 M Na citrate pH 5.6, 1.0 M $LiSO_4$, 0.5 M $NH_4SO_4$ buffer (SG1-Eco Screen, Molecular Dimensions). Crystals were cryoprotected in 25% glycerol and flash-cooled in liquid nitrogen. Diffraction data was collected at the European Synchrotron Radiation Facility MASSIF-3 and ID30B beamlines, Grenoble, France at a temperature of 100 K. Crystallographic data was processed using the autoPROC package[49]. Phases were obtained by molecular replacement using Phaser[50]. Atomic model refinement was completed using COOT[51] and Phenix.refine[50]. The quality of refined models was assessed using MolProbity[52]. Structural figures were generated using ChimeraX[53].

*CryoEM structure determination*

SpCas9 was mixed with a 3-fold excess of either binder3 or binder10, and the complex was purified using S200 10/300 gel filtration column (GE Healthcare) in 20 mM Tris-HC pH 7.5, 250 mM KCl. The purified complex was applied to a glow discharged 300-mesh holey carbon grid 300-mesh holey carbon grid (Au 1.2/1.3 QuantifoilMicro Tools), blotted for 4 seconds at 95% humidity, 10 ˚C, plunge frozen in liquid ethane (Vitrobot Mark IV, FEI) and stored in liquid nitrogen. Data collection was performed on a 300 kV Titan Krios G4 microscope equipped with a FEI Falcon IV detector and SelectrisX energy filter. Micrographs were recorded at a magnification of 165kx, pixel size of 0.726 Å, and a nominal defocus ranging from -0.8 mm to -2.2 mm.

Acquired cryo-EM data was processed using cryoSPARC v4.5.3[54]. Micrographs were patch motion corrected, and micrographs with a resolution estimation worse than 5 Å were discarded after patch CTF estimation. Initial particles were picked using blob picker with 90-135 Å. Particles were extracted with a box size of 360x360 pixels, down-sampled to 220x220 pixels. After 2D classification, clean particles were used for *ab initio* 3D reconstruction and initial non-uniform 3D reconstruction[55]. This model was used for additional template-based picking of particles. Following several rounds of 3D classification, where classes containing unbound Cas9 were excluded, the class with most detailed binder features was re-extracted using full box size and subjected to non-uniform and local refinement to generate final reconstructions. The local resolution was calculated and visualized using ChimeraX[53]. The *in silico* models were docked into density using ChimeraX[53].

*Birch allergen blocking assay*

Anti-Bet v1 binder blocking capacity was assessed by first coating NuncSorp (Thermofisher) plates with 2 µg/ml of anti-human IgE monoclonal antibody (NBS-C BioScience, Vienna, Austria; clone Le27; Cat#0908-1-010) in coating buffer (15 mM $Na_2CO_3$, 34.87 mM $NaHCO_3$) and incubating overnight at 4 ºC. The plates were washed

with PBS+0.05% Tween and blocked using PBS+1% BSA for 2 hours at room temperature. Then, sera of birch allergic patients were added at a concentration of 4 ng/ml of anti-Betv1 IgE. Biotinylated Bet v1 allergen at 1 nM concentration was preincubated for 2 hours at room temperature with 4-fold serial dilutions of the BetV1_binder2 starting at 2 μM or with 5-fold serial dilutions of the cocktail of REGN5713, REGN5714 and REGN5715 (starting at 50 nM each) and then added to the IgE coated plate. After two-hour incubation at room temperature, the plates were washed with PBS+0.05% Tween and streptavidin horseradish peroxidase (BD Pharmigen, USA; Cat#554066; 1:1000 dilution) was added and incubated for 1 hour. Plates were washed and tetramethylbenzidine substrate (BD Biosciences, San Diego, USA; Cat#555214) was added and incubated further 20 minutes. The reaction was stopped with 2 M sulfuric acid. Absorbance was measured on a spectrophotometer at 450 nm with a 630 nm reference, and blocking percentage was measured by subtracting the absorbance of the sample in the absence of the binder.

*SpCas9 gene editing*

For SpCas9-sgRNA plasmid cloning, lentiCRISPR v2 (Addgene #52961, a gift from Feng Zhang) was digested with BsmBI (NEB). Oligonucleotides encoding for the sgRNA targeting the NSD2 gene were annealed and ligated into the digested lentiCRISPR v2 plasmid. All binders were human codon optimized using the GenSmart Codon Optimization tool and ordered as inserts with homology overhangs for cloning from Twist bioscience. Final binder plasmids were generated by isothermal assembly (NEBuilder® HiFi DNA Assembly Cloning Kit, NEB).

HEK293T (ATCC CRL-3216) were maintained in DMEM plus GlutaMax (Thermo Fisher Scientific), supplemented with 10% (vol/vol) fetal bovine serum (FBS, Sigma-Aldrich) and 1 × penicillin-streptomycin (Thermo Fisher Scientific) at 37 °C and 5% CO2. Cells were maintained at confluency below 90% and passaged every 2-3 days. For testing inhibitor efficiency, HEK293T cells were seeded in 48-well cell culture plates (Greiner) and transfected at 70% confluency using 300 ng Cas9+sgRNA plasmid, 500 ng inhibitor plasmid, and 5 uL Lipofectamine 2000 according to the manufacturer's instructions (Thermo Fisher Scientific). The next day, cells were split and selected with either Puromycin, Blasticidin or both. Three days post-transfection, cells were harvested and genomic DNA was isolated by direct lysis.

The DNA from the cell lysate was prepped for next-generation sequencing as previously described[56]. In the first PCR round, genomic regions of interest were amplified using GoTaq Green Master Mix (Promega) and primers that included Illumina forward and reverse adaptor sequences. A second PCR round, also using GoTaq Green Master Mix (Promega), introduced p5-p7 barcodes into the products from the first round. The resulting amplified amplicons were pooled and quantified using a Qubit 3.0 fluorometer (Invitrogen). The libraries were then sequenced using a MiSeq platform (Illumina, 150 bp,

paired-end). Sequencing data and resulting gene editing insertion-deletion rates were analyzed using CRISPResso2[57].

## References


1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
2. Cao, L. et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
3. Berger, S. et al. Preclinical proof of principle for orally delivered Th17 antagonist miniproteins. *Cell* (2024) doi:10.1016/j.cell.2024.05.052.
4. Yang, H. et al. Design of cell-type-specific hyperstable IL-4 mimetics via modular de novo scaffolds. *Nat. Chem. Biol.* **19**, 1127–1137 (**20**23).
5. Sahtoe, D. D. et al. Design of amyloidogenic peptide traps. Nat. Chem. Biol. 20, 981–990 (2024).
6. Edman, N. I. et al. Modulation of FGF pathway signaling and vascular differentiation using designed oligomeric assemblies. Cell (2024) doi:10.1016/j.cell.2024.05.025.
7. Cao, L. et al. Design of protein-binding proteins from the target structure alone. Nature **605**, 551–560 (2022).
8. Gainza, P. et al. De novo design of protein interactions with learned surface fingerprints. Nature **617**, 1–9 (2023).
9. Chevalier, A. et al. Massively parallel de novo protein design for targeted therapeutics. Nature **550**, 74–79 (2017).
10. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. Nature 1–3 (2023) doi:10.1038/s41586-023-06415-8.
11. Bennett, N. R. et al. Improving de novo protein binder design with deep learning. *Nat Commun* **14**, 2625 (2023).
12. Baek, M. et al. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* 1–5 (2023) doi:10.1038/s41592-023-02086-5.
13. Dauparas, J. et al. Robust deep learning–based protein sequence design using ProteinMPNN. Science 0, eadd2**187** (2022).
14. Wicky, B. I. M. et al. Hallucinating symmetric protein assemblies. Science 0, eadd1964 (2022).
15. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. *Biorxiv* 2021.10.04.4**630**34 (2022) doi:10.1101/2021.10.04.463034.
16. Goverde, C. A. et al. Computational design of soluble and functional membrane protein analogues. Nature **631**, 449–458 (2024).
17. Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. Science **359**, 1350–1355 (2018).
18. Walter, M. R. The Role of Structure in the Biology of Interferon Signaling. *Front. Immunol.* **11**, 606489 (2020).
19. Hermiston, M. L., Xu, Z. & Weiss, A. CD45: A Critical Regulator of Signaling Thresholds in Immune Cells. *Annu. Rev. Immunol.* **21**, 107–137 (2003).
20. Gönczy, P. & Hatzopoulos, G. N. Centriole assembly at a glance. *J. Cell Sci.* **132**, jcs228833 (2019).

21. Hatzopoulos, G. N. et al. Tuning SAS-6 architecture with monobodies impairs distinct steps of centriole assembly. *Nat. Commun.* **12**, 3805 (2021).
22. Bousquet, J. et al. Allergic rhinitis. *Nat. Rev. Dis. Prim.* **6**, 95 (2020).
23. Dall'Antonia, F., Pavkov-Keller, T., Zangger, K. & Keller, W. Structure of allergens and structure based epitope predictions☆. *Methods (San Diego, Calif)* **66**, 3–21 (2014).
24. Ipsen, H. & Løwenstein, H. Isolation and immunochemical characterization of the major allergen of birch pollen (Betula verrucosa). *J. Allergy Clin. Immunol.* **72**, 150–159 (1983).
25. Tai, H.-Y. et al. The different modes of binding of the dust mite allergens, Der f 7 and Der p 7, on a monoclonal antibody WH9 contribute to the differential reactivity. *J. Microbiol., Immunol. Infect.* **51**, 478–484 (2018).
26. Pang, S. L. et al. Crystal structure and epitope analysis of house dust mite allergen Der f 21. *Sci. Rep.* **9**, 4933 (2019).
27. Kofler, S. et al. Crystallographically Mapped Ligand Binding Differs in High and Low IgE Binding Isoforms of Birch Pollen Allergen Bet v 1. *J. Mol. Biol.* **422**, 109–123 (2012).
28. Atanasio, A. et al. Targeting immunodominant Bet v 1 epitopes with monoclonal antibodies prevents the birch allergic response. J. Allergy Clin. Immunol. **149**, 200–211 (2022).
29. Bushweller, J. H. Targeting transcription factors in cancer — from undruggable to reality. *Nat. Rev. Cancer* **19**, 611–624 (2019).
30. Pacesa, M., Pelea, O. & Jinek, M. Past, present, and future of CRISPR genome editing technologies. Cell 187, 1076–1100 (2023).
31. Koonin, E. V. & Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions Royal Soc B Biological Sci* **374**, 20180087 (2019).
32. Wiegand, T., Karambelkar, S., Bondy-Denomy, J. & Wiedenheft, B. Structures and Strategies of Anti-CRISPR-Mediated Immune Suppression. *Annu Rev Microbiol* **74**, 1–17 (2020).
33. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. Science **348**, 1477–1481 (2015).
34. Dong, D. et al. Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. Nature **546**, 436–439 (2017).
35. Zhu, Y. et al. Diverse Mechanisms of CRISPR-Cas9 Inhibition by Type IIC Anti-CRISPR Proteins. *Mol Cell* **74**, 296-309.e7 (2019).
36. Liu, L., Yin, M., Wang, M. & Wang, Y. Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race. Mol Cell **73**, 611-620.e3 (2019).
37. Zambaldi, V. et al. De novo design of high-affinity protein binders with AlphaProteo. *arXiv* (2024) doi:10.48550/arxiv.2409.08022.
38. Gainza, P. et al. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods* **17**, 184–192 (2020).
39. Pak, M. A. et al. Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLOS ONE* **18**, e0282689 (2023).

40. Baryshev, A. et al. Massively parallel measurement of protein–protein interactions by sequencing using MP3-seq. Nat. Chem. Biol. 1–10 (2024) doi:10.1038/s41589-024-01718-x.
41. Yin, R., Feng, B. Y., Varshney, A. & Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci. : A Publ. Protein Soc.* **31**, e4379 (2022).
42. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 6**30**, 493–500 (2024).
43. Leman, J. K. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nat Methods 17, 665–680 (2020).
44. Coventry, B. & Baker, D. Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLoS Comput. Biol.* **17**, e1008061 (2021).
45. Nudelman, I., Akabayov, S. R., Scherf, T. & Anglister, J. Observation of Intermolecular Interactions in Large Protein Complexes by 2D-Double Difference Nuclear Overhauser Enhancement Spectroscopy: Application to the 44 kDa Interferon–Receptor Complex. *J. Am. Chem. Soc.* **133**, **14**755–14764 (2011).
46. Chang, V. T. et al. Initiation of T cell signaling by CD45 segregation at 'close-contacts.' *Nat. Immunol.* **17**, 574–582 (2016).
47. Tan, K. W. et al. Crystal Structure of Der f 7, a Dust Mite Allergen from Dermatophagoides farinae. *PLoS ONE* **7**, e44850 (2012).
48. Pacesa, M. et al. R-loop formation and conformational activation mechanisms of Cas9. Nature **609**, 191–196 (2022).
49. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr Sect D Biological Crystallogr* **67**, 293–302 (2011).
50. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr Sect D Struct Biology* **75**, 861–877 (2019).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr Sect D Biological Crystallogr **66**, 486–501 (2010).
52. Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci **27**, 293–315 (2018).
53. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci 30, 70–82 (2021).
54. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296 (2017).
55. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat Methods 17, 1214–1221 (2020).
56. Marquart, K. F. et al. Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. Nat. Commun. 12,

5114 (2021).
57. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).