De novo design of miniprotein agonists and antagonists targeting G protein-coupled receptors

Edin Muratspahić[1,2,#], David Feldman[1,2,3#], David E. Kim[1,2,4,#], Xiangli Qu[1,2,#], Ana-Maria Bratovianu[3,#], Paula Rivera-Sánchez[3,#], Federica Dimitri[3], Jason Cao[5,6], Brian P. Cary[5,6], Matthew J. Belousoff[5,6], Peter Keov[5,6], Qingchao Chen[7], Yue Ren[7], Justyn Fine[8], Isaac Sappington[1,2], Thomas Schlichthaerle[1,2], Jason Z. Zhang[1,2], Arvind Pillai[1,2], Ljubica Mihaljević[1,2,4], Magnus Bauer[1,2], Susana Vázquez Torres[1,2], Amir Motmaen[1,2], Gyu Rie Lee[1,2,4], Long Tran[2,9], Xinru Wang[1,2], Inna Goreshnik[1,2], Dionne K. Vafeados[1,2], Justin E. Svendsen[10,11], Parisa Hosseinzadeh[10,11], Nicolai Lindegaard[12], Matthäus Brandt[12], Yann Waltenspühl[12], Kristine Deibler[12], Luke Oostdyk[13], William Cao[13], Lakshmi Anantharaman[13], Lance Stewart[1,2], Lauren Halloran[14], Jamie B. Spangler[14], Patrick M. Sexton[5,6], Bryan L. Roth[15], Brian E. Krumm[15], Denise Wootten[5,6], Christopher G. Tate[7], Christoffer Norn[1,2,3*], David Baker[1,2,4*]

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195.
[2]Institute for Protein Design, University of Washington, Seattle, WA 98195.
[3]BioInnovation Institute, DK2200 Copenhagen N, Denmark.
[4]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195.
[5]Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, VIC Australia.
[6]ARC Centre for Cryo-electron Microscopy of Membrane Proteins, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, VIC Australia.
[7]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, UK.
[8]Program in Molecular Biophysics, Johns Hopkins University, Baltimore, MD 21208, USA; Translational Tissue Engineering Center, Johns Hopkins University, Baltimore, MD 21231, USA
[9]Department of Chemical Engineering, University of Washington, Seattle, WA 98195.
[10]Department of Bioengineering, Knight Campus for Accelerating Scientific Impact, University of Oregon, 6231 University of Oregon, Eugene, OR, 97403, USA.
[11]Department of Chemistry and Biochemistry, University of Oregon, 1253 University of Oregon, Eugene, OR, 97403, USA.
[12]Novo Nordisk A/S, Novo Nordisk Park 1, 2760 Måløv, Denmark.
[13]LeadHunter Services, Eurofins DiscoverX, LLC, CA-94583, California, USA.
[14]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA; Translational Tissue Engineering Center, Johns Hopkins University, Baltimore, MD 21231, USA; Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.
[15]Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA and NIMH Psychoactive Drug Screening Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[#]These authors contributed equally.

*Correspondence: chris.norn@skape.bio, dabaker@uw.edu

Abstract

G protein-coupled receptors (GPCRs) play key roles in physiology and are central targets for drug discovery and development, yet the design of protein agonists and antagonists has been challenging as GPCRs are integral membrane proteins and conformationally dynamic. Here we describe computational *de novo* design methods and a high throughput "receptor diversion" microscopy based screen for generating GPCR binding miniproteins with high affinity, potency and selectivity, and the use of these methods to generate MRGPRX1 agonists and CXCR4, GLP1R, GIPR, GCGR and CGRPR antagonists. Cryo-electron microscopy data reveals atomic-level agreement between designed and experimentally determined structures for CGRPR-bound antagonists and MRGPRX1-bound agonists, confirming precise conformational control of receptor function. Our *de novo* design and screening approach opens new frontiers in GPCR drug discovery and development.

Main

G protein-coupled receptors (GPCRs) are the largest and most diverse family of membrane receptors in the human genome[1], and play critical roles in many physiological processes. GPCRs are implicated in a wide array of diseases including cancer, cardiovascular and metabolic diseases, and neurological disorders[1], and hence are at the forefront of drug discovery and development[2]. Over the past decades, biologics including antibodies, nanobodies, and peptides have gained momentum as GPCR therapeutics and tools[3]. However, the design of biologics modulating GPCR signaling remains an outstanding challenge, often requiring a combination of strategies such as the insertion of peptide fragments from native proteins or screening of random libraries[4]. It has been particularly difficult to generate GPCR agonists, which has necessitated considerable antibody and receptor engineering efforts[5,6].

Advances in computational protein design have enabled the design of miniprotein binders with atomic-level accuracy for many targets of biological interest[7]. Methodologies such as RFdiffusion[7] and Rosetta[8] enable the design of miniprotein binders with desirable properties, including exceptional selectivity[9], high protease stability[10], and extended biological half-life[11]. Despite these advances, formidable challenges remain, particularly for functional miniproteins targeting membrane-embedded binding pockets such as flexible, recessed GPCR epitopes, which need to be conformationally specific to induce function. We reasoned that specialized computational design and new high-throughput experimental screening methods would be needed to tackle these challenges, and set out to develop appropriate methods.

Development of computational and experimental methods to target diverse GPCR epitopes

To enable targeting of the deeply recessed orthosteric binding site epitopes — critical for modulating class A GPCR function — we implemented two complementary design methods to generate functional miniproteins. First, we developed a “motif-directed” RFdiffusion approach that rather than diffusing an entire binding protein, starts with just a five-residue peptide (the motif) to interact with target hot spot residues within the recessed binding pocket (**Fig. 1a**). The short peptide can more readily penetrate into the deep pocket, and once good solutions for a binding peptide are found, the interacting peptide is kept in a fixed position and full miniproteins are generated using the motif-scaffolding capabilities of RFdiffusion[7]. To increase diversity of designs for library-scale experimental screening, we developed an iterative partial diffusion approach which generates new designs in the vicinity of the most promising *in silico* solutions at each stage of the process. Second, we developed an approach, MetaGen, that employs structurally diverse scaffolds from the AlphaFold generated metaproteome[12,13] (**Supplementary Fig. 1**) in Rosetta RifDock calculations. In contrast to traditional *de novo* miniprotein backbone libraries[8], often composed of straight helices and short loops ill-suited for engaging deeply recessed epitopes, these scaffolds feature protruding elements, such as kinked helices and beta-hairpin loops, but are still confidently predicted from a single sequence — a key criterion for designability[14]. Following backbone design with either RFdiffusion or MetaGen we used ProteinMPNN for sequence design[15] and AlphaFold2 (AF2) *initial guess*[12] as well as Rosetta metrics for filtering designs[14]. To design class B receptor antagonists, we similarly deployed the MetaGen backbone library or generated backbones from scratch using RFdiffusion.

Due to the challenging nature of class A GPCR epitopes, we reasoned that high-throughput screening (HTS) methods would be necessary to complement computational design for robust identification of functional binders. To address this, we developed Receptor Diversion (RD), a purification-free HTS assay that operates directly in human cells (**Fig. 1b**). In this assay, both the membrane protein target and the candidate binder are expressed in a human cell line, with the binder localized within the secretory pathway using a genetic tag (e.g., an endoplasmic reticulum (ER) retention signal). This allows the binder to interact with the extracellular face of the membrane protein target. High-affinity interactions cause "diversion" of the target from its normal trafficking pattern, which can be visualized as an increased binder-target colocalization (**Fig. 1c**). Across 7 diverse GPCRs, we observed a robust binding signal suitable for high-throughput screening (**Fig. 1d, e**) with a cross-GPCR Z' average of 0.47 when sampling 100 cells per binder (**Supplementary Table 1**). RD has the advantages that (i) the target can be expressed at near-endogenous levels in a relevant cell line and does not have to be produced as a stable soluble protein (challenging for GPCRs) as required for display methods, (ii) binders discovered through the screen must be efficiently translated into ER in human cells, be soluble and function in the molecularly crowded environment of the secretory pathway, and (iii) the binder must specifically bind the target in order to induce receptor diversion. To deploy the assay at library scale, we use optical pooled screening (OPS), where individual designs are encoded together with a DNA barcode, and optically genotyped using *in situ* sequencing (**Fig. 1f-i**). The RD platform enables screening of up to 100,000 designs through imaging of up to 10^7 cells providing expression and co-localization data at the single-cell level. As we were unsure how well RD screening would work in practice for library screening at the beginning of this study, we also explored the use of yeast display paired with either soluble GPCRs in nanodiscs[16] or GPCRs displayed on mammalian cells (biofloating)[17].

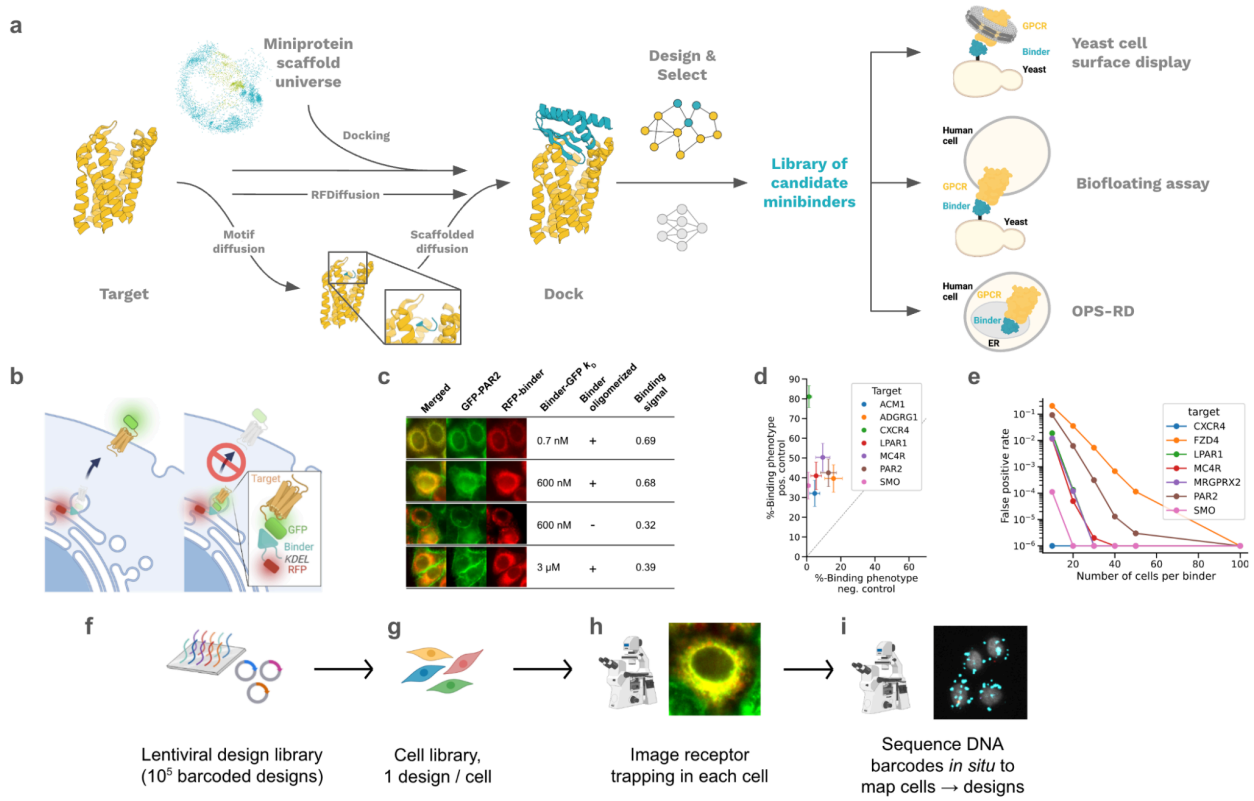

Fig. 1. GPCR binder computational design and screening methods. a Designed backbones targeting GPCRs of interest were generated either *de novo* using constrained or scaffold-guided RFdiffusion (bottom), or by docking a library of 7,000 native miniproteins (top). Following sequence assignment and selection of most promising designs based on *in silico* metrics, class A and class B GPCR binders were screened either directly in functional assays or first by high-throughput binding assays including yeast cell surface display using nanodiscs, biofloating assay or a newly developed Optical Pooled Screening-Receptor Diversion (OPS-RD) assay in mammalian cells in which designed binders retained in the ER retain fluorescently tagged wild-type receptors. Binding is detected by converting binder-receptor interactions to an optical phenotype: in the absence of binding, fluorescently tagged receptors traffic to the cell surface while the design is retained separately in the secretory pathway (**b,** *left*), whereas a successful binder colocalizes with the receptor in the secretory pathway (**b**, *right*). **c** Using nanobodies with known affinities[18] targeting a GFP-fused protease-activated receptor 2 (PAR2), the binding signal (GFP-RFP pixel cross-correlation) is proportional with the binding affinity, can be enhanced using oligomerized binding constructs with increased avidity. **d** The binding phenotype is robust across seven GFP-fused GPCRs, with positive controls (C5-oligomerized 0.7 nM anti-GFP nanobody) showing significantly higher binding signals compared to negative controls (non-binding miniproteins). The fraction of cells with the binding phenotype is computed from ≥80 cells, and SEM is scaled to N = 50 cells (a scale suitable for HTS). **e** False-positive rate at a fixed false-negative rate (5%) as a function of the number of cells imaged across GPCR targets based on the same controls as **d**. To deploy OPS-RD at scale, **f** designed binders are synthesized on oligo arrays and cloned into a lentiviral library, **g** low multiplicity of infection (MOI) transduction creates a cell library with one binder design per cell, **h** binding is quantified by receptor trapping, and **i** *in situ* sequencing of a DNA barcode reveals the identity of the binder in each cell.

Pharmacology and biophysical characterization of receptor penetrating class A MRGPRX1 agonists

To explore the potential of computational design to create GPCR agonists, we focused on the Mas-related G protein-coupled receptor X1 (MRGPRX1), an emerging target for itch and pain[19]. Using the MetaGen approach, we targeted a large epitope within the orthosteric binding pocket spanning transmembranes 2-7 (TM2–TM7) of three active-state structures, reasoning that active-state stabilization alone would be sufficient to generate agonists[20]. We screened a library of 13,000 designs using OPS-RD, and succeeded in mapping optical binding phenotypes for 800,000 cells to their design genotypes. Averaging optical phenotypes across cells, we ranked each design and selected 64 designs (**Fig. 2a**), then generated an additional 27 designs using partial diffusion, resulting in a complete set of 91 designs selected for further characterization.

Of these, 50 were highly expressed in E. coli and subsequently screened in a calcium mobilization assay to explore their ability to stimulate intracellular signaling. Consistent with the design strategy, seven miniproteins demonstrated agonistic activity at 10 μM (Supplementary **Fig. 2**). Next, we generated concentration-response curves for the seven hits and obtained two full agonists with EC_{50} values of 390 nM and 1 μM, respectively, while additionally discovering a partial agonist which displayed an EC_{50} of 1.4 μM (**Fig. 2b, Supplementary Table 2**). The three hits were structurally diverse (**Fig. 2c**), highly expressed, monomeric by SEC (**Fig. 2d**), and have CD spectra consistent with the expected molecular structure (**Fig. 2e**) as well as high thermal stability (**Fig. 2f**).

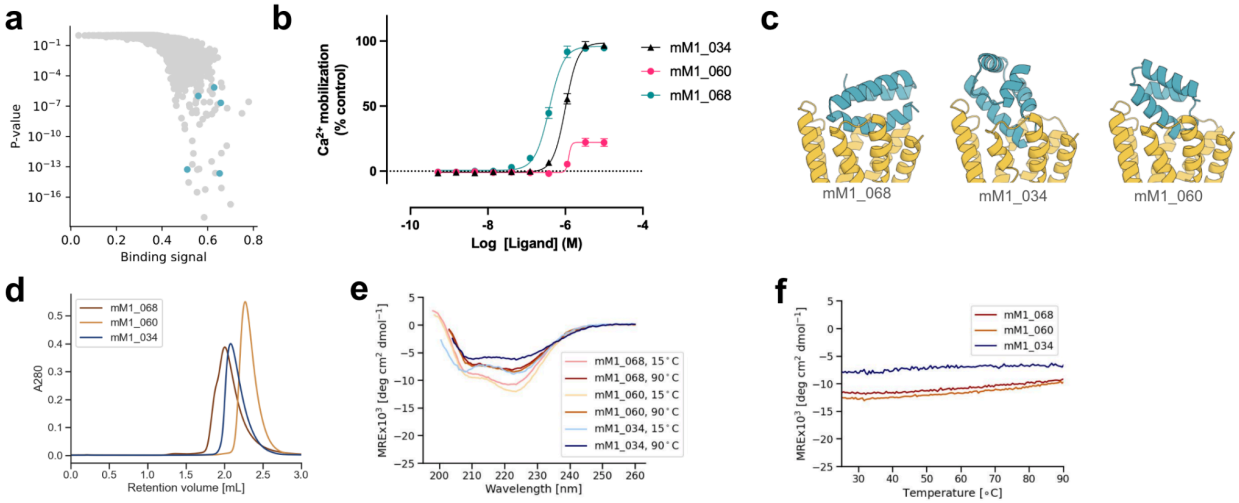

Fig. 2. Biophysical characterization and pharmacological properties of MRGPRX1 binders. a 13,000 miniprotein binders were designed as agonist with MetaGen and tested using OPS-RD. The colocalization (binding signal) induced in cells with the same binder was compared to the colocalization distribution across all imaged cells (>2.5 million), and P-values were computed using a Kolmogorov–Smirnov (K-S) test. **b** Concentration-response curves of three agonist hits measured in a calcium flux assay (n=3). **c** Computational models, **d** size-exclusion chromatography (SEC) traces**, e** circular dichroism (CD) spectra of the top three agonist hits, and **f** melting curves. Receptor structures are truncated for clarity.

Cryo-EM structures of agonists bound to MRGPRX1

We determined the cryo-electron microscopy (cryo-EM) structure of mM1_068 and mM1_060 miniprotein binders bound to hMRGPRX1 in complex with a mini-G_q protein[20] at global resolutions of 3.29 Å and 3.13 Å, respectively (**Fig. 3a, Supplementary Fig. 3 and Fig. 4**). The mM1_068 agonist adopts a proline-kinked three-helical bundle fold nearly identical to the designed model (0.7 Å Cα-RMSD), stabilizing the receptor in an active-state conformation closely resembling the target receptor structure (0.7 Å Cα-RMSD across the top half of the receptor compared to 8DWG). Similarly, the cryo-EM structure of the complex between mM1_060 and hMRGPRX1 are very close to the design model, both over the design alone (0.7 Å Cα-RMSD) and over the top half of the (active-state) receptor structure (0.8 Å Cα-RMSD). The local resolution for the miniprotein alpha helices were lower with higher B-factors compared to the transmembrane bundle (**Supplementary Table 3**) which is to be expected given the size of the miniproteins and their partial protrusion from the MRGPRX1 orthosteric site. The lower resolution of the helices enabled only backbone modelling of residues exposed to the lipid environment. However, EM density was clearly observed in the cryo-EM maps for the alpha helices within the orthosteric site or residues which made close interactions with MRGPRX1 residues or extracellular loops. For the MRGPRX1:mM1_068 complex, mM1_068 contributes 15 residue sidechains and 1102 $Å^2$ of buried surface area (BSA) to the interface while MRGPRX1 contributes 23 residue sidechains and 1004 $Å^2$ of BSA. For the MRGPRX1:mM1_060 complex, despite being smaller in size, mM1_060 contributes 22 residue sidechains and 1269 $Å^2$ of BSA to the interface while MRGPRX1 contributes 31 residue sidechains and 1292 $Å^2$ of BSA. Together, mM1_068 and mM1_060 share 17 residue sidechains despite having nearly opposite orientations in the MRGPRX1 orthosteric site (**Supplementary Table 4**).

Previous structure and functional studies identified critical residues within the orthosteric site of MRGPRX1 necessary for the endogenous peptide, bovine adrenal medulla (BAM) 8-22 to activate the receptor for signaling[20,21]. Both miniproteins overlap with the BAM 8-22 site (**Supplementary Table 4**). Several sidechains ($E157^{4.60}$, $L240^{6.59}$, $F236^{6.55}$, $W241^{6.60}$, and $F250^{7.31}$) have positions which differ in the determined structures, which may relate to the observed partial and full agonism of MRGPRX1. Overall, the cryo-EM data are in close agreement with the computational design models and confirm that the miniprotein binders sterically occlude the hMRGPRX1 orthosteric site (**Fig. 3b-3d**).

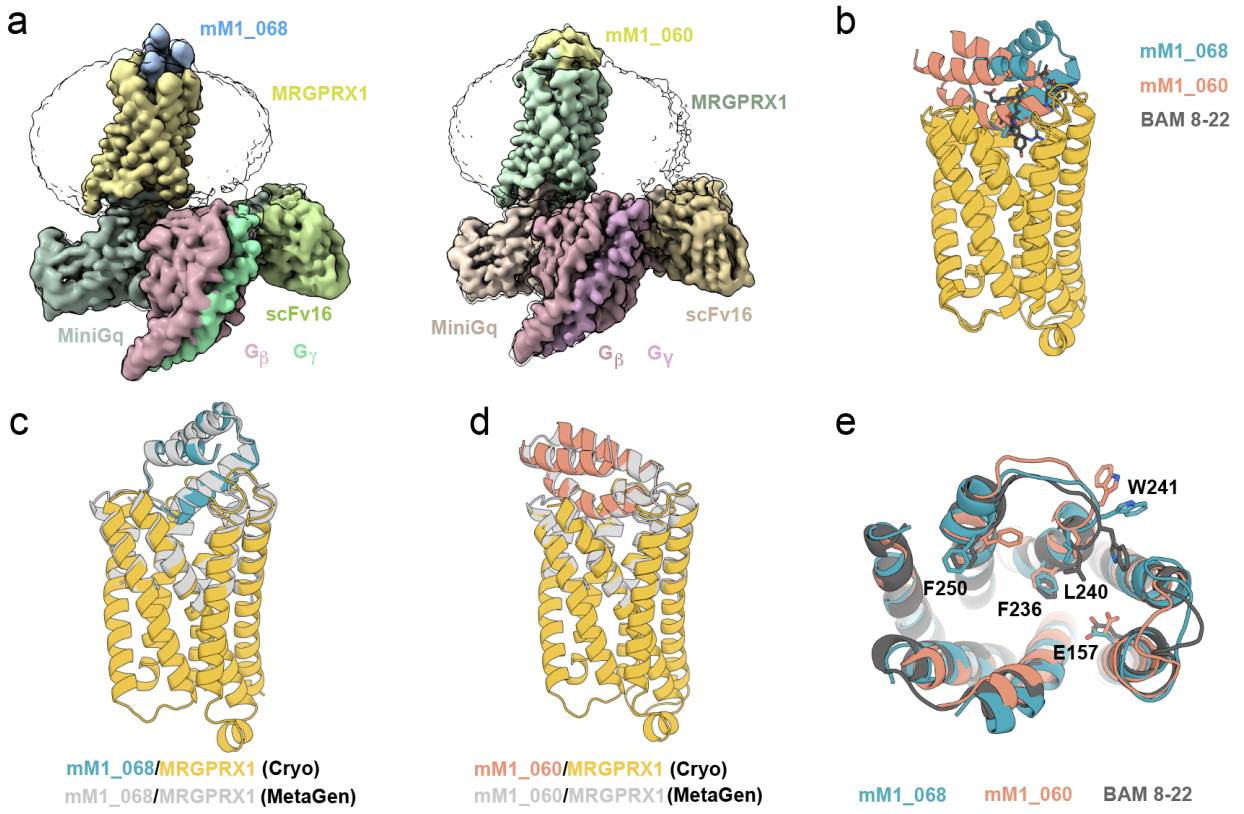

Fig. 3. Cryo-EM structures of MetaGen-designed MRGPRX1 binders. a Cryo-EM maps of hMRGPRX1 bound to miniprotein mM1_068 (left) and hMRGPRX1 bound to mM1_060 (right). The silhouettes show the map at low threshold to enable visualization of the detergent micelles. **b** Aligned cryo-EM models of mM1_068, mM1_060, and BAM 8-22 bound to hMRGPRX1. **c** Alignment of the experimental structure of mM1_068 + hMRGPRX1 complex with the designed model. **d** Alignment of the experimental structure of mM1_060 + hMRGPRX1 complex with the designed model. **e** Key residues involved in MRGPRX1 activation and signaling from the cryo-EM structures of MRGPRX1 in complex with mM1_068 and mM1_060 reveals significant differences compared to the MRGPRX1–BAM 8-22 structure.

Pharmacology and biophysical characterization of receptor penetrating class A CXCR4 antagonists

We sought to design receptor-penetrating antagonists for the C-X-C chemokine receptor type 4 (CXCR4), a class A target receptor that has been implicated in cancer and viral infection[22]. We reasoned that interacting with the receptor at both extracellular loops and deeper within the transmembrane region would be required to stabilize the inactive receptor conformation and to yield a potent CXCR4 antagonist. 26,000 designs generated with scaffold-guided RFdiffusion (**Fig. 1a**, **Fig. 3a**) were screened using the biofloating approach (**Fig. 1a**, **Supplementary Fig. 5a**)[17], which uses CXCR4 expressed at the plasma membrane of mammalian cells[17] to probe yeast display libraries. We identified two hits, both from the same backbone (**Fig. 3b**, **Supplementary Fig. 5a-e, Supplementary Fig. 6a**). The proteins were expressed in E. coli, and eluted from SEC in a single peak (**Fig. 3c**, **Supplementary Fig. 6b**). While the binder dCX1_002 had an IC_{50} in the μM range (**Supplementary Fig. 6c**), the binder dCX1_001 had an IC_{50} of 24 nM (**Supplementary Fig. 7a, Supplementary Table 5**) and was highly thermostable (**Fig. 3d**, **e**). dCX1_001 antagonized CXCL12-mediated signaling through G_i coupled CXCR4 with a pA_2 value of 7.6 (25 nM) (**Fig. 3f**) and no agonistic activity was observed (**Supplementary Fig. 7b**). dCX1_001 was also identified from the 26,000-design library using yeast display with nanodisc stabilized CXCR4 (**Supplementary Fig. 8a-d**).

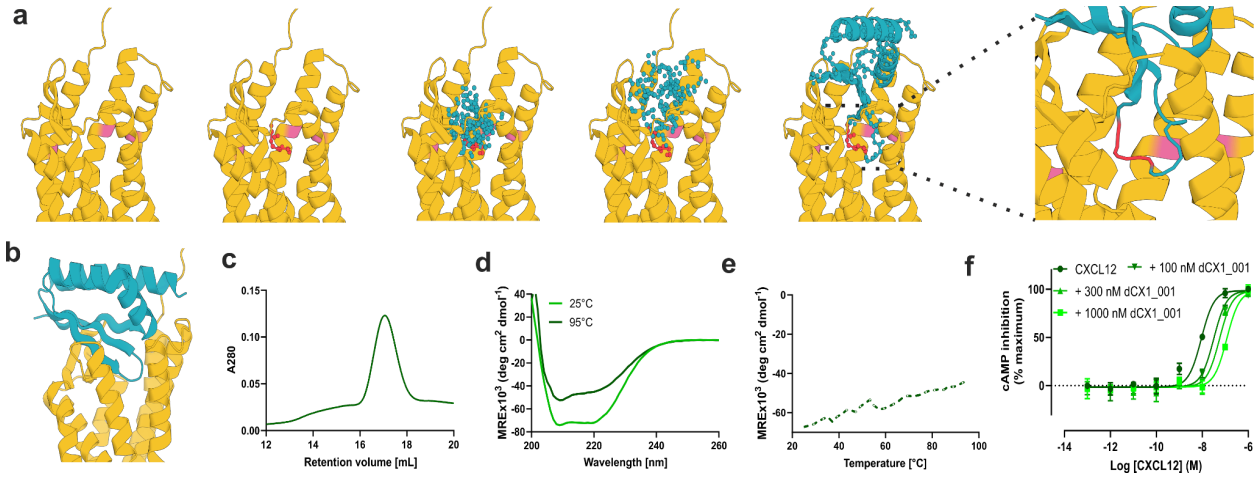

Fig. 4. Biophysical characterization and pharmacological properties of CXCR4 binder dCX1_001. a A representative RFdiffusion trajectory for generating binders (blue) against the CXCR4 (yellow, PDB ID: 4RWS). Selected hot spots are highlighted in pink and *de novo* pentamer motifs used for scaffolding are shown in red. Inset shows deep insertion of the motif (red) and resulting binder. Receptor structure was truncated for clarity. **b** Computational model of the most potent CXCR4 binder, dCX1_001 (miniprotein in blue, receptor in yellow). **c** Size-exclusion chromatography (SEC) traces**, d** circular dichroism (CD) spectra and **e** melting curves of the dCX1_001 binder. **f** Functional cAMP assay of dCX1_001 binder in CHO cells stably expressing CXCR4. Data are shown as mean ± SEM (n=4). Schild regression analysis indicates dCX1_001 is an antagonist with *p*A2 of 7.6 ± 0.3 (25 nM) and slope of 0.68 ± 0.13).

Pharmacology and biophysical characterization of multiple class B GPCR antagonists

To explore the design of antagonists of class B GPCRs, which include many therapeutic targets[23], we applied our design methods to glucagon-like peptide 1 receptor (GLP1R), gastric inhibitory polypeptide receptor (GIPR), glucagon receptor (GCGR) and the calcitonin gene-related peptide receptor (CGRPR). We targeted the soluble extracellular domain (ECD) of these receptors, reasoning that binding to the ECD should induce steric hindrance and prevent peptide interaction with the receptor, thereby resulting in antagonism (**Fig. 4a**). For GLP1R, GIPR and GCGR, we used yeast display and soluble ECDs of receptors to identify miniprotein binders. After testing a yeast library of approximately 10,000 designs for the GLP1R (**Supplementary Fig. 9a-e**) we expressed 96 designs and identified binders dGl1_024 and mGl1_008 with an affinity of 27 nM and 5.3 nM, respectively (**Supplementary Fig. 10a-e**). Similarly, after probing yeast display libraries of about 18,000 and 12,000 designs for GIPR (**Supplementary Fig. 11a-d)** and GCGR**,** (**Supplementary Fig. 12a-d**), respectively, we expressed 96 designs for each target receptor and obtained miniprotein binders with nanomolar and picomolar affinities (**Supplementary Fig. 13a-j**, **Supplementary Fig. 14a-j**).

CGRPR is a heterodimer consisting of calcitonin-receptor like receptor (CLR) and receptor activity-modifying protein 1 (RAMP1) and is an established target for developing migraine therapeutics[24]. After closely inspecting the nature of the CGRPR epitope, we hypothesized that we could achieve a >1% hit rate, given the high epitope hydrophobicity and absence of loops, and did not attempt high-throughput screening techniques. We screened designs in a functional, one-point cAMP assay using a SK-N-MC cell line (**Supplementary Fig. 15**). Out of 96 ordered RFdiffusion designs, 67 expressed, and a three-helix bundle miniprotein dC1_021 was identified with an IC_{50} of 447 nM (**Supplementary Fig. 16, Supplementary Table 6**). Out of 89 MetaGen-derived backbones, we identified the competitive antagonist mC1_023 of mixed αβ-topology (**Fig. 4a**) with an IC_{50} of 37 nM (**Supplementary Fig. 17a, Supplementary Table 6**) and a pA_2 of 5 nM (**Fig. 4b**). Disulfide stapling[25] of a second MetaGen hit, mC2_022 (**Fig. 4a**), yielded an antagonist binder with an IC_{50} of 420 nM (**Supplementary Fig. 17b, Supplementary Table 6**) and a pA_2 of 13 nM (**Fig. 4a, c**). To increase the potency of the RFdiffusion hit dC1_021, we performed partial diffusion[6]. Out of 78 designs that expressed, 36 had measurable antagonistic activity against CGRPR in a one-point cAMP assay (**Supplementary Fig. 18**). Concentration-response curves of the 20 most promising binders identified the competitive antagonist dC2_049 with an IC_{50} value of 4.5 nM (**Supplementary Fig. 19**, **Supplementary Table 6**) and a pA_2 of 3.9 nM (**Fig. 4a, d**). All three antagonists migrated as single peaks in size exclusion chromatography (SEC), with mC1_023 eluting as a monomer, whereas mC2_022 and dC2_049 eluted as dimers, had CD spectra consistent with their structures, and high thermal stability (**Fig. 4e-g**). dC2_049 exhibited high selectivity for the CGRPR with little or no significant cross-reactivity at the related adrenomedullin receptors 1 (AM_1) and 2 (AM_2), calcitonin receptor (CTR) and amylin 1 receptor (AMY_1R) as assessed by the ability of miniproteins to inhibit a single concentration of the endogenous agonists for each receptor (**Fig. 4h-k**). A sequence-similar (sequence identity 59%) RFdiffusion-designed binder, dC2_050, derived by partial diffusion starting from the same parent structure as dC2_049, had similar pharmacological properties and biophysical characteristics (**Supplementary Fig. 20a-e**,

Supplementary Fig. 21a-d, Supplementary Table 6). Both dC2_049 and dC2_050 were also CGRPR antagonists in COS-7 cells transiently expressing CLR and RAMP1 (**Supplementary Fig. 22a**, **b**).

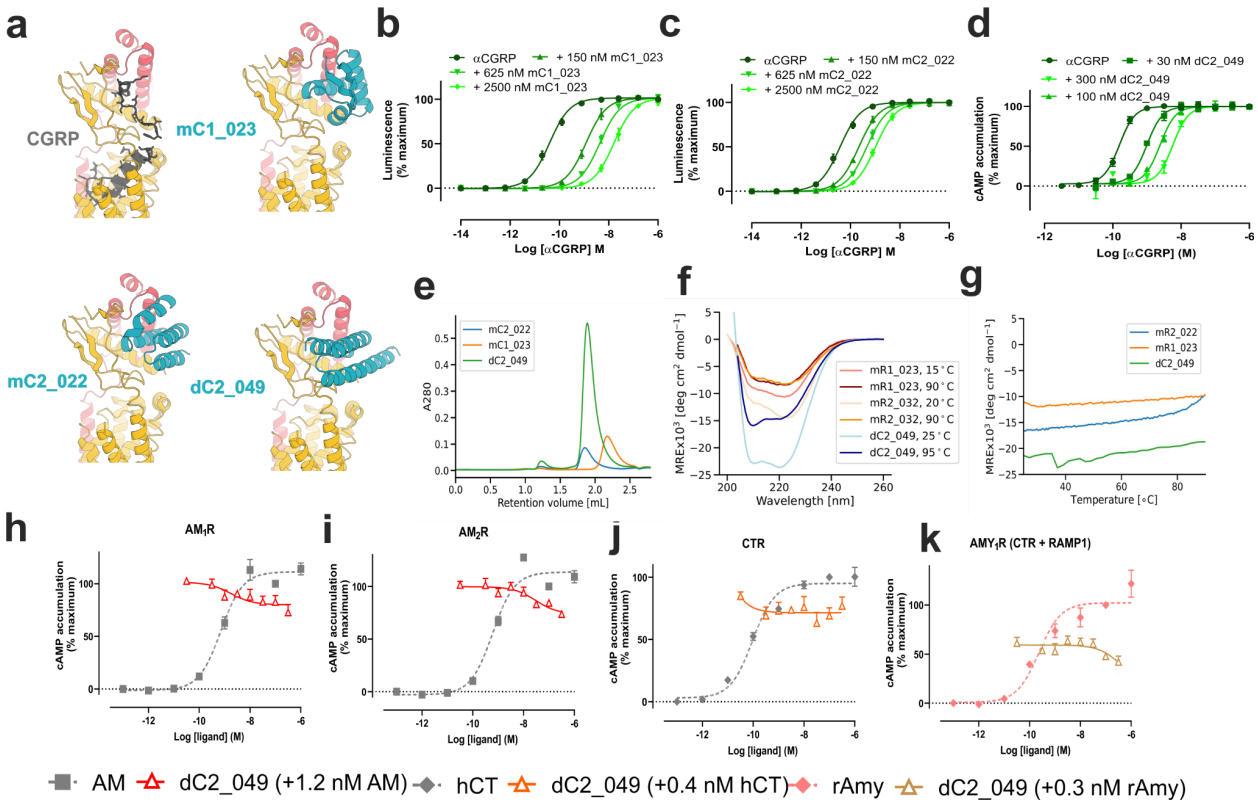

Fig. 5. Biophysical and pharmacological characterization of CGRPR binders. a, Structure of the CGRPR (yellow) bound to αCGRP (gray, PDB ID: 6E3Y) and computational design models (blue) of MetaGen (mC1_023, mC2_022) and RFdiffusion (dC2_049) generated antagonists. Receptor structures are truncated for clarity. Schild regression analysis and functional estimates of **b** mC1_023 (pA_2 of 8.3 ± 0.1 (5 nM) and slope 0.95 ± 0.04, n=4), **c** mC2_022 (pA_2 = 7.9 ± 0.1 (13 nM), slope 0.61 ± 0.04, n=4) and **d** dC2_049 (pA_2 = 8.4 ± 0.1 (3.9 nM), slope = 0.82 ± 0.05, n=4). **e** Size-exclusion chromatography (SEC) traces, **f** circular-dichroism (CD) spectra and **g** melting curves of mC2_022, mC1_23 and dC2_049 binders. Selectivity profile of dC2_049 binder at the **h** adrenomedullin receptor 1 (AM1), **i** adrenomedullin receptor 2 (AM2), **j** calcitonin receptor (CTR) and **k** amylin receptor 1 (AMY1R). Data in figures are shown as mean ± SEM (n=4).

Cryo-EM structures of antagonists bound to CGRPR

We determined the cryo-EM structure of dC2_049 and dC2_050 miniprotein binders bound to CGRPR (**Fig. 5a-e**) with global resolutions of 3.2 Å and 4.1 Å, respectively (**Supplementary Fig. 23**, **24**). The cryo-EM structures (**Supplementary Fig. 23, 24, Supplementary Table 7**) are in good agreement with computational design models (~ 1 Å Cα RMSD) (**Fig. 5a-e**) and confirm that the binders sterically occlude binding of the C-terminal portion of the CGRP (**Fig. 5a-e**). Local resolutions for the ECD were lower than for the transmembrane bundle, which is commonly observed among class B1 GPCR cryo-EM structures[26], enabling only backbone modelling of most residues in the ectodomain. However, density was observed in the cryo-EM

map for some larger, interfacial sidechains along helix-1 and helix-3 of dC2_049. For instance, TRP72^{ECD}, a key residue for CGRP activity[27], forms hydrophobic contacts with Met12 of dC2_049 (**Supplementary Fig. 25**).

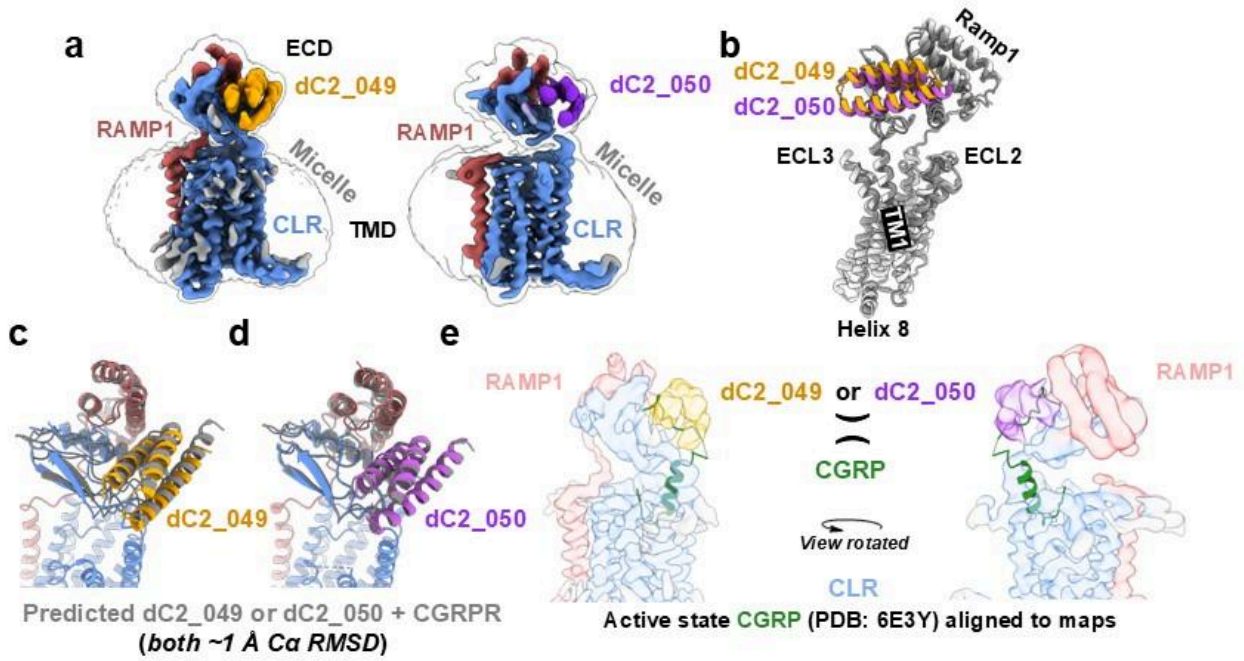

Fig. 6. Cryo-EM structures of RF-diffusion designed CGRPR binders. a Cryo-EM maps of CGRPR bound to dC2_049 (left) and CGRPR bound to dC2_050 (right). The silhouettes show the map at low threshold to enable visualization of the detergent micelles. **b** Aligned models of dC2_049 (gold) and dC2_50 (purple) bound to CGRPR (colored white and gray, respectively). **c** Alignment of the experimental structure of dC2_049 + CGRPR with the predicted structure (colored gray) of dC2_049 and the CGRPR ectodomain. **d** Alignment of the experimental structure of dC2_050 + CGRPR with the predicted structure (colored gray) of dC2_050 and the CGRPR ectodomain. **e** Maps shown as translucent surfaces of CGRPR bound to dC2_049 (left) and dC2_050 (right) with the active state structure of CGRP bound to CGRPR (receptor and G protein not shown for clarity) aligned to the ectodomains. The densities for dC2_049 and dC2_050 sterically occlude binding of the C-terminal section of CGRP.

Discussion

GPCRs have been longstanding challenges for drug discovery and development of protein-based ligands owing to their structural complexity and dynamic character. We show that *de novo* design can address these challenges by generating miniprotein binders targeting MRGPRX1, CXCR4, GLP1R, GIPR, GCGR and CGRPR with diverse affinity, potency, and selectivity profiles. Agonists have been particularly challenging to obtain due to the need for conformational selectivity, requiring discrimination between subtle structural differences in the orthosteric binding site that distinguish active from inactive states[28]. Here, we demonstrate the *de novo* design of two atomically accurate binders for MRGPRX1, capable of precisely controlling the receptor's conformation (within 0.7 Å) to induce agonism, including both full and partial miniprotein agonists. These findings establish *de novo* design as a viable strategy for engineering GPCR-targeting ligands that not only recognize but also precisely control a conformational epitope to achieve a defined and desired pharmacological outcome. Complementing our computational design approaches, our in-cell OPS-RD platform enables high-throughput screening for difficult GPCR targets by circumventing the need for engineering of soluble receptor preparations in artificial nanodiscs, liposomes or mutant receptor proxies, which can potentially alter sampling of receptor conformations and functional properties[29].

The therapeutic potential of *de novo* designed GPCR antagonists and agonists is considerable given the central roles GPCRs play in cellular function and disease. The ability to computationally design binders interacting with specific receptor regions in specific conformations – difficult or impossible to control in screens of immune repertoire libraries – is a step change in methodology for obtaining functional biologics targeting integral membrane receptors. Beyond therapeutics, designed binders have considerable potential as versatile tools for drug discovery, uncovering novel pharmacological insights into receptor function, and stabilizing receptor conformations for structural studies. This work paves the way for transformative GPCR-related applications in both basic research and drug discovery.

Methods

Binder design using RFdiffusion and metaproteomic scaffolds

The cryo-EM structures of GLP1R, (PDB ID: 5VAI), GIPR (PDB ID: 7FIN, 7FIY), GCGR (PDB ID: 5XEZ), CGRPR (PDB ID: 6E3Y) and the crystal structure of CXCR4 (PDB ID: 4RWS) were used as targets for designing binders with RFdiffusion. Additionally, cryo-EM structures of CGRPR (PDB IDs: 3N7S, 7KNU, 6E3Y), MRGPRX1 (PDB IDs: 8DWC, 8DWG, 8DWH), GLP1R (PDB IDs: 6VCB, 6X18, 7DUQ), GIPR (PDB IDs: 2QKH, 4HJ0, 7FIN) and GCGR (PDB IDs: 6WPW, 8JIT, 8JIU) served as targets for binder design using metaproteomic scaffolds. All target structures were truncated to the region containing the binding epitope.

Backbone generation using motif-scaffolded RFdiffusion targeting GLP1R, GIPR, GCGR or free RFdiffusion against CGRPR was performed as previously described[7]. For the GLP1R, GIPR and GCGR, 50,000-100,000 backbones were created using following hot spot residues chosen within the ECD of the receptor GLP1R L95, GIPR M32 and GCGR F33, W36 and W87. For the CGRPR, three hydrophobic hotspot residues (L33, W72, F92) were chosen within the ECD of the receptor and approximately 50,000 backbones were generated. Sequences were designed using ProteinMPNN (10 sequences per backbone)[15], followed by FastRelax and AF2 *initial guess*[12]. Designs generated by RFdiffusion were selected based on pAE_interacion < 8, pLDDT_binder > 85, Rosetta ddG < -45, spatial_aggregation_propensity (sap) < 60 for GLP1R, pAE_interaction < 6, pLDDT_binder > 90, Rosetta ddG < -45 and sap < 60 for GIPR, pAE_interaction < 8, pLDDT_binder > 90, Rosetta ddG < -50 and sap < 45 for GCGR and pAE_interaction < 8, pLDDT_binder > 90 and Rosetta ddG < -45 for CGRPR.

Metaproteome-derived designs targeting CGRPR, MRGPRX1, GLP1R, GIPR, and GCGR were generated using the RIFdock, motif extraction, and recycling strategy outlined in Cao et al.[8]. Following sequence design and prediction. Selection criteria varied by target: CGRPR designs were chosen based on pAE_interaction < 8, binder_RMSD < 2, and scaffold_pLDDT > 90; MRGPRX1 designs met pAE_interaction < 10 or (sap < 40 & contact_molecular_surface > 600 & membrane_insertion_energy > 4 & Rosetta ddG < −51); GLP1R designs satisfied pAE_interaction < 12, sap < 40, ddG < −30, scaffold_pLDDT > 85, membrane_insertion_energy > 4, and binder_RMSD < 2; GIPR designs were selected based on pAE_interaction < 6, binder_RMSD < 2, ddG < −40, scaffold_pLDDT > 90, membrane_insertion_energy > 4, and sap < 35; and GCGR designs met pAE_interaction < 12, binder_RMSD < 2, ddG < −40, scaffold_pLDDT > 85, sap < 40, and membrane_insertion_energy > 4.

Partial diffusion was performed on the AF2 model of the most promising CGRPR hit (dC1_022). Roughly 3,000 backbones were designed by applying 10, 15, and 20 noising timesteps out of a total of 50 timesteps in the noising schedule followed by denoising steps (diffuser.partial_T input values of 10, 15 and 20). The resulting backbone libraries after free and partial RFdiffusion were subjected to sequence design using ProteinMPNN (10 sequences per backbone)[15], followed by FastRelax and AF2 *initial guess*[12]. The resulting libraries were filtered based on AF2 pAE_interaction < 4, pLDDT_binder > 90, and Rosetta ddG < −45.

For the CXCR4 binder design, we used RFdiffusion first to generate penetrating pentamers using hotspot residues W94, I259, I284. About 1,000 pentamers were designed and 50 of them with the deepest insertion within the binding pocket of the receptor based on their distance to the hotspot residues were selected for subsequent scaffolding. Per selected pentamer, 1,000 scaffolds were generated by building 0-70 residues on the N and C-termini of the central three residues. The lengths of the termini were randomly sampled but were restricted to a final total length range of 65-75 residues for each design. To reduce the likelihood of diffusing scaffolds that would cross the extracellular membrane surface and interact with the transmembrane portion of the target, prior to scaffolding, hydrophobic cell membrane-facing residues of the receptor were mutated to glutamines. Following backbone design, mutated residues were reverted to native sequences and the backbones were sequence designed using ProteinMPNN (10 sequences per backbone) in combination with FastRelax. The structures of these designs were then predicted by AF2 *initial guess*. Designs that passed *in silico* criteria (AF2 pAE_ineraction < 15, pLDDT_binder > 75, and Rosetta ddG < −45) were next subjected to an iterative partial diffusion approach. For each iteration, the receptor backbone and sequence were kept fixed and the designed complex was subjected to 20 partial diffusions (diffuser.partial_T = 15). Backbones from the last 10 denoising timesteps of each diffusion trajectory and the final design at T=0 were sequence designed using ProteinMPNN, and the resulting 220 designs were predicted using AF2 *initial guess*. The AF2 prediction with the lowest pAE_interaction was chosen as the input for the next iteration for a total of 10 iterations. Designs that passed more stringent *in silico* criteria (pAE_interaction < 8, pLDDT_binder > 80, Rosetta ddG < −45 and sap < 60) were selected for library construction and high throughput screening. Sequences were designed using the receptor template that contained a C mutation in the binding pocket, previously used for structural stabilization of the receptor[30], however, prior to the final AF2 prediction, designs were mutated to the native D of the receptor.

Cloning, expression and purification of protein binders

Protein binder designs were ordered as synthetic genes (eBlocks, Integrated DNA Technologies) with compatible BsaI overhangs to the target cloning vector, LM0627 for Golden Gate assembly[31]. Subcloning into LM0627 resulted in the following product: MSG-[protein]-GSGSHHWGSTHHHHHH, with the C-terminal SNAC cleavage tag and 6xHis affinity tag. Briefly, Golden Gate subcloning reactions of designs were performed in 96-well PCR plates in 1 µL volume. Reaction mixtures were then transformed into a chemically competent expression strain (BL21(DE3)) and 10 mL of these split directly into four 96-deep well plates containing 990 uL of auto-induction media (autoclaved TB-II media supplemented with kanamycin, 2 mM $MgSO_4$, 1X 5052). Designs generated using the MetaGen pipeline were plated to single colonies and sequence verified before inoculating expression media. Post overnight incubation at 37°C (20-24 hours), cells were harvested, lysed, and clarified lysates applied to a 75 µL bed of Ni-NTA agarose resin in a 96-well fritted plate equilibrated with a Tris wash buffer. After sample application, the resin was washed, and samples were eluted in 200 µL of a Tris elution buffer containing 300 mM imidazole. Proteins were then purified via SEC using an AKTA FPLC equipped with an ALIAS autosampler capable of running samples from two 96-well source plates. A Superdex75 Increase 5/150 GL column was used (Cytiva 29148722). CXCR4 binder hits identified by yeast display were ordered as fully cloned genes

(Integrated DNA Technologies), transformed into chemically competent *E. coli* strain BL21(DE3) and expressed in 50 mL of auto-induction media with reagents described above. Purification was conducted analogous to other binders, but using a S75 10/300 GL column (Cytiva 29148721). To verify the identity of MetaGen designed proteins, intact mass spectra were obtained via reverse-phase LC/MS on an Agilent G6230B TOF on an AdvanceBio RP-Desalting column, and subsequently deconvoluted by way of Bioconfirm using a total entropy algorithm. RFdiffusion designed binders identified as hits in screens were confirmed by sequencing.

Circular dichroism

For circular dichroism (CD) measurements, diffusion-derived designs were diluted to 0.4 mg/ml in 20 mM Tris (pH 8.0) and 100 mM NaCl, while metaproteome-derived designs were analyzed at 50 μM in PBS (pH 7.4). Spectra were acquired on a JASCO J-1500 CD Spectrophotometer. Thermal melt analyses were performed between 25℃ and 95℃, measuring CD at 222 nm. All reported measurements were acquired within the linear range of the instrument.

Cell culture

CHO-K1/CRE-Luc/CGRPR cells were cultured in Ham's F-12K (Kaighn's) Medium (Gibco) containing 10% FBS. RBL-2H3 cells were cultured as per standard procedures. HEK293T and SK-N-MC cells were cultured in Dulbecco's modified Eagle's medium (DMEM) medium (Thermo Fisher) containing 10% FBS and penicillin-streptomycin (500 U/mL). COS-7 cells were cultured in DMEM containing 10% FBS only. CHO-CXCR4 stable cell line was cultured in DMEM/F12 medium supplemented with 10% FBS, penicillin-streptomycin (500 U/mL), 4 μg/ml puromycin and 100 μg/ml hygromycin B. LentiX 293T cells (Takara #632180) and HeLa cells, the latter optimized for optical pooled screening and kindly gifted by Iain Cheeseman, were cultured in D10 media (DMEM with GlutaMAX, 10% (v/v) FBS, and 100 U/mL penicillin–streptomycin). All cells were grown at 37°C with a humidified atmosphere and 5% CO_2.

Generation of CXCR4-expressing cell line via lentiviral transduction for yeast display in mammalian cells

The full-length CXCR4 gene was cloned into the pCDH lentiviral expression plasmid (Addgene). Viruses were prepared using the pPACKH1 HIV Lentivector Packaging Kit (System Bioscience)[32]. Briefly, 3×10^6 HEK 293T cells were plated on 10 cm dishes and cultured in Iscove's Modified Dulbecco's Media (IMDM, Thermo Fisher) supplemented with 10% FBS overnight. The next day, 2 μg of pCDH plasmids encoding the CXCR4 genes were separately transfected into HEK 293T cells, along with the pPACK packaging plasmid mix. GeneJuice (Sigma) was used as the transfection reagent. GPCR lentivirus was collected from the media after two days and filtered through 0.45 μm filters. Approximately 1×10^5 HEK 293T cells cultured in a 24-well plate were transduced with GPCR lentivirus in the presence of 8 μg/mL polybrene (Sigma) in 500 L complete DMEM culture media. Immediately after transduction, HEK 293T cells were centrifuged at 800×g for 30 min at 32°C. Cells were then incubated overnight at 37°C in a humidified 5% CO_2 incubator. The culture media was replaced with fresh complete DMEM culture media on the day after transduction, and transduced cells were harvested 10 days post-transduction for assessment of GPCR expression via flow cytometry.

DNA library preparation for yeast display

The DNA library was prepared as previously described[8]. All protein sequences were padded to a uniform length by adding a (GGGS)n linker at the C terminal of the designs, to avoid the biased amplification of short DNA fragments during PCR reactions. The protein sequences were reversed translated and optimized using DNAworks2.0 with the *S. cerevisiae codon* frequency table. Homologous to the pETCON plasmid, oligo libraries encoding the designs were ordered from Twist Bioscience. Combinatorial libraries were ordered as IDT (Integrated DNA Technologies) ultramers with the final DNA diversity ranging from 1×10^6 to 1×10^7. All libraries were amplified using Kapa HiFi Polymerase (Kapa Biosystems) with a qPCR machine (BioRAD CFX96). In detail, the libraries were firstly amplified in a 25 µL reaction, and PCR reaction was terminated when the reaction reached half the maximum yield to avoid over-amplification. The PCR product was loaded to a DNA agarose gel. The band with the expected size was cut out and DNA fragments were extracted using QIAquick kits (Qiagen, Inc.). Then, the DNA product was re-amplified as before to generate enough DNA for yeast transformation. The final PCR product was cleaned up with a QIAquick Clean up kit (Qiagen, Inc.). For the yeast transformation, 2-3 µg of digested modified pETcon vector (pETcon3) and 6 µg of insert were transformed into EBY100 yeast strain using the protocol as described before. DNA libraries for deep sequencing were prepared using the same PCR protocol, except the first step started from yeast plasmid prepared from 5×10^7 to 1×10^8 cells by Zymoprep (Zymo Research). Illumina adapters and 6-bp pool-specific barcodes were added in the second qPCR step. Gel extraction was used to get the final DNA product for sequencing. All libraries include the native library and different sorting pools were sequenced using Illumina NextSeq/MiSeq sequencing.

Yeast display

General yeast display methodologies were carried out with EBY-100 yeast cells, as previously described[33,34]. Yeast clones for biofloating assay were grown in SD-CAA medium at 30ºC while shaking at 200 rpm. Yeast cultures were induced in SG-CAA medium at 20ºC while shaking at 200 rpm at an initial optical density (OD) of 1.0 (1×10^7 cells/mL). For soluble receptor-based approach, yeast EBY-100 strain cultures were grown in C-Trp-Ura media and induced in SG-CAA. Cells were washed with PBSF (PBS with 1% BSA) and incubated with the Flag-tagged CXCR4 target (DIMA Biotech, SKU:FLP100074) or biotinylated GLP1R (SinoBiological 13944-H49H-B, GIPR (SinoBiological, 18774-H49H-B) and GCGR (Acro Biosystems, GCR-H82E3) , respectively. For the first round of sorting, cells were incubated with the Flag-tagged CXCR4 nanodisc target or biotinylated ECDs of GLP1R, GIPR and GCGR and labelled with corresponding antibodies simultaneously for 20 minutes whereas for the sorting rounds thereafter, cells were first pre-incubated with the target for 20 minutes and then labelled with corresponding antibodies for additional 20 minutes. Anti-c-Myc fluorescein isothiocyanate (FITC, Miltenyi Biotech) antibody was used for labeling cells and either anti-Flag-phycoerythrin (PE anti-DYKDDDDK, BioLegend) for recognizing Flag-tagged CXCR4 nanodisc target or anti-streptavidin phycoerythrin (SAPE, Thermo Fisher). The concentration of FITC was used at 1/4 concentration of the Flag-tagged or biotinylated target. For the first round of sorting 1 µM concentration of the receptor target was used. The remaining subsequent sorts were performed with varying concentrations (10 pm - 1 µM) of the target. The final sorting pools of the library

were sequenced using Illumina NextSeq/MiSeq sequencing. All FACS data were analyzed in FlowJo.

Biofloating-based library binding assays

Mammalian cells were grown to 60-90% confluency, detached with trypsin-EDTA, and quenched via addition of culture medium. Dissociated cells were washed and centrifuged at 400×g for 5 min twice with PBS and stained with CellTraceTM Violet dye (Thermo Fisher Scientific) via 30 min incubation at 4°C with 2.5 μM dye in PBS at 1×10^6 cells/mL. Following incubation, the mammalian cells were washed three times with PBSA and then resuspended to a concentration of 2.5×10^6 cells/mL in PBSA (PBS with 0.1% BSA) containing Alexa647-conjugated anti-cmyc antibody (Cell Signaling Technology, clone 9B11) (1:100 dilution). Induced yeast cells were washed and centrifuged at 3,500×g for 3 min and aliquoted into a 96-well plate at 5×10^5 yeast cells/well. The plate was centrifuged at 3,500×g for 3 min and resuspended in 20 μL/well of the mammalian cell stock solution to achieve a final ratio of 10:1 yeast:mammalian cells. Incubation proceeded at 4°C for 1 hr with rotation. The cells were then pelleted, washed, and resuspended in PBSA for analysis on a CytoFLEX flow cytometer. No forward/side scatter gating was implemented. The 'yeast cells/complex' metric was computed as described previously[17]. Experiments were performed in triplicate.

Suspension-cell based FACS selections

Target-null and target-expressing mammalian cells were grown to 70-90% confluency, detached with trypsin-EDTA, and quenched via addition of FBS-containing culture medium. Cells were pelleted at 400×g for 5 min and washed three times with PBS. Target-null cells were biotinylated using EZ-Link Sulfo-NHS-SS-Biotin (Thermo Fisher Scientific). The target-null cells were resuspended at 2.5×10^7 cells/mL in PBS pH 8 containing 13 μM of EZ-Link Sulfo-NHS-SS-Biotin reagent and incubated at 4°C for 30 min with rotation. Three washes were then conducted using PBSA (pH 7.3) to quench the reaction and remove excess byproducts. Non-biotinylated target-null cells were also washed twice using PBSA. Target-expressing cells were stained with CellTrace™ Violet dye (Thermo Fisher Scientific). 1×10^7 induced yeast were pelleted at 3,500×g for 3 min, washed twice with PBSA, and resuspended in 300 mL of PBSA containing 1×10^6 biotinylated target-null cells to achieve a yeast:mammalian cell ratio of 10:1. The yeast/mammalian cell mixture was then incubated for 45 min at 4°C with rotation (negative selection). After 45 min, 100 μL of streptavidin-coated magnetic beads per 1×10^6 biotinylated cells were added to the cell mixture and incubation proceeded for 15 min at 4°C with rotation. The cell mixture was then washed once with PBSA and centrifuged at 400×g for 5 min. The pellet was gently resuspended in 5 mL of PBSA and cells were separated over an LS magnetic column (Miltenyi Biotec), according to the manufacturer's protocol. The flow-through solution, depleted of target-null cell-binding yeast, was pelleted at 3,500×g for 5 min. The pellet was then resuspended in 300 μL of PBSA containing 2×10^6 non-biotinylated target-null cells, and incubated for 30 min at 4°C with rotation (pre-block). PBSA (300 L) containing 1×10^6 CellTrace™ Violet dye-labeled target-expressing cells was then added to the yeast/mammalian cell mixture, and incubation proceeded for 45 min at 4°C with rotation (2:1 target-null:target-expressing cell ratio). After 45 min, anti-cmyc Alexa647 antibody was added to the mixture at a dilution of 1:100 and incubated for 15 min. The cell mixture was then washed

once with PBSA and centrifuged at 400×g for 5 min. The pellet was gently resuspended in 1 mL of PBSA and separated via FACS using a SONY SH800 Cell Sorter. The dual positive population was gated, representing yeast labeled with the fluorescent anti-cmyc antibody bound to CellTrace™ dye-labeled target-expressing cells. The sorted cells were collected in 3 mL of SD-CAA and grown for 1-2 days. The yeast were then induced in SG-CAA for analysis or further rounds of sorting.

Individual yeast clone characterization

The enriched yeast mixture from the final round of sorting was plated after FACS selection (~600 yeast cells) on SD-CAA plates and grown for 2 days. Individual clones were inoculated in 1 mL of liquid SD-CAA media for 1-2 days, and subsequently induced for 1-2 days in 1 mL of SG-CAA using a 96-well deep-well plate. On the day of clone characterization, 5×10^5 cells of each yeast clone were transferred to each well of a 96-well V-bottom plate for analysis. Each clone was represented twice on the 96-well plate to enable binding analysis against both target-null and target-expressing mammalian cells, enabling analysis of 48 clones per plate. The yeast cells were washed twice with PBSA and centrifuged at 3,500×g for 3 min. Target-null and target-expressing cells were independently stained with CellTrace™ dye as described above. Each of the mammalian cell line stocks were resuspended at 1.25×10^6 cells/mL in PBSA containing Alexa647-conjugated anti-cmyc antibody (Cell Signaling Technology, clone 9B11) at a dilution of 1:100. Each yeast clone in the 96-well plate was then resuspended separately with 20 µL of target-null cells or 20 µL of target-expressing mammalian cells (yeast:mammalian cell ratio of 20:1). Incubation proceeded for 1 hr at 4°C with rotation. The cell mixtures were then washed once with PBSA and centrifuged at 400×g for 5 min. The cell pellets were gently resuspended in 100 µL of PBSA and analyzed on a CytoFLEX flow cytometry instrument (Beckman Coulter).

Optical screen

Plasmids:

A GFP reporter vector for MRGPRX1 was generated by cloning full-length human MRGPRX1 into a lentiviral entry vector encoding an N-terminal signal peptide, FLAG tag, and GFP, as well as a C-terminal BFP fused to a c-myc tag followed by a 2A peptide and blasticidin selection marker (pLenti/mIGK-FLAG-eGFP-BFP-myc-P2A-blast) using NEBridge Golden Gate Assembly Kit (BsmBI-v2) (New England Biolabs #E1602L). The lentiviral entry vector for binder library cloning (pLenti/puro-T2A-RUSH-C5-mCherry) was prepared by replacing the U6 promoter in lentiguide-BC-plasmid (Addgene #127168) with an EF1a promoter, puromycin resistance marker, T2A peptide, RUSH secretion tag (Addgene #65294), C5 oligomerization domain (PDB 2B98) and mCherry, followed by an entry site for cloning of barcoded binders. Barcoded binders were synthesized as Twist oligo pools containing the designed binder, a C-terminal KDEL endoplasmic reticulum retention tag, stop codon, and a 10-nt barcode suitable for *in situ* sequencing. Thus, the final binder library construct encoded puromycin resistance separated by a 2A peptide from a protein fusion comprising a secretion tag, oligomerization domain, mCherry tag, designed binder, and ER retention tag, followed immediately by a non-coding barcode. Barcoded binder libraries were cloned into the lentiviral entry vector as reported previously[35,36]. Briefly, oligo pools were amplified with KAPA HiFi HotStart Ready Mix (Roche #KK2601), 1X

EvaGreen qPCR dye (Biotium #31000), 500 nM forward and reverse primers (dialout primer FW: TCTGAACAGGCTcgtctct, dialout primer RV: CTATCGCCAAGTcgtctct) (Integrated DNA Technologies), and 80 pg/μL of template in 30 μL reactions. PCRs were conducted with the following thermal cycling protocol: 95 °C for 3 min, 14-16 cycles of (98 °C for 20 s, 65 °C for 20 s, 72 °C for 45 s), then 72 °C for 1 min. Following amplification, reactions were gel purified using Zymoclean Gel DNA Recovery Kit (Zymo #D4007) and quantified with Qubit Broad Range dsDNA Quantitation assay (Thermo Fisher Scientific #Q32853). Plasmid libraries were then constructed using NEBridge Golden Gate Assembly Kit (BsmBI-v2) (New England Biolabs #E1602L), with a 3:1 molar ratio of insert:vector for 0.3 kb inserts. Assembly reactions were incubated at 42 °C for 1 h, and heat inactivated at 60 °C for 5 min. Reactions were purified using DNA Clean and Concentrator-5 kit (Zymo #D4014) and electroporated in Endura™ Competent Cells (Biosearch Technologies #60242-2) using a Gene Pulser Xcell (Biorad 1652662) set to 1.8 kV, 600 ohms, and 10 μF, and recovering for 60 minutes at 37 °C, 250 rpm in 1 mL of Endura recovery media (Biosearch Technologies #60242-2). Cultures were incubated for 6-14 h at 37 °C in 50 mL of LB media with 100 μg/mL of carbenicillin. Assembly and transformation efficiency were assessed, observing around 10^8 colony forming units per μg of transformed DNA. The resulting plasmid library was validated via Illumina MiSeq sequencing (500-cycle Nano v2 kit) with a target coverage of 30–100X.

Generation of reporter cell lines for OPS-RD:

Lentivirus was generated for the MRGPRX1 GFP reporter and binder libraries as described previously. Reporter cell lines overexpressing the target receptor-GFP fusions were established using lentiviral transduction, as described in Feldman et. al.[37].

Isogenic reporter cell lines were generated by single-cell sorting of GFP+ cells into 96-well plates. After outgrowth of clones, replicate plates were imaged and the final clones were selected based on the expression level and subcellular localization of target receptor-GFP fusions.

The binder lentiviral library was prepared as previously described[35], with the exception that lentivirus were first titered, and transduction of reporter cell lines targeted an MOI of 5-10%. Libraries were transduced in three biological replicates.

In situ sequencing:

Screening was conducted as described previously, with the following modifications. Cells were plated at a density of 15×10^4/well in 6-well glass-bottom plates (Cellvis #P06-1.5H-N) 72 h prior to in situ sequencing to promote optimal adhesion and spreading. After rolling circle amplification, but prior to the first sequencing cycle, cells were stained with DAPI and imaged in the DAPI, GFP and mCherry channels to measure localization of the MRGPRX1 reporter and binder. A total of 9 cycles of in situ sequencing were conducted.

Image analysis and ranking of designs:

Localization and in situ sequencing images were analyzed using a Python-based pipeline, as previously described[35]. Segmentation was done with Cellpose v2, using the DAPI stain and non-specific background from in situ sequencing as nuclear and cytoplasmic inputs, respectively. Each cell was assigned a binding score based on pixelwise cross-correlation of MRGPRX1-GFP with binder-mCherry. Cells containing a common barcode were then clustered by spatial proximity. Each binder was then scored based on the average binding score amongst its cell clusters. For each of the top 100 binders, the corresponding clusters were inspected by eye, and 60 candidates were selected based on the strength of the localization phenotype and reproducibility across biological replicates.

***In vitro* GPCR pharmacology**

cAMP assay for CGRPR and CXCR4 was carried out as previously described using commercially available G_s and G_i Cisbio kits[38]. To measure antagonism of CGRPR binders, a concentration-response curve of the endogenous CGRP was first generated using a SK-N-MC cell line (ATCC, HTB-10). The G_s-mediated cAMP accumulation was measured in a final volume of 40 uL. The stimulation buffer containing 0.5 mM IBMX (Sigma-Aldrich) was used for serial dilutions of tested ligands. Approximately 10 uL of 2,500 cells per well was used to seed cells into a white 384-well plate. The reaction mixture was incubated at 37 °C for 30 min and the reaction was terminated by adding 10 µL of cryptate-labeled cAMP and cAMP d2-labeled antibody, respectively. Following an incubation for 1 hour at room temperature, cellular cAMP levels were quantified by homogeneous time-resolved fluorescence resonance energy transfer (HTRF, ratio 665/620 nm) on a Neo2 plate reader (Agilent). Screening of CGRPR antagonist binders was conducted by analogy except for pre-incubating binders for 30 min at 37°C followed by CGRP incubation for an additional 30 minutes under the same conditions. Screening for antagonism of CXCR4 binders was performed by measuring a G_i-mediated cAMP inhibition using a commercially available CHO-CXCR4 stable cell line (GenScript, M00556). 3,500 cells per well in 10 uL were mixed with 5 uL of 4X CXCL12 and forskolin (5 µM final concentration), respectively. The reaction mixture was incubated for 30 minutes at 37°C and then 10 uL of cryptate-labeled cAMP and cAMP d2-labeled antibody, respectively, were added. The antagonistic profile of CXCR4 binders was measured in a manner similar to CGRPR antagonists. Binders were first pre-incubated (4X, 5 µL) with 5 µL of 3,500 cells for 30 minutes at 37°C followed by addition of an EC_{80} of CXCL12 (4X, 5 µL) and forskolin (4X, 5 µL).

Receptor-mediated cAMP production was also determined using COS-7 cells transiently expressing each target receptor. COS-7 cells were transfected using polyethylenimine (PEI Max, mol. wt. 40,000; Polysciences, Warrington, PA) and pcDNA3 DNA plasmids containing CLR and RAMP1 (CGRPR), CLR and RAMP2 (AM_1R), CLR and RAMP3 (AM_2R), CTR and RAMP1 (AMY_1R), or CTR alone. Receptor and RAMP DNA constructs were transfected at a 1:1 ratio using 10 ng/well per plasmid, for a total of 20 ng of DNA per well; pcDNA3 plasmid was used as a control to equalize the total amount of transfected DNA in the case of CTR alone. DNA and PEI Max were each prepared in 150 mM NaCl, then combined to yield a 1:6 DNA:PEI Max ratio and incubated for 15 minutes at room temperature. The DNA/PEI mixtures were added COS-7 cells in suspension, then 13,000 cells per well were seeded into 96-well clear

plates (Corning) and incubated at 37°C in 5% CO_2 for 48 h, before performing the cAMP assay. On the day of assay, the culture media was replaced with stimulation buffer (phenol red–free DMEM containing 25mM HEPES, 0.1% w/v bovine serum albumin (Sigma-Aldrich) and 0.5 mM 3-isobutyl-1-methylxanine, pH 7.4) and incubated for 30 minutes at 37°C in 5% CO_2. Cells were then stimulated for 30 min with varying combinations of agonist peptides in the presence and absence of varying concentrations of dC2_049 and dC2_050 binders. The reaction was terminated by aspiration of the stimulation buffer and addition of ice-cold ethanol. After evaporation of ethanol, the cells were lysed with 60 µL/well lysis buffer (5 mM HEPES, 0.1% w/v bovine serum albumin, 0.3% Tween 20, pH 7.4). The concentration of cAMP in the lysates was detected with the cAMP Gs HiRange homogeneous time-resolved Forster Resonance Energy Transfer (HTRF) kit (CisBio). The plates were read on a PHERAstar plate reader (BMG LABTECH).

For the luciferase assay, CHO-K1/Cre-Luc/CGRPR cells (M00187, GenScript) were seeded at a density of 10,000 cells per well in 20 µL of growth medium in a white 384-well plate (Cat. No.: 3570, Corning). The cells were incubated overnight (approximately 16 hours) at 37°C with 5% CO_2. A concentration-response curve of the agonist αCGRP was generated to determine its EC_{50}. The antagonistic activity of the CGRPR binders was assessed in the presence of EC_{80} of αCGRP. After the overnight incubation, 4x working solutions of the ligands were prepared by serially diluting the antagonist or agonist in growth medium. Subsequently, 10 µL of the 4x antagonist or growth medium was added to each well. After a 30-minute incubation at 37°C, 10 µL of the 4x agonist working solution was added to each well, and the cells were further incubated for 6 hours at 37°C with 5% CO_2. Following treatment, 40 µL of Bio-Glo™ Luciferase Assay Detection Solution (Cat. No.: G7941, Promega) was added to each well to initiate the luminescent reaction. Luminescence was then measured using a SpectraMax iD5 Multimode Plate Reader (Molecular Devices).

For calcium mobilization assay, RBL-2H3 cells (Eurofins) were seeded in a total volume of 20 µL/well, in black, clear-bottom, Poly-D-lysine coated 384-well microplates and incubated at 37°C. Subsequently, media was replaced with 20 µL of Dye Loading Buffer, consisting of 1X Dye, 1X Additive A, 2.5 mM Probenecid (freshly prepared) in HBSS / 20 mM HEPES, and incubated for 30-60 minutes at 37°C. For agonism, cells were incubated with 10 µL of HBSS / 20 mM HEPES. Vehicle (prepared at 3X concentration) was included in the buffer when generating agonist concentration response curves to obtain the EC_{80} for subsequent antagonist screening. Cells were incubated in the dark for 30 minutes at room temperature. The agonist activity of ligands was measured on a FLIPR Tetra (MDS). 10 µL of the sample (prepared at 4X concentration in HBSS / 20 mM HEPES) was added to the cells 5 seconds before calcium mobilization was monitored for 2 minutes. For antagonist measurements, after dye loading, 10 µL of the sample (prepared 3X) was added and cells were incubated for 30 minutes at room temperature. 10 µL of an EC_{80} of the agonist, prepared in HBSS / 20 mM HEPES, was added to the cells 5 seconds before calcium mobilization was monitored for 2 minutes.

Surface plasmon resonance spectroscopy (SPR)

Kinetic measurements for GLP1R binders

Binding studies were executed on a Biacore™ T200 or a Biacore™ 8K (Cytiva) instrument. The experiments were conducted at 25°C. Anti-human IgG monoclonal antibody (Human antibody capture kit, Cytiva) was immobilized onto both flow cells of a sensor chip (Series S, CM 5) using the Amine Coupling Kit (Cytiva) following the manufacturer's guidelines. Subsequently, GLP1R-Fc (R&D systems) was captured by injecting it over flow cell 2. Subsequently, the binding of de novo binders was probed by injecting them as analytes in increasing concentrations at a rate of 50 µl min-1 for 120 seconds and allowing them to dissociate for 300 seconds. After each analyte injection cycle, the anti-human IgG surface was regenerated via 3 M $MgCl_2$ pH 2.3 (Cytiva) injections. Binding curves underwent processing, which involved subtraction of reference surface signals as well as blank buffer injections. The binding rate constants were extracted by globally fitting a 1:1 Langmuir model to the data using Biacore T200 Evaluation Software (version 3.2) or Biacore Insight Evaluation Software (version 5.0.18.22102). Data was plotted using GraphPad Prism (version 10.4.1). Three independent experiments were conducted in duplicates.

Competition SPR for GLP1R binders

Competition experiments were executed analogously to the kinetic measurements described above using the Dual injection command. Briefly, GLP1R-Fc (R&D systems) was captured by on the active flow cell of an anti-human IgG sensor surface. Subsequently, the bins of the binders were assessed by injecting them immediately after another using the dual injection command. Analytes were injected with a concentration of 1 µM. Sensogram were aligned and extracted using the Biacore Insight Evaluation Software (version 5.0.18.22102) and plotted using GraphPad Prism (version 10.4.1).

Kinetic measurements for GIPR and GCGR binders

Binding studies were executed on a Biacore™ 8K (Cytiva) instrument. The experiments were conducted at 25°C. Biotinylated GIPR (SinoBiological, 18774-H49H-B) and GCGR (Acro Biosystems, GCR-H82E3) ectodomain proteins were captured by Streptadvidin using Biotin CAPture Kit (Cytiva #28920234) following the manufacturer's guidelines. The GIPR or GCGR ectodomain samples at concentration of 0.125 µg/mL were injected at a flow rate of 10 µL/min in HBS-EP+ (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, Cytiva #BR100669) aiming for a capture level of ~150 response units. The kinetic measurements of the best 96 designs from yeast library screening were performed by injecting them as analytes in increasing concentrations ranging from 0.0128nM, 0.064nM, 0.32nM, 1.6nM, 8nM, 40nM, 200nM, 1000nM to 5000nM in a single cycle with 9 steps. Analytes were diluted in HBS-EP+ and injected at a flow rate of 30 µL/min to monitor association. HBS-EP+ was used as a running buffer during dissociation at a flow of 30 µL/min. Binding kinetics were determined by global fitting of curves assuming a 1:1 Langmuir interaction using the Cytiva evaluation software.

Purification of the CGRPR-binder complex

The CLR and RAMP1 constructs used for this study were validated and used for structural determination previously[39]. The CLR construct contained an N-terminal FLAG tag and a C-terminal 8x histidine tag, flanked by 3C protease cleavage sites. RAMP1 contained an N-terminal FLAG tag epitope. To increase the recombinant expression, the native signal peptides of CLR and RAMP1 were replaced with a hemagglutinin signaling peptide. The heterodimeric CGRPR was formed by the co-expression of CLR and RAMP1 in *Trichoplusia ni* insect cells (Expression systems) using baculovirus as reported previously[40].

The purification of CGRPR was conducted as described previously[40]. In brief, after removal of tags from CLR by addition of 3C protease (10 ug/mL, home-made), the CGRPR was solubilized using detergent (1% w/v LMNG and 0.06% w/v CHS) for 1 h at 4 °C and purified by binding to M1 anti-FLAG affinity resin. The crude eluate containing apo CGRPR was semi-quantified using nanodrop and 5-fold molar excess of dC2_049 was added and incubated on ice for 2 h to enable formation of the ternary complex. The mixture of CGRPR and dC2_049 was subjected to SEC on a Superdex 200 Increase 10/300 column (GE Healthcare) that was pre-equilibrated with the SEC buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 2mM $MgCl_2$). The eluted complex was concentrated to 11 mg/mL.

For the dC2_050-CGRPR complex, the purification was conducted using a similar protocol but with addition of dC2_050 during solubilization and throughout the purification. The formation of dC2_050-CGRPR complex was initiated by the addition of 100 nM dC2_050-CGRPR during solubilization. The solubilized CGRPR complex was immobilized by batch binding to M1 anti-FLAG affinity resin. The resin was sequentially washed in the presence of 25 nM dC2_050-CGRPR and eluted using a calcium-free buffer supplemented with 500 nM dC2_050. The eluted complexes were profiled by SEC in the SEC buffer with 50 nM dC2_050, and concentrated to 6 mg/mL.

Vitrified specimens and cryo-EM data collection

Gold-coated[41] Quantifoil r1.2/1.3 grids were glow discharged using a GloQube Plus (air chamber, 15 mA, 140 s, negative polarity). Thawed sample (3 µL, 5.5 mg/mL of C8-CGRPR and 6 mg/mL of C10-CGRPR) was applied to the grid, the grid was blotted (blot force 17, blot time 7 s, 100% humidity, 4 °C, Vitrobot Mk IV), and the sample was vitrified in liquid ethane.

Images of dC2_049-CGRPR complexes (9815 compressed TIFF movies, 50 fractions/movie) were collected on a ThermoFisher Scientific Titan Krios G4 microscope fitted with a cold-FEG, Selectris-X energy filter, and Falcon 4i direct electron detector. The microscope was operated at 300 kV and 165 kx indicated magnification, with a pixel size of 0.75 Å. The energy filter was operated with a slit width of 10 eV. Images were recorded using aberration free image shift with an exposure time of 9.42 s, a dose rate of 2.99 e^-/px/s, and a total dose of 50 e^-/Å.

Data from dC2_050-CGRPR was collected on a Thermo Fisher Scientific Glacios microscope, operated at an accelerating voltage of 200 kV with a C2 aperture in nanoprobe EFTEM mode, spot size 5, fitted with a Falcon 4 direct electron detector. Movies were recorded as compressed TIFFs in normal-resolution mode yielding a physical pixel size of 0.86 Å/pixel with an exposure time of 4.89 s amounting to a total exposure of 50 $^{e-}/A^2$. Defocus was varied in the range

between -0.7 and -1.1 µm. Beam-image shift was used to acquire data from 21 surrounding holes after which the stage was moved to the next collection area using EPU software package.

Data processing

For the 300 kV dataset of dC2_049-CGRPR, fractionated TIFF files were pre-processed into 69 optics groups using the EPU_group_AFIS.py script (https://github.com/DustinMorado/EPU_group_AFIS) for import into RELION 5.0[42]. Patch (4 x 4 patches) motion correction[43] was performed with MotionCor3 (https://github.com/czimaginginstitute/MotionCor3). Contrast transfer function (CTF) estimation was performed with CTFFIND (version 4.1.14)[44]. Micrographs with estimated maximum resolution values of >5 Å were discarded, leaving 8396 micrographs. Particle picking was performed using a Laplacian-of-Gaussian algorithm, as implemented in RELION-5, with 90 Å and 190 Å minimal and maximal diameters, respectively. From this, 3,942,967 initial particles were extracted with a box size of 256^2 px (binned to 64 px). Reference-free 2D-classification was performed in cryoSPARC 4.6.0, and 731,523 particles were selected and re-extracted without binning. An additional round of 2D-classification and *ab initio* reconstruction was performed in cryoSPARC (version 4.6.0). 428,662 selected particles were refined using RELION 5.0 and subjected to particle polishing[45,46]. Using cryoSPARC, 285,458 particles were selected, following a final 2D-classification step, and non-uniform refinement was used to generate a map with a 3.18 Å global resolution (0.143 Fourier shell correlation cutoff). To further improve map quality, the RELION 5.0 polished particle stack (428k particles) was subjected to 3D classification within RELION 5.0 and the best class (194k particles) was further refined in RELION 5.0 with the BLUSH regularization. This particle stack was analyzed in cryoDRGN (version 3.3.3)[45,46]. Initially after Variability Autoencoder (VAE) training, particles with high magnitude latent space vectors were excluded and subjected to a further round of higher-resolution VAE training. This was then analyzed using the analyze_landscape functionality within cryoDRGN, and particles (175k) belonging to the most populated 3D volume were re-exported back into RELION 5.0 for further rounds of 3D refinement and CTF refinement, yielding a more interpretable map for the extracellular domain and dC2_049 binding position.

For the 200 kV dataset of dC2_050-CGRPR, dose-fractionated TIFF movies were preprocessed into their corresponding beam-image shift optics groups using the EPU_group_AFIS.py script (https://github.com/DustinMorado/EPU_group_AFIS), imported into RELION 5.0[42] and motion corrected using MotionCor3[43] (4 x 4 patch tracking) and had their CTF parameters estimated using CTFFIND 4.1.14[43]. Micrographs with robust CTF information beyond 5 Å were selected for further processing. Particles were picked using crYOLO (1.9.9[47]), yielding 4.8M particle positions. This stack of particles was extracted and Fourier scaled to 64 pix^2 and subjected to rounds of 2D classification, multiple class *ab initio* and heterogeneous refinement in cryoSPARC (4.6.0)[48], resulting in a homogenized particle stack of 1.3M particles. This set of particles was then re-centered and re-extracted at their native pixel sampling and underwent 3D classification and 3D refinement in RELION 5.0, resulting in 490k particles undergoing Bayesian Particle Polishing. These higher signal-to-noise particles were then further refined in cryoSPARC (4.6.0), by a 2D classification, non-uniform refinement followed by a local refinement with a 3D-mask excluding any density from the detergent micelle. This yielded a 4.06 Å (0.143 FSC) map that was used for model building.

PDB models were refined into both maps using a combination of Molecular Dynamics Flexible Fitting (MDFF) as implemented in iSOLDE[49] followed by rounds of manual refinement in Coot[50] and real-space refinement in PHENIX[51].

Generation of MRGPRX1 constructs for cryo-EM

For the expression of MRGPRX1–Gαq protein complex, the full-length DNA of human MRGPRX1 (UniProtKB: Q96LB2) was subcloned into a modified version of pFastBac1 (Invitrogen) baculovirus expression vector. Specially, the N-terminal of MRGPRX1 sequence was incorporated with a string of hemagglutinin (HA) signal peptide, followed by a Flag-tag, a 10× His-tag and a TEV protease site. Then, a thermostabilized apocytochrome b562RIL (BRIL) and HRV3C protease sites were fused to the N-terminus of MRGPRX1 to facilitate the protein expression and purification. For the $G\alpha_q$ protein, the same mini-GαqiN heterotrimer construct used for the expression of HT2A–G_q–NBOH complex was introduced to facilitate the formation of the receptor complex.

Expression of MRGPRX1–$G\alpha_q$ protein complex

Recombinant baculovirus containing the MRGPRX1 and mini-$G\alpha_q$iN heterotrimer were generated using the Bac-to-Bac Baculovirus Expression System (Invitrogen). In brief, the constructs were transformed into DH10Bac competent cells (Invitrogen), recombinant bacmid was purified according to manufacturer's protocol. For the generation of virus, Spodoptera frugiperda (Sf9) insect cells (Expression Systems) were plated into a 12-well plate at a concentration of 5×10^5 cells per well and transfected with 5 µg of purified bacmid using cellfectin reagent to obtain recombinant baculovirus. After 96 hours of incubation at 27^0C, the supernatant was collected as the P0 viral stock and used to generate high-titer baculovirus P1 stock by infection with 40 ml of 2×10^6 Sf9 cells per milliliter and incubation for 96 hours. Viral titers were determined by flow cytometric analysis of Sf9 cells stained with 1:200 diluted gp64-PE monoclonal antibody (Thermo Fisher Scientific). For the expression of the MRGPRX1–Gαq complex, Sf9 cells were grown to a density of 2.0×10^6 cells per milliliter and then co-infected with the baculoviruses of MRGPRX1 and mini-$G\alpha_q$iN heterotrimer at a multiplicity of infection (MOI) ratio of 3.5:2. After 48 hours of infection, the cells were harvested by centrifugation, washed in HN buffer (10 mM HEPES and 100 mM NaCl, pH 7.5) and stored at −80 °C for future use.

Purification of MRGPRX1–$G\alpha_q$ protein complex

For MRGPRX1–$G\alpha_q$ protein complex purification, Sf9 cell pellets were thawed on ice and resuspended in buffer containing 20 mM HEPES, pH 7.5, 50 mM NaCl, 10mM $MgCl_2$, 5mM $CaCl_2$ and 3 units of Apyrase (NEB) supplemented with complete Protease Inhibitor Cocktail tablets (Roche). After stirring for 1.5 hours at room temperature, the cell suspension was dounced to homogeneity and subsequently ultracentrifuged at 100,00 x g (Ti45 rotor, Beckman) for 30 minutes to collect the membrane. Membrane material was solubilized in buffer containing 50 mM HEPES, pH 7.5, 100 mM NaCl, 5% (w/v) glycerol, 0.5% (w/v) lauryl maltose neopentyl glycol (LMNG), 0.05% (w/v) cholesteryl hemisuccinate (CHS), and 500 µg of scFv16 for 6 hours at 4 °C. Solubilized proteins were isolated by ultracentrifugation at 100,000 x g (Ti70 rotor, Beckman) for 45 minutes and then incubated with Talon IMAC resin (Clontech) and 20 mM

imidazole overnight at 4 °C. The following day, the Talon resin with immobilized protein complex was collected with a gravity flow column and washed with 25 column volumes of buffer containing 20 mM HEPES, pH 7.5, 100 mM NaCl, 20 mM imidazole, 0.01% (w/v) LMNG, 0.001% (w/v) CHS and 5% glycerol. The protein complex was eluted with the same buffer supplemented with 250 mM imidazole. Released proteins were further concentrated to 0.5 ml and subjected to size-exclusion chromatography on a Superdex 200 10/300 GL Increase column (GE Healthcare) that was pre-equilibrated with 20 mM HEPES, pH 7.5, 100 mM NaCl, 100 μM TCEP, 0.00075% (w/v) LMNG, 0.00025 (w/v) glyco-diosgenin (GDN) and 0.00075% (w/v) CHS. Peak fractions were pooled and incubated with 15 μl of His-tagged PreScission protease (GenScript) and 2 μl of PNGase F (NEB) at 4 °C overnight to remove the N-terminal BRIL and potential glycosylation. The proteins were concentrated and further purified by size-exclusion chromatography using the same buffer. Peak fractions were pooled and concentrated to 5 mg ml^{-1}. To ensure a full binding of MRGPRX1 ligands, 50 μM of adducts F8 and E12 were added to the concentrated sample and incubated overnight at 4 °C before grid-making.

Expression and purification of scFv16

Expression and purification of scFv16 was performed as previously described. In brief, the scFv16 gene was cloned into a modified pFastBac1 vector, expressed from insect Sf9 cells using the baculovirus method and purified by size-exclusion chromatography. Supernatant containing secreted scFv16 was pH balanced to pH 7.8 by the addition of Tris base powder. Media chelating agents were quenched by the addition of 1 mM nickel chloride and 5 mM calcium chloride and stirred for 1 hour at room temperature. The supernatant was collected by centrifugation and incubated with 1 ml of His60 Ni Superflow Resin (Takara) overnight at 4 °C. The following day, the His60 Ni Superflow Resin was collected by a gravity flow column and washed with 20 column volumes of buffer containing 20 mM HEPES, pH 7.5, 500 mM NaCl and 10 mM imidazole, scFv16 was eluted with the same buffer supplemented with 250 mM imidazole. scFv16 protein was further purified by size-exclusion chromatography using a Superdex 200 10/300 GL (GE Healthcare), peak fractions were collected and concentrated to 2 mg ml^{-1} for future use.

Cryo-EM grid preparation, data collection and three-dimensional reconstitution

For the preparation of cryo-EM grid, 3.2 μl of each MRGPRX1 complex was applied individually onto glow-discharged Quantifoil R1.2/1.3 Au300 holey carbon grids (Ted Pella) in a Vitrobot chamber (FEI Vitrobot Mark IV). The Vitrobot chamber was set to 4 °C and 100% humidity with a blot time range from 3 seconds to 6 seconds. The grids were flash frozen in a liquid ethane/propane (40/60) mixture and stored in liquid nitrogen for further screening and data collection. Cryo-EM imaging was performed on a 200 keV G3 Talos Arctica. Micrographs were recorded using a Gatan K3 direct electron detector at a physical pixel size of 0.876 Å. Movies were automatically collected using SerialEM using a multi-shot array as previously described. Data were collected at an exposure dose rate of ~15 electrons per pixel per second as recorded from counting mode. Images were recorded for ~2.7 seconds in 60 subframes to give a total exposure dose of ~50 electrons per $Å^2$. All subsequent classification and refinement steps were performed with cryoSPARC using previously described workflow. In brief, merged curated non-duplicate particles from multiple picking regimes were subjected to multi-reference refinement. This generated a final stack of particles that created a map with respective

resolutions as reported in Supplementary Table 5 (by Fourier shell correlation (FSC) using the 0.143-Å cutoff criterion) after local contrast transfer function (CTF) refinement and post-processing in cryoSPARC. Alternative post-sharpening was performed using deepEMhancer and EMready.

Model building and refinement

For the models of the MRGPRX1:G_q:adduct complexes, we used the structures of the MRGPRX1, G_q trimer, and scFv16 adopted from the MRGPRX1:G_q complex (Protein Data Bank (PDB): 8DWC) and the predicted adduct structure. Each complex subunit was docked into the cryo-EM maps using Chimera and Phenix. The models were manually adjusted in Coot and then subjected to several rounds of real-space refinement refinement in Phenix. The model statistics were validated using Molprobity. Refinement statistics are provided in Supplementary Table X. Structure figures were prepared by either ChimeraX or PyMOL (https://pymol.org/).

Pharmacological data analysis

In vitro pharmacological analysis was carried out with GraphPad Prism (GraphPad Software, San Diego). Data are presented as means ± SEM over technical sample averaged biological replicates. For the luciferase assay, Relative Light Units (RLU) values were obtained by subtracting the luminescence values of the background (media alone + Luciferase Assay Detection Solution) to the ones of each sample. RLU values from the luciferase assay, relative fluorescence units (RFU) from calcium mobilization assay and data from cAMP were fitted to three-parameter nonlinear regression curves, a slope of one and logarithmic scale. Responses were then normalized using the following equation: (signal of test sample - signal of vehicle control) / (positive control ligand - signal of vehicle control), with positive control ligands being CGRP for CGRPR (100 nM in a cAMP assay with COS-7 cells, 1 μM in a cAMP assay with SK-N-MC cells when assaying RFdiffusion designs, or 1 μM in a CRE-Luc assay with CHO-K1/CGRPR cells when assaying MetaGen designs), CXCL12 for CXCR4 (1 μM), and BAM 8-22 for MRGPRX1 (0.1 μM).

For CGRPR, data were normalized to 100%, i.e. the saturating concentration of CGRP in the assay (either 100 nM or 1 μM), and fitted to three-parameter nonlinear regression curves using Global Gaddum-Schild regression analysis. For CXCR4, data were normalized to 100%, i.e. the saturating concentration of CXCL12 (1 μM), and fitted to three-parameter nonlinear regression curves using Global Gaddum-Schild regression analysis. Data from calcium flux were fitted to four-parameter nonlinear regression curves. To measure antagonism, percentage inhibition was calculated by normalizing the RFU of the test sample relative to the response achieved with the EC_{80} of BAM 8-22 control.

Acknowledgements

We thank Luki Goldschmidt and Kandise VanWormer, respectively, for maintaining the computational and wet laboratory resources at the Institute for Protein Design. E.M. is Erwin Schrödinger Postdoctoral Fellow (J-4663). This research was supported by Defense Threat Reduction Agency Grant HDTRA1-21-1-0038 (GR018007, D.B.), Gift from Microsoft (GF117374: Microsoft Protein Prediction Research, D.B.), Howard Hughes Medical Institute (GR020267, G.R.L., D.B.), Novo Nordisk (GR018355, E.M., D.B.), The Audacious Project at the Institute for Protein Design, (PG117878, PG117879, PG117866, D.B.), The Nordstrom Barrier Institute for Protein Design Directors Fund (GF124659, D.B.), The Open Philanthropy Project Improving Protein Design Fund (GF129460, G.R.L., S.V.T., D.B.), The Open Philanthropy Project Universal Flu Vaccine Fund (GF129461, D.B.), The Wu Tsai Protein Innovation Fund (GF151772, T.S., D.B.), Cancer Research Grand Challenge grant provided by Cancer Research UK (GR050755, A.M.). The project or effort depicted was or is sponsored by the Department of the Defense, Defense Threat Reduction Agency grant HDTRA1-21-1-0007 (GR013444, D.B.). This research was also supported by the National Institutes of Health's National Cancer Institute, grant R01CA240339 (GR009231, D.B.) and grant K99-CA293001 (J.Z.Z.), and National Institutes of Health's National Institute on Aging, grant R01AG063845. (GR009173, D.B.). This study used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award BER-ERCAP0022018. C.N, D.F., A.B., P.R.S. and F.D. were supported by grants from the BioInnovation Institute Foundation (BII22SG1021010, BII24SG1021475, BII24SG1022030). We acknowledge Angeli Tongson, Jason Walters, Gordon Leung, Quishi Wang, Paolo Gonzales for supporting pharmacological characterization of MRGPRX1 designs. Esperanza Rivera de Torre for assisting in circular dichroism studies. B.E.K. and B.L.R. were supported by the National Institute of Health NIDA award (R01DA055656). P.M.S. and D.W. were supported by National Health and Medical Research Council of Australia Investigator grants (2025694 and 2026300, respectively). J.B.S. was supported by NSF CAREER Award (2143160), Department of Defense Award (W81XWH-21-1-0891), NIH NIDCR Award R21DE031436 and CureSearch for Children's Cancer Award.

Competing interests

E.M., D.F., D.E.K., X.Q., A.B., P.R.S., F.D., J.F., L.J.S., J.B.S., C.N., and D.B are listed as inventors or major contributors on records of innovation at the University of Washington and associated provisional patent applications that relate discoveries described in this manuscript. The Baker lab has received sponsored research funding from Novo Nordisk in support of the GLP1R research described in this manuscript. E.M., D.F., D.E.K., A.B., P.R.S., L.J.S., C.N., and D.B. are shareholders of Skape Bio Aps. All other authors declare no competing interests.

Contributions

E.M., D.F., K.D., C.J.T., C.N. and D.B. initiated the project. D.E.K., T.S., M.B., I.S., X.W., developed computational design pipelines using RFdiffusion. C.N. and D.F. developed computational design pipelines using MetaGen and partial RFdiffusion. E.M.., D.F., D.E,K., X.Q., A.P. and C.N. designed binders. D.F. and C.N. conceived the OPS-RD HTS assay. D.F, A.B., and P.R.S. developed cell lines and performed the OPS-RD assay. E.M., X.Q., A.B., D.E.K., T.S., L.M., S.V.T., Y.R. and C.N. expressed and purified binders. E.M., X.Q., F.D., P.K., L.O.,

W.C., N.L., M.B., Y.W. and L.A. pharmacologically characterized binders. J.F. performed biofloating assay. E.M., X.Q., J.Z.Z., A.M., L.T., G.R.L., I.G. and D.K.V. performed yeast display experiments. J.C., B.P.C., and M.J.B. determined cryo-EM structure of CGRPR binders and generated associated figures. Q.C. determined cryo-EM structure of CXCR4 binders. B.L.R. and B.E.K. determined cryo-EM structure of MRGPRX1 binders. J.E.S. and P.H. provided reagents. E.M. and C.N. wrote the manuscript and prepared figures and all authors edited the manuscript. K.D., J.B.S., B.E.K., B.L.R., P.M.S., D.W., C.G.T., C.N., and D.B. provided research support and supervision.

References

1. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
2. Congreve, M., De Graaf, C., Swain, N. A. & Tate, C. G. Impact of GPCR Structures on Drug Discovery. *Cell* **181**, 81–91 (2020).
3. Zhang, M. *et al.* G protein-coupled receptors (GPCRs): advances in structures, mechanisms, and drug discovery. *Signal Transduct. Target. Ther.* **9**, 88 (2024).
4. Ren, H. *et al.* Function-based high-throughput screening for antibody antagonists and agonists against G protein-coupled receptors. *Commun. Biol.* **3**, 146 (2020).
5. Ma, Y. *et al.* Structure-guided discovery of a single-domain antibody agonist against human apelin receptor. *Sci. Adv.* **6**, eaax7379 (2020).
6. Fontaine, T. *et al.* Structure elucidation of a human melanocortin-4 receptor specific orthosteric nanobody agonist. *Nat. Commun.* **15**, 7029 (2024).
7. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
8. Cao, L. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
9. Roy, A. *et al.* De novo design of highly selective miniprotein inhibitors of integrins αvβ6 and αvβ8. *Nat. Commun.* **14**, 5660 (2023).
10. Berger, S. *et al.* Preclinical proof of principle for orally delivered Th17 antagonist miniproteins. *Cell* **187**, 4305-4317.e18 (2024).
11. Case, J. B. *et al.* Ultrapotent miniproteins targeting the SARS-CoV-2 receptor-binding domain protect against infection and disease. *Cell Host Microbe* **29**, 1151-1161.e5 (2021).
12. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
13. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
14. Bennett, N. R. *et al.* Improving de novo protein binder design with deep learning. *Nat. Commun.* **14**, 2625 (2023).
15. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
16. Ju, M.-S. *et al.* A human antibody against human endothelin receptor type A that exhibits antitumor potency. *Exp. Mol. Med.* **53**, 1437–1448 (2021).
17. Krohl, P. J. *et al.* Discovery of antibodies targeting multipass transmembrane proteins using a suspension cell-based evolutionary approach. *Cell Rep. Methods* **3**, 100429 (2023).
18. Fridy, P. C. *et al.* A robust pipeline for rapid production of versatile nanobody repertoires. *Nat. Methods* **11**, 1253–1260 (2014).
19. Cao, C. *et al.* Structure, function and pharmacology of human itch GPCRs. *Nature* **600**, 170–175 (2021).
20. Liu, Y. *et al.* Ligand recognition and allosteric modulation of the human MRGPRX1 receptor. *Nat. Chem. Biol.* **19**, 416–422 (2023).
21. Guo, L. *et al.* Ligand recognition and G protein coupling of the human itch receptor

MRGPRX1. *Nat. Commun.* **14**, 5004 (2023).
22. Shi, Y., Riese, D. J. & Shen, J. The Role of the CXCL12/CXCR4/CXCR7 Chemokine Axis in Cancer. *Front. Pharmacol.* **11**, 574667 (2020).
23. Liang, Y.-L. *et al.* Toward a Structural Understanding of Class B GPCR Peptide Binding and Activation. *Mol. Cell* **77**, 656-668.e5 (2020).
24. Edvinsson, L., Haanes, K. A., Warfvinge, K. & Krause, D. N. CGRP as the target of new migraine therapies — successful translation from bench to clinic. *Nat. Rev. Neurol.* **14**, 338–350 (2018).
25. Yao, S. *et al.* De novo design and directed folding of disulfide-bridged peptide heterodimers. *Nat. Commun.* **13**, 1539 (2022).
26. Cary, B. P. *et al.* New Insights into the Structure and Function of Class B1 GPCRs. *Endocr. Rev.* **44**, 492–517 (2023).
27. Booe, J. M. *et al.* Structural Basis for Receptor Activity-Modifying Protein-Dependent Selective Peptide Recognition by a G Protein-Coupled Receptor. *Mol. Cell* **58**, 1040–1052 (2015).
28. Hauser, A. S. *et al.* GPCR activation mechanisms across classes and macro/microscales. *Nat. Struct. Mol. Biol.* **28**, 879–888 (2021).
29. Lavington, S. & Watts, A. Lipid nanoparticle technologies for the study of G protein-coupled receptors in lipid environments. *Biophys. Rev.* **12**, 1287–1302 (2020).
30. Qin, L. *et al.* Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **347**, 1117–1122 (2015).
31. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).
32. Cheng, Z. *et al.* Inhibition of BET bromodomain targets genetically diverse glioblastoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 1748–1759 (2013).
33. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
34. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768 (2006).
35. Feldman, D. *et al.* Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **17**, 476–512 (2022).
36. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17 (2019).
37. Feldman, D. *et al.* Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **17**, 476–512 (2022).
38. Muratspahić, E. *et al.* Design and structural validation of peptide–drug conjugate ligands of the kappa-opioid receptor. *Nat. Commun.* **14**, 8064 (2023).
39. Liang, Y.-L. *et al.* Cryo-EM structure of the active, Gs-protein complexed, human CGRP receptor. *Nature* **561**, 492–497 (2018).
40. Josephs, T. M. *et al.* Structure and dynamics of the CGRP receptor in apo and peptide-bound forms. *Science* **372**, eabf7258 (2021).
41. Russo, C. J. & Passmore, L. A. Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens. *J. Struct. Biol.* **193**, 33–44 (2016).

42. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
43. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
44. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
45. Kinman, L. F., Powell, B. M., Zhong, E. D., Berger, B. & Davis, J. H. Uncovering structural ensembles from single-particle cryo-EM data using cryoDRGN. *Nat. Protoc.* **18**, 319–339 (2023).
46. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
47. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
48. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
49. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 519–530 (2018).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
51. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 531–544 (2018).