

# De novo design of high-affinity protein binders with AlphaProteo

Vinicius Zambaldi<sup>\*,1</sup>, David La<sup>\*,1</sup>, Alexander E. Chu<sup>\*,1</sup>, Harshnira Patani<sup>\*,1</sup>, Amy E. Danson<sup>\*,1</sup>, Tristan O. C. Kwan<sup>\*,1</sup>, Thomas Frerix<sup>\*,1</sup>, Rosalia G. Schneider<sup>\*,1</sup>, David Saxton<sup>\*,1</sup>, Ashok Thillaisundaram<sup>\*,1</sup>, Zachary Wu<sup>\*,1</sup>, Isabel Moraes<sup>2</sup>, Oskar Lange<sup>2</sup>, Eliseo Papa<sup>1</sup>, Gabriella Stanton<sup>1</sup>, Victor Martin<sup>1</sup>, Sukhdeep Singh<sup>1</sup>, Lai H. Wong<sup>1</sup>, Russ Bates<sup>2</sup>, Simon A. Kohl<sup>2</sup>, Josh Abramson<sup>1</sup>, Andrew W. Senior<sup>1</sup>, Yilmaz Alguel<sup>3</sup>, Mary Y. Wu<sup>4</sup>, Irene M. Aspalter<sup>5</sup>, Katie Bentley<sup>5,6</sup>, David L.V. Bauer<sup>7</sup>, Peter Cherepanov<sup>3</sup>, Demis Hassabis<sup>1</sup>, Pushmeet Kohli<sup>1</sup>, Rob Fergus<sup>1,†</sup> and Jue Wang<sup>1,†</sup>

<sup>\*</sup>Equal contributions, <sup>†</sup>Equal supervision, <sup>1</sup>Google DeepMind, <sup>2</sup>Work performed while at Google DeepMind, <sup>3</sup>The Chromatin Structure and Mobile DNA Laboratory, The Francis Crick Institute, London, UK, <sup>4</sup>COVID Surveillance Unit, The Francis Crick Institute, London, UK, <sup>5</sup>Cellular Adaptive Behaviour Laboratory, The Francis Crick Institute, London, UK., <sup>6</sup>Department of Informatics, King's College London, London, UK. K.B. performed the work at the Cellular Adaptive Behaviour Laboratory, The Francis Crick Institute, London, UK, <sup>7</sup>RNA Virus Replication Laboratory, The Francis Crick Institute, London, UK

Computational design of protein-binding proteins is a fundamental capability with broad utility in biomedical research and biotechnology. Recent methods have made strides against some target proteins, but on-demand creation of high-affinity binders without multiple rounds of experimental testing remains an unsolved challenge. This technical report introduces AlphaProteo, a family of machine learning models for protein design, and details its performance on the *de novo* binder design problem. With AlphaProteo, we achieve 3- to 300-fold better binding affinities and higher experimental success rates than the best existing methods on seven target proteins. Our results suggest that AlphaProteo can generate binders "ready-to-use" for many research applications using only one round of medium-throughput screening and no further optimization.

## Experimental highlights

- We introduce the AlphaProteo protein design system and experimentally test binders designed against eight structurally diverse target proteins.
- For seven of the targets, between 9% and 88% of the designs tested in the wet lab were experimentally verified as successful binders. These figures are higher than the best existing method and 5- to 100-fold higher than other methods. For one of these targets we report the first computationally designed binders.
- The *in silico* performance of AlphaProteo on hundreds of target proteins from the PDB is comparable to these seven targets, suggesting that the method can potentially generalize widely. We chose one of the most challenging targets from this PDB screen as an 8th target but failed to obtain binders.
- We obtain binders with 80-960 picomolar affinities to four targets and low-nanomolar affinities to another three without needing high-throughput screening or experimental affinity optimization. For the seven targets, our designs have 3- to 300-fold better binding affinities than the best previous designed binder.
- We test binders for two of our targets for biological function, demonstrating inhibition of VEGF signaling in human cells and SARS-CoV-2 neutralisation in Vero monkey cells.
- Cryo-EM and X-ray crystallography confirm the designed binder and binder-target complex structures.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Results</b>	<b>3</b>
2.1	Sub-nanomolar-affinity binders from medium-throughput screening	3
2.1.1	Multiple binding hits within one 96-well plate of designs per target	6
2.1.2	State-of-the-art binding affinities on 7 targets	7
2.1.3	Designs bind the target epitope as intended	9
2.1.4	Designs have specific binding within our target set and are structurally diverse	9
2.2	Functional and structural validation of binders	11
2.2.1	Binders neutralize SARS-CoV-2 variants in live virus neutralization assays	11
2.2.2	Binders inhibit VEGF receptor downstream signaling in cells	11
2.2.3	Experimental structures of binder-target complexes confirm binding mode and structure	11
<b>3</b>	<b>Conclusion</b>	<b>14</b>
	<b>References</b>	<b>15</b>
	<b>Supplementary information</b>	<b>18</b>
<b>S1</b>	<b>Experimental methods</b>	<b>18</b>
S1.1	Target protein expression and purification	18
S1.2	Yeast surface display and flow cytometry	19
S1.2.1	Primary binding screen	19
S1.2.2	Interface mutation, competitive inhibition, and specificity experiments	19
S1.3	Designed binder expression and purification	20
S1.4	Measurement of binding affinity / binding dissociation constants ( $K_D$ )	20
S1.4.1	Homogeneous Time Resolved Fluorescence (HTRF)	20
S1.4.2	Bio-Layer Interferometry (BLI)	21
S1.5	Circular dichroism (CD) spectroscopy	22
S1.6	Western blot analysis of VEGF-A signaling in HUVECs	22
S1.7	SARS-CoV-2 virus neutralization assay	23
S1.8	Cryo-EM sample preparation, data collection and image processing	23
S1.9	X-ray crystallography sample preparation, data processing and structure solving	24
<b>S2</b>	<b>Iterative development and <i>in silico</i> benchmarking of AlphaProteo</b>	<b>24</b>
S2.1	AF2-based benchmark	25
S2.2	AF3-based benchmark	25
<b>S3</b>	<b><i>In silico</i> screening of PDB targets</b>	<b>27</b>
<b>S4</b>	<b>Comparison to other design methods</b>	<b>27</b>
S4.1	Comparison of experimental success rates to Rfdiffusion	27
S4.2	Comparison of binding affinity ( $K_D$ ) to other methods	27
	<b>Supplementary figures</b>	<b>28</b>
	<b>Supplementary tables</b>	<b>38</b>
	<b>Supplementary references</b>	<b>44</b>

## 1. Introduction

Protein-protein interaction is a fundamental aspect of protein function, and protein-binding proteins are a basic building block for therapeutics, diagnostics, and biomedical research [19, 29]. Traditionally, antibodies, nanobodies, and other scaffolds such as DARPins are developed into binders against a wide range of targets by immunization or directed evolution [36, 33, 12]. However, experimental selection does not afford control over the target epitope and is often too laborious for routine research applications. Computational design of binders *de novo*, without using a natural protein as a starting point, can target pre-specified epitopes and generate binders that are smaller, more thermostable, and easier to express than antibodies [10, 39, 6].

Recently, deep-learning based models have achieved major advances in biomolecular structure prediction [21, 2, 28, 24, 1] and protein design [18, 43, 37, 14, 7, 34]. This has enabled progress on key scientific and societal challenges [22], including the prediction and design of protein-protein interactions [9, 17, 4, 43, 11, 13, 8]. It is now possible to obtain computationally designed binders to some targets without high-throughput screening [43, 13, 11]. High binding affinity without experimental optimization has also been achieved in some cases, such as for small peptides or disordered targets [41, 44]. However, success rates remain low against convex or polar epitopes, the affinity of the initial designs is usually poor, and many targets remain intractable [45, 3].

In this technical report focusing solely on experimental validation, we present the AlphaProteo protein design system and show that it can design *de novo* protein-binding proteins with the following advantages:

1. **High success rate:** stable, highly expressed, and specific binders can be obtained from screening tens of design candidates, alleviating the need for high-throughput methods.
2. **High affinity:** for every target tested except one, the best binders have sub-nanomolar or low-nanomolar binding affinity ( $K_D$ ), minimizing the labor needed for downstream affinity optimization.
3. **General:** binders are successfully obtained against a range of targets with diverse structural and biochemical properties, using a single design method without complex manual intervention.

## 2. Results

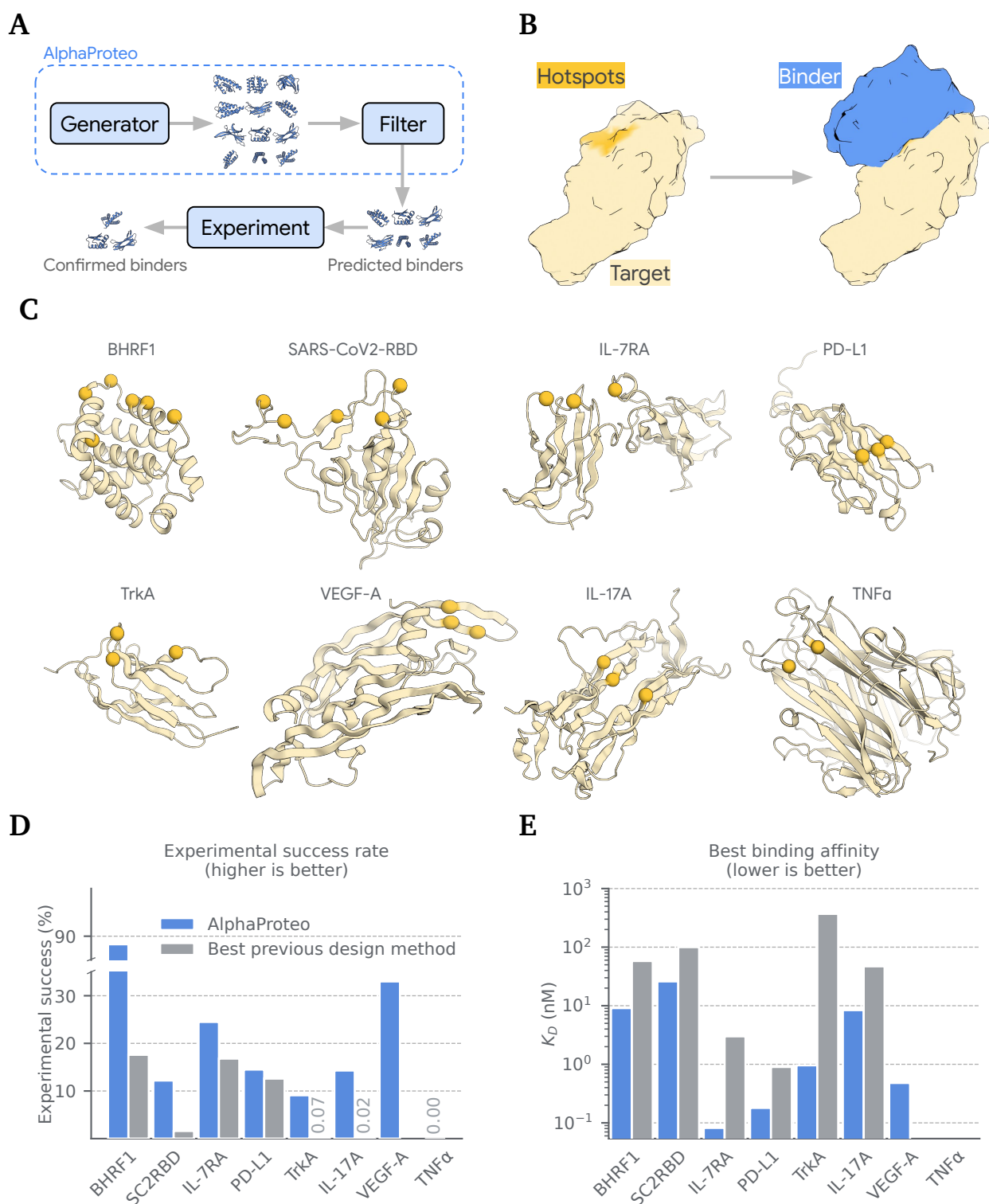
AlphaProteo comprises two components (Figure 1A): a generative model trained on structure and sequence data from the Protein Data Bank (PDB) and a distillation set of AlphaFold predictions, as well as a filter which scores generated designs to predict whether they will succeed experimentally. To design binders, we input a structure of the "target" protein and optionally designate "hotspot" residues representing the target epitope; the generative model outputs a structure and sequence of a candidate binder for that target (Figure 1B). We generate a large number of design candidates and then filter them to a smaller set prior to experimental testing. The generative model compares favorably to the best existing method on *in silico* benchmarks (Figure S1, Section S2).

### 2.1. Sub-nanomolar-affinity binders from medium-throughput screening

To validate AlphaProteo experimentally, we designed binders against eight target proteins with diverse structural properties, of which two are viral proteins involved in infection and six are therapeutically important human proteins (Figure 1C, Table S1):

1. **BHRF1**, an oncogenic protein from Epstein-Barr virus; inhibition via binding can kill cancer cells and slow tumor growth [35]. It has a hydrophobic groove that perfectly accommodates a helix on its binding partner, facilitating binding.
2. **SARS-CoV-2 spike protein receptor-binding domain (SC2RBD)**, a protein domain required for COVID-19 infection. We targeted its interface to the human ACE2 receptor as disrupting this interaction is known to block SARS-CoV-2 from infecting human cells [42]. Previous design efforts have succeeded against this polar and convex site but required experimental optimization to achieve high affinity [5, 11].
3. **Interleukin-7 Receptor- $\alpha$  (IL-7RA)**, a cell-surface receptor involved in lymphocyte development and a therapeutic target for acute lymphoblastic leukemia and HIV. We targeted the binding site of the native interleukin-7 ligand, which is moderately hydrophobic and subject to high success rates in previous design efforts [6, 43].
4. **Programmed Death-Ligand 1 (PD-L1)**, a cell-surface receptor that controls immune cell proliferation and is an important therapeutic target for cancer. The target site is flat and difficult to bind by small molecules and smaller proteins [11, 45].
5. **Tropomyosin Receptor Kinase A (TrkA)**, a nerve growth factor receptor involved in autoimmune disease and an analgesic target for treating chronic pain. We targeted a hydrophobic pocket addressed by previous design efforts. Previous binding affinities were poor without experimental optimization [6].
6. **Interleukin-17A (IL-17A)**, a secreted protein that triggers inflammation and a therapeutic target in autoimmune disease. We targeted the interface of IL-17A with its native receptor, which comprises two chains of a homodimer and has a large polar pocket. Existing designed binders to IL-17A have poor unoptimized affinities and required screening large libraries to obtain [3].
7. **Vascular Endothelial Growth Factor A (VEGF-A)**, a secreted growth factor controlling angiogenesis and a therapeutic target for cancer and diabetic retinopathy. We targeted a small hydrophobic patch bound by the native VEGF receptor [32]. No designed binders to this target have been published despite its biomedical importance.
8. **Tumor Necrosis Factor Alpha (TNF $\alpha$ )**, a pro-inflammatory cytokine produced during inflammation and a therapeutic target for inflammatory disease [16, 31, 30]. We targeted a polar region between two subunits of the TNF $\alpha$  homotrimer where it interacts with the native TNF receptor. No computationally designed binders against this target have been reported.

We chose the above targets for their biological importance, to span a range of design problem difficulty, and to allow comparison to existing design methods. To compare to RFdiffusion [43], we selected the target where it had the highest experimental success rate (IL-7RA) and the two targets where it had the lowest (PD-L1, TrkA), omitting the other 2 tested targets to conserve our experimental bandwidth. We chose BHRF1 and SC2RBD as an additional easy and difficult target, respectively, which have precedent in the computational design literature. IL-17A and VEGF-A were selected as difficult targets that had no confirmed computationally designed binders at the time of the work. After experimental testing on the above 7 targets was completed, TNF $\alpha$  was chosen as an 8th very difficult target based on *in silico* analysis (Section 2.1.1, Section S3). No additional targets beyond these 8 were experimentally evaluated during the course of this work.



**Figure 1 | Overview and experimental performance of AlphaProteo.**

(A) Schematic of design system. The generative model outputs designed structures and sequences of binder candidates and the filter is a model or procedure that predicts whether a design will bind. (B) Schematic of target-structure-conditioned binder design as performed by the generative model. (C) Crystal structures (light yellow) and hotspot residues (dark yellow spheres) of seven target proteins for binder design experiments in this work. VEGF-A and IL-17A are both disulfide-linked homodimers. See Table S1 for PDB IDs and hotspot residue numbers. (D) Percent of all tested designs with measured binding, from AlphaProteo (blue) or the best previous binder design method (gray). (E) Binding affinities of the best per-target  $K_D$  values from AlphaProteo (blue) or the best previous method. These represent the affinities of non-optimized computational designs – see Table 1 for  $K_D$  values of the best optimized computational designs from the literature. The exact values plotted in (D) and (E) are also shown in Table 1 with data sources (see also Section S4).

	BHRF1	SC2RBD	IL-7RA	PD-L1	TrkA	IL-17A	VEGF-A	TNF $\alpha$
<b>Experimental success rate (%)</b> (higher is better)								
AlphaProteo	<b>88</b> (94)	<b>12</b> (172)	<b>25</b> (94)	15 (159)	<b>9</b> (131)	<b>14</b> (63)	<b>33</b> (94)	<b>0</b> (54)
RFdiffusion	–	–	17 (95)	13 (95)	0.0 (95)	–	–	–
Other design methods	18 <sup>a</sup> (17)	1.6 <sup>b</sup> (63)	0.15 <sup>c</sup> (14,912)	13 <sup>b</sup> (16)	0.07 <sup>c</sup> (14,982)	0.02 <sup>d</sup> (15,000)	–	–
<b>Binding K<sub>D</sub> (nM)</b> (lower is better)								
AlphaProteo	<b>8.5</b> (94)	26 (172)	<b>0.082</b> (94)	<b>0.18</b> (159)	<b>0.96</b> (131)	8.4 (63)	<b>0.48</b> (94)	–
RFdiffusion	–	–	14* (95)	1.6* (95)	370* (95)	–	–	–
Other design methods	58 <sup>a</sup> (17)	100 <sup>e</sup> (100,000)	3 <sup>c</sup> (14,912)	0.9 <sup>b</sup> (16)	3000 <sup>c</sup> (14,982)	47 <sup>d</sup> (15,000)	–	–
Other design methods, optimized	16*, <sup>a</sup>	16*, <sup>e</sup>	0.31 <sup>c</sup>	0.65 <sup>f</sup>	1.4 <sup>c</sup>	<b>0.01</b> <sup>d</sup>	–	–

<sup>a</sup> Procko et al. [35] <sup>b</sup> Gainza et al. [11] <sup>c</sup> Cao et al. [6] <sup>d</sup> Berger et al. [3] <sup>e</sup> Cao et al. [5] <sup>f</sup> Yang et al. [45]

**Table 1 | Experimental success rates and affinities of AlphaProteo and other methods.**

Percentage of designs with measured binding and best per-target binder affinity for AlphaProteo, RFdiffusion (as measured by us using yeast display, see Section S4), and other computational design methods. Number of designs tested are in parentheses. "Other design methods, optimized" lists the best affinity after experimental optimization of any computationally designed binder. Binders derived from selection-based methods, such as antibodies and nanobodies, are not considered here. K<sub>D</sub> values from the literature come from biolayer interferometry (BLI) or surface plasmon resonance (SPR) assays, except where noted by asterisks (\*), where we measured the K<sub>D</sub> ourselves using HTRF (Section S4). Some targets used for method development (Section S2) have more detailed results in Table S2.

### 2.1.1. Multiple binding hits within one 96-well plate of designs per target

For each target, we generated a large set of *in silico* designs 50-140 amino acids long (Table S1) and used an automated filtering procedure to choose between 47 and 172 binder candidates to test for binding by yeast surface display. We tested designs for the initial set of seven targets and observed experimental success rates, or the fraction of designs with measurable binding (Section S1.2), ranging from 9%, on TrkA, to 88%, on BHRF1 (Table 1). Per-target success rates were >5% for 7 targets, >10% for 6 targets and >20% for 5 targets (Figure 1D, Table 1).

Our success rates are higher than the best alternative current method on 7 targets (Figure 2B, Table 1). On VEGF-A, AlphaProteo is the first computational design method, to our knowledge, to obtain successful binders, although antibodies have been developed using traditional methods [27]. On BHRF1, SC2RBD, and IL-17A, AlphaProteo has, respectively, 5-, 8-, and 700-fold higher success rates than the next-best method (Figure 1D, Table 1).

To compare AlphaProteo quantitatively to RFdiffusion [43], the current state-of-the-art (SoTA) binder design method, we tested published RFdiffusion binder designs for IL-7RA, PD-L1, and TrkA

alongside AlphaProteo designs in the same yeast display assay (Section S4). In this direct comparison, AlphaProteo had higher overall experimental success rates on all three targets (Figure 1D, Table 1). These results indicate that AlphaProteo is strongly competitive to SoTA in terms of success rates.

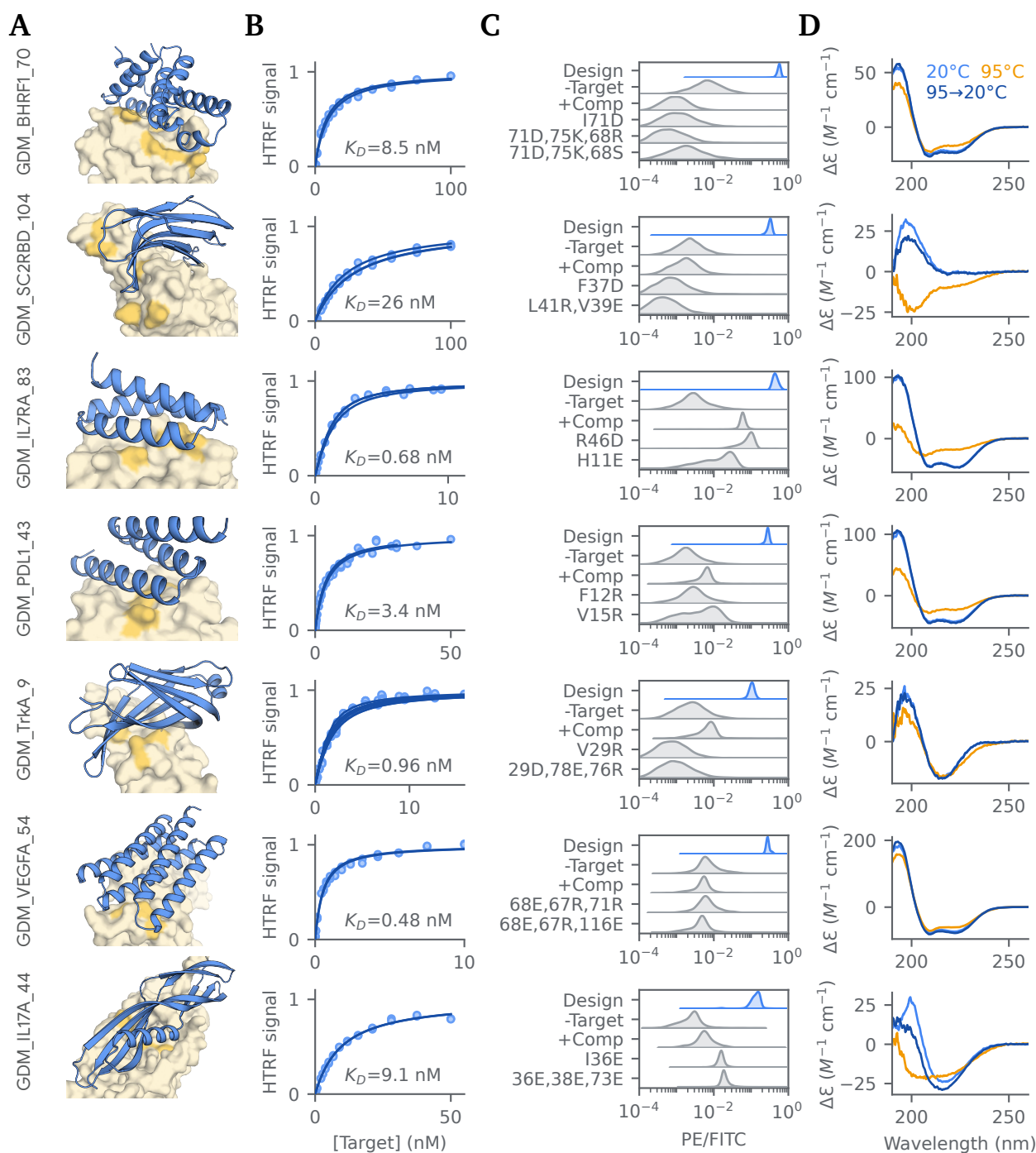
We note that SC2RBD, PD-L1, and TrkA were used to develop AlphaProteo (Section S2), so these success rates may overestimate performance on novel targets. However, for BHRF1, IL-7RA, VEGF-A, and IL-17A, we only performed a single round of medium-throughput testing, showing that high success rates can be obtained prospectively for even quite challenging targets.

After obtaining results on these seven targets, we investigated the potential target range of AlphaProteo by computing its *in silico* success rate for 3 epitopes on each of 200 randomly selected target proteins from the PDB (Section S3). The above 7 targets spanned a similar range of *in silico* success rate as this wider list of targets, confirming that they are representative of the difficulty of most potential targets. The screening also identified several particularly challenging targets, including TNF $\alpha$ , with *in silico* success rates very close to 0. Given TNF $\alpha$ 's unusual *in silico* difficulty and high biomedical importance, we designed and experimentally tested binders to this target, but failed to obtain hits. This is consistent with the low *in silico* performance on this target, and is likely due to a flat, highly polar binding site at an interface between 2 subunits in a homotrimer. Encouragingly, however, 80% of the sampled PDB targets have higher *in silico* success rates than the most difficult target where we successfully obtained binders, IL-17A (Figure S2). This suggests that AlphaProteo can generalize to a wide range of biologically important binder design problems.

### 2.1.2. State-of-the-art binding affinities on 7 targets

High experimental success rates can reduce the labor and cost of obtaining binders, but once hits have been found, a far more important metric is binding affinity ( $K_D$ ) to the target. Most therapeutic antibodies have low-picomolar  $K_D$  values [15, 40], which is achieved by many rounds of experimental affinity maturation. For binders used as research tools, low-nanomolar  $K_D$  values or better are also typical [26]. To measure how strongly our designed binders bound their target, we recombinantly expressed and purified yeast screening hits in *E. coli* to measure their  $K_D$  values *in vitro*. Overall, 93% of designs chosen for follow up successfully expressed in *E. coli* (Table S3), and the majority were monodisperse by size-exclusion chromatography (Figure S4). A subset of designs assayed by circular dichroism (CD) spectroscopy all exhibited the expected secondary structures (Figure 2D, Figure S5). Furthermore, the designs exhibited partial or no unfolding up to 95°C in CD thermal melts, indicating that they are extremely thermally stable with  $T_m$  values > 95 °C (Figure S5). For the recombinantly produced designs, we measured  $K_D$  values using a homogeneous time-resolved fluorescence (HTRF) equilibrium saturation binding assay (Section S1.4).

AlphaProteo's best per-target  $K_D$  values were <1 nM for 4 targets, <10 nM for 6 targets, and <30 nM for 7 targets (Figure 1E, Table 1, Figure S3). The best  $K_D$  overall was 82 pM, for the design IL7RA\_70 (Table S7, Figure S6). We identified 9 total binders with sub-nanomolar  $K_D$  values: 4 for IL-7RA, 2 for PD-L1, 1 for TrkA, and 2 for VEGF-A (Table S7). Compared to the best unoptimized binders from other design methods, AlphaProteo  $K_D$  values were better on all targets, by margins of 7-, 4-, 37-, 5-, 380-, and 5-fold, for BHRF1, SC2RBD, IL-7RA, PD-L1, TrkA, and IL-17A, respectively (Figure 1E, Table 1). Even compared to previous designed binders that have been optimized experimentally through multiple rounds of mutation and selection, the best AlphaProteo  $K_D$  values were still better on BHRF1, IL-7RA, PD-L1, and TrkA (Table 1, "Other design methods, optimized"). Taken together, the success rates and affinities achieved by AlphaProteo suggest that it can generate binders for many research applications after screening one round of 10-100 designs and no further experimentation.



**Figure 2 | Biochemical characterization of representative binders for each target.**

(A) Design models, (B) HTRF equilibrium saturation binding and  $K_D$  values fitted from 1:1 binding models, (C) Yeast display on interface mutants and competitive inhibition, and (D) Circular dichroism spectra before (20 °C) and after thermal melting (95 °C and 95 → 20 °C). Note that the designs here were chosen to showcase all 4 measurement types and therefore may not be the highest-affinity binder for each target. A list of the best binders per target and their  $K_D$  values can be found in Table S7 (also see Figure S6). HTRF y-axis is normalized to the fitted maximal signal (additional HTRF data in Figure S6).



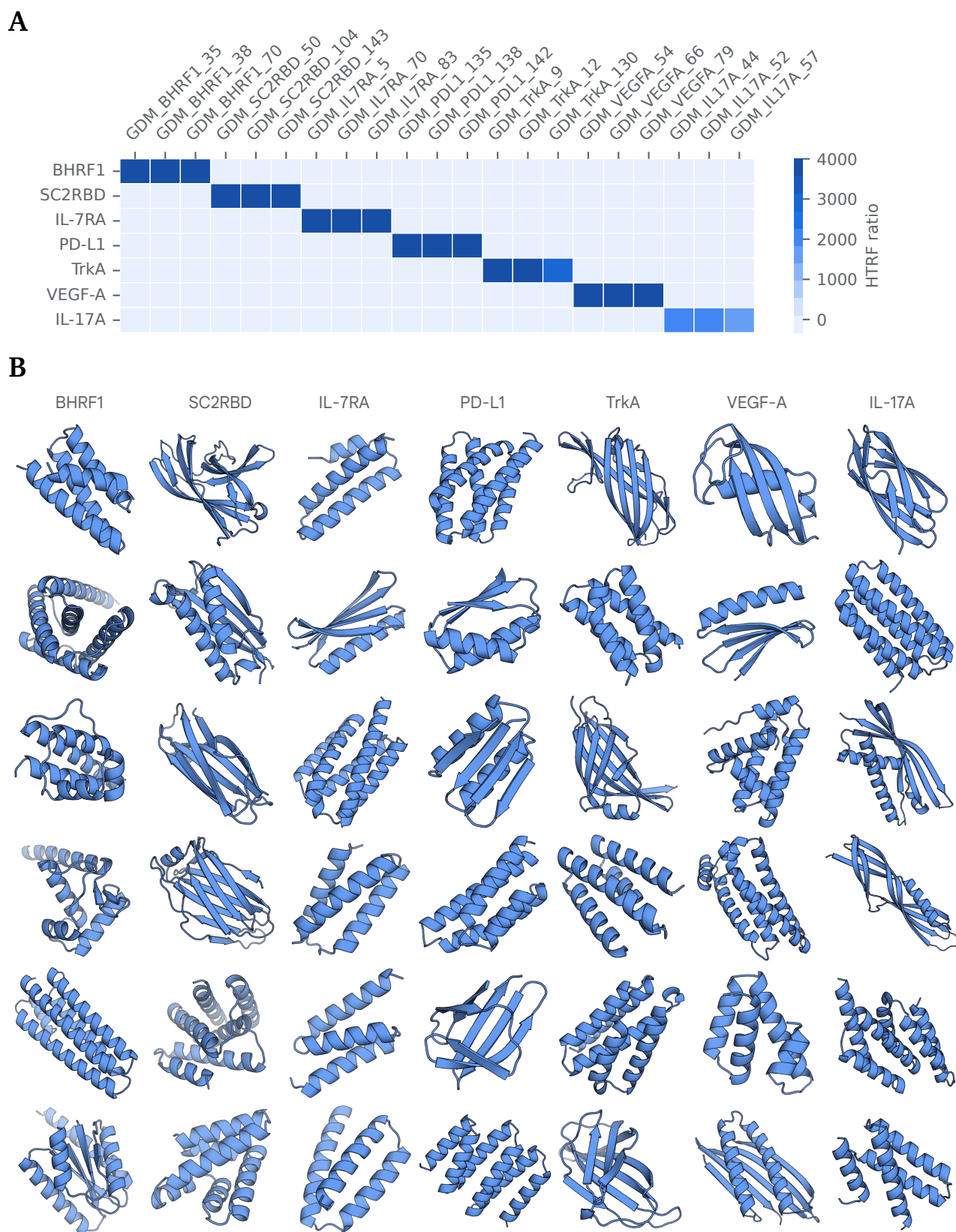
### 2.1.3. Designs bind the target epitope as intended

To test whether the designs bind the intended epitope on the target, we measured binding in the presence of a known competitive binder with the same target site (Figure 2C, Section S1). As expected, this reduced binding signal in all cases, with the reduction being smaller where our binders had a much higher affinity than the competitor. To test whether our designs bind their targets via the intended interactions, we measured binding of our top binders after mutating 1-3 residues at the target-binding interface in their design models (Figure 2C, Figure S8, Figure S9). Almost all mutants had lower binding signal than their parent, suggesting successful disruption of the binding interface by the mutations. A small number of mutants had higher binding signal than the parent. This is not surprising given that we chose the mutations by visual intuition, which likely did not fully account for structural subtleties that could lead to improved binding (Section S1.2.2). Overall, these results indicate that both the binder and target interact with each other via the interfaces that were intended by design.

### 2.1.4. Designs have specific binding within our target set and are structurally diverse

To test the specificity of a subset of our top binders, we measured their binding against all 7 targets. All binders tested exhibit observable binding only to the intended target (Figure 3A), although it is important to note that for many downstream applications a more thorough test of specificity, such as against all proteomic targets, would need to be carried out.

We analyzed the structural diversity of our successful designs to gain insight into how many independent solutions our method is able to generate for each design problem. Diversity is also practically important as it maximizes the chance that one of the designs will satisfy downstream requirements that are not known in advance. We looked at the distribution of pairwise TM-scores (Figure S10A) and secondary structure content (Figure S10B) across binding hits for each target. Compared to the active designs from RFdiffusion, AlphaProteo designs were consistently lower in structural similarity to each other and had a higher frequency of all-beta structures. These observations are consistent with visual inspection of our experimentally confirmed binder designs, which reveals a variety of all-alpha, mixed alpha/beta, and all-beta folds (Figure 3B).



**Figure 3 | Specificity and diversity of designed binders.**

(A) Specificity: HTRF binding signal of a subset of top binders (1 nM) measured against each target (100 nM). All binders show on-target binding signal, and none of the binders show any non-specific binding signal against any of the off-targets tested. (B) Diversity: Examples of experimentally confirmed AlphaProteo binders from different structural clusters at a TM-score cutoff of 0.6.

## 2.2. Functional and structural validation of binders

### 2.2.1. Binders neutralize SARS-CoV-2 variants in live virus neutralization assays

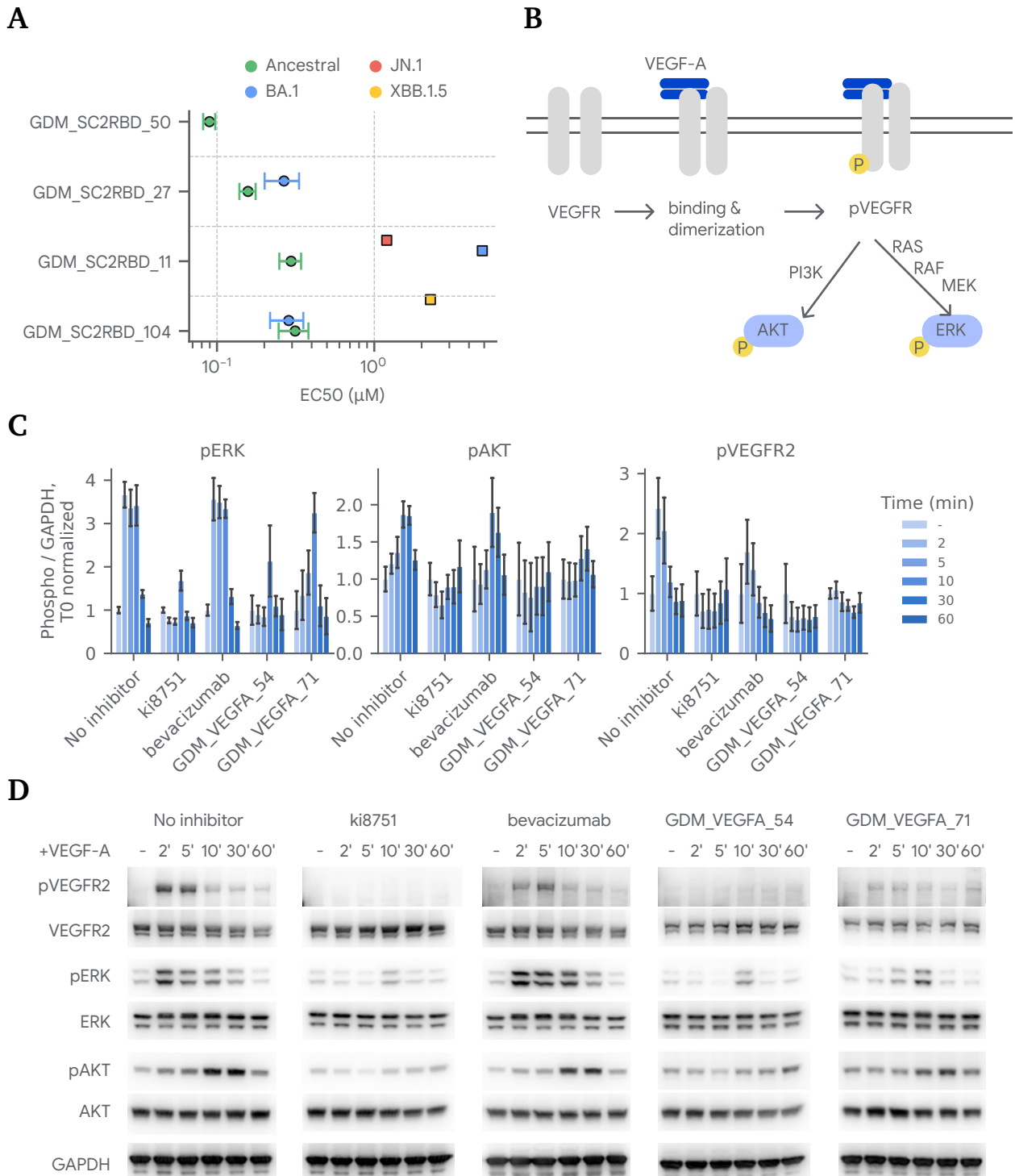
To determine if our binders exhibit the intended biological activity, we tested their ability to bind and neutralize live SARS-CoV-2. We tested four of our binders (GDM\_SC2BRD\_11, GDM\_SC2BRD\_27, GDM\_SC2RBD\_104 and GDM\_SC2BRD\_50) for the ability to neutralize four variants of SARS-CoV-2 that circulated globally from 2020 and 2024 and prevent them from infecting Vero cells [38]. All four binders successfully neutralized an ancestral strain (hCoV19/England/02/2020) with 50% inhibitory concentrations ( $EC_{50}$ ) of 89-300 nM (Figure 4A, Figure S11). This variant has an identical spike protein to the virus first identified in 2019 and is the source of the target structure used for design. These  $EC_{50}$  values are 2- to 10-fold higher than our measured *in vitro* binding affinities (Table S7), consistent to what has been observed in the same assay for clinical monoclonal antibodies such as sotrovimab ( $K_D=0.21$  nM,  $EC_{50}=0.67$  nM against a single SARS-CoV-2 isolate) [38]. Interestingly, two of the binders (GDM\_SC2RBD\_11 and GDM\_SC2RBD\_129) were able to neutralize three of the tested variants. The binder which showed the highest potency and lowest  $EC_{50}$  (GDM\_SC2RBD\_50) only inhibited the ancestral variant. All four variants were neutralized by at least one designed binder.

### 2.2.2. Binders inhibit VEGF receptor downstream signaling in cells

We also tested our designed binders GDM\_VEGFA\_54 and GDM\_VEGFA\_71 for their ability to inhibit VEGF signaling. We measured phosphorylation of VEGF receptor 2 (VEGFR2) and downstream ERK and AKT kinases in primary human umbilical vein endothelial cells (HUVECs) stimulated with human VEGF-A (Figure 4C). Incubation with GDM\_VEGFA\_54 leads to substantially reduced phosphorylation of ERK, AKT, and VEGFR2 compared to a VEGF-A-only control (Figure 4, "no inhibitor"). This effect is similar to that of ki8751 [25], a potent small-molecule VEGFR2 kinase inhibitor. The effect is more potent than that of the anti-VEGF-A monoclonal antibody bevacizumab, the active component of the clinically approved drug Avastin [23], which we tested at an equimolar concentration to our binders in this experiment. This concentration of bevacizumab is 1000-fold lower than that usually tested *in vitro* on HUVECs [20], suggesting that GDM\_VEGFA\_54 is a more potent VEGF-A inhibitor than bevacizumab in HUVECs. The second binder tested, GDM\_VEGFA\_71, leads to a weaker, although still visible reduction in phosphorylation of ERK, AKT, and VEGFR2. These results are consistent with our relative *in vitro* binding affinities of GDM\_VEGFA\_54 and GDM\_VEGFA\_71 for VEGF-A, which are 0.48 and 4.7 nM, respectively.

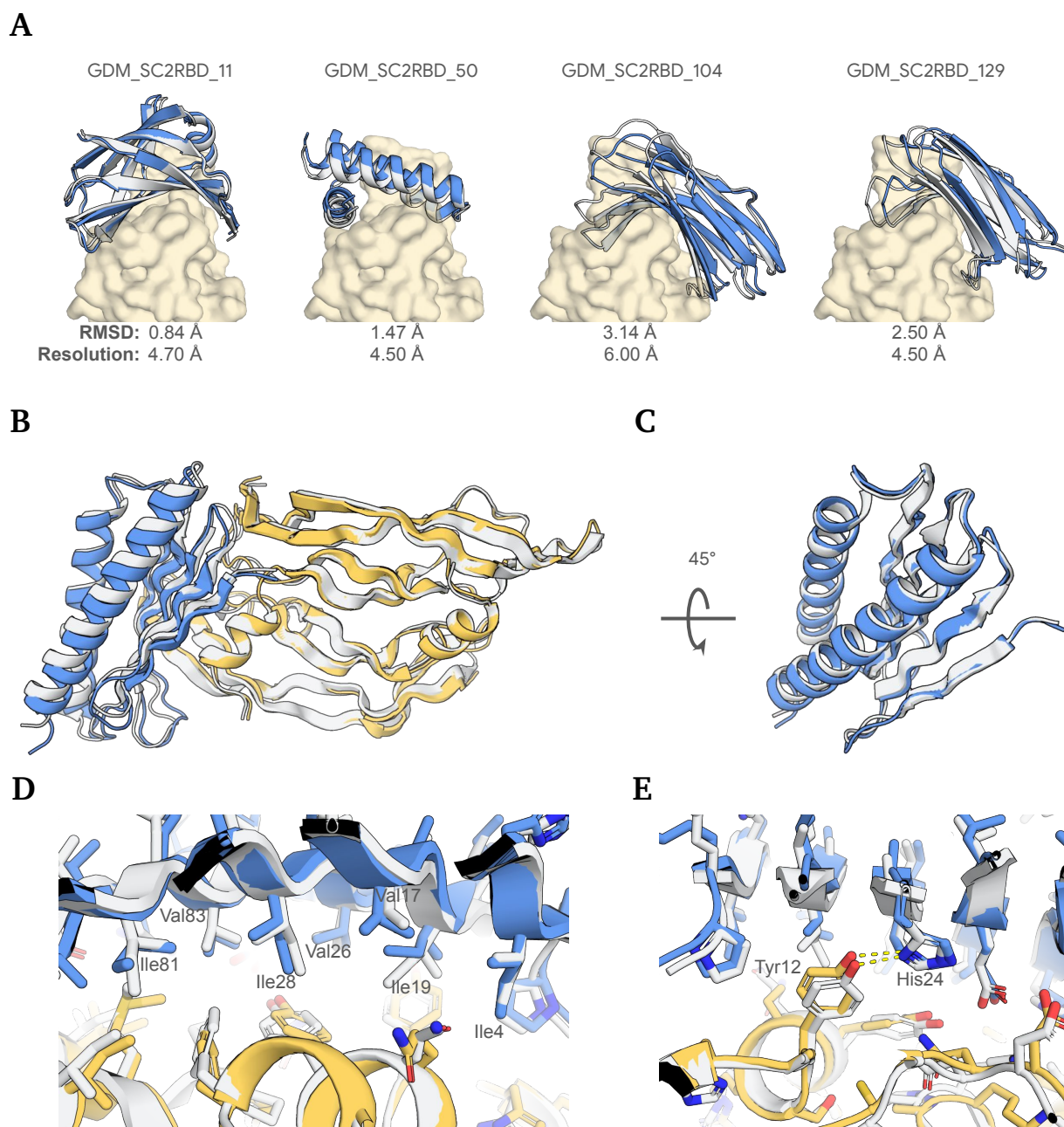
### 2.2.3. Experimental structures of binder-target complexes confirm binding mode and structure

To validate the structures and binding modes of our designs, we used cryo-electron microscopy (cryo-EM) to obtain structures of GDM\_SC2RBD\_11, GDM\_SC2RBD\_50, GDM\_SC2RBD\_104, and GDM\_SC2RBD\_129 in complex with the SARS-CoV-2 spike S1 protein at 4.5 - 6.0 Å resolution (Figure 5A and Figure S12). The experimental structures closely recapitulate the designed binder-target complexes, with binder  $C\alpha$  RMSDs of 0.84 - 3.14 Å using the target S1 protein as an alignment reference. We additionally obtained an X-ray crystal structure of GDM\_VEGFA\_71 in complex with VEGF-A, at 2.65 Å resolution (Figure 5B). The binder folded extremely closely to its designed structure, a mixed alpha-beta fold with a 5-strand beta sheet interacting with VEGF-A, demonstrating atomic level accuracy that shows a  $C\alpha$  RMSD of 0.78 Å between AF3 model and experimental structure. The designed binding orientation was also highly accurate, with a target-aligned binder  $C\alpha$  RMSD of 1.65 Å. Most sidechains of the binder interacting with the target also had the correct rotamer, including a buried hydrogen bond between a histidine of the binder and a tyrosine of VEGF-A which was recapitulated almost perfectly in the experimental structure (Figure 5E).



**Figure 4 | Inhibition of SARS-CoV-2 viral infection and VEGF signaling by designed binders.**

(A) 50% inhibitory concentration ( $EC_{50}$ ) of 4 designed SC2RBD binders in a virus neutralization assay against 4 SARS-CoV-2 variants (Figure S11, Section S1). Error bars show the standard error on the underlying dose-response curve. Binders with low affinity, where complete neutralisation (0% infection) could not be observed, are displayed with square symbols. In these cases the error on the  $EC_{50}$  estimate for the dose-response curves could not be meaningfully determined and error bars are omitted. (B) Schematic representation of the VEGF-A signaling pathway. VEGF-A binding leads to dimerization of VEGFR, phosphorylation of VEGFR and downstream signaling cascade leading to ERK and AKT phosphorylation. (C) Ratio of phosphorylated to total ERK, AKT, and VEGFR2 western blot band intensities before (-) and 2, 5, 10, 30, and 60 minutes after treatment with small-molecule VEGFR2 inhibitor ki8751, monoclonal antibody bevacizumab, or designed VEGF-A binders. Values are normalized to pre-treatment values. Shown are the mean and S.E.M of 3 (for binders) or 6 (for controls) biological replicates. (D) Western blot of phosphorylated and total ERK, AKT, and VEGFR2 from HUVEC cells after 2 to 60 minutes of treatment with VEGF-A and binders GDM\_VEGFA\_54, GDM\_VEGFA\_71, ki8751, or bevacizumab. Inhibition of VEGF-A signaling is observed by a reduction in pERK, pAKT, and pVEGFR2 band intensity relative to VEGF-A-only ("no inhibitor") control.



**Figure 5 | Experimental structures of binders to SARS-CoV-2 spike and VEGF-A.**

(A) Cryo-EM structures of designed binders (blue) in complex with SARS-CoV-2 spike protein (yellow), aligned to AF2-multimer prediction (gray) on spike protein. Values are shown for the cryo-EM structure resolution and target-aligned binder  $\text{C}\alpha$  RMSDs between AF2-multimer and experimental structures. (B) Crystal structure of complex between VEGF-A homodimer (yellow) and design GDM\_VEGFA\_71 (blue), aligned to AF2-multimer prediction (gray) on VEGF-A (binder  $\text{C}\alpha$  RMSD = 1.65 Å). (C) Rotated view of binder monomer (binder-aligned binder  $\text{C}\alpha$  RMSD = 0.78 Å). (D-E) Closeup of the binder-target interface showing close agreement of sidechains between experimental structure and AF2-multimer prediction of design. (D) Packing of hydrophobic sidechains of the binder at the interface. Most have near-perfect agreement between design and structure, except Val17, Ile19, and Ile81, which have slight deviations. (E) A designed hydrogen bond between His24 of the binder and Tyr12 of VEGF-A.

### 3. Conclusion

Our results show that AlphaProteo is capable of generating low- to sub-nanomolar binders for a diverse range of targets after a single round of medium-throughput testing. The binders are small (5-15 kDa), thermostable, and highly expressed, and therefore potentially already suitable for use in some research applications without further optimization. However, it is important to note that we have experimentally validated relatively few targets in this work, and all our binders are designed using a target crystal structure as input. We hope to further improve AlphaProteo's performance and expand its capabilities to address a wider range of binder design problems, including challenging targets such as TNF $\alpha$  as well as those which lack experimental structures or a single well-defined conformation. We believe that AlphaProteo will unlock new solutions for many biological applications, such as controlling cell signaling, imaging proteins, cells, and tissues, conferring target specificity to various effector systems, and beyond.

### Additional notes

The contents of this report are intended for research purposes only, and not for clinical use. This report does not include machine learning methods due to biosecurity and commercial considerations. We are looking to develop a safe and responsible protein design offering for the community, informed by our work and consultations on biosecurity and safety.

### Acknowledgements

The authors would like to thank the following people for their input and feedback: Jonas Adler, Andy Ballard, Charlie Beattie, David Belanger, Lucy Colwell, Andrew Cowie, Sarah Elwes, Richard Evans, Conor Griffin, John Jumper, Svend Kjær, Antonia Paterson, Matteo Perino, Francesca Pietra, Uchechi Okereke, Olaf Ronneberger, Freyr Sverrisson, Nick Swanson, Kathryn Tunyasuvunakool, Augustin Židek. We would also like to thank Dane Wittrup (Dept. of Chemical Engineering, Massachusetts Institute of Technology) for his generous gift of yeast vector pCTcon2 and Svend Kjær (Structural Biology Science Technology Platform, The Francis Crick Institute) for his production of the SARS-CoV-2 spike protein.

### Contributions

Machine learning model development, generation of design candidates, experimental success rate, experimental binding affinity measurements, and VEGF-A binder crystal structure determination were performed by Google DeepMind.

Cell-based assays and cryo-EM structure determination were performed by research groups at The Francis Crick Institute, London, UK.

## References

- [1] Josh Abramson et al. “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. In: *Nature* 630.8016 (2024), pp. 493–500. DOI: 10.1038/s41586-024-07487-w.
- [2] Minkyung Baek et al. “Accurate prediction of protein structures and interactions using a three-track neural network”. In: *Science* 373.6557 (2021), pp. 871–876. DOI: 10.1126/science.abj8754.
- [3] Stephanie Berger et al. “Preclinical proof of principle for orally delivered Th17 antagonist miniproteins”. In: *Cell* 187.16 (2024), 4305–4317.e18. DOI: 10.1016/j.cell.2024.05.052.
- [4] Patrick Bryant, Gabriele Pozzati, and Arne Elofsson. “Improved prediction of protein-protein interactions using AlphaFold2”. In: *Nat. Commun.* 13.1 (2022). DOI: 10.1038/s41467-022-28865-w.
- [5] Longxing Cao et al. “De novo design of picomolar SARS-CoV-2 miniprotein inhibitors”. In: *Science* 370.6515 (2020), pp. 426–431. DOI: 10.1126/science.abd9909.
- [6] Longxing Cao et al. “Design of protein-binding proteins from the target structure alone”. In: *Nature* 605.7910 (2022), pp. 551–560. DOI: 10.1038/s41586-022-04654-9.
- [7] Alexander E Chu, Tianyu Lu, and Po-Ssu Huang. “Sparks of function by de novo protein design”. In: *Nat. Biotechnol.* 42.2 (2024), pp. 203–215. DOI: 10.1038/s41587-024-02133-2.
- [8] J Dauparas et al. “Robust deep learning-based protein sequence design using ProteinMPNN”. In: *Science* 378.6615 (2022), pp. 49–56. DOI: 10.1126/science.add2187.
- [9] Richard Evans et al. “Protein complex prediction with AlphaFold-Multimer”. In: *bioRxiv* (2021). DOI: 10.1101/2021.10.04.463034.
- [10] Sarel J Fleishman et al. “Computational design of proteins targeting the conserved stem region of influenza hemagglutinin”. In: *Science* 332.6031 (2011), pp. 816–821. DOI: 10.1126/science.1202617.
- [11] Pablo Gainza et al. “De novo design of protein interactions with learned surface fingerprints”. In: *Nature* 617.7959 (2023), pp. 176–184. DOI: 10.1038/s41586-023-05993-x.
- [12] Michaela Gebauer and Arne Skerra. “Engineered protein scaffolds as next-generation therapeutics”. In: *Annu. Rev. Pharmacol. Toxicol.* 60.1 (2020), pp. 391–415. DOI: 10.1146/annurev-pharmtox-010818-021118.
- [13] Odessa J Goudy et al. “In silico evolution of autoinhibitory domains for a PD-L1 antagonist using deep learning models”. In: *Proc. Natl. Acad. Sci. U. S. A.* 120.49 (2023). DOI: 10.1073/pnas.2307371120.
- [14] Thomas Hayes et al. “Simulating 500 million years of evolution with a language model”. In: *bioRxiv* (2024). DOI: 10.1101/2024.07.01.600583.
- [15] Hennie R Hoogenboom. “Selecting and screening recombinant antibody libraries”. In: *Nat. Biotechnol.* 23.9 (2005), pp. 1105–1116. DOI: 10.1038/nbt1126.
- [16] Shi Hu et al. “Comparison of the inhibition mechanisms of adalimumab and infliximab in treating tumor necrosis factor alpha-associated diseases from a molecular view”. In: *J. Biol. Chem.* 288.38 (2013), pp. 27059–27067. DOI: 10.1074/jbc.M113.491530.
- [17] Ian R Humphreys et al. “Computed structures of core eukaryotic protein complexes”. In: *Science* 374.6573 (2021). DOI: 10.1126/science.abm4805.
- [18] John B Ingraham et al. “Illuminating protein space with a programmable generative model”. In: *Nature* 623.7989 (2023), pp. 1070–1078. DOI: 10.1038/s41586-023-06728-8.

- [19] Joël Janin, Ranjit P Bahadur, and Pinak Chakrabarti. “Protein–protein interaction and quaternary structure”. In: *Q. Rev. Biophys.* 41.2 (2008), pp. 133–180. DOI: 10.1017/S0033583508004708.
- [20] Yanan Jia et al. “Effect of bevacizumab on the tight junction proteins of vascular endothelial cells”. In: *Am. J. Transl. Res.* 11.9 (2019), pp. 5546–5559.
- [21] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.
- [22] Oleg Kovalevskiy, Juan Mateos-Garcia, and Kathryn Tunyasuvunakool. “AlphaFold two years on: Validation and impact”. In: *Proc. Natl. Acad. Sci. U. S. A.* 121.34 (2024), e2315002121. DOI: 10.1073/pnas.2315002121.
- [23] I Krämer and H-P Lipp. “Bevacizumab, a humanized anti-angiogenic monoclonal antibody for the treatment of colorectal cancer”. In: *J. Clin. Pharm. Ther.* 32.1 (2007), pp. 1–14. DOI: 10.1111/j.1365-2710.2007.00800.x.
- [24] Rohith Krishna et al. “Generalized biomolecular modeling and design with RoseTTAFold All-Atom”. In: *Science* 384.6693 (2024), ead12528. DOI: 10.1126/science.ad12528.
- [25] Kazuo Kubo et al. “Novel potent orally active selective VEGFR-2 tyrosine kinase inhibitors: synthesis, structure-activity relationships, and antitumor activities of N-phenyl-N’-{4-(4-quinolyloxy)phenyl}ureas”. In: *J. Med. Chem.* 48.5 (2005), pp. 1359–1366. DOI: 10.1021/jm030427r.
- [26] J P Landry et al. “Measuring affinity constants of 1450 monoclonal antibodies to peptide targets with a microarray-based label-free assay platform”. In: *J. Immunol. Methods* 417 (2015), pp. 86–96. DOI: 10.1016/j.jim.2014.12.011.
- [27] S Lien and H B Lowman. “Therapeutic Anti-VEGF Antibodies”. In: *Therapeutic Antibodies*. Ed. by Yuti Chernajovsky and Ahuva Nissim. Berlin, Heidelberg: Springer, 2008, pp. 131–150. DOI: 10.1007/978-3-540-73259-4\_6.
- [28] Zeming Lin et al. “Evolutionary-scale prediction of atomic-level protein structure with a language model”. In: *Science* 379.6637 (2023), pp. 1123–1130. DOI: 10.1126/science.ad2574.
- [29] Anthony Marchand, Alexandra K Van Hall-Beauvais, and Bruno E Correia. “Computational design of novel protein–protein interactions – An overview on methodological approaches and applications”. In: *Curr. Opin. Struct. Biol.* 74.102370 (2022), p. 102370. DOI: 10.1016/j.sbi.2022.102370.
- [30] David McMillan et al. “Structural insights into the disruption of TNF-TNFR1 signalling by small molecules stabilising a distorted TNF”. In: *Nat. Commun.* 12.1 (2021), p. 582. DOI: 10.1038/s41467-020-20828-3.
- [31] Yohei Mukai et al. “Solution of the structure of the TNF-TNFR2 complex”. In: *Sci. Signal.* 3.148 (2010), ra83. DOI: 10.1126/scisignal.2000954.
- [32] Yves A Muller et al. “VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 Å resolution and mutational analysis of the interface”. In: *Structure* 6.9 (1998), pp. 1153–1167. DOI: 10.1016/S0969-2126(98)00116-6.
- [33] Serge Muyldermans. “Applications of Nanobodies”. In: *Annu. Rev. Anim. Biosci.* 9.1 (2021), pp. 401–421. DOI: 10.1146/annurev-animal-021419-083831.
- [34] Pascal Notin et al. “Machine learning for functional protein design”. In: *Nat. Biotechnol.* 42.2 (2024), pp. 216–228. DOI: 10.1038/s41587-024-02127-0.



- [35] Erik Procko et al. “A computationally designed inhibitor of an Epstein-Barr viral bcl-2 protein induces apoptosis in infected cells”. In: *Cell* 157.7 (2014), pp. 1644–1656. DOI: 10.1016/j.cell.2014.04.034.
- [36] Linghui Qian et al. “The dawn of a New Era: Targeting the “undruggables” with antibody-based therapeutics”. In: *Chem. Rev.* 123.12 (2023), pp. 7782–7853. DOI: 10.1021/acs.chemrev.2c00915.
- [37] Jeffrey A Ruffolo et al. *Design of highly functional genome editors by modeling the universe of CRISPR-Cas sequences*. 2024. DOI: 10.1101/2024.04.22.590591.
- [39] Daniel-Adriano Silva et al. “De novo design of potent and selective mimics of IL-2 and IL-15”. In: *Nature* 565.7738 (2019), pp. 186–191. DOI: 10.1038/s41586-018-0830-7.
- [40] William R Strohl. “Structure and function of therapeutic antibodies approved by the US FDA in 2023”. In: *Antib. Ther.* 7.2 (2024), pp. 132–156. DOI: 10.1093/abt/tbae007.
- [41] Susana Vázquez Torres et al. “De novo design of high-affinity binders of bioactive helical peptides”. In: *Nature* 626.7998 (2024), pp. 435–442. DOI: 10.1038/s41586-023-06953-1.
- [42] Alexandra C Walls et al. “Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein”. In: *Cell* 181.2 (2020), 281–292.e6. DOI: 10.1016/j.cell.2020.02.058.
- [43] Joseph L Watson et al. “De novo design of protein structure and function with RFdiffusion”. In: *Nature* (2023). DOI: 10.1038/s41586-023-06415-8.
- [44] Kejia Wu et al. “Sequence-specific targeting of intrinsically disordered protein regions”. In: *bioRxiv* (2024). DOI: 10.1101/2024.07.15.603480.
- [45] Wei Yang et al. “Design of high affinity binders to convex protein target sites”. In: *bioRxiv* (2024). DOI: 10.1101/2024.05.01.592114.

## Supplementary information

### S1. Experimental methods

#### S1.1. Target protein expression and purification

Purified protein stocks for IL-7RA(21-239), TrkA(34-423), PD-L1(19-239), VEGF-A(27-191), and IL-17A(24-155) were purchased from BioTechne, with catalog numbers AVI10317, AVI11378, AVI156, AVI293, and BT7955, respectively. IL-7RA, PD-L1, and TrkA have C-terminal Fc and biotinylated Avi tags, while VEGF-A has a biotinylated C-terminal Avi tag and IL-17A is biotinylated via sugars. VEGF-A and IL-17A are disulfide-linked homo-dimers. For X-ray crystallography, VEGF165 (Uniprot P15692-4) was purchased from Qkine, with catalog number Qk048.

For BHRF1, a recombinant protein construct (Uniprot P03182, residues 2-160) was produced with an N-terminal Twin-Strep tag and a 3C protease cleavage site. Transformed BL21 (DE3) (Thermo Scientific) cultures were grown in Terrific Broth (TB) medium (Melford) supplemented with carbenicillin (50 µg/mL) at 37 °C with shaking. At  $OD_{600} = \sim 0.6$ , protein expression was induced with 0.1 mM IPTG, the temperature reduced to 21 °C and cultures were grown overnight.

Cells were harvested and resuspended in 20 mM Tris pH 8.0, 300 mM NaCl supplemented with 0.5 mg/mL lysozyme, 100 U DNase I, 1 mM  $MgCl_2$  and a cOMplete EDTA-free protease inhibitor tablet (Roche) at a 1:5 cell weight to buffer ratio. Cell lysis was achieved by sonicating the cell suspension at 40% amplitude (15 seconds on / 45 seconds off) for 24 cycles on ice. Lysate was centrifuged at 48,000 x g for 45 min at 4 °C and the supernatant was recovered and filtered through a 0.45 µm filter (Sartorius). The sample was applied to a 5 mL StrepTrap XT column (Cytiva) pre-equilibrated with Strep binding buffer (100 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM TCEP) using an AKTA Pure 25 M. Following sample application, the column resin was washed with 10 column volumes (CV) of the same buffer before the protein was eluted with 10 CV of 1x BXT elution buffer (IBA Lifesciences) supplemented with 0.5 mM TCEP. 1 CV fractions were collected and assessed via SDS-PAGE to confirm presence of the protein of interest. BHRF1 was pooled and concentrated using a 10 kDa MWCO concentrator (Vivaspin). The sample was further purified by size exclusion chromatography (SEC) using a Superdex 75 increase 10/300 GL column pre-equilibrated with 20 mM sodium phosphate pH 7.5, 0.5 mM TCEP. Fractions were confirmed by SDS-PAGE and the concentration was measured by absorbance at 280 nm using a NanoDrop One (Thermo Scientific) and the BHRF1 construct's theoretical extinction coefficient [S27]. Purified protein was aliquoted and stored at -80 °C.

For SC2RBD, a recombinant protein construct (NCBI reference NC\_045512, residues 319-541) of SARS-CoV-2 Spike S1 glycoprotein corresponding to the receptor binding domain was produced with a C-terminal Twin-Strep tag. The signal peptide from immunoglobulin kappa gene product (METDTLLLWVLLLWVPGSTGD) was used to direct secretion of the construct. The corresponding codon-optimized DNA fragment was cloned into mammalian expression vector pQ-3C-2xStrep for expression in Expi293F cells. Expi293F cells grown at 37 °C in 5% CO<sub>2</sub> in shake flasks containing FreeStyle 293 medium were transfected with endotoxin free plasmid preparation using ExpiFectamine reagent (Thermo Fisher Scientific). Conditioned medium was harvested 4 and 8 days post-transfection. Recombinant protein was captured on Streptactin XT (IBA LifeSciences) affinity resin. Following extensive washes in TBSE buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA), the protein was eluted in 1x BXT buffer (IBA LifeSciences) and further purified by SEC using a Superdex 200 16/600 column (GE Healthcare) in TBSE buffer. The purified protein was concentrated using a 10 kDa MWCO concentrator (Sartorius), aliquoted, snap-frozen in liquid nitrogen and stored at -80 °C.

## S1.2. Yeast surface display and flow cytometry

### S1.2.1. Primary binding screen

Binder design sequences were codon-optimized by DNAworks [S14] and most were synthesized by Twist as gene fragments flanked by *BsaI* restriction sites as well as homology regions to a modified pETcon vector (pCTcon2, a generous gift from K. Dane Wittrup, MIT). *Saccharomyces cerevisiae* strain EBY100 cells (50  $\mu$ L) were transformed using a modified lithium acetate method without using the single-strand carrier DNA [S11] with 50 ng of linearised plasmid and a minimum of 10 ng of gene fragment insert in a 96-well plate. Cells were grown at 30 °C shaking at 1,000 rpm in complete synthetic medium -Trp -Ura + 2% glucose for 48-72 hours. For protein expression, a volume of yeast cells were centrifuged at 1,800 x g for 5 minutes at 20 °C and resuspended in 1 mL complete synthetic medium + 0.1% glucose + 2% galactose (SGCAA) to OD<sub>600</sub> = 1.0. Cells were incubated at 30 °C overnight and a volume of cells at OD<sub>600</sub> = 0.4 were washed twice with 200  $\mu$ L of 1x PBS + 0.1% BSA (PBSF), centrifuged at 1,800 x g for 3 minutes at 20 °C and the supernatant was removed.

To screen for binding, yeast cells were then incubated with biotinylated target proteins (diluted in PBSF) for 1 hour, washed twice with PBSF and incubated with 25  $\mu$ g/mL fluorescein isothiocyanate (FITC)-conjugated anti-Myc antibody (FITC-Ab) (Abcam) and 30  $\mu$ g/mL streptavidin-phycoerythrin (SAPE, Thermo Fisher Scientific) for 30 minutes. For VEGF-A and IL-17A, an avidity method with increased sensitivity was used; target proteins were pre-incubated with 25  $\mu$ g/mL FITC-Ab and 30  $\mu$ g/mL SAPE for 30 minutes before incubating with cells. Following binding, cells were washed once with PBSF and resuspended in 200  $\mu$ L of PBSF. Cells were analyzed on the CytoFlex LX (Beckman Coulter) or ZE5 Cell Analyzer (Bio-Rad) flow cytometers by measuring fluorescence of FITC and phycoerythrin (PE) to detect binder expression and target binding respectively.

Flow cytometry data were analyzed to compute a "binding signal", defined as:

$$\text{signal} = (\log_{10} \text{PE}_{\text{FITC}+, +\text{target}} - \log_{10} \text{PE}_{\text{FITC}-, +\text{target}}) - (\log_{10} \text{PE}_{\text{FITC}+, -\text{target}} - \log_{10} \text{PE}_{\text{FITC}-, -\text{target}})$$

where  $\text{PE}_{\text{FITC}+, +\text{target}}$  is the mean PE (binding) signal of the FITC+ (binder-expressing) subpopulation in a well where target protein has been added (Figure S3A, Figure S3B). "FITC-" indicates the non-binder-expressing cell population, and "-target" indicates a control well containing the same binder but to which no target has been added. FITC+ and FITC- cells are identified by k-means clustering. This metric captures the shift in PE signal due to binder expression and target binding in excess of the PE shift due to binder expression alone or target binding alone, thus controlling for experiment artifacts which could lead to false positives. Designs with a binding signal > 0.2 were considered successful binders, except in the case of IL-17A, where this threshold was set to 1.3 to account for background binding. These thresholds were calibrated manually by visually inspecting scatterplots of the raw yeast data.

### S1.2.2. Interface mutation, competitive inhibition, and specificity experiments

Interface mutations were selected by manual visual inspection of predicted structures of the designed binder-target complexes. We generated single-mutants with a hydrophobic residue (alanine, valine, leucine, and isoleucine) on the target-facing interface of the binder changed to a charged residue (aspartate, glutamate, arginine, lysine), as well as a small number of multiple-mutants with combinations of the single mutations. Mutants were screened following the same method as the primary binding screen.

For competition assays, the competitor protein used for BHRF1, SC2RBD, IL-7RA, PD-L1, VEGF-A, and IL-17A, respectively, are BINDI [S21], LCB1 [S5], RFD\_IL7RA\_55, RFD\_PDL1\_76, RFD\_TrkA\_88 [S28],

VEGFR1 (ACROBiosystems, VE1-H52H9) and IL-17R (Biotechne, 11234-IR-100). Yeast cells were incubated with biotinylated target proteins with or without a competitor protein for 1 hour (the competitor protein was added to the biotinylated target protein master mix just before adding to the cells). The cells were then washed twice with PBSF and incubated with 25  $\mu\text{g}/\text{mL}$  FITC-Ab (Abcam) and 30  $\mu\text{g}/\text{mL}$  SAPE (Thermo Fisher Scientific) for 30 minutes. For VEGF-A and IL-17A, an avidity method with increased sensitivity was used, similar to the primary binding screen; target proteins were pre-incubated with 25  $\mu\text{g}/\text{mL}$  FITC-Ab and 30  $\mu\text{g}/\text{mL}$  SAPE for 30 minutes before adding competitor protein and incubating with cells. Following binding, cells were washed once with PBSF and resuspended in 200  $\mu\text{L}$  of PBSF.

To test for specificity, 1 nM of each binder was tested for binding against 100 nM target using a homogeneous time resolved fluorescence (HTRF) assay readout and similar methods to those described in the HTRF methods section below.

### **S1.3. Designed binder expression and purification**

Designed binders with the highest binding signal by yeast display (Figure S3) were selected for *E. coli* expression and follow up experiments. Designs purchased as gene fragments were cloned into a modified pTriEx-4 vector containing an N-terminal 8-His tag and a 3C protease cleavage site using NEBridge Golden Gate cloning (NEB) at *BsaI* sites, transformed into DH5- $\alpha$  competent cells (Thermo Scientific), miniprep (Qiagen), and verified by Sanger sequencing (Azenta). A small number of designs were purchased from Twist Bioscience directly as cloned plasmids in pTriEx-4 or pET-29b. For expression, plasmids were transformed into BL21 (DE3) cells and the entire transformation mix inoculated into autoinduction medium consisting of TB medium, 0.05% glucose, 0.2% alpha-lactose, and 50  $\mu\text{g}/\text{mL}$  carbenicillin or 50  $\mu\text{g}/\text{mL}$  kanamycin. Cultures were incubated at 37  $^{\circ}\text{C}$  with shaking (220 or 1000 rpm) for 24 hours, harvested at 2,568 x *g* for 10 minutes, and pellets stored at -80  $^{\circ}\text{C}$  until purification. Cell pellets were chemically lysed using BugBuster Master Mix (Novagen) supplemented with cComplete EDTA-free protease inhibitor (Roche) with shaking for 20 minutes at room temperature. Lysates were clarified by centrifugation for 1 hour at 2,568 x *g*, then purified by immobilized metal affinity chromatography (IMAC) using Ni-NTA in either 0.1 mL spin columns (Cytiva) or in HisPur™ Ni-NTA 96-well Spin Plates (Thermo Scientific), followed by SEC on an AKTA Pure 25 M (Cytiva) equipped with an ALIAS autosampler (Spark Holland) using a Superdex 75 increase 10/300 GL column equilibrated in 20 mM sodium phosphate pH 7.5. Protein samples were analyzed by SDS-PAGE, and where required, concentrated using a 3 kDa MWCO Vivaspinn concentrator (Cytiva). Protein concentrations were measured in triplicate by absorbance at 280 nm with a NanoDrop One (Thermo Scientific) using theoretical extinction coefficients [S27]. Binders that have no theoretical extinction coefficients were assayed (BCA assay, Thermo Scientific) or Bradford assay (Thermo Scientific). Purified proteins were aliquoted and stored at -80  $^{\circ}\text{C}$  until further use.

Where larger quantities of designed binders were required, for example for CD and X-ray crystallography experiments, expression was scaled up to 100-1000 mL BL21 (DE3) cultures and the above protocol followed with minor modifications. Cells were lysed using sonication on ice, and lysates were clarified by centrifugation at 48,000 x *g* for 45 minutes at 4  $^{\circ}\text{C}$  before being applied to 5 mL Ni-NTA column (Cytiva), followed by SEC.

### **S1.4. Measurement of binding affinity / binding dissociation constants ( $K_D$ )**

#### **S1.4.1. Homogeneous Time Resolved Fluorescence (HTRF)**

Binding affinities ( $K_D$ s) were measured in equilibrium saturation-binding experiments with fixed binder design concentration and target titration. The total assay volume was 16  $\mu\text{L}$  and all proteins

and reagents were diluted in PPI europium detection buffer (Revvity). Target protein was premixed with HTRF acceptor reagent Streptavidin-d2 (Revvity), serially diluted, and transferred to a white ProxiPlate 384-shallow well microplate ('assay plate', Revvity). Subsequently, 1 nM of each binder was added to the assay plate in duplicate (binders with  $K_D < 0.5$  nM were later re-assayed with 0.1 nM binder to ensure robust data fitting). The assay plate was centrifuged at 500 x g for 30 seconds, sealed and incubated at room temperature for between 30 minutes and 1 hour. HTRF donor mAb Anti-6HIS-Eu Gold (Revvity) was then added to a final concentration of 2 nM (1x), using a Mantis microfluidic liquid dispenser (Formulatrix) running software version 5.1.1 on Windows 10. The assay plate was centrifuged, sealed and incubated for a further 1 hour at room temperature. HTRF signal was measured using a PHERAstar FSX (BMG) plate reader equipped with an HTRF 337/665/620 optic module running software version 5.70 R6 on Windows 10. The measurement conditions were as follows; 60  $\mu$ s integration delay, 400  $\mu$ s integration time, 60 flashes. The optimal focal (Z) height was determined using channel B for each experiment. HTRF ratios were calculated by dividing the acceptor signal at 665 nm by the donor signal at 620 nm, and multiplying by a factor of 10,000. Mean background signal for each target-acceptor concentration (0 nM binder) was subtracted, and data were analyzed using custom Python code by fitting to the general 1:1 binding equation.

$$R = \frac{R_{\max}}{2B} \left( B + A + K_D - \sqrt{(B + A + K_D)^2 - 4AB} \right)$$

where R is the measured equilibrium HTRF signal, A and B are the titrated and fixed binding partner concentrations, respectively, and  $R_{\max}$  and  $K_D$  are the fitted maximal HTRF signal and binding dissociation constants, respectively. We used this equation because some of our binders had  $K_D$  values close to or lower than the fixed binder concentration used in the experiment, which causes the more common hyperbolic equation of 1:1 binding to overestimate the true  $K_D$  [S15]. To ensure reliable model fitting, we always used a fixed binder concentration no more than 2-fold higher (and usually much lower) than the estimated  $K_D$  [S15].

#### S1.4.2. Bio-Layer Interferometry (BLI)

For selected controls and designs, we measured  $K_D$ s by kinetic BLI assays to establish confidence in the HTRF results (see "Comparison of binding affinity ( $K_D$ ) to other methods" above). Data were collected on the Octet R8 (Sartorius AG, Göttingen, Germany) using the integrated Octet Discovery software version 12.2.2.20. Recombinant proteins were diluted from concentrated frozen stocks in 20 mM sodium phosphate pH 7.5, 0.05% Tween-20 (BLI buffer). A seven-point dilution series of the analyte protein was also prepared in BLI buffer to create a titration curve. Ni-NTA biosensors (Sartorius, catalog number 18-5102) were pre-equilibrated in BLI buffer for at least 10 minutes prior to starting the experiment. A fixed concentration of "ligand" (8His-tagged binder) was loaded onto sensors for 120-240 seconds, briefly washed for 10 seconds, followed by a 60 second baseline. Association of a titration series of "analyte" (target protein) was then performed for 90-420 seconds, followed by dissociation for 600-1200 seconds. All steps were performed at 25 °C and with shaking at 1000 rpm. Loading, association, and dissociation durations were optimized for each binder-target pair. Data were processed using Octet Analysis Studio (version 12.2.2.26). Measurements from reference sensors not loaded with ligand, as well as a reference well with 0 nM analyte, were subtracted from the final data to account for non-specific binding of analyte to the sensors and baseline drift due to unloading of ligand from sensors, respectively. Baseline (pre-association) signal was aligned to 0 before final analysis, where kinetic constants were obtained by nonlinear regression of 1:1 or 2:1 binding equations to the data. Fits were performed globally, over both association and dissociation, with a shared  $R_{\max}$  for all analyte concentrations.

### S1.5. Circular dichroism (CD) spectroscopy

Data were collected on a Jasco J-815 circular dichroism spectrometer, running software Spectra Manager Version 2.15.20, equipped with a PTC-348 temperature control device. Far-UV spectra (260-190 nm) and thermal unfolding measurements were recorded in 1 mm quartz glass cuvettes (Hellma) containing protein solutions at 10  $\mu$ M in 20 mM sodium phosphate pH 7.5. Baselines containing 20 mM sodium phosphate pH 7.5 were collected prior to sample analysis.

Spectra were recorded in the far-UV (260-190 nm) at 20 °C with a scanning speed of 200 nm/min and a digital integration time (DIT) of 0.25 seconds. 25 accumulations (spectral scans) were recorded and automatically averaged by the software. Thermal unfolding data were recorded at 222 nm between 2-95 °C at a ramp rate of 2 °C/min, with measurements recorded at 0.2 °C intervals. The DIT was set to 4 seconds. Following thermal unfolding measurements, spectra in the far-UV were collected at 95 °C to measure CD spectra changes post thermal unfolding. Additional CD spectra were then collected following cooling of the same samples to 20 °C, to observe refolding.

Data processing: For spectra, the buffer baseline scans were subtracted from each sample dataset, the final 15 nm of measurements (between 260-245 nm) were normalized, and the CD signals in mdeg were converted to  $\Delta\epsilon$  ( $M^{-1}cm^{-1}$ ).

### S1.6. Western blot analysis of VEGF-A signaling in HUVECs

Human umbilical vein endothelial cells (HUVEC, PromoCell #C-12008) were cultured in ECG Medium 2 KIT (PromoCell, #C-22111) according to the manufacturer's instructions. Cells were plated at passage 6 in a six well format. The following day cells were starved for 4 hours in Endothelial Cell Basal Medium 2 (PromoCell, #C-22211). Control inhibitors or binder proteins were added after 3 hours of starvation for the remaining hour at the following concentration: 1  $\mu$ M ki8751 (Bio-Techne, #2542/10), 1  $\mu$ M bevacizumab (Biosynth, #FB76708), or 1  $\mu$ M binders GDM\_VEGFA\_54, GDM\_VEGFA\_71. Following the 1 hour treatment, cells were stimulated with 30 ng/mL hVEGF-A (Peprotech, #100-20-10UG) for 2, 5, 10, 30 or 60 minutes (0 minute refers to untreated). Subsequently, cells were washed with ice cold PBS and frozen at -80 °C.

Cells were lysed in 60  $\mu$ l/well D0.4 lysis buffer (20 mM HEPES pH 7.5, 0.4 M NaCl, 10% glycerol, 0.4% Triton X-100, 10 mM EGTA and 5 mM EDTA, 1x HALT protease inhibitor (Thermo Fisher Scientific, #87786), 1x HALT phosphatase inhibitor (Thermo Fisher Scientific, #78420), 1 mM DTT, 25 mM NaF and 25 mM sodium-b-glycerophosphate. The protein concentration of each sample was measured using the Pierce™ BCA assay (Thermo Fisher Scientific, #10678484) and the protein concentration from each replicate was adjusted to the same concentration per sample.

Western blots were performed using 4-12% Bis-Tris SDS-PAGE gels (Thermo Fisher Scientific, #NP0336BOX) and blotted on a 0.2  $\mu$ m NC2 nitrocellulose membrane. Membranes were cut at 100 kDa and blocked with 5% BSA in TBS-Tween (0.1%). The following antibodies were used at 1:500 concentration and incubated overnight at 4 °C: phospho-VEGF Receptor 2 (Tyr1175) (Cell Signaling Technologies [CST], #3770), VEGF Receptor 2 (CST, #2479), phospho-p44/42 MAPK (Erk1/2) (CST, #4377), p44/42 MAPK (Erk1/2) (CST, #9102), phospho-Akt (Ser473) (CST, #4060), Akt (CST, #4691) and GAPDH (Novus Biologicals, #NB300-221). The following HRP-conjugated secondary antibodies were used at 1:5000 concentration: donkey anti-rabbit IgG (Abcam, #ab16284) and donkey anti-mouse IgG (Abcam, #ab98799). Blots were developed using HRP substrate (Millipore, #11556345) and were developed on an Amersham Imagequant 800. The phosphorylated version of each protein was detected on a different membrane than the non-phosphorylated protein. The mean intensity for each band was measured using the same quantification area and the ratio of phosphorylated to non-phosphorylated protein was calculated.

### S1.7. SARS-CoV-2 virus neutralization assay

Experiments were performed by the Francis Crick Institute COVID Surveillance Unit following the protocol outlined in [S26]. Briefly, 10-point binder dose response series were generated by serially diluting each binder in duplicate in 20 mM sodium phosphate buffer before diluting further to achieve final testing concentrations of 1.7-11,200 ng/mL in 10% fetal bovine serum (FBS). With appropriate positive and negative controls, binder dose response series were then run through the standard live-virus neutralization assay against 2 variants of concern (VOC) and 2 variants of interest. Duplicate assay plates were run, so each biological repeat contained 4 technical replicates. Two biological repeats were run on separate days using different flasks of cells, vials of virus, and bottles of media. Thus each plot of Supplementary Figure S11 consists of 160 independent data points. The data points were generated from 4 replicates of 40 independent titrations. EC<sub>50</sub> values were calculated using nonlinear regression with a 4-parameter dose response curve fit.

### S1.8. Cryo-EM sample preparation, data collection and image processing

The Spike ectodomain construct used in the cryo-EM experiments was based on the Wuhan SARS-CoV-2 isolate. The protein (spanning residues 1-1208 from UniProt ID YP\_009724390) harbored point mutations K986P and V987P stabilizing the pre-fusion conformation, disrupted furin cleavage site, C-terminal T4 fibrin trimerization domain, and a hexa-histidine affinity tag [S30]. The protein was produced by expression in stably transformed Expi293F cells and purified by capture onto immobilized Ni affinity resin, followed by SEC, as previously described [S30, S22].

Four  $\mu$ L freshly isolated trimeric SARS-CoV-2 Spike ectodomain (1.2 mg/mL), supplemented with 0.2 mg/mL GDM\_SC2RBD\_104, GDM\_SC2RBD\_50, GDM\_SC2RBD\_11, or GDM\_SC2RBD\_129 and 0.1% n-octyl glucoside in 150 mM NaCl, 20 mM Tris-HCl, pH 8.0, was spotted onto fresh 400-mesh R1.2/1.3 C-flat holey carbon grids (Electron Microscopy Sciences product CF413-50-Au for 1 minute, under 100% humidity at 20 °C, prior to blotting and plunge-freezing in liquid ethane-propane using Vitrobot Mark IV (Thermo Fisher Scientific). Cryo-EM data were acquired on a Titan Krios G2 cryo-electron microscope equipped with a Falcon 4i direct electron detector (Thermo Fisher Scientific). Selectris energy filter (Thermo Fisher Scientific) with a slit width of 10 eV was used for imaging complexes containing GDM\_SC2RBD\_11 and GDM\_SC2RBD\_129. A total of 4500, 8342, 6728, and 8482 micrograph movies were recorded from grids containing GDM\_SC2RBD\_104, GDM\_SC2RBD\_50, GDM\_SC2RBD\_11, and GDM\_SC2RBD\_129, respectively. Data collections proceeded with a defocus range set to -1.5 to -3.5  $\mu$ m and a magnification corresponding to calibrated pixel size of 1.08 Å (GDM\_SC2RBD\_104 and GDM\_SC2RBD\_50) or 0.95 Å (GDM\_SC2RBD\_11 and GDM\_SC2RBD\_129) (Table S5).

1,674 EER frames recorded per micrograph movie were processed into 31 fractions, with an exposure dose of 1.04 e/Å<sup>2</sup> (GDM\_SC2RBD\_104 and GDM\_SC2RBD\_50) or 1.25 e/Å<sup>2</sup> (GDM\_SC2RBD\_11 and GDM\_SC2RBD\_129) per fraction. The micrograph frames were aligned, summed and weighted as implemented in Relion-5.0beta [S18, S32], and contrast transfer function parameters were estimated using Gctf-v1.18 [S31]. Reference-free 2D classification of an initial subset of particles picked using Gaussian blob function in Relion revealed 2D averages belonging to monomeric S1 protein, due to dissociation of the trimeric Spike. Particles belonging to well-defined 2D classes were used to train Topaz [S4], which was used to pick the entire datasets. The particles, extracted with 4-fold binning, were subjected to three rounds of 2D classification in Relion, using 400 classes in each round; the regularization parameter T was increased from 2 during the first round to 8 in the last round of 2D classification. Particles contributing to well-defined 2D classes, re-extracted with 2-fold binning, were used to generate initial 3D models and subjected to 3D classification into 4-7 classes in Relion, with the regularization parameter T set to 8 (Table S5, Figure S12). The best particle sets were used for

3D reconstruction, followed by Bayesian polishing [S32]. The final reconstructions were obtained using soft masks in conjunction with Blush regularization, as implemented in Relion-5.0beta [S17]. Resolution metrics reported in this work were according to the gold-standard Fourier shell correlation (FSC) 0.143 criterion [S23, S24]. For illustration purposes, cryo-EM maps were locally filtered using EMReady [S13]. Rigid body docking of S1 protein chain (from PDB ID 7ZBU) [S25] and binder models into the final cryo-EM maps was done in Coot [S10], and the figures were prepared using PyMOL Molecular Graphics System, Version 3.0 (Schrödinger, LLC). Final cryo-EM maps will be deposited with the Electron Microscopy Data Bank (EMDB); the raw data will be available upon request.

### **S1.9. X-ray crystallography sample preparation, data processing and structure solving**

GDM\_VEGFA\_71 and VEGF-A were mixed in a molar ratio of 2.5:1, and incubated at room temperature for 1 hour with shaking at 1000 rpm. The GDM\_VEGFA\_71/VEGF-A complex was purified by SEC using a Superdex 200 Increase 10/300 GL column (Cytiva), equilibrated with 20 mM Tris pH 7.5, 150 mM NaCl, and verified by SDS-PAGE. The GDM\_VEGFA\_71/VEGF-A complex was concentrated to 12 mg/mL using a 10 kDa MWCO concentrator (Vivaspin). Crystallisation was performed using a Mosquito crystallization robot (SPT Labtech) by sitting-drop vapor diffusion (50 nL complex + 50 nL crystallization solution) in 3-well crystallization plates (SWISSCI) containing 25  $\mu$ L of crystallization solutions in each reservoir. Crystals of the protein complex grew within two weeks at 20 °C in mother liquor containing 0.1 M phosphate/citrate pH 4.2 and 40% v/v ethanol. Crystals were harvested with 10  $\mu$ m Micromount loops (MiTeGen) and snap-frozen in liquid nitrogen prior to data collection. X-ray data were collected from a single crystal at 100 K on the I04 beamline at Diamond Light Source (Harwell, UK) with a wavelength of 0.9537 Å. All data were automatically processed by xia2 [S12]. Initial phases for the GDM\_VEGFA\_71/VEGF-A complex were obtained by maximum-likelihood molecular replacement using Phaser (version 2.8.3) [S19] from the CCP4 Suite (version 9.0.002) [S2] using the AF3-predicted structure as a search model. The structure solution was subjected to repetitive rounds of restrained refinement using Refmac5 (version 5.8.0430) [S20] and interactive manual building in COOT (version 0.9.8.95) [S9]. NCS and Jelly Body restraints were also used throughout the refinement. The final structure quality at 2.56 Å was assessed using Molprobit (version 0.9.8.95) [S29]. Data collection and refinement statistics are provided in Table S6.

## **S2. Iterative development and *in silico* benchmarking of AlphaProteo**

During development of AlphaProteo, we trained two versions of the generative model (Figure 1A), referred to here as "v1" and "v2". To evaluate these models, we used two different *in silico* benchmarks, each consisting of a set of targets along with a definition of success rate, or the fraction of designs satisfying certain computational success criteria. We compared the v1 and v2 generative models to the current best binder-design method RFdiffusion [S28].

First, we used an existing binder design benchmark based on AlphaFold 2 (AF2), where a designed binder against each of 5 target proteins is considered a success if its AF2 prediction has interchain predicted aligned error < 10, binder-aligned binder RMSD < 1 Å, and pLDDT > 80 (see detailed methods below). These criteria were shown to be highly predictive of experimental binding success [S3]. On this benchmark, the v2 generative model has higher success rates than RFdiffusion on 4 out of 5 targets (Figure S1A). The v1 model outperformed RFdiffusion on 4 of 5 targets when RFdiffusion is run at noise level 1 but underperformed RFdiffusion at noise level 0.

Given that AlphaFold 3 (AF3) is more accurate than AF2 on protein complex prediction, we developed a second benchmark based on AF3 [S1]. On a set of 9 targets, we considered a design a success



if its minimum interchain predicted aligned error  $< 1.5$ , predicted TM-score  $> 0.8$ , and complex RMSD  $< 2.5$ . We found these optimized criteria to be a better proxy of experimental success than AF2 on a published *de novo* binder dataset (Figure S1B, Figure S1C) [S6]. On this benchmark, the v2 model had higher *in silico* success rates than RFdiffusion on all targets and the v1 model outperforms both RFdiffusion variants on six of nine targets (Figure S1C). These conclusions do not change when success rate is adjusted to account for diversity (pairwise TM-score clustering at various thresholds) or novelty (pHMMER bit-score  $< 50$  against Uniref50) (Section S1, Figure S1D). Taken together, these results show that the *in silico* performance of AlphaProteo is at or above the SoTA, consistent with our experimental results.

We experimentally tested a design system containing the v1 model against all 7 targets, a v1-based system with an improved filter against SC2RBD and PD-L1, and a v2-based system on PD-L1 and TrkA (Table S2). Both improving the filter and the model resulted in increased experimental success rates. For simplicity, the results in Table 1 are pooled over all tested designs for each target. Importantly, all designs tested in this work were generated in a "zero-shot" manner, without using any known binder as a starting point.

### S2.1. AF2-based benchmark

We followed published procedures [S28] to run a previously proposed AF2 benchmark [S3] on AlphaProteo designs. This includes generating designs for IL-7RA, PD-L1, TrkA, Insulin receptor, and hemagglutinin H1 using published input specifications (Table S1) [S28], inputting them to the "AlphaFold2 initial guess" script (AF2ig) [S28, S3], and computing the fraction of successful designs, defined as those with interchain AF2 pAE  $< 10$ , binder-aligned binder RMSD  $< 1 \text{ \AA}$ , and pLDDT  $> 80$ . All mentions of "AlphaFold 2" in the context of benchmarks refer to the AF2ig method; we did not run unmodified AF2 for any analyses in this work. We generated 200 designs per target per model. Additionally, we downloaded RFdiffusion<sup>1</sup> and ran it on the benchmark using both noise=0 and noise=1 settings to ensure we could reproduce the published performance of RFdiffusion. As in [S28], we redesigned sequences with ProteinMPNN [S8] with low sampling temperature 0.0001. We present RFdiffusion success rates from both the original publication and from our reproduction.

### S2.2. AF3-based benchmark

To create an AF3-based *in silico* benchmark for binder design, we looked for metrics and thresholds that most enrich for experimental success across 640,000 previously characterized *de novo* binder designs against 11 targets [S6]. We used this dataset because it was not filtered on any AF2- or AF3-based metrics prior to experimental testing and therefore any filters we derive from it have the best chance of generalizing to future design methods. We predicted the structure of each binder-target complex using AF3 while inputting the structure of the target as a template and using only a single sequence (no multiple sequence alignment). We selected the best out of five diffusion head samples using a ranking confidence of  $(0.8 \text{ iptm} + 0.2 \text{ ptm})$ , the individual components of which are described below. We computed the retrospective success rate (fraction of designs with observed binding at 4000 nM target) among the top 1% of designs according to a panel of AF2- and AF3-based metrics (Figure S1B):

---

<sup>1</sup><https://github.com/RosettaCommons/RFdiffusion>

**AF3** (see Supplementary Information of [S1] for details)

1. **ptm**: prediction aligned error (PAE) matrix reduction, maximum average error across aligning on individual residues.
2. **ptm binder / ptm target**: intra-chain reduction of the PAE matrix, aligning on binder / target chain residues and considering errors on the same chain.
3. **iptm**: interchain reduction of the PAE matrix, taking into account only those PAE entries for TM computation that are not on the chain that is being aligned on.
4. **min pae interaction**: minimum value across all interchain terms in the PAE matrix.
5. **rmsd**: root mean squared error between the designed and predicted complex structures.

**AF2 initial guess**

1. **pae binder / pae target**: average of the PAE matrix when only considering the binder / target chain.
2. **pae interaction**: average PAE of the interchain residues.
3. **plddt total**: average plddt over the predicted complex structure.
4. **monomer rmsd**: root mean squared error between the binder design and prediction when aligning on the binder chain.

We developed a new definition of *in silico* success by performing a combinatorial sweep over the following grid of filtering metric thresholds (start, stop, step):

<b>AF3</b>		<b>AF2 initial guess</b>	
min pae interaction	(0, 7, 0.5)	af2 monomer rmsd	(0, 3, 0.5)
ptm binder	(0, 1, 0.05)	pae interaction	(0, 11, 0.5)
rmsd	(0, 3, 0.5)	plddt	(60, 95, 5)

For each target, we ranked the different filter settings according to the binding success rate among passing examples from the data collected by Cao et al. [S6]. We optimized the average per-target rank subject to the constraint that at least 10 designs have to pass filters. We chose to aggregate performance across targets by rank rather than by pass rates due to large variability in the latter. This yielded the following "optimized" filtering thresholds for both AF3 and AF2:

<b>AF3</b>		<b>AF2 initial guess</b>	
min pae interaction	< 1.5	af2 monomer rmsd	< 1.5
ptm binder	> 0.8	pae interaction	< 7.0
rmsd	< 2.5	plddt	> 90

Per-target retrospective success rates based on these filters are shown in Figure S1C. As the AF3 optimized filters slightly outperform the AF2 optimized ones across all targets, we used the AF3 filters to define *in silico* success for a new benchmark. We then computed success rates (using the optimized AF3 thresholds) on designs from AlphaProteo models v1 and v2, as well as our local installation of RFdiffusion. As targets, we selected the original RFdiffusion design targets as well as new targets that we addressed experimentally in this work (Table S1). To account for structural diversity, we used TM-align to compute pairwise TM-scores separately for designs sampled for each target and from each model, and we used a greedy algorithm to cluster these designs at a given TM-score threshold. To account for novelty, we searched each design sequence against Uniref50 using Jackhmmer [S16] and considered it novel if its maximum bit-score is less than 50.

### S3. *In silico* screening of PDB targets

In order to estimate and compare the difficulty of potential future binder design problems to the 8 targets we experimentally evaluated, we computed *in silico* success rates for a random subset of target proteins from the PDB. Starting from 45k clusters derived at 40% sequence homology [S1], we filtered for PDB entries containing up to 5 protein chains and ranging from 30 to 400 residues in length, to exclude those too large to process efficiently. After removing singleton clusters, we sampled 1 protein representative from 200 randomly selected clusters (out of the remaining 6000 final clusters). For each of the 200 proteins, we sampled 1 chain at random to serve as the target, and sampled 3x distinct regions on the protein surface to serve as binding hotspots. Finally, we generated 5000 binders for each target:hotspot combination and computed their *in silico* success rate, or fraction of designs predicted to bind according to the AlphaFold 3 benchmark criteria (Section S2.2). Overall, 600 target epitopes were tested; these were pooled by target, for a total of 200 *in silico* success rates plotted in Figure S2.

### S4. Comparison to other design methods

#### S4.1. Comparison of experimental success rates to RFdiffusion

Published experimental success rates of RFdiffusion were based on 96-well biolayer interferometry (BLI) measurements [S28], which are impractical for us to perform on our own designs. Therefore, to quantitatively compare success rates, we downloaded sequences of the published RFdiffusion designs<sup>2</sup> and synthesized and tested them by yeast display (Section S1.2) alongside our own designs. Our results matched the published success rates for PD-L1 but were 2-fold lower for IL-7RA (16% versus 33%) and 0% for TrkA (versus 6%) (Table S2). This is potentially due to differences between the yeast surface environment versus purified proteins, as well as our use of 10- to 20-fold lower target concentrations (0.5-1  $\mu\text{M}$  versus 10  $\mu\text{M}$  in [S28]), which may exclude lower-affinity hits.

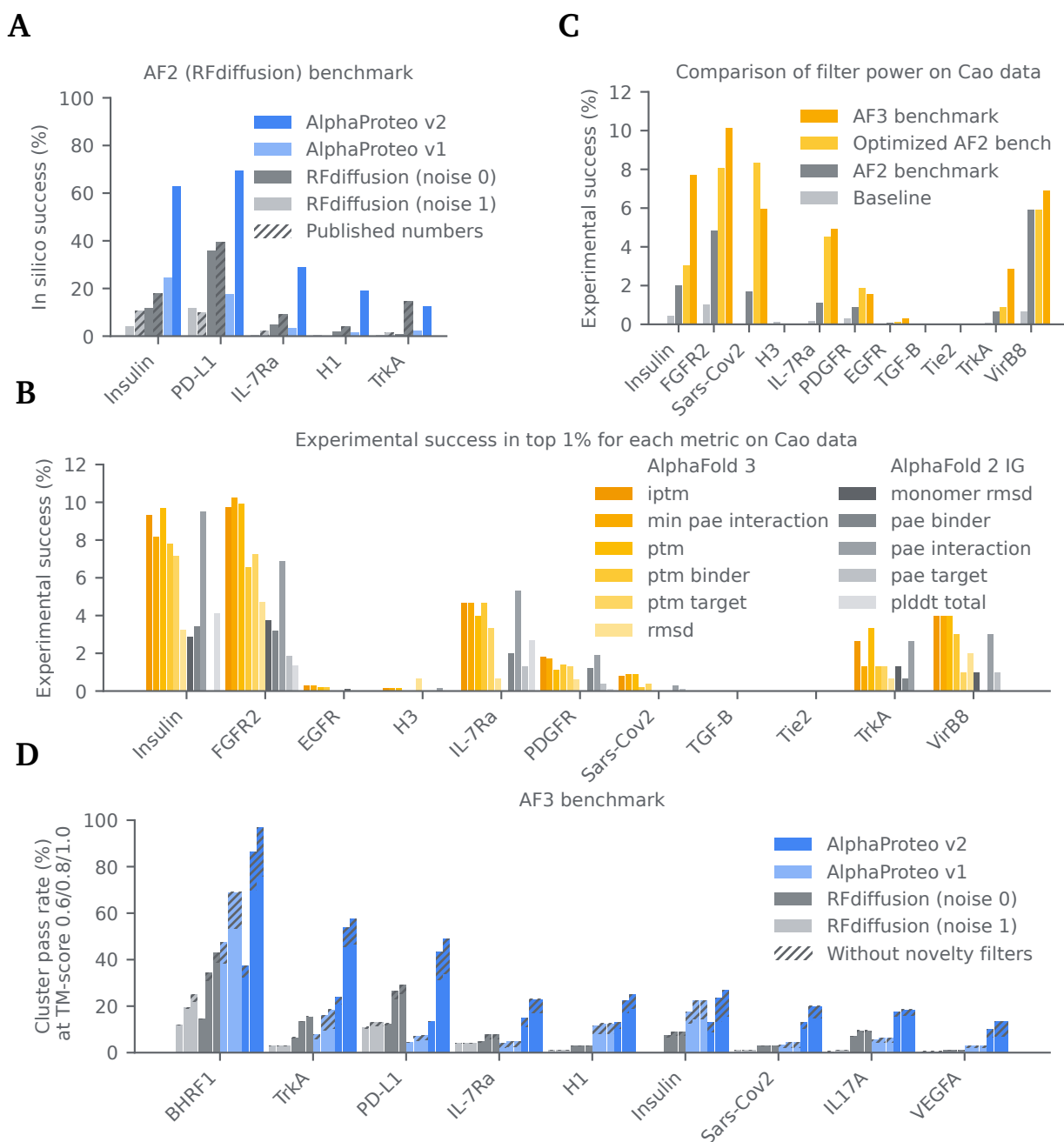
#### S4.2. Comparison of binding affinity ( $K_D$ ) to other methods

We included the best designed binders from the literature for BHRF1 [S21], SC2RBD [S5], IL-7RA, PD-L1, and TrkA [S28] as controls in our HTRF  $K_D$  measurements (Table 1, Table S2, Table S3; Figure S6). For the TrkA and IL-7RA binders we successfully reproduced the literature values, obtaining 10% higher and 2-fold lower  $K_D$  values, respectively, than previously reported. For PD-L1, because we also screened the original set of RFdiffusion designs, we found a design with 1000-fold better  $K_D$  (RFD\_PDL1\_76,  $K_D=1.6$  nM) than the published "best" design (RFD\_PDL1\_77,  $K_D=1.4$   $\mu\text{M}$ ) [S28]. Therefore, we compared our results to the higher-affinity PD-L1 binder (Table 1, Table S2). For the BHRF1 and SC2RBD control binders ("BINDI" and "LCB1", respectively), we obtained  $K_D$  values  $\sim$ 10-fold higher than what was reported previously from biolayer interferometry (BLI) experiments [S21, S5]. A possible explanation is that proteins are in solution and in equilibrium in HTRF while in BLI, one species is immobilized on a 2-dimensional surface. This may allow the mobile species to rebind another molecule of the immobilized species without dissociating from the surface, increasing the apparent binding affinity relative to the solution-phase [S7]. We verified that BINDI and LCB1, as well as AlphaProteo's best binders for these targets, have  $K_D < 1$  nM in our own BLI experiments (Figure S7). As the exact  $K_D$  values from BLI are not quantitative in this regime, we compare these binders on the basis of our HTRF measurements (Table 1).

---

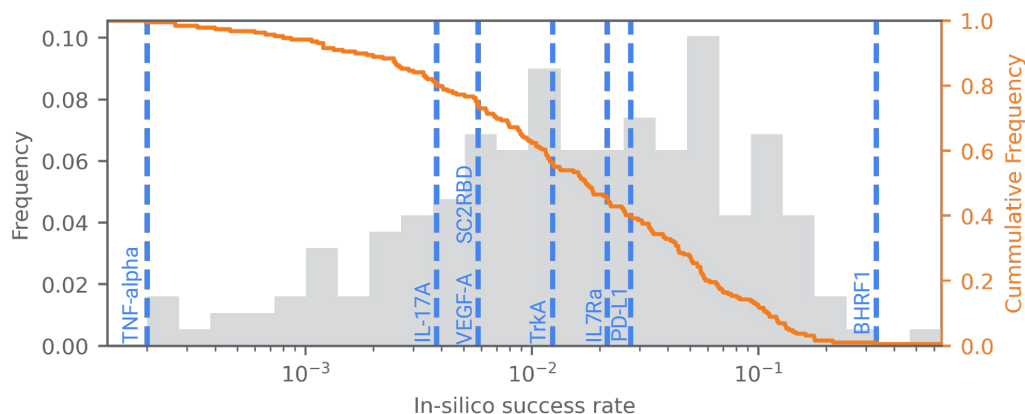
<sup>2</sup><https://figshare.com/s/439fdd59488215753bc3>

## Supplementary figures



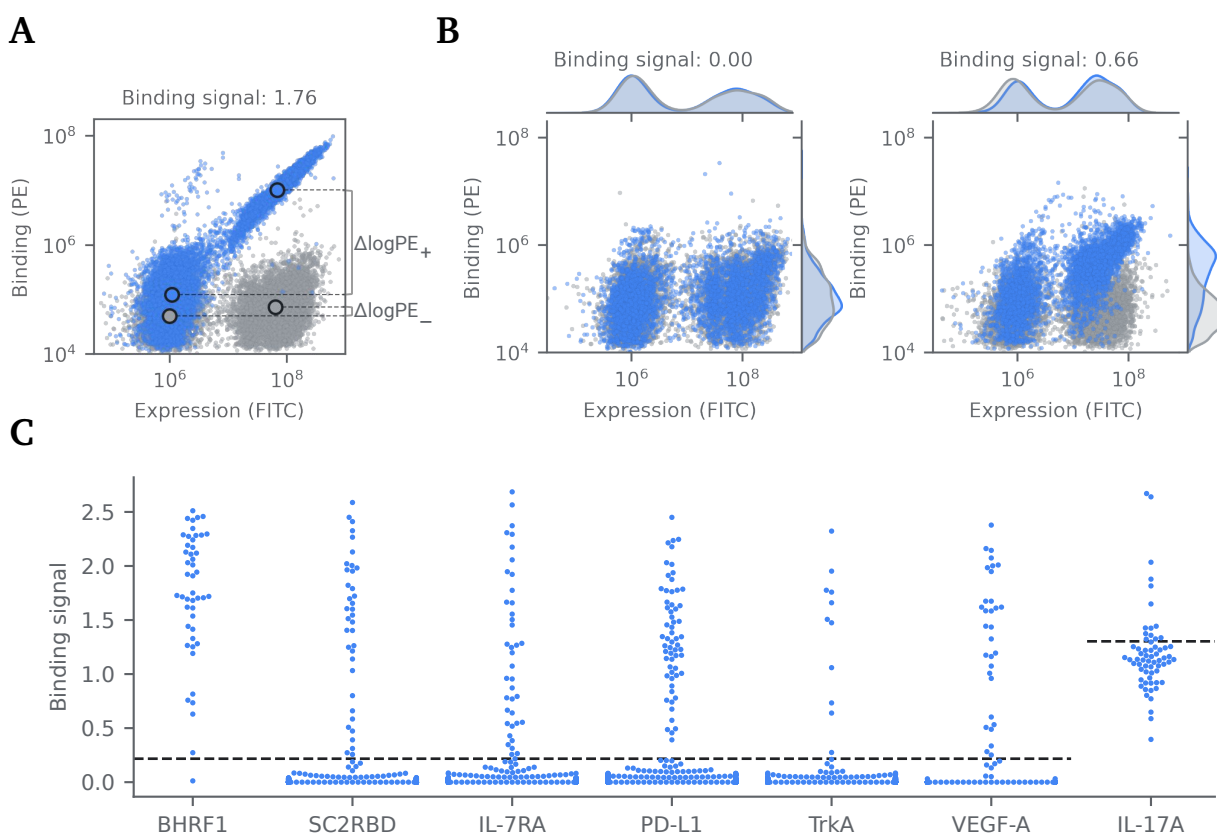
**Figure S1 | In-silico performance of AlphaProteo and development of an AF3-based binder design benchmark.**

See Section S2 for full details. **(A)** *In silico* success rates of AlphaProteo and RFdiffusion under the "AF2 (RFdiffusion) benchmark", which consists of the AF2 initial guess prediction method, scoring thresholds, and design targets described in [S28]. For RFdiffusion, both the published values and our own reproduction of its performance are shown. **(B)** Retrospective experimental success rate of designs from [S6] with the top 1% values of each AF2- or AF3-derived metric. This identifies the metrics that individually have the strongest predictive value for experimental success. **(C)** Retrospective success rate of designs from [S6] after filtering by different definitions of *in silico* success: "Baseline": fraction of successful binders in the unfiltered data from [S6]; "AF2 benchmark": metrics and filtering thresholds used in [S28] and **(A)**; "Optimized AF2 benchmark": optimized thresholds on the same metrics used in the AF2 benchmark; "AF3 benchmark": optimized thresholds on a small set of the most predictive AF3 metrics from **(B)**. The "AF3 benchmark" filtering criteria enrich most strongly for experimental success. **(D)** *In silico* success rates of AlphaProteo and RFdiffusion under the "AF3 benchmark", consisting of both the targets in this work and the previous AF2 benchmark targets, along with optimized AF3 metrics and thresholds as shown in **(C)**. Clustered bars of the same color show diversity-adjusted success rates via pairwise TM-score clustering at different thresholds (0.6/0.8/1.0). Hatched bars show the reduction in success rate after excluding designs with sequence bit-score > 50 in pMMER search against the Uniref50 dataset.



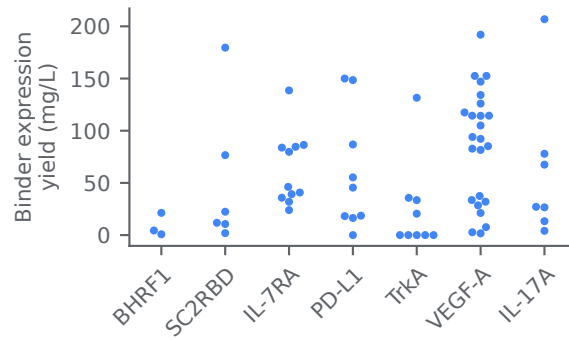
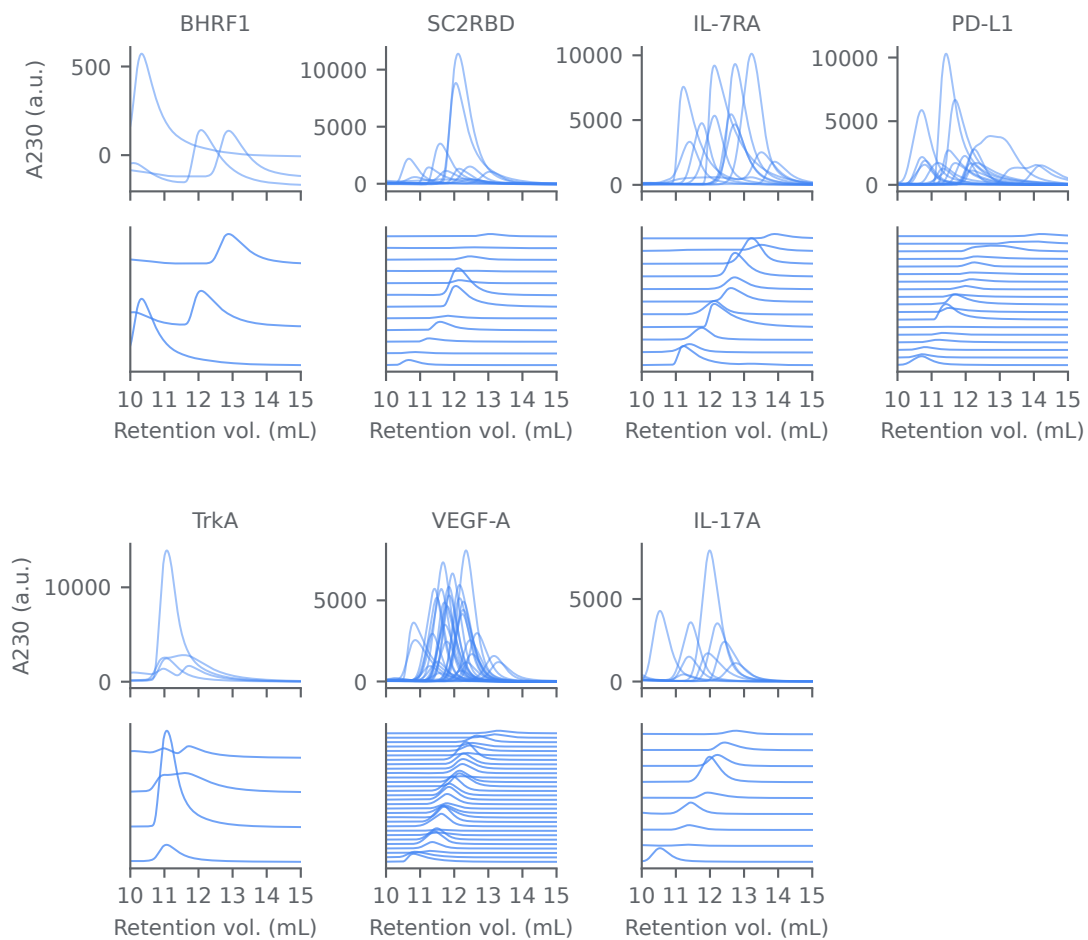
**Figure S2 | Distribution of *in silico* success rates.**

Histogram (gray) and complementary cumulative density (orange line) of *in silico* success rates for AlphaProteo binder design against 200 randomly sampled target proteins from the PDB (Section S3). The 7 targets for which we successfully obtained binders (labeled blue dotted lines) cover a broad range of *in silico* success rates. TNF $\alpha$ , where we failed to obtain binders, is among the most challenging *in silico* targets, while IL-17A, where we succeeded experimentally, is more difficult than 80% of the *in silico* targets.

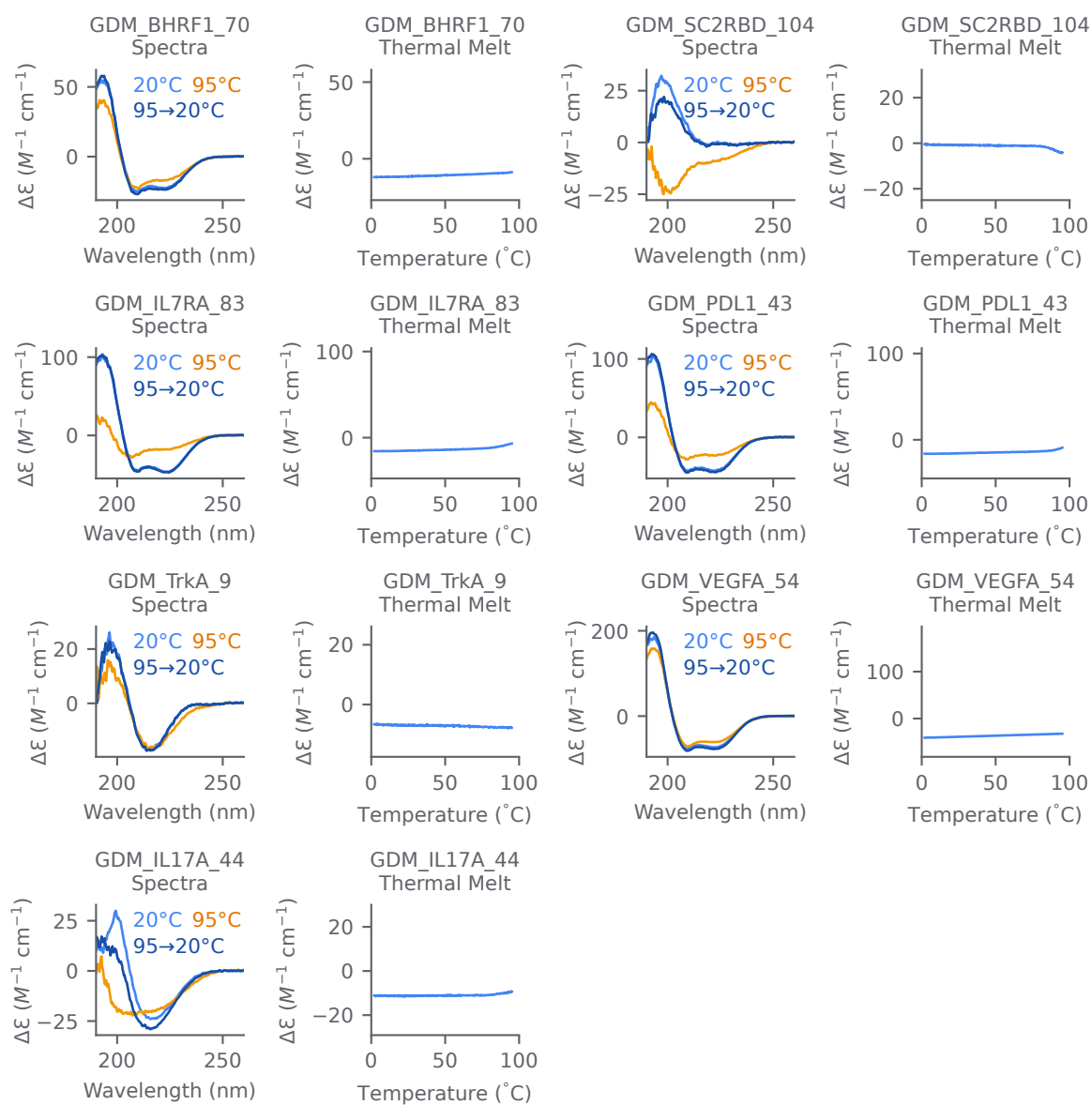


**Figure S3 | Yeast display screening of binder designs.**

(A) Binding signal computed as  $(\Delta \log(\text{PE}_+) - \Delta \log(\text{PE}_-))$  was used to systematically determine binding success. This metric captures the shift in PE signal for the positive population (binding in the presence of the target) in excess of PE shift in the negative population (in the absence of the target), which factors out experiment artifacts which could lead to false positives. (B) Examples of FITC/PE scatterplots for no binding (left) and weak binding (right). (C) Binding signal distribution by target for designs tested via yeast surface display. Dotted lines denote the binary binding threshold: for the first 6 targets the cut-off is set to 0.2 and was empirically determined by FITC/PE plot analysis, whereas for IL-17A the cut-off was set more stringently (due to anomalous yeast display behavior), based on its positive binding control.

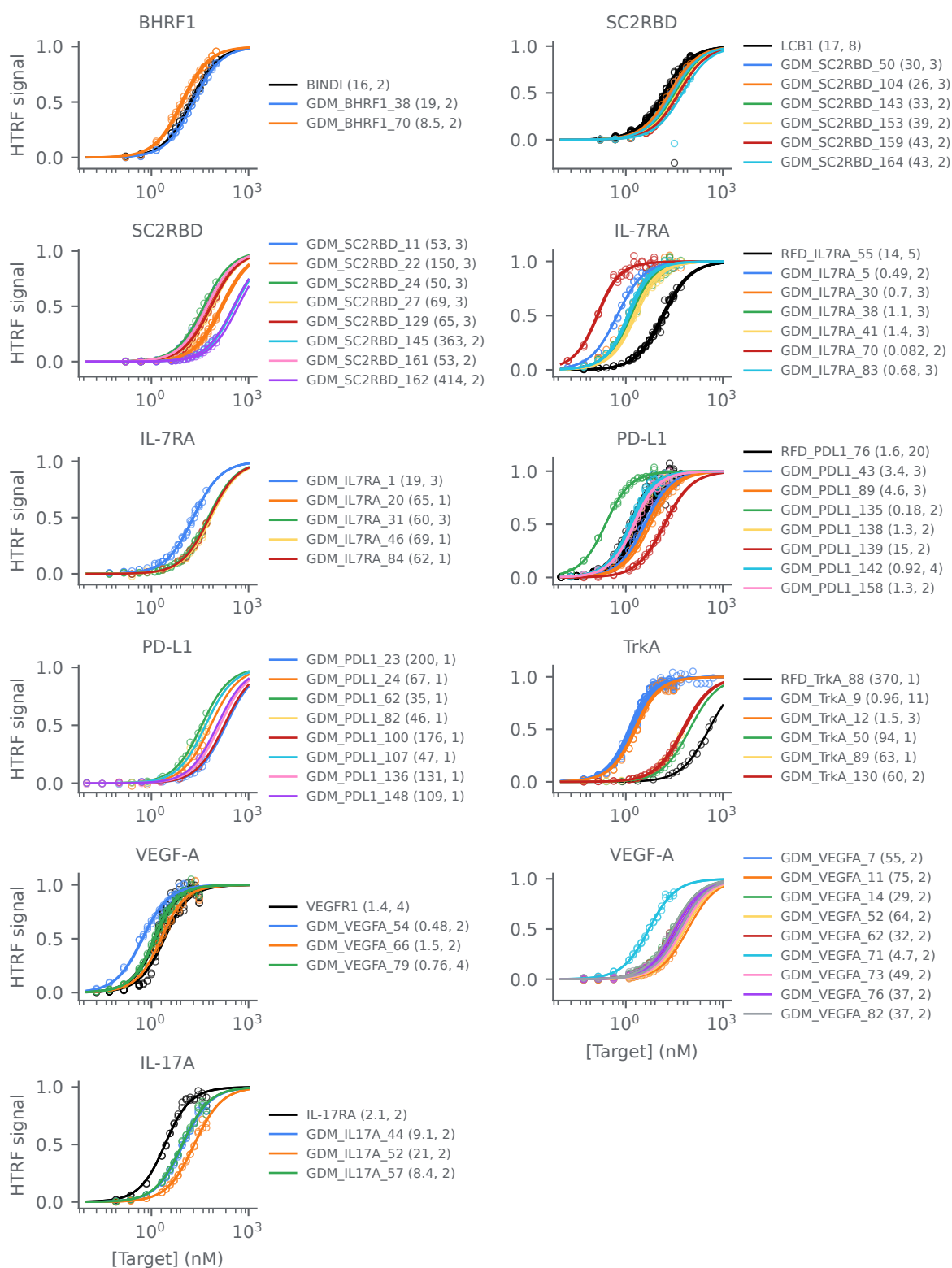
**A****B****Figure S4 | Expression yield and size-exclusion chromatography of binder hits.**

(A) Protein yield from 10 mL *E. coli* expression of yeast display binding hits, showing that most binders are highly expressed.  
(B) Size-exclusion chromatography of binding hits, showing that most binders are monodisperse and likely monomeric.



**Figure S5 | Circular dichroism spectra and thermal melts.**

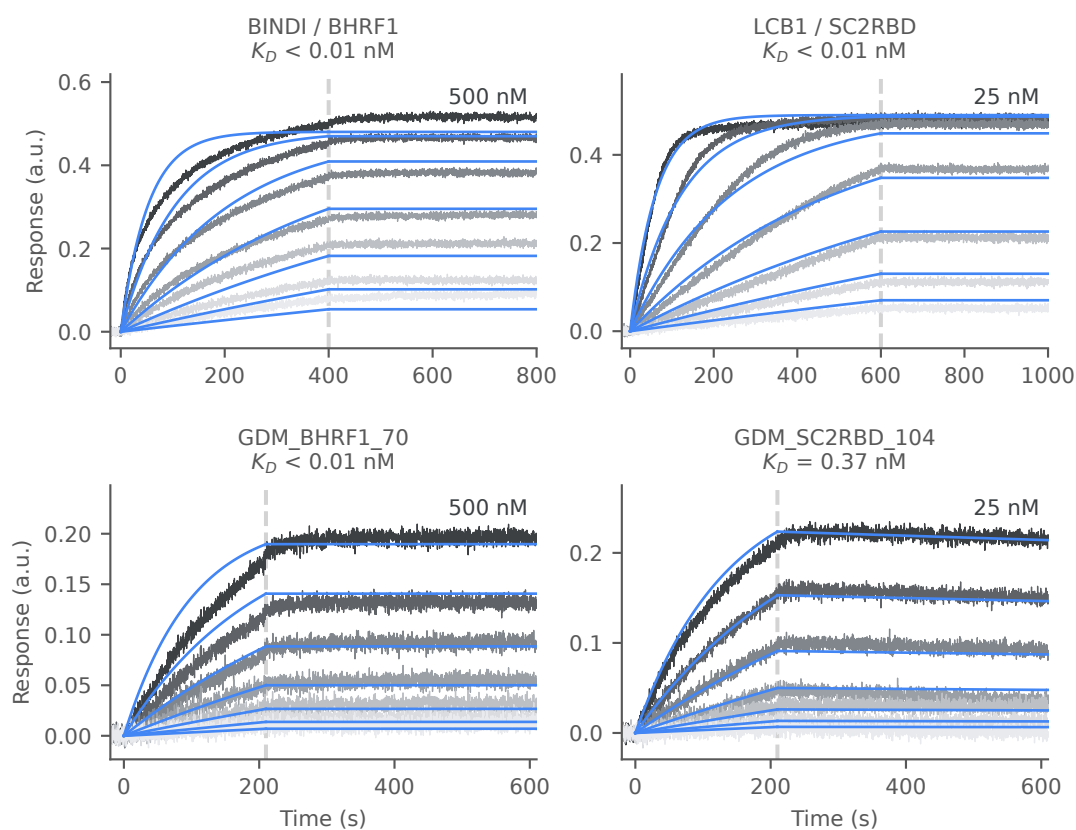
Circular dichroism spectra before (20  $^{\circ}\text{C}$ ) and after melting (95  $^{\circ}\text{C}$  and 95  $\rightarrow$  20  $^{\circ}\text{C}$ ). For most designs tested, spectra show expected secondary structure and refolding.



**Figure S6 | Equilibrium saturation binding by homogeneous time-resolved fluorescence (HTRF).**

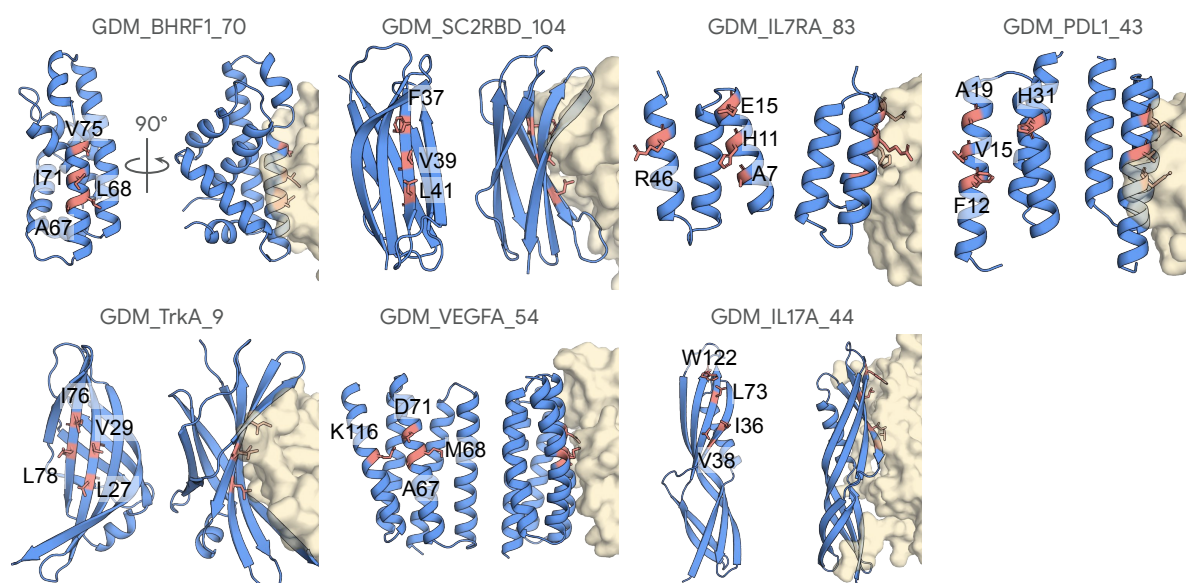
HTRF equilibrium saturation binding affinity measurement data for all designs that expressed and displayed observable binding. Control binders from the literature are shown in black. The binder design was held at a fixed concentration of either 0.1 nM or 1 nM and the target protein concentration was titrated. Parentheses indicate the fitted  $K_D$  value from a generalized (square-root form) 1:1 binding equation (Section S1.4.1) and the number of replicates for each design. Note that the x-axis is in log-scale to highlight order-of-magnitude  $K_D$  differences, but this gives the appearance that data is less saturated than it is. We report the inter-replicate variation as well as the fitting uncertainty in (Table S7).





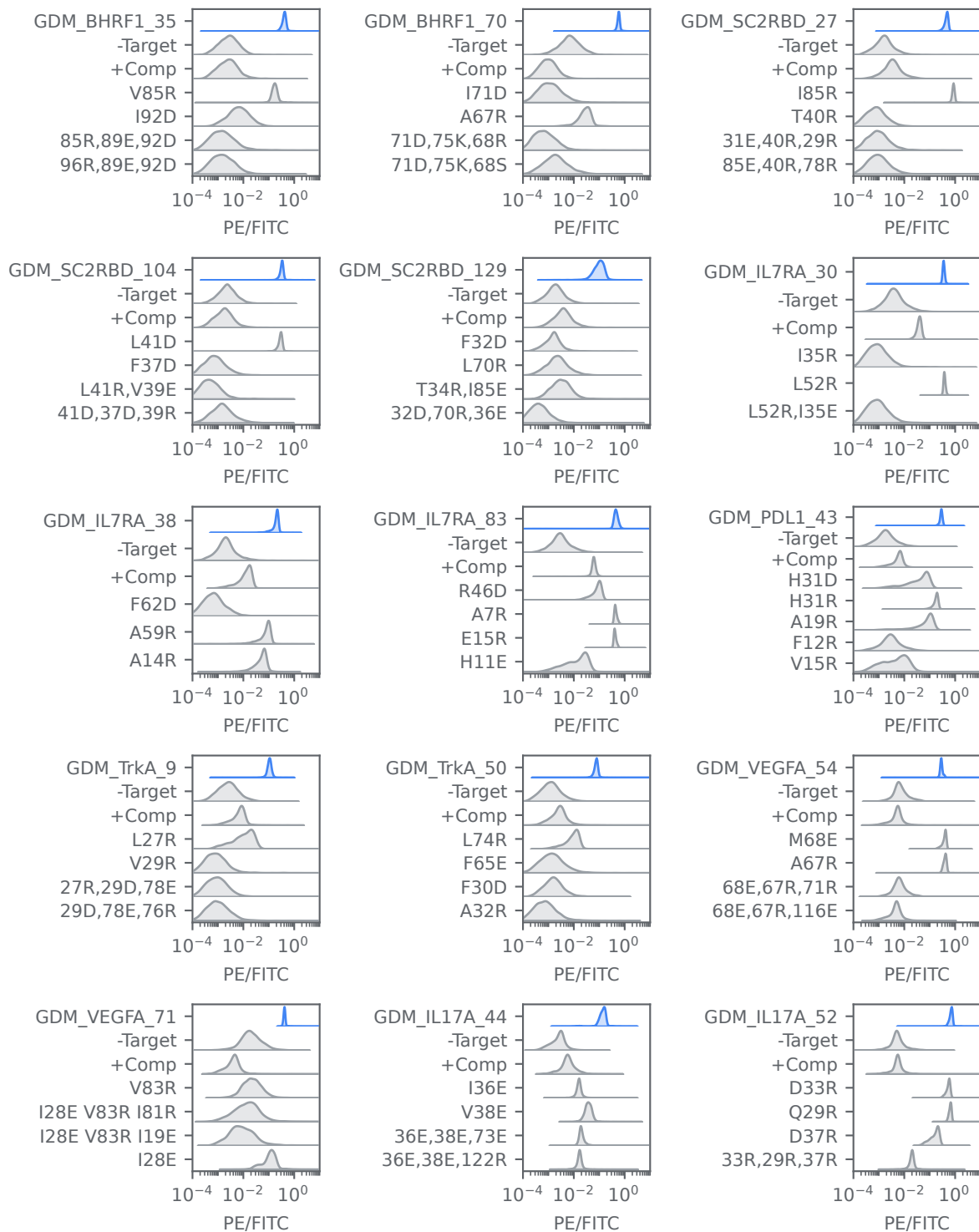
**Figure S7 | Binding kinetics by biolayer interferometry (BLI).**

BLI kinetic data (gray gradient), fitted 1:1 binding models with shared global  $R_{max}$  (blue lines), and resulting  $K_D$  estimates for selected controls and designs (Section S1.4.2). Analyte concentrations are in a 2-fold serial dilution from the maximum concentration indicated. " $K_D < 0.01$ " indicates a fitted  $K_D$  value that is too low to interpret reliably due to lack of apparent dissociation.



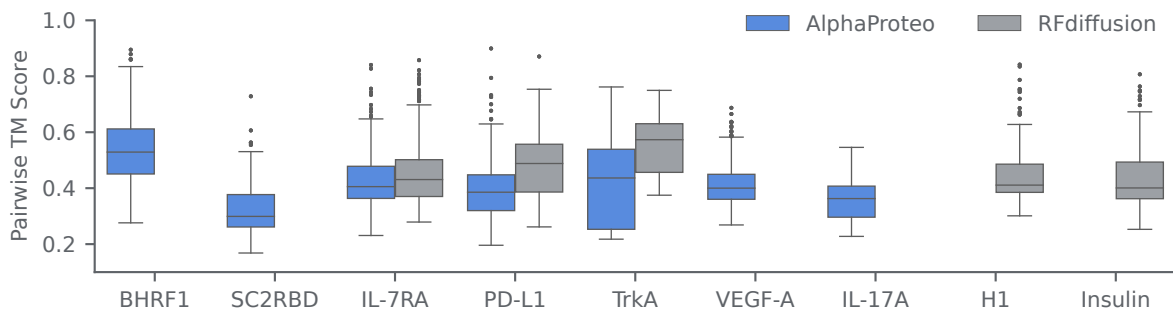
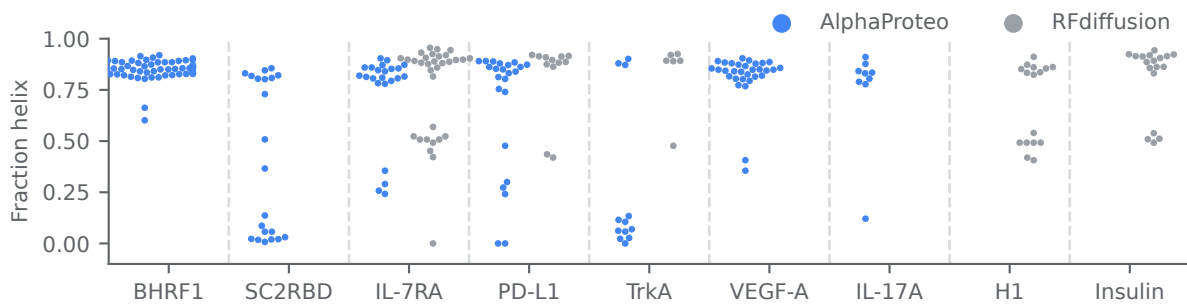
**Figure S8 | Structural locations of interface mutations.**

Examples of interface residues (red) that were mutated for one representative binder (blue) per target (pale yellow).

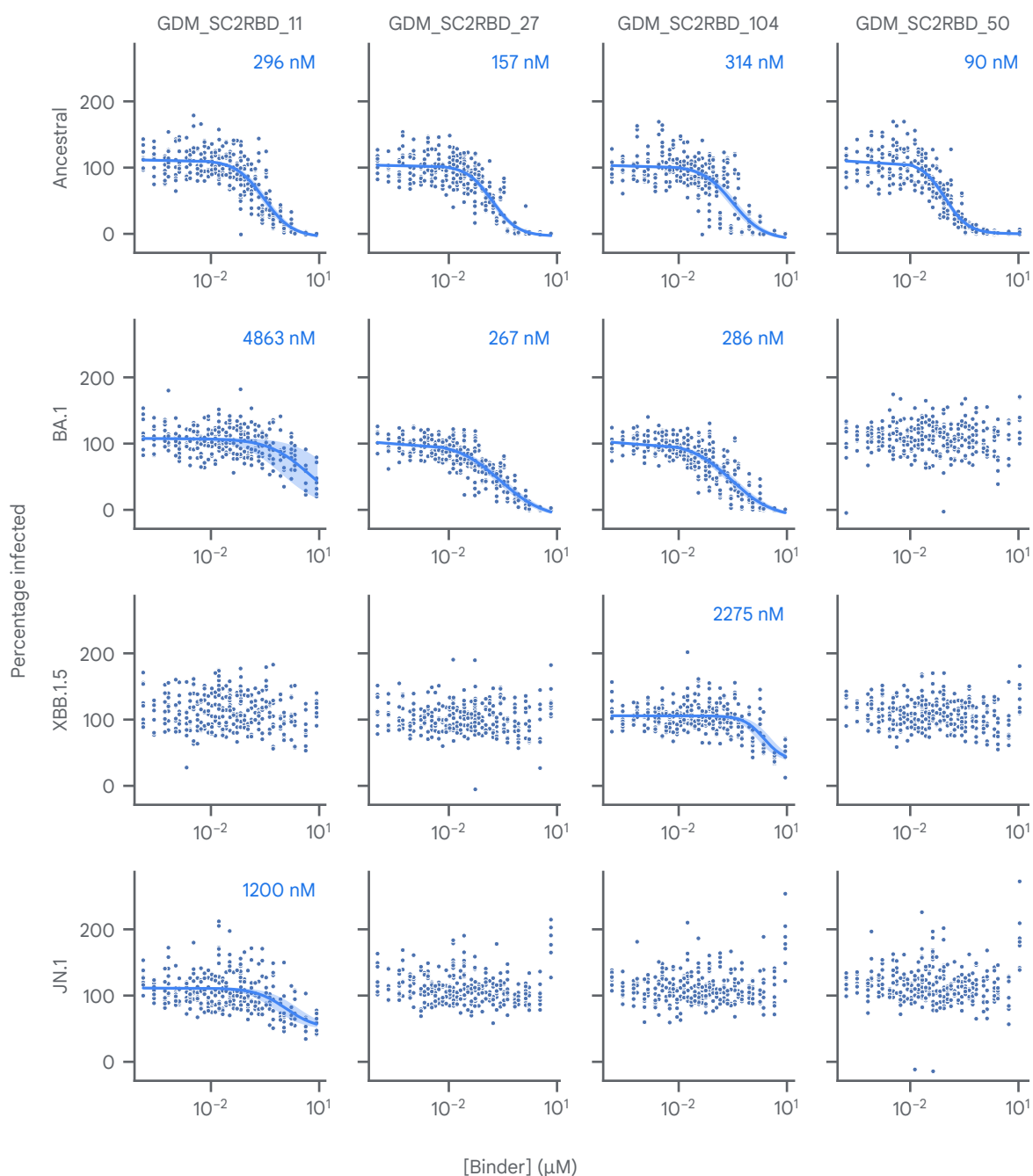


**Figure S9 | Interface mutation and competition experiments on the best binders.**

Distributions of PE (binding) normalized to FITC (expression) from flow cytometry yeast display for interface mutants of selected binders as well as competitive inhibition by a previous known binder (Section S1.2.2). In most cases, mutating the interface residue or adding a competitive binder decreases the binding signal, confirming that the design binds via the intended interface residues to the intended target site.

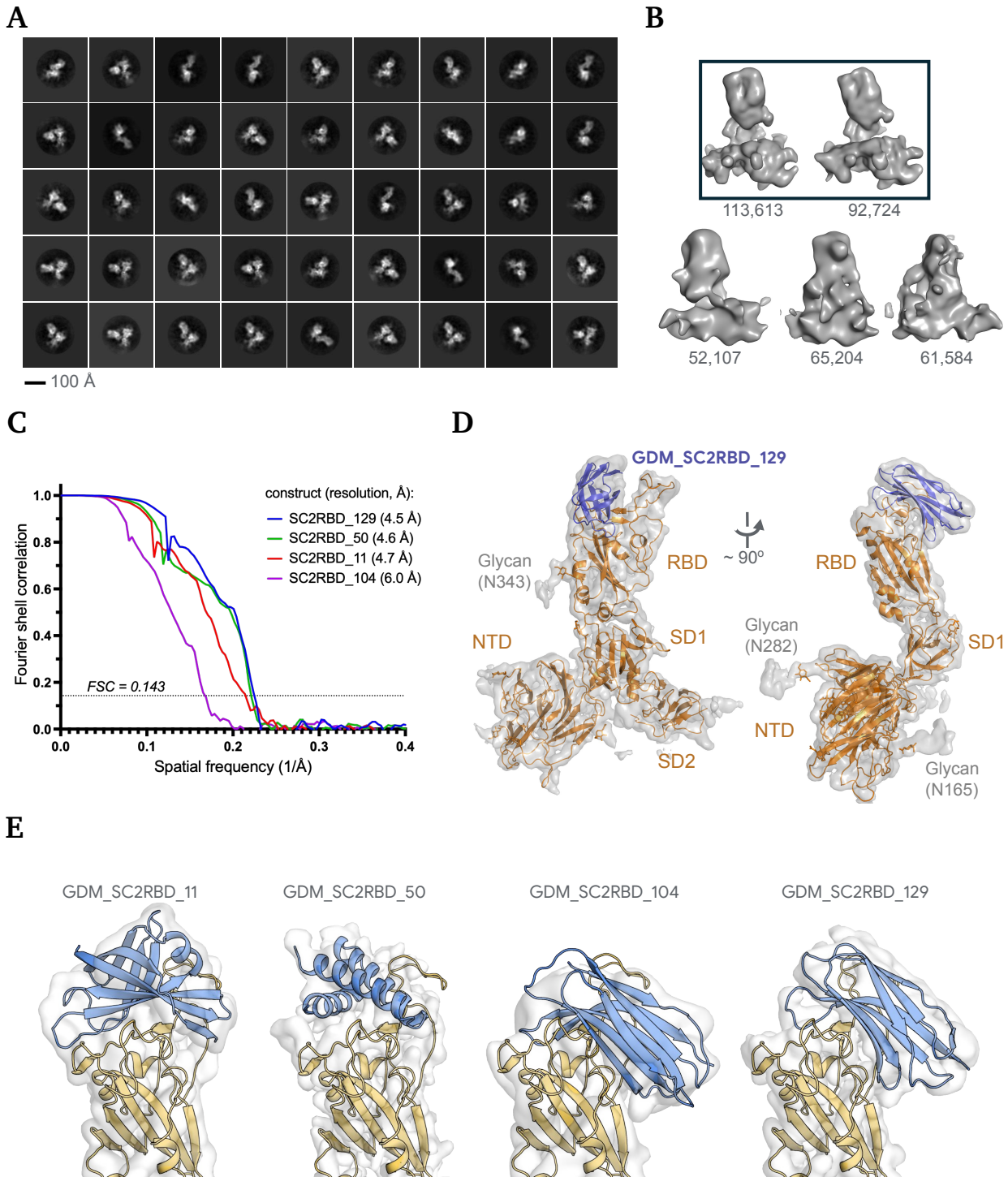
**A****B****Figure S10 | Diversity of experimentally tested binder designs.**

(A) Pairwise TM score among designs from AlphaProteo and RFdiffusion with experimentally confirmed binding activity, including additional RFdiffusion targets H1 and Insulin [S28]. Experimental success of RFdiffusion designs in this plot is based on the published results rather than our yeast display measurements, as our assay did not observe binding for any TrkA designs (Section S4). (B) Fraction of binder residues annotated by DSSP as helix. The fraction of loop-annotated residues is low and relatively constant for all designs, so the fraction of beta sheets is  $1 - \text{fraction helix}$ .



**Figure S11 | SARS-CoV-2 neutralization assay for four selected binders over four variants of interest.**

SARS-CoV-2 virus neutralization assay was performed in Vero cells by the Francis Crick Institute COVID Surveillance Unit following the protocol outlined in [S26]. Each plot consists of 160 independent data points (4 technical replicates, 2 biological replicates, 40 independent titrations).  $EC_{50}$  values were calculated using nonlinear regression with a 4-parameter dose response curve fit. Fits are shown only when standard error on  $EC_{50}$  was within one order of magnitude and the percentage infected is reduced to at least 80%. Standard errors at each dilution are shown as shaded areas. All four binders tested successfully neutralize the ancestral variant of the SARS-CoV-2 virus.



**Figure S12 | SARS-CoV-2 cryo-EM data processing.**

(A) 2D class averages of particle images from the SC2RBD\_129 dataset corresponding to dissociated spikes; the scale bar is 100 Å. (B) The result of 3D classification of the same particles into 5 classes; the number of particles belonging to each class is indicated underneath. The two best 3D classes collectively comprising 206337 particles used for the final 3D reconstruction are boxed. (C) Half-map Fourier shell correlation (FSC) for each of the final reconstructions. Dotted line indicates the gold-standard cut-off at FSC of 0.143. (D) Final reconstruction of S1 in complex with GDM\_SC2RBD\_129 ligand in two orthogonal orientations. The cryo-EM map is shown as a transparent gray surface with docked S1 (from PDB ID 7ZBU, orange) and the ligand (blue) chains as cartoons. Locations of the ligand, individual S1 domains (RBD, NTD, SD1 and SD2) as well as select glycans (attached to Asn residues 165, 282, and 343) are indicated. (E) SARS-CoV-2 receptor binding domain (yellow) bound to the binder design (blue) was fitted together into the Cryo-EM density map (transparent gray surface).

## Supplementary tables

Design target	PDB ID	Target chain and residue numbers	Target hotspot residues	Natural binding partner	Binder length range (for benchmarks)
BHRF1	2wh6	A2-158	A65, A74, A77, A82, A85, A93	BH3 helix	80-120
SC2RBD	6m0j	E333-526	E485, E489, E494, E500, E505	ACE2 receptor	80-120
IL-7RA	3di3	B17-209	B58, B80, B139	IL-7	50-120
PD-L1	5o45	A17-132	A56, A115, A123	PD-1	50-120
TrkA	1www	X282-382	X294, X296, X333	Nerve growth factor	50-120
Insulin	4zxb	E6-155	E64, E88, E96	Insulin receptor	40-120
H1	5vli	A1-50, A76-80, A107-111, A258-322, B1-68, B80-170	B21, B45, B52	None	40-120
VEGF-A	1bj1	V14-107, W14-107	W81, W83, W91	VEGFR1, VEGFR2	50-140
IL-17A	4hsa	A17-131, B19-127	A94, A116, B67	IL-17 receptor alpha	50-140
TNF $\alpha$	1tnf	A12-157, B12-157, C12-157	A113, C73	None	50-120

**Table S1 | Binder design problem specifications for *in silico* benchmarking and experimental testing.**

Input target structures and hotspot residues for the design targets addressed here. For IL-7RA, PD-L1, TrkA, we used the same definitions as RFdiffusion [S28]. For other targets, we used the highest resolution crystal structure available, and chose hotspot residues at the binding site of a natural interaction partner. When making designs for experimental validation, we used a binder length range of 50-140 amino acids.

	BHRF1	SC2RBD	IL-7RA	PD-L1	TrkA	IL-17A	VEGF-A	TNF $\alpha$
<b>Success (%)</b> (higher is better)								
AlphaProteo v1	88 (94)	8.5   8.5 <sup>†</sup> (94)   (47)	24.5 (94)	9.6 (94)	9.6 (94)	14.3 (63)	33 (94)	0 (30)
AlphaProteo v1 (improved filters)	–	29 (31)	–	16.1 (31)	–	–	–	–
AlphaProteo v2	–	–	–	26.5 (34)	8.1 (37)	–	–	0 (24)
RFdiffusion (our measurement)	–	–	16.8 (95)	12.6 (95)	0.0 (95)	–	–	–
RFdiffusion (published)	–	–	33.7 (95)	12.6 (95)	6.3 (95)	–	–	–
<b>Binding K<sub>D</sub> (nM)</b> (lower is better)								
AlphaProteo v1	8.5 (94)	30   26 <sup>†</sup> (94)   (47)	0.08 (94)	3.4 (94)	0.96 (94)	8.4 (63)	0.48 (94)	–
AlphaProteo v1 (improved filters)	–	33 (31)	–	47 (31)	–	–	–	–
AlphaProteo v2	–	–	–	0.18 (34)	60 (37)	–	–	–
RFdiffusion (our measurement)	–	–	14.1 (95)	1.56 (95)	370 (95)	–	–	–
RFdiffusion (published)	–	–	30 (95)	1400 (95)	328 (95)	–	–	–

**Table S2 | Experimental binding success rates and affinities of AlphaProteo model variants and RFdiffusion.**

Percentage of tested designs with observable binding for different AlphaProteo variants (Section S2) and RFdiffusion. Number of tested designs is in parentheses. Empty fields (–) are method/target combinations that were not tested. For success rate, AlphaProteo and "RFdiffusion (our measurement)" values come from yeast display assays performed by us, while "RFdiffusion (published)" shows BLI screening results from [S28] (Section S4). For AlphaProteo v1 on SC2RBD, a subset of 47 designs were made without hotspot conditioning. Their results are indicated separately with a dagger (†). All other designs used hotspot conditioning. For K<sub>D</sub> values, AlphaProteo and "RFdiffusion (our measurement)" values come from HTRF assays performed by us, while "RFdiffusion (published)" shows BLI titration results from Watson et al. [S28] (Section S4).

	BHRF1	SC2RBD	IL-7RA	PD-L1	TrkA	IL-17A	VEGF-A
Yeast surface display hits selected for <i>E. coli</i> expression	2	18	11	21	10	9	31
Successfully subcloned and expressed	2	17	11	20	5	9	31
Tested by HTRF	2	17	11	18	5	7	31
Observed HTRF binding	2	14	11	16	5	3	20

Table S3 | Number of yeast hits successfully expressed in *E. coli* and tested for HTRF binding.

Replicate	Variant	Binder	EC <sub>50</sub> (μM)	SE(EC <sub>50</sub> )
A	ancestral	GDM_SC2RBD_104	0.308	0.106
	ancestral	GDM_SC2RBD_11	0.273	0.057
	ancestral	GDM_SC2RBD_27	0.171	0.029
	ancestral	GDM_SC2RBD_50	0.089	0.011
	BA.1	GDM_SC2RBD_104	0.203	0.062
	BA.1	GDM_SC2RBD_11	3.053	4.702
	BA.1	GDM_SC2RBD_27	0.167	0.057
	XBB.1.5	GDM_SC2RBD_104	1.958	1.186
B	ancestral	GDM_SC2RBD_104	0.317	0.085
	ancestral	GDM_SC2RBD_11	0.324	0.078
	ancestral	GDM_SC2RBD_27	0.144	0.023
	ancestral	GDM_SC2RBD_50	0.090	0.011
	BA.1	GDM_SC2RBD_104	0.405	0.134
	BA.1	GDM_SC2RBD_27	0.382	0.096
	XBB.1.5	GDM_SC2RBD_104	2.544	1.309
	JN.1	GDM_SC2RBD_11	1.019	0.936

Table S4 | EC<sub>50</sub> concentrations measured in live virus inhibition assays for SARS-CoV-2.



	SC2RBD_104	SC2RBD_50	SC2RBD_11	SC2RBD_129
<b>Data collection</b>				
Microscope, operating voltage	Titan Krios G2, 300 keV			
Detector	Falcon 4i			
Automation software	EPU			
Energy filter (slit width)	None		Selectris (10 eV)	
Magnification (nominal)	75,000		130,000	
Pixel size (Å)	1.08		0.95	
Underfocus range (nominal, μm)	1.5 - 3.5		1.5 - 3.3	
Number of EER frames per movie		1,674		
Total electron fluence ( $e/\text{Å}^2$ )	32.2		41	
Total number of micrograph movies acquired	4,500	8,342	6,728	8,482
<b>Reconstruction</b>				
Software for 2D classification	Relion-5.0beta			
Software for 3D classification	Relion-5.0beta			
Software for final reconstruction	Relion-5.0beta, with Blush regularization			
Symmetry	C1			
Number of initially extracted particles	2,217,923	5,790,093	1,937,854	1,872,411
Number of particles used in 3D classification	375,214	960,875	126,130	385,262
Number of classes in 3D classification	7	6	4	5
Number of particles in final reconstruction	92,321	265,173	118,794	206,337
Global resolution (FSC 0.143, Å)	6.0	4.6	4.7	4.5

Table S5 | Cryo-EM data processing.

<b>Data collection</b>	
Space group	P 41 21 2
Temperature	100 K
Number of crystals	1
Cell dimensions ((a, b, c (Å)), ( $\alpha$ , $\beta$ , $\gamma$ , (°)))	(87.802 87.802 185.73) (90 90 90)
Wavelength (Å)	0.9537
<b>Refinement</b>	
Resolution range (Å)	87.80 - 2.56 (2.68 - 2.56)
No. of reflections	24246 (2883)
Completeness for the range (%)	100 (99.8)
Redundancy	26.6
R <sub>merge</sub>	0.122
CC1/2	0.994 (0.205)
Mean I/ $\sigma$ (I)	4.4 (0.3)
Wilson B factor (Å <sup>2</sup> )	47.460
Resolution range (Å)	79.50 – 2.56
No. observations (total/test set)	22961 / 1110
Completeness (%)	94.99
Rwork/Rfree (%)	0.225 / 0.227
No. of atoms	
Protein (NON-HYDROGEN)	2386
Ligand/ion	0
Waters	0
Average B all atoms (Å <sup>2</sup> )	51.701
R.m.s. deviations	
Bond lengths (Å)	0.014
Bond angles (°)	2.06
Ramachandran	
outliers (%)	1
favored (%)	89.63
<sup>b</sup> Numbers in parentheses refer to the highest-resolution shell	

Table S6 | Crystallographic refinement statistics for the GDM\_VEGFA\_71/VEGF-A complex structure.



## Supplementary references

Some references are listed in both the main bibliography and the supplementary bibliography section.

- [S1] Josh Abramson et al. “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. In: *Nature* 630.8016 (2024), pp. 493–500. DOI: 10.1038/s41586-024-07487-w.
- [S2] Jon Agirre et al. “The CCP4 suite: integrative software for macromolecular crystallography”. In: *Acta Crystallogr. D Struct. Biol.* 79.6 (2023), pp. 449–461. DOI: 10.1107/S2059798323003595.
- [S3] Nathaniel R Bennett et al. “Improving de novo protein binder design with deep learning”. In: *Nat. Commun.* 14.1 (2023). DOI: 10.1038/s41467-023-38328-5.
- [S4] Tristan Bepler et al. “Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs”. In: *Nat. Methods* 16.11 (2019), pp. 1153–1160. DOI: 10.1038/s41592-019-0575-8.
- [S5] Longxing Cao et al. “De novo design of picomolar SARS-CoV-2 miniprotein inhibitors”. In: *Science* 370.6515 (2020), pp. 426–431. DOI: 10.1126/science.abd9909.
- [S6] Longxing Cao et al. “Design of protein-binding proteins from the target structure alone”. In: *Nature* 605.7910 (2022), pp. 551–560. DOI: 10.1038/s41586-022-04654-9.
- [S7] Camille Daniel et al. “Solution-phase vs surface-phase aptamer-protein affinity from a label-free kinetic biosensor”. In: *PLoS One* 8.9 (2013), e75419. DOI: 10.1371/journal.pone.0075419.
- [S8] J Dauparas et al. “Robust deep learning-based protein sequence design using ProteinMPNN”. In: *Science* 378.6615 (2022), pp. 49–56. DOI: 10.1126/science.add2187.
- [S9] P Emsley et al. “Features and development of *coot*”. In: *Acta Crystallogr. D Biol. Crystallogr.* 66.Pt 4 (2010), pp. 486–501. DOI: 10.1107/S0907444910007493.
- [S10] Paul Emsley and Kevin Cowtan. “Coot: model-building tools for molecular graphics”. In: *Acta Crystallogr. D Biol. Crystallogr.* 60.Pt 12 Pt 1 (2004), pp. 2126–2132. DOI: 10.1107/S0907444904019158.
- [S11] R Daniel Gietz and Robert H Schiestl. “High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method”. In: *Nat. Protoc.* 2.1 (2007), pp. 31–34. DOI: 10.1038/nprot.2007.13.
- [S12] Richard J Gildea et al. “*xia 2. multiplex* : a multi-crystal data-analysis pipeline”. In: *Acta Crystallogr. D Struct. Biol.* 78.6 (2022), pp. 752–769. DOI: 10.1107/S2059798322004399.
- [S13] Jiahua He, Tao Li, and Sheng-You Huang. “Improvement of cryo-EM maps by simultaneous local and non-local deep learning”. In: *Nat. Commun.* 14.1 (2023), p. 3217. DOI: 10.1038/s41467-023-39031-1.
- [S14] David M Hoover and Jacek Lubkowski. “DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis”. In: *Nucleic Acids Res.* 30.10 (2002), e43. DOI: 10.1093/nar/30.10.e43.
- [S15] Inga Jarmoskaite et al. “How to measure and evaluate binding affinities”. In: *Elife* 9 (2020). Ed. by Sebastian Deindl and John Kuriyan, e57264. DOI: 10.7554/eLife.57264.
- [S16] L Steven Johnson, Sean R Eddy, and Elon Portugaly. “Hidden Markov model speed heuristic and iterative HMM search procedure”. In: *BMC Bioinformatics* 11.1 (2010), p. 431. DOI: 10.1186/1471-2105-11-431.

- [S17] Dari Kimanius et al. “Data-driven regularization lowers the size barrier of cryo-EM structure determination”. In: *Nat. Methods* 21.7 (2024), pp. 1216–1221. DOI: 10.1038/s41592-024-02304-8.
- [S18] Dari Kimanius et al. “New tools for automated cryo-EM single-particle analysis in RELION-4.0”. In: *Biochem. J.* 478.24 (2021), pp. 4169–4185. DOI: 10.1042/BCJ20210708.
- [S19] Airlie J McCoy et al. “Phaser crystallographic software”. In: *J. Appl. Crystallogr.* 40.Pt 4 (2007), pp. 658–674. DOI: 10.1107/S0021889807021206.
- [S20] Garib N Murshudov et al. “REFMAC5 for the refinement of macromolecular crystal structures”. In: *Acta Crystallogr. D Biol. Crystallogr.* 67.4 (2011), pp. 355–367. DOI: 10.1107/s0907444911001314.
- [S21] Erik Procko et al. “A computationally designed inhibitor of an Epstein-Barr viral bcl-2 protein induces apoptosis in infected cells”. In: *Cell* 157.7 (2014), pp. 1644–1656. DOI: 10.1016/j.cell.2014.04.034.
- [S22] Annachiara Rosa et al. “SARS-CoV-2 can recruit a heme metabolite to evade antibody immunity”. In: *Sci. Adv.* 7.22 (2021), eabg7607. DOI: 10.1126/sciadv.abg7607.
- [S23] Peter B Rosenthal and Richard Henderson. “Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy”. In: *J. Mol. Biol.* 333.4 (2003), pp. 721–745. DOI: 10.1016/j.jmb.2003.07.013.
- [S24] Sjors H W Scheres and Shaoxia Chen. “Prevention of overfitting in cryo-EM structure determination”. In: *Nat. Methods* 9.9 (2012), pp. 853–854. DOI: 10.1038/nmeth.2115.
- [S25] Jeffrey Seow et al. “A neutralizing epitope on the SD1 domain of SARS-CoV-2 spike targeted following infection and vaccination”. In: *Cell Rep.* 40.8 (2022). DOI: 10.1016/j.celrep.2022.111276.
- [S26] Marianne Shawe-Taylor et al. “Divergent performance of vaccines in the UK autumn 2023 COVID-19 booster campaign”. In: *Lancet* 403.10432 (2024), pp. 1133–1136. DOI: 10.1016/S0140-6736(24)00316-7.
- [S27] C M Stoscheck. “Quantitation of protein”. In: *Methods Enzymol.* Methods in enzymology 182 (1990), pp. 50–68. DOI: 10.1016/0076-6879(90)82008-p.
- [S28] Joseph L Watson et al. “De novo design of protein structure and function with RFdiffusion”. In: *Nature* (2023). DOI: 10.1038/s41586-023-06415-8.
- [S29] Christopher J Williams et al. “MolProbity: More and better reference data for improved all-atom structure validation”. In: *Protein Sci.* 27.1 (2018), pp. 293–315. DOI: 10.1002/pro.3330.
- [S30] Antoni G Wrobel et al. “SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects”. In: *Nat. Struct. Mol. Biol.* 27.8 (2020), pp. 763–767. DOI: 10.1038/s41594-020-0468-7.
- [S31] Kai Zhang. “Gctf: Real-time CTF determination and correction”. In: *J. Struct. Biol.* 193.1 (2016), pp. 1–12. DOI: 10.1016/j.jsb.2015.11.003.
- [S32] Jasenko Zivanov et al. “A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0”. In: *Elife* 11 (2022). DOI: 10.7554/eLife.83724.