nature

https://doi.org/10.1038/s41586-026-10656-8

Accelerated Article Preview

De novo design of miniproteins targeting GPCRs

Received: 7 April 2025

Accepted: 13 May 2026

Accelerated Article Preview
Published online: 21 May 2026

Cite this article as: Muratspahić, E. et al. De novo design of miniproteins targeting GPCRs. *Nature* https://doi.org/10.1038/s41586-026-10656-8 (2026)

Edin Muratspahić, David Feldman, David E. Kim, Xiangli Qu, Ana-Maria Bratovianu, Paula Rivera-Sánchez, Jan Hendrik Voss, Emil P. T. Hertz, Mads Jeppesen, Federica Dimitri, Kensuke Sakamoto, Amrita Nallathambi, Pia Peceli, Jianjun Cao, Brian P. Cary, Matthew J. Belousoff, Peter Keov, Phuc N. H. Trinh, Qingchao Chen, Yue Ren, Justyn Fine, Sudha Mishra, Annu Dalal, Shachie Sinha, Ramanuj Banerjee, Manisankar Ganguly, Karthik Varappalayam Karuppusamy, Isaac Sappington, Thomas Schlichthaerle, Jason Z. Zhang, Arvind Pillai, Brian Coventry, Ljubica Mihaljević, Magnus Bauer, Susana Vázquez Torres, Amir Motmaen, Gyu Rie Lee, Long Tran, Xinru Wang, Inna Goreshnik, Dionne K. Vafeados, Justin E. Svendsen, Parisa Hosseinzadeh, Nicolai Lindegaard, Matthäus Brandt, Yann Waltenspühl, Kristine Deibler, Lukas Deweid, Anja Bennett, Jendrik Schöppe, Tiantang Dong, Xiaoli Yan, Luke Oostdyk, William Cao, Lakshmi Anantharaman, Johan J. Weisser, Jesper Frank Bastlund, Christoffer Bundgaard, Ayodeji A. Asuni, Justin G. English, Lance Stewart, Lauren Halloran, Jamie B. Spangler, André Lieber, Arun K. Shukla, Patrick M. Sexton, Bryan L. Roth, Brian E. Krumm, Denise Wootten, Christopher G. Tate, Christoffer Norn & David Baker

De novo design of miniproteins targeting GPCRs

Edin Muratspahić[1,2,#], David Feldman[1,2,3#], David E. Kim[1,2,4,#], Xiangli Qu[1,2,#], Ana-Maria Bratovianu[3,5,#], Paula Rivera-Sánchez[3,5,#], Jan Hendrik Voss[5], Emil P. T. Hertz[5], Mads Jeppesen[5], Federica Dimitri[3], Kensuke Sakamoto[6,7], Amrita Nallathambi[5], Pia Peceli[5], Jianjun Cao[8,9], Brian P. Cary[8,9], Matthew J. Belousoff[8,9], Peter Keov[8,9], Phuc N.H. Trinh[8,9], Qingchao Chen[10], Yue Ren[10], Justyn Fine[11,12], Sudha Mishra[13], Annu Dalal[13], Shachie Sinha[13], Ramanuj Banerjee[13], Manisankar Ganguly[13], Karthik Varappalayam Karuppusamy[14], Isaac Sappington[1,2], Thomas Schlichthaerle[1,2], Jason Z. Zhang[1,2], Arvind Pillai[1,2], Brian Coventry[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[2,15], Xinru Wang[1,2], Inna Goreshnik[1,2], Dionne K. Vafeados[1,2], Justin E. Svendsen[16,17], Parisa Hosseinzadeh[16,17], Nicolai Lindegaard[18], Matthäus Brandt[18], Yann Waltenspühl[18], Kristine Deibler[18], Lukas Deweid[18], Anja Bennett[18], Jendrik Schöppe[18], Tiantang Dong[19], Xiaoli Yan[19], Luke Oostdyk[20], William Cao[20], Lakshmi Anantharaman[20], Johan J. Weisser[21], Jesper Frank Bastlund[21], Christoffer Bundgaard[21], Ayodeji A. Asuni[21], Justin G. English[1,2], Lance Stewart[1,2], Lauren Halloran[12,22,23], Jamie B. Spangler[12,22,23], André Lieber[14], Arun K. Shukla[13], Patrick M. Sexton[8,9], Bryan L. Roth[6,7], Brian E. Krumm[6,7], Denise Wootten[8,9], Christopher G. Tate[10], Christoffer Norn[1,2,3,5*], David Baker[1,2,4]

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.
[2]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA.
[3]BioInnovation Institute, DK-2200 Copenhagen N, Denmark.
[4]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.
[5]Skape Bio ApS, DK-2200 Copenhagen N, Denmark.
[6]Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
[7]NIMH Psychoactive Drug Screening Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
[8]Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, VIC, Australia.
[9]ARC Centre for Cryo-electron Microscopy of Membrane Proteins, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, VIC, Australia.
[10]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, UK.
[11]Program in Molecular Biophysics, Johns Hopkins University, Baltimore, MD 21208, USA.
[12]Translational Tissue Engineering Center, Johns Hopkins University, Baltimore, MD 21231, USA.
[13]Department of Biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur 208016, India.
[14]University of Washington, Department of Medicine, Division of Medical Genetics, Seattle, WA 98195, USA.
[15]Department of Chemical Engineering, University of Washington, Seattle, WA 98195, USA.
[16]Department of Bioengineering, Knight Campus for Accelerating Scientific Impact, University of Oregon, Eugene, OR 97403, USA.
[17]Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403, USA.
[18]Novo Nordisk A/S, Novo Nordisk Park 1, 2760 Måløv, Denmark.
[19]Novo Nordisk Research Centre China, Changping District, Beijing 102206, China.
[20]LeadHunter Services, Eurofins DiscoverX, LLC, Fremont, CA94583, USA.
[21]H. Lundbeck A/S, Global Research, 9 Ottiliavej DK-2500 Copenhagen-Valby, Denmark.
[22]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.
[23]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

[#]These authors contributed equally.
*Correspondence: edinm2@uw.edu, chris.norn@skape.bio

Abstract

G protein-coupled receptors (GPCRs) play key roles in physiology and are central targets for drug discovery and development[1,2], but the design of protein agonists and antagonists has been challenging as GPCRs are integral membrane proteins and conformationally dynamic[3–6]. 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. We design miniprotein agonists that activate receptors involved in itch and pain, as well as antagonists that inhibit receptors implicated in cancer, metabolic disorders such as diabetes and obesity, and migraine. Cryo-electron microscopy (cryo-EM) structures of five receptor-bound designs are close to the computational design models. A designed chemokine receptor antagonist mobilizes hematopoietic stem and progenitor cells *in vivo* at a level comparable to a clinically used drug, with fewer adverse effects.

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 are therefore central to 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 strategies such as 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, that need to be engaged to induce function. We reasoned that specialized computational design and high-throughput experimental screening would be needed to tackle these challenges, and set out to develop appropriate methods.

Computational and experimental methods to target diverse GPCR epitopes

To enable targeting of deeply recessed orthosteric binding site epitopes, critical for modulating GPCR function, we implemented two complementary design methods to generate functional miniproteins. First, we developed a "motif-guided" RFdiffusion approach that rather than diffusing an entire binding protein, starts with a five-residue peptide (the motif) to interact with target hot spot residues within the recessed binding pocket (**Fig. 1a**). The short peptide more readily penetrates the deep pocket, and once receptor penetrating solutions are found, the peptide is fixed and full miniproteins are generated using motif-guided RFdiffusion[7]. To increase design diversity for library-scale screening, we developed an iterative partial diffusion approach that generates new designs near the most promising *in silico* solutions. Second, we developed MetaGen, which employs structurally diverse scaffolds from the AlphaFold (AF) predicted structural metaproteome[12,13] (**Supplementary Fig. 1**) in Rosetta RifDock calculations. Unlike 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, while remaining confidently predicted from a single sequence — a key criterion for designability[14]. Following backbone design we used ProteinMPNN for sequence design[15] and AF2 *initial guess*[12] along with Rosetta metrics for filtering designs[14]. Both methods yielded designs engaging the orthosteric pocket of five GPCRs with several designs reaching the lower portion of the orthosteric ligand-binding cavity, up to 12 Å below the outer membrane surface (**Supplementary Fig. 2**). To design class B receptor antagonists, we used MetaGen backbone library or generated backbones using RFdiffusion.

Across all targets, RFdiffusion and MetaGen sampled distinct structures and topologies, each contributing experimentally validated binders (**Supplementary Fig. 3-6)**.

Due to the challenging nature of class A GPCR epitopes, we reasoned that high-throughput screening methods would be necessary to complement computational design for identification of functional binders. To address this, we developed Receptor Diversion (RD), a purification-free microscopy-based assay that operates directly in human cells (**Fig. 1b**). In this assay, both the membrane protein target and candidate binder are expressed in a human cell line, with the binder localized within the secretory pathway using a genetic tag (e.g., KDEL; an endoplasmic reticulum (ER) retention signal). This allows the binder to interact with the extracellular face of the target during transit through the secretory pathway. High-affinity interactions cause "diversion" of the target from its normal trafficking pattern, which can be visualized as increased binder-target subcellular colocalization, i.e. binding signal (**Supplementary Fig. 7**). Across 7 diverse GPCRs, we observed a robust binding signal suitable for high-throughput screening (**Fig. 1c, d**) 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 the 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. 1i-f**). The OPS-RD platform enables screening of up to 100,000 designs through imaging of up to 10^7 cells providing expression and colocalization data at the single-cell level.

To benchmark the OPS-RD platform against yeast display[16], we used the MetaGen approach to generate a library of 4,036 designs targeting the extracellular domain (ECD) of the class B GPCR pituitary adenylate cyclase-activating polypeptide type I receptor (PAC1R) and screened the library alongside 200 negative controls using yeast display (**Supplementary Fig. 8a-g**) and OPS-RD (**Fig. 1j**). Yeast display hits were identified and ranked by *in silico* binding affinity (SC_{50}) estimation[8] using a cut-off of 4 μM. Binding affinity was determined by surface plasmon resonance (SPR) for hits that expressed in *E. coli* (**Supplementary Table 2**, **Supplementary Fig. 8h**). The OPS-RD binding signal was highly predictive of SPR-validated binders from yeast display (ROC AUC = 0.92; **Fig. 1g**) demonstrating concordance between the two screening technologies. Moreover, the binding signal obtained by OPS-RD was statistically significant for >75% SPR-validated yeast display hits and <2% of the negative controls at a false discovery rate (FDR) threshold of 5%. We tested the top 60 hits found by OPS-RD (FDR <5%, ranked on binding signal) in a two-point SPR screen and found 33 out of 60 proteins had binding affinity below 4 μM (**Supplementary Table 3**). OPS-RD performance is tunable by threshold choice: tightening the BH-adjusted p-value cutoff and/or requiring binding signal, increases SPR confirmation (for example, from 55% at FDR 5% to ~70% at an F1-optimal cutoff, while retaining ~91% of confirmed binders; **Supplementary Fig. 9**). As we were unsure of OPS-RD performance at the beginning

of this study, we also used yeast display paired with either soluble GPCRs in nanodiscs[17] or GPCRs displayed on mammalian cells (biofloating)[18].

Characterization of MRGPRX1 and NK1R miniprotein 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, we targeted a large epitope within the orthosteric binding pocket spanning transmembranes 2-7 (TM2–TM7) of three active-state structures[20]. We screened a library of 13,125 designs using OPS-RD, and mapped phenotypes for >800,000 cells to their design genotypes. Averaging optical phenotypes across cells, we ranked designs and selected 64 (**Fig. 2a**), then generated 27 additional designs using partial diffusion, resulting in 91 designs for further characterization. Of these, 50 were highly expressed in E. coli.

We screened designs in a calcium mobilization assay and identified seven with agonistic activity at 10 μM (**Supplementary Fig. 10**). Concentration-response curves showed that two were full agonists - similar activity to the endogenous BAM (8-22) with EC_{50} values of 390 ± 70 nM (mean ± s.e.m., n=3) and 1000 ± 100 nM (mean ± s.e.m., n=3), respectively, and one was a partial agonist with an EC_{50} of 1400 ± 200 nM (mean ± s.e.m., n=3) (**Fig. 2b**, **Supplementary Table 4**). These agonists are structurally diverse (**Fig. 2c**), highly expressed, monomeric by SEC (**Supplementary Fig. 11a**), and have CD spectra consistent with the expected molecular structure (**Supplementary Fig. 11b**) as well as high thermal stability (**Supplementary Fig. 11c**).

We carried out computational site-saturation mutagenesis using the cryo-EM structures (see following section) and models of the top three designs, calculating ΔΔG values for all single, double, and triple mutants of the three miniproteins in complex with MRGPRX1. Experimental characterization identified 21 variants with greater than 75% activity of BAM (8-22) (**Supplementary Fig. 12**). The top variants achieve better hydrophobic packing in the miniprotein-receptor interface. The most potent, mM1_034_F12W_Y27F exhibited an EC_{50} of 42 ± 7 nM (mean ± s.e.m., n=4), similar to BAM (8-22) and a 12-fold increase over the parent mM1_034 (**Fig. 2i**, **Supplementary Fig. 13**). A second variant, mM1_034_F12W_A58M with an EC_{50} of 110 ± 20 nM (mean ± s.e.m., n=5), displayed roughly a four-fold improvement over the parental mM1_034 (mean ± s.e.m., n=4), but had decreased efficacy (68 ± 3%, mean ± s.e.m., n=5) compared to BAM (8-22) (mean ± s.e.m., n=5) and mM1_034 (**Fig 2i**, **Supplementary Fig. 13**). Using TRUPATH (BRET-based G-protein dissociation) and β-arrestin-2 recruitment assays, that have been developed as an alternative to second messenger assays to minimize signal overamplification[21,22], both variants displayed low micromolar potency with reduced efficacy relative to BAM (8–22) (**Supplementary Fig. 14a–c**, **see legend for discussion**). Alanine mutations of key receptor residues decreased efficacy of both miniproteins and BAM (8–22) (**Supplementary Fig. 15**) indicating that optimized miniproteins engage orthosteric binding site yet act as partial agonists in less amplified signaling assays.

To evaluate receptor selectivity, we measured activity across 318 GPCRs using the PRESTO-Tango platform, which provides a genome-scale readout of GPCR activation[23]. Both mM1_034_F12W_Y27F and mM1_034_F12W_A58M elicited activities on only a small subset of receptors, indicating minimal off-target agonism under these conditions, and were

confirmed to be selective for MRGPRX1 in follow-up experiments **(Fig. 2j, Supplementary Fig. 14d-g**).

Next we focused on the neurokinin 1 receptor (NK1R), involved in pain, addiction and anxiety[24]. We designed miniproteins against the active state of NK1R using RFdiffusion (N=8,328) and MetaGen (N=16,675) and probed yeast display libraries using detergent-stabilized NK1R (**Supplementary Fig. 16a-g**). Of 71 RFdiffusion hits, 63 were expressed while for MetaGen, 22 hits were identified and 20 expressed (**Supplementary Fig. 17**). In a functional cAMP assay, 20 RFdiffusion designs acted as agonists with EC_{50} values from 1 nM to 231 nM and E_{max} values from 24% to 91% (**Fig 2k**, **Supplementary Fig. 18a**, **Supplementary Table 5**). The RFdiffusion agonists span 11 helical topologies (**Fig 2l**, **Supplementary Fig. 18b**), and occupy the orthosteric binding pocket, engaging NK1R with a loop motif structurally similar to the endogenous substance P (**Fig. 2l**). No MetaGen hits were active or recapitulated this motif. Furthermore, four RFdiffusion designs displayed antagonism when testing 1 μM concentration (**Supplementary Fig. 19a-b**) while remaining designs were inactive.

Cryo-EM structure of miniprotein agonists bound to MRGPRX1

We determined the cryo-electron microscopy (cryo-EM) structure of mM1_068 and mM1_060 bound to MRGPRX1 in complex with a mini-G_q protein[20] at global resolutions of 3.29 Å and 3.13 Å, respectively (**Fig. 2d, Supplementary Fig. 20-21**). The mM1_068 agonist adopts a proline-kinked three-helical bundle nearly identical to the design 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 is 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 was lower with higher B-factors compared to the transmembrane bundle (**Extended Data Table 1**) as expected given the size of the miniproteins and their partial protrusion from the orthosteric site. The lower resolution of the helices enabled only backbone modelling of residues exposed to the lipid environment. However, density was clearly observed for the helices within the orthosteric site and residues making close interactions with MRGPRX1 residues or extracellular loops. Both complexes show extensive interfaces with substantial buried surface area contributions from both miniprotein and receptor. Despite differences in size and orientation within the orthosteric site, mM1_068 and mM1_060 engage overlapping interaction networks, sharing multiple contact residues (**Supplementary Table 6**).

Previous structural and functional studies identified residues within the orthosteric site of MRGPRX1 necessary for BAM (8-22) to activate the receptor for signaling[20,25]. Both miniproteins overlap with the BAM (8-22) site (**Supplementary Table 6**). Several sidechains (E157$^{4.60}$, L240$^{6.59}$, F236$^{6.55}$, W241$^{6.60}$, and F250$^{7.31}$) differ in the determined structures, which may relate to the observed partial and full agonism. Overall, the cryo-EM data agree closely with the design models and confirm that the miniprotein binders sterically occlude the MRGPRX1 orthosteric site (**Fig. 2e-h**).

Characterization of CXCR4, CCR5 and OXTR miniprotein antagonists

We sought to design antagonists for the C-X-C chemokine receptor type 4 (CXCR4), a class A target receptor implicated in cancer and viral infection[26]. To prevent binding of CXCL12, we targeted the orthosteric site of CXCR4, generating 25,713 designs with motif-guided RFdiffusion (**Fig. 1a**, **Fig. 3a**). Yeast display libraries were probed using the biofloating approach (**Fig. 1a**, **Supplementary Fig. 22a-e**)[18], which uses CXCR4 expressed at the plasma membrane of mammalian cells[18], and nanodisc-stabilized CXCR4 (**Supplementary Fig. 23a-d**)**.** We identified five hits (**Supplementary Fig. 24a-e**) which eluted from SEC in a single peak (**Supplementary Fig. 24a-e**). While dCX1_003, dCX1_004 and dCX1_005 had IC_{50} values in the micromolar range (**Supplementary Fig. 24c-e**, **Supplementary Table 7**), dCX1_001 and dCX1_002 had IC_{50} values of 24 ± 5 nM (mean ± s.e.m., n=4) and 120 ± 30 nM (mean ± s.e.m., n=3), respectively (**Supplementary Fig. 24a-b, Supplementary Table 7**). dCX1_001 was highly thermostable (**Supplementary Fig. 25a-b**), antagonized CXCL12-mediated CXCR4 signaling with a pA_2 value of 7.6 ± 0.3 (25 nM, mean ± s.e.m., n=4) (**Fig. 3b**) and displayed no agonistic activity (**Supplementary Fig. 25c**). dCX1_001 antagonized CXCR4 and was selective across multiple assays (cAMP, $G_{\alpha i}$ dissociation and β-arrestin-1 recruitment) with little or no measurable activity at related CXCR7 and CCR5 (**Fig. 3c-h**).

The closely related chemokine receptor CCR5 is implicated in cancer and viral infection[26]. We designed and screened 26,398 binders using a yeast display with nanodisc-stabilized CCR5 (**Supplementary Fig. 26a-d**). 33 miniproteins were tested in $G_{\alpha oA}$ protein dissociation (**Supplementary Fig. 27a-b**) and β-arrestin-1 recruitment assays (**Supplementary Fig. 27c-d**), yielding three functional hits. dCC1_005 and dCC1_042 acted as antagonists in both assays with IC_{50} values in the low micromolar to mid-nanomolar range (**Supplementary Fig. 28a-c**, **Supplementary Table 7**). dCC1_002 acted as an agonist in the cAMP assay with EC_{50} and E_{max} values of 814 ± 257 nM (mean ± s.e.m., n=6) and 101 ± 2% (mean ± s.e.m., n=6), respectively, and in the $G_{\alpha oA}$ protein dissociation assay with an EC_{50} of 561 ± 161 nM (mean ± s.e.m., n=3) (**Supplementary Fig. 29a-c**, **Supplementary Table 7**). In the NanoBiT-based assay, dCC1_002 partially recruited β-arrestin-1 and β-arrestin-2 in the low micromolar range (**Supplementary Fig. 29d-e**). Both dCC1_042 and dCC1_002 were selective for CCR5 over CXCR4 in the cAMP assay (**Supplementary Fig. 30a-c**).

Using a similar nanodisc-based screen we next evaluated a library of 9,163 designs against the oxytocin receptor (OXTR) (**Supplementary Fig. 31a-d**) involved in cancer, pain, psychiatric and neuropsychiatric disorders and reproduction[27], and identified antagonist dOX1_003 with an IC_{50} of 330 ± 80 nM (mean ± s.e.m., n=4) (**Supplementary Fig. 32a**, **Supplementary Table 7**). Further optimization with partial diffusion improved potency, yielding antagonist dOX2_003 with a $pA2$ of 6.7 ± 0.1 (251 nM, mean ± s.e.m., n=4) (**Supplementary Fig. 32b**).

Cryo-EM structure of dCX1_001 antagonist bound to CXCR4

The antagonist dCX1_001 was designed against monomeric CXCR4. We attempted to determine the structure of dCX1_001, with and without a rigid fusion, bound to monomeric or dimeric CXCR4 (**Supplementary Fig. 33a-c**), but only a small fraction of well-behaved particles were observed, which were insufficient for reliable 3D reconstruction (**Supplementary Fig. 34a-j**). Instead, we determined the cryo-EM structure of dCX1_001 bound to homotrimeric CXCR4 at 3.28 Å resolution (**Supplementary Fig. 35a-d**, **Extended Data Table 2**). dCX1_001 adopts a complex α/β fold with two helices packing against three β-strands, closely matching the design model (0.6 Å Cα RMSD) (**Fig. 3i-l**). The dCX1_001-CXCR4 complex closely matches the RFdiffusion model, reproducing the binding mode and key hydrophobic contacts between the β-sheet and ECL2 (**Fig. 3k**). The receptor adopts a conformation that was not anticipated during design, but is similar to other CXCR4 structures, in which helices 6 and 7 shift outward relative to the reference structure (1.8 Å Cα RMSD over the extracellular half)[28]. The miniprotein preserves the designed hydrophobic interactions with ECL2 but pivots around the β-sheet, yielding a ligand Cα RMSD of 3.9 Å in the target-aligned frame while retaining 70% of the designed contacts (**Supplementary Fig 36)**. Although the homotrimeric structure supports the overall binding mode of dCX1_001, its relevance to pharmacological activity remains unclear.

To support the cryo-EM structure, antagonist activity of dCX1_001 was measured in a $G_{\alpha i}$ dissociation assay using cells expressing the wild-type or alanine-mutated CXCR4 at key interface residues. Mutation of the E179A^{ECL2}, Y190A^{ECL2} and E268A$^{6.64}$ significantly decreased activity (**Supplementary Fig. 37a-e**, **Supplementary Fig. 38a-b**). We also mutated key dCX1_001 residues interacting with the binding pocket and measured antagonistic activity in a cAMP assay (**Supplementary Fig. 39a**). These mutations abolished their inhibitory activity further supporting the proposed binding mode of dCX1_001 (**Supplementary Fig. 39b**).

Characterization of antagonists targeting ECD of multiple class B GPCRs

To explore design of antagonists of class B GPCRs, which include many therapeutic targets[29], we targeted glucagon-like peptide 1 receptor (GLP1R), gastric inhibitory polypeptide receptor (GIPR), glucagon receptor (GCGR), calcitonin gene-related peptide receptor (CGRPR) and parathyroid hormone 1 receptor (PTH1R). We targeted soluble extracellular domain (ECD), reasoning that ECD binding should induce steric hindrance and prevent peptide interaction, resulting in antagonism (**Fig. 4a-o**). For GLP1R, GIPR and GCGR, we used yeast display and soluble ECDs of receptors to identify binders. Probing a yeast library of 10,080 designs generated with RFdiffusion and MetaGen against GLP1R (**Supplementary Fig. 40a-e**), and expressing 49 hits with an estimated SC_{50}[8] below 400 nM, yielded dGl1_024 and mGl1_008 antagonists with IC_{50} values of 61 nM (mean, n=2) and 39 nM (mean, n=2), respectively, (**Fig. 4a**) and affinity of 27 nM ± 6 nM (mean ± s.e.m., n=3) and 5.3 nM ± 1.8 nM (mean ± s.e.m., n=3), respectively (**Supplementary Fig. 41a-e**, **Supplementary Fig. 42**). Similarly, probing yeast display libraries of 17,508 and 12,399 designs for GIPR (**Supplementary Fig. 43a-d)** and GCGR**,** (**Supplementary Fig. 44a-d**), respectively, and expressing 96 designs per target yielded binders with picomolar and nanomolar affinities (**Fig. 4b**, **Supplementary Fig. 45a-d**, **Supplementary Fig. 46a-d**). We confirmed that GIPR binders are antagonists in a cAMP assay with dGP1_035 and dGP1_040 having IC_{50} values of 7.9 ± 0.3 nM (mean ± s.e.m, n=4) and 13 ± 7 nM (mean ±

s.e.m., n=4), respectively (**Fig. 4b**, **Supplementary Fig. 47a**, **Supplementary Table 8**). All ten GIPR antagonists were selective for GIPR with minimal or no detectable activity at GLP1R or GCGR (**Fig. 4b, Supplementary Fig. 47b-c**). For PTH1R, we used OPS-RD to screen a library of 8,966 designs targeting the ECD (**Supplementary Fig. 48**). Of 92 screening hits, 50 bound the ECD (**Supplementary Table 9**). Four most promising binders displayed antagonism in a β-arrestin-2 recruitment assay and had IC_{50} values ranging from 810 ± 250 nM (mean ± s.e.m., n=3) to 450 ± 270 pM (mean ± s.e.m., n=3) (**Fig. 4c**, **Supplementary Fig. 49**, **Supplementary Table 8**).

CGRPR is a heterodimer consisting of calcitonin-receptor like receptor (CLR) and receptor activity-modifying protein 1 (RAMP1) and an established migraine target[30]. Structural inspection suggested that we could achieve a >1% hit rate, given the high epitope hydrophobicity and absence of loops, and thus we tested only 96 RFdiffusion and 89 Meta-Gen derived designs in a one-point cAMP assay using SK-N-MC cells (**Supplementary Fig. 50a-c**). Out of the 96 RFdiffusion designs, 67 expressed, and a three-helix bundle miniprotein dC1_021 antagonized CGRPR signaling with an IC_{50} of 440 ± 40 nM (mean ± s.e.m., n=3) (**Supplementary Fig. 51**, **Supplementary Table 10**). Of the 89 MetaGen-derived designs, 83 expressed and four showed antagonistic activity in a one-point luciferase assay (**Supplementary Fig. 52a-c**). We identified a competitive antagonist mC1_023 of mixed αβ-topology (**Fig. 4d**) with an IC_{50} of 37 ± 2 nM (mean ± s.e.m., n=4) (**Supplementary Fig. 53**, **Supplementary Table 10**) and a pA_2 of 8.3 ± 0.1 (5 nM, mean ± s.e.m., n=4) (**Fig. 4d**). Disulfide stapling[31] of the hit mC1_044 with micromolar potency (**Supplementary Fig. 53**) yielded antagonist mC2_022 with a pA_2 of 7.9 ± 0.2 (13 nM, mean ± s.e.m., n=4) (**Fig. 4e**) and IC_{50} of 420 ± 60 nM (mean ± s.e.m., n=4) (**Supplementary Fig. 53**, **Supplementary Table 10**). To improve potency of dC1_021, we performed partial diffusion[6]. Out of 78 designs that expressed, 36 had antagonistic activity in a one-point cAMP assay (**Supplementary Fig. 54a-c**). Concentration-response curves of the 20 most promising binders identified competitive antagonist dC2_049 with an IC_{50} of 4.5 ± 0.9 nM (mean ± s.e.m., n=4) (**Supplementary Fig. 55a-d**, **Supplementary Table 10**) and a pA_2 of 8.4 ± 0.1 (3.9 nM, mean ± s.e.m., n=4) (**Fig. 4f**). 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 **Supplementary Fig. 56a-c**). dC2_049 exhibited high selectivity for CGRPR with little or no significant cross-reactivity at related adrenomedullin receptors 1 (AM_1) and 2 (AM_2), calcitonin receptor (CTR) and amylin 1 receptor (AMY_1R) (**Fig. 4g-i**). A sequence-similar RFdiffusion binder dC2_050, derived by partial diffusion from the same parent structure as dC2_049, had similar pharmacological and biophysical properties (**Supplementary Fig. 57a-e**, **Supplementary Fig. 58a-d, Supplementary Table 10**). Both dC2_049 and dC2_050 were also CGRPR antagonists in COS-7 cells transiently expressing CLR and RAMP1 (**Supplementary Fig. 59a-b**). To examine drug-like properties *in vivo*, we selected mC2_022, and compared pharmacokinetics of unmodified and Fc-fused forms. Following a single intravenous dose in mice (3 mg/kg), plasma levels were quantified by UPLC-MS/MS over 24 h. The unmodified miniprotein was rapidly cleared ($t_{½}$ ≈ 0.4 h), whereas Fc-fusion extended exposure ($t_{½}$ ≈ 25 h) with >200-fold lower clearance (1.2 vs 0.0051 L/h/kg) (**Supplementary Fig. 60a-e**).

Cryo-EM structure of miniprotein antagonists bound to CGRPR

We determined the cryo-EM structure of dC2_049 and dC2_050 bound to CGRPR (**Fig. 4j-o**) with global resolutions of 3.2 Å and 4.1 Å, respectively (**Supplementary Fig. 61-62**, **Extended Data Table 3**). The cryo-EM structures (**Supplementary Fig. 61-62**) agree closely with the design models for dC2_049 and dC2_050. The dC2_049 binder superimposes at 0.9 Å in the ECD-aligned frame and 0.5 Å after binder-only superposition, while the dC2_050 binder displays RMSDs of 1.0 Å and 0.7 Å, respectively (Fig. 4j-o). In both structures, the binders sterically occlude the binding site for the C-terminal portion of αCGRP (Fig. 4j-o). Local resolutions for the ECD were lower than for the transmembrane bundle, as commonly observed for class B1 GPCR structures[32], enabling only backbone modelling of most ectodomain residues. However, density was observed for some larger, interfacial sidechains along helix-1 and helix-3 of dC2_049. For instance, $Trp72^{ECD}$, a key residue for αCGRP activity[33], forms hydrophobic contacts with Met12 of dC2_049 (**Supplementary Fig. 63**). To complement cryo-EM structures, antagonistic activity of dC2_049 and dC2_050 was measured in a cAMP assay using cells expressing wild-type or alanine-mutated RAMP1 or CLR at key interface interfaces. Mutation of several residues abolished activity, further supporting the proposed binding modes (**Supplementary Fig. 64-65**).

We retrospectively analyzed AF2 and Rosetta metrics between binders and non-binders designed by RFdiffusion and MetaGen across class A and class B targets, but there were no consistent trends across most targets (**Supplementary Fig. 66**). The experimentally characterized designs had already been filtered using these same metrics, and lack of discrimination within the post-filter window does not mean that the metrics fail to separate binders from non-binders across the full design space.

In vivo pharmacology of dCX1_001 antagonist

The CXCR4-CXCL12 interaction regulates retention and migration of hematopoietic stem and progenitor cells (HSPCs) within the bone marrow niche[34]. Inhibition of this axis mobilizes HSPCs into peripheral blood for collection and autologous transplantation, a clinically established approach for treating hematological malignancies[35]. We evaluated the efficacy and tolerability of dCX1_001 in CD46 transgenic mice[36]. Subcutaneous administration increased circulating Lin^-, $Sca1^+$, $cKit^+$ (LSK) cells and colony forming units (CFU), with efficacy comparable to AMD3100 (plerixafor), an approved small molecule drug for non-Hodgkin and multiple myeloma (**Fig. 5a-b**, **Supplementary Fig. 67a-b**). Using a helper-dependent adenovirus (HDAd) vector system, we confirmed GFP^+ cells and vector genomes in hematopoietic compartments demonstrating successful re-engraftment of transduced HSPCs mobilized by dCX1_001 (**Fig. 5c-e**, **Supplementary Fig. 67c-d**). Notably, dCX1_001 had fewer unwanted side effects than AMD3100: it induced lower levels of circulating leukocytes (**Fig. 5f**, **Supplementary Fig. 68a**) than AMD3100 and did not lead to adverse hematologic effects or elevated systemic cytokines (**Supplementary Fig. 68b-i**). Thus, dCX1_001 achieves HSPC mobilization comparable to AMD3100 with a lower risk of side effects.

Discussion

GPCRs have been longstanding challenges for drug discovery and development of protein-based ligands owing to their structural complexity, recessed binding pockets, and dynamic character.

We show that *de novo* design can address these challenges by generating miniprotein binders targeting eleven receptors with diverse affinity, potency, and selectivity. Agonists are particularly challenging to obtain due to the need for conformational selectivity, requiring discrimination between subtle structural differences in the orthosteric site that distinguish active from inactive states[37]. Here, we demonstrate the *de novo* design of MRGPRX1 binders, two of them atomically accurate (within 0.7 Å), and NK1R binders, capable of inducing full and partial agonism. While these findings establish *de novo* design as a viable strategy for engineering GPCR-targeting miniprotein antagonists and agonists, we note that using active and inactive receptor states alone is not sufficient to ensure design of an agonist or antagonist due to the subtle differences between states and the complexities of GPCR signaling dynamics; indeed some of our designs may act as allosteric modulators. We focused here on peptide- and protein-binding GPCRs that comprise more open and shallow orthosteric binding pockets as proof-of-concept for the *de novo* protein design in engineering GPCR agonists and antagonists. Given the success rates future work should examine the potential of these methods to design miniprotein agonists, antagonists or allosteric modulators against GPCRs with small molecules as endogenous ligands – aminergic GPCRs are more challenging targets as they typically have narrow and partially occluded orthosteric pockets buried within the transmembrane domain. Complementing computational design, our OPS-RD platform enables high-throughput screening for difficult GPCR targets by circumventing the need for soluble receptor preparations in nanodiscs, liposomes or mutant receptor proxies, which may alter receptor conformations and function[38]. While the maximal capacity (<100,000 designs) is lower than display technologies, screening against the native receptor in the membrane environment and bypass solubilization provides a major benefit.

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 represents a step change in methods for obtaining functional biologics targeting integral membrane receptors. Combined with their smaller size for improved tissue penetration, high stability, rapid design and optimization, and potential for modifications to enhance metabolic stability[11,12], miniproteins represent an attractive therapeutic modality. Beyond therapeutics, designed GPCR binders could serve as tools for drug discovery, probing pharmacology and receptor function, and stabilizing receptor conformations for structural studies.

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. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768 (2006).
17. 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).
18. 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).
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. Olsen, R. H. J. *et al.* TRUPATH, an open-source biosensor platform for interrogating the GPCR transducerome. *Nat. Chem. Biol.* **16**, 841–849 (2020).
22. DiBerto, J. F., Olsen, R. H. J. & Roth, B. L. TRUPATH: An Open-Source Biosensor Platform for Interrogating the GPCR Transducerome. *Methods Mol. Biol. Clifton NJ* **2525**, 185–195 (2022).
23. Kroeze, W. K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).
24. Schank, J. R. Neurokinin receptors in drug and alcohol addiction. *Brain Res.* **1734**, 146729 (2020).
25. Guo, L. *et al.* Ligand recognition and G protein coupling of the human itch receptor MRGPRX1. *Nat. Commun.* **14**, 5004 (2023).
26. Shi, Y., Riese, D. J. & Shen, J. The Role of the CXCL12/CXCR4/CXCR7 Chemokine Axis in Cancer. *Front. Pharmacol.* **11**, 574667 (2020).
27. Jurek, B. & Neumann, I. D. The Oxytocin Receptor: From Intracellular Signaling to Behavior. *Physiol. Rev.* **98**, 1805–1908 (2018).
28. Saotome, K. *et al.* Structural insights into CXCR4 modulation and oligomerization. *Nat. Struct. Mol. Biol.* **32**, 315–325 (2025).
29. Liang, Y.-L. *et al.* Toward a Structural Understanding of Class B GPCR Peptide Binding and Activation. *Mol. Cell* **77**, 656-668.e5 (2020).
30. 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).
31. Yao, S. *et al.* De novo design and directed folding of disulfide-bridged peptide heterodimers. *Nat. Commun.* **13**, 1539 (2022).
32. Cary, B. P. *et al.* New Insights into the Structure and Function of Class B1 GPCRs. *Endocr. Rev.* **44**, 492–517 (2023).
33. 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).
34. Karpova, D. & Bonig, H. Concise Review: CXCR4/CXCL12 Signaling in Immature Hematopoiesis—Lessons From Pharmacological and Genetic Models. *Stem Cells* **33**, 2391–2399 (2015).
35. Karpova, D. *et al.* Continuous blockade of CXCR4 results in dramatic mobilization and expansion of hematopoietic stem and progenitor cells. *Blood* **129**, 2939–2949 (2017).
36. Li, C. *et al.* Single-dose MGTA-145/plerixafor leads to efficient mobilization and in vivo transduction of HSCs with thalassemia correction in mice. *Blood Adv.* **5**, 1239–1249 (2021).
37. Hauser, A. S. *et al.* GPCR activation mechanisms across classes and macro/microscales. *Nat. Struct. Mol. Biol.* **28**, 879–888 (2021).
38. Lavington, S. & Watts, A. Lipid nanoparticle technologies for the study of G protein-coupled receptors in lipid environments. *Biophys. Rev.* **12**, 1287–1302 (2020).

Fig 1. Computational design and screening of GPCR binders. a Designed backbones targeting GPCRs were generated *de novo* by free or motif-guided RFdiffusion, or by docking 7,000 metaproteome-derived miniproteins. After sequence assignment and *in silico* prioritization, class A and B GPCR designs were screened in functional assays or by high-throughput binding assays, including yeast surface display with soluble receptors, biofloating assay and Optical Pooled Screening-Receptor Diversion (OPS-RD). **b** OPS-RD converts binder-receptor interactions into an optical phenotype. Without binding, fluorescently tagged receptors traffic to the cell surface while the design is retained separately in the secretory pathway. Upon binding, ER-retained binders trap fluorescently tagged wild-type receptors, causing colocalization in the secretory pathway. **c** Across seven GFP-fused GPCRs, positive controls (oligomerized 0.7 nM anti-GFP nanobody C5) produced significantly higher binding signals than negative controls (non-binding miniproteins). The fraction of cells with the binding phenotype was calculated from at least 80 cells, and s.e.m. was scaled to n=50 cells, compatible with high-throughput screening. **d** False-positive rate at a fixed false-negative rate of 5% as a function of the number of cells imaged across GPCR targets, based on the same controls as in **c**. **e**, To deploy OPS-RD at scale, designed binders were synthesized on oligonucleotide arrays, cloned into a lentiviral library, introduced at low multiplicity of infection so each cell carried one design, quantified by receptor trapping, and identified by *in situ* barcode sequencing. **f** 4,036 miniprotein antagonists targeting the PAC1R ECD were designed and screened using OPS-RD. The GFP-RFP pixel cross-correlation, corr(GFP,RFP), for cells carrying the same design was compared with the distribution across more than 2 million imaged cells. *P* values were computed by a one-sided Kolmogorov-Smirnov test and adjusted by the Benjamini-Hochberg procedure (Data availability). **g** ROC curve for classifying yeast display SPR-validated hits by OPS-RD binding signal; AUC, area under the curve.

Fig. 2. Characterization of MRGPRX1 and NK1R agonists. a 13,000 MetaGen-designed MRGPRX1 agonists were 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). P values were computed using a one-sided Kolmogorov–Smirnov (K-S) test and adjusted with the Benjamini-Hochberg procedure. P values are provided in the MRGPRX1 dataset (Data availability). **b** Concentration-response curves of agonist hits and positive control BAM (8-22) in a calcium flux assay. Data are shown as mean ± s.e.m, n=3 biological replicates (mM1_034, mM1_060, mM1_068) or n=4 biological replicates (BAM (8-22)). **c** Computational models of three agonist hits (receptors truncated for clarity). **d** Cryo-EM maps of mM1_068 (left) and mM1_060 (right) bound to MRGPRX1 (low threshold map indicates detergent micelles). **e** Alignment of mM1_068, mM1_060, and BAM (8-22) cryo-EM models bound to MRGPRX1. Alignment of cryo-EM structures of **f** mM1_068 and **g** mM1_060 bound to MRGPRX1 with the design models. **h** Key activation and signaling residue differences in MRGPRX1 between miniprotein- and BAM (8-22)-bound cryo-EM structures. **i** Concentration-response curves of optimized agonists compared to mM1_034 and BAM (8-22) in CHO-K1 cells overexpressing MRGPRX1. Data represent the mean ± s.e.m., n=4 biological replicates (mM1_034, mM1_034_F12W_Y27F) or n=5 biological replicates (BAM (8-22), mM1_034_F12W_A58M)). **j** 320-GPCR selectivity screen of improved agonists at a concentration of 3 μM. Black dots represent activated GPCRs (≥3-fold over the vehicle indicates potential positives). Dopamine D2 receptor agonist quinpirole (100 nM) served as positive control (red dot). **k** cAMP assay in CHO cells stably expressing NK1R. Substance P served as positive control. Data are shown as mean ± s.e.m., n=3 biological replicates. **l** Computational design models of three representative RFdiffusion-designed agonists (blue) bound to NK1R (yellow, PDB ID: 7P02) compared to Substance P (grey, PDB ID: 7P02).

Fig. 3. Pharmacology and cryo-EM structure of a designed CXCR4 antagonist. 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. The receptor structure was truncated for clarity. **b** Functional cAMP assay of dCX1_001 binder in CHO cells stably expressing CXCR4. Data are shown as mean ± s.e.m., n=4 biological replicates. 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). **c** Inhibition of G protein activation by dCX1_001 binder at the CXCR4 in a NanoBiT-based $G_{\alpha i3}$ dissociation assay (mean ± s.e.m., n=3 biological replicates). β-arrestin-1 recruitment assay of dCX1_001 binder at **d** CXCR4 and **e** CXCR7 in a NanoBiT-based luminescence assay (mean ± s.e.m., n=3 biological replicates). **f** GloSensor cAMP, **g** NanoBiT-based $G_{\alpha i3}$ dissociation and **h** β-arrestin-1 recruitment of dCX1_001 binder at the CCR5 (mean ± s.e.m., n=3 biological replicates). **i** Side and **j** top view of cryo-EM maps and cryo-EM models of dCX1_001 bound to homotrimeric CXCR4. Lipids at the central axis are shown in green. **k** Alignment of CXCR4 and dCX1_001 cryo-EM structure with RFdiffusion design model (Cα RMSD=1.8 Å). Shown is alignment to one protomer of the homotrimeric CXCR4. **l** Key receptor residues involved in the antagonistic activity of dCX1_001 miniprotein.

Fig. 4. Characterization of class B GPCR miniprotein antagonists. a, mGI1_008 antagonism in BHK21/GLP1R/Cre-luc cells in the presence of 15 pM semaglutide. mGI1_008 antagonized GLP1R signaling (IC_{50} = 39 nM). Exendin 9-39 was used as control (IC_{50} = 13.9 nM). Data are mean, n=2 biological replicates. **b** cAMP assays of dGP1_035 and dGP1_040 antagonizing GIPR, GLP1R and GCGR signaling, respectively, (IC_{50} 7.9 ± 0.3 nM (dGP1_035) and 13 ± 7 nM (dGP1_040)). Both show only little or no significant activity at the GLP1R and GCGR. Data are mean ± s.e.m., n=4 biological replicates (GIPR, GLP1R), n=6 biological replicates (GCGR). **c** β-arrestin-2 recruitment assays of mPT_084 and mPT_095 antagonizing PTH1R signaling (IC_{50} = 450 ± 270 pM and 163 ± 44 nM, respectively). Data are mean ± s.e.m., n=3 biological replicates. Design models of MetaGen **d** mC1_023 **e** mC2_022 and RFdiffusion **f** dC2_049 bound to CGRPR (PDB ID: 6E3Y, receptor truncated). Concentration response curves of antagonists in the presence of αCGRP. Functional estimates (mean ± s.e.m.) are mC1_023 (pA_2 of 8.3 ± 0.1 (5 nM) and slope 0.96 ± 0.04, n=4 biological replicates), mC2_022 (pA_2 = 7.9 ± 0.2 (13 nM), slope 0.61 ± 0.04, n=4 biological replicates), dC2_049 (pA_2 = 8.4 ± 0.1 (3.9 nM), slope = 0.82 ± 0.05, n=4 biological replicates). Selectivity of dC2_049 at **g** calcitonin receptor (CTR) **h** adrenomedullin receptors 1 and 2 (AM1 and AM2) and **i** amylin 1 receptor (AMY1R). Data are mean ± s.e.m., n=4 biological replicates. Cryo-EM maps of CGRPR-bound **j** dC2_049 and **k** dC2_050 (low threshold maps represent detergent micelles). **l** Alignment of dC2_049 and dC2_50 design models bound to CGRPR, respectively. Alignment of CGRPR-bound **m** dC2_049 and **n** dC2_050 cryo-EM structures with the design models. **o** Map densities indicate miniproteins sterically occlude binding of the αCGRP C-terminus.

Fig. 5. Mobilization of HSPCs by dCX1_001 and AMD3100. Human CD46-transgenic mice were injected subcutaneously with PBS (control), AMD3100 (5 mg/kg) or dCX1_001 (5 mg/kg). Blood samples were collected at indicated time points. **a** Flow cytometry quantification of Lin-/Sca1+/cKit+ (LSK) cells (mean ± s.e.m., n=5 mice per group). Statistical analysis was performed using two-sided two-way ANOVA with Sidak's multiple comparisons test. Asterisks (control vs. ligand) and hashes (dCX1_001 vs. AMD3100) indicate statistical significance (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$, $^{\#\#}p \leq 0.01$, $^{\#\#\#}p \leq 0.001$, ns, not significant). **b** Colony-forming assay of white blood cells (WBC) plated in cytokine-supplemented, semi-solid medium; colonies counted after 10 days (mean ± s.e.m., n=5 mice per group). Statistical analysis as in **a**, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$, $^{\#}p \leq 0.05$, $^{\#\#}p \leq 0.01$, ns, not significant). **c** Schematic of *in vivo* mobilization and HSPC transduction. Mice received dexamethasone (16 h and 1.45 h) followed by AMD3100 or dCX1_001 (45 min) and intravenous HDAd; tissues were collected after 72 h. **d** GFP expression in total mononuclear cells (MNCs) from bone marrow (BM), spleen, peripheral blood (PB), and bone marrow LSK cells (mean ± s.e.m., n=5 mice per group). Two-sided Kruskal-Wallis with Dunn's multiple comparisons test; asterisks (control vs. ligand) denote statistical significance, *$p \leq 0.05$, **$p \leq 0.01$, ns, not significant. **e** Vector copy number (VCN) per cell in BM MNCs at 72 h after HDAd (mean ± s.e.m, n=5 mice (control, dCX1_001), n=4 mice (AMD3100)). Hashes (control vs. ligand) indicate statistical significance, *$p \leq 0.05$. **f** Blood cell counts at 30 min, 60 min and 72 h (mean ± s.e.m.; n=5 mice per group). Statistical analysis as in **d**. Hashes (dCX1_001 vs. AMD3100) represent statistical significance, $^{\#}p \leq 0.05$, $^{\#\#}p \leq 0.01$, ns, not significant. Neutrophils (NE); lymphocytes (LY); monocytes (MO); eosinophils (EO), basophils (BA). Exact *p* values and detailed statistics for **a**, **b** and **d-f** are provided in the Source data. Statistics were performed only for indicated comparisons; other comparisons were not statistically analyzed.

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), PAC1R (PDB ID: 8E3X), OXTR (PDB ID: 7RYC) and NK1R (PDB ID: 7P02) the crystal structures of CXCR4 (PDB ID: 4RWS) and CCR5 (PDB ID: 5UIW) 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), PAC1R (7JQD, 2JOD, 6M1I, 6M1H, 8E3X and AlphaFold model AF-P41586-F1-model_v4), NK1R (7P02, 7RMG, and 7RMH), PTH1R (7UZO, 8FLU, 7Y36, 7VVL, 6FJ3, and AF model AF-Q03431-F1-model_v4) 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-guided RFdiffusion targeting GLP1R, GIPR, GCGR or constrained 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 and sampling temperature of <0.1)[15], followed by FastRelax and AF2 *initial guess*[12]. *In silico* cutoffs goals were defined a priori from our previous benchmarking of AF2 and Rosetta metrics across multiple targets[7,14]. Where feasible, we targeted pLDDT_binder > 90, pAE_interaction < 8, binder_RMSD < 2, and low spatial aggregation propensity (SAP) (≤ 35) to reduce aggregation risk while maintaining confident local structure. We treated pAE_interaction and pLDDT_binder as primary acceptance filters, then set Rosetta ddG and SAP to filter remaining candidates. When very few designs met these strict criteria, thresholds were relaxed within the target specific ranges reported above, guided by interface chemistry. Targets that can be engaged via larger hydrophobic contact patches generally support more stringent Rosetta ddG cutoffs.

Designs generated by RFdiffusion were selected based on pAE_interaction < 8, pLDDT_binder > 85, Rosetta ddG < -45, 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, pAE_interaction < 4, pLDDT_binder > 90, Rosetta ddG < -45 and SAP < 35 for PAC1R, and pAE_interaction < 8, pLDDT_binder > 90 and Rosetta ddG < -45 for CGRPR. Scaffolds for MetaGen were selected from AF generated metaproteome[12,13] containing 200 million protein structures to as follows: longest_loop <= 7 & longest_strand <= 10 & longest_helix <= 20 & fraction_loop <= 0.35 & length <= 90. Following this, we designed all the scaffolds using ProteinMPNN, and computed structural metrics with Rosetta and AF2. We then filtered for globularity (percent_core_SCN > 20%) & SAP <= 40 & energy (score_per_res <= -2.3) & pLDDT_scaffold > 90. Finally, we clustered all scaffolds by sequence with mmseqs[39] to 80% sequence identity and picked the best scaffold within each cluster by pLDDT_scaffold. This resulted in a final set of 8,920 scaffolds.

Metaproteome-derived designs targeting CGRPR, MRGPRX1, GLP1R, GIPR, PAC1R, PTH1R, NK1R 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; PAC1R designs passed SAP < 35 & RMSD_binder < 2 & membrane_insertion_energy > 4 & scaffold_plddt > 85 & scaffold_rmsd < 2 & ddg < -40 & pAE_interaction < 6; PTH1R designs passed SAP < 35 & RMSD_binder < 2 & membrane_insertion_energy > 4 & ddg < -30 & pAE_interaction < 12; NK1R designs passed pAE_interaction < 12, binder_RMSD < 2, ddG < −30, 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. For MRGPRX1, additional designs were generated after OPS-RD hit identification using partial RFdiffusion, filtered as described above, and the top 27 designs were selected based on pAE_interaction.

Partial RFdiffusion was performed on the AF2 model of the most promising CGRPR design hit (dC1_021). 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_interaction < 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[40], however, prior to the final AF2 prediction, designs were mutated to the native D of the receptor. Binders against CCR5 and OXTR were designed using the same design strategy as the CXCR4 campaign. The initial *in silico* metrics prior to iterative partial diffusion were CCR5 (pAE_interaction < 15 and pLDDT_binder > 60) and OXTR (pAE_interaction < 15 and pLDDT_binder > 70). More stringent *in silico* criteria were used following iterative partial diffusion in order to obtain libraries on a chip scale: CCR5 (pAE_interaction < 8, pLDDT_binder > 85, Rosetta ddG < −60 and SAP < 60) and OXTR (pAE_interaction < 10, pLDDT_binder > 85, Rosetta ddG < −45 and SAP < 60). The initial OXTR hit was further optimized using partial diffusion and genes encoding designs that passed *in silico* metrics ((pAE_interaction < 10, pLDDT_binder > 90, Rosetta ddG < −80 and SAP < 60) were obtained for subsequent expression, purification and functional characterization. For the NK1R binder design, the initial *in silico* metric was pAE_interaction < 6 and pLDDT_binder > 90. Designs were then partially diffused and filtered on pAE_interaction < 6, pLDDT_binder > 90 and Rosetta ddG < −45.

Rigid fusions between the CXCR4-bound dCX1_001 and *de novo* helical repeat (DHR) scaffolds[41] were designed with WORMS[42]. We enumerated rigid-body splices between the binder and a range of DHRs that vary in repeat number, length, and curvature, and discarded configurations that produced steric clashes with the CXCR4 complex. Junction residues at the binder–DHR interfaces were subsequently redesigned with ProteinMPNN to maximize rigidity and solubility. Candidate designs were evaluated with AlphaFold2 and Rosetta, and designs that passed *in silico* metrics (pAE_interaction < 7, pLDDT_binder > 90, Rosetta ddG < −45 and SAP < 45) were selected for characterization. Data were analyzed using Python v3.1, NumPy v1.26.4 and pandas v2.2.2 and visualized using seaborn v0.13.2 and matplotlib v3.9.2. Structures were visualized and images were generated using PyMOL v3.1.

Cloning, expression and purification of protein binders

Protein binder designs were obtained as synthetic genes (eBlocks, Integrated DNA Technologies) with compatible BsaI overhangs to the target cloning vector, LM0627 for Golden Gate assembly[43]. 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 obtained 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. Sequences of binders characterized in more detail are shown in the **Supplementary Table 11**.

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°C and 95°C, measuring CD at 222 nm. All reported measurements were acquired within the linear range of the instrument.

Cell culture

Cell lines were not tested for mycoplasma contamination or authenticated, but were routinely monitored for normal morphology, growth characteristics, and contamination. CHO-K1/CRE-Luc/CGRPR (Genscript, M00350) cells were cultured in Ham's F-12K (Kaighn's) Medium (Gibco) containing 10% FBS. CHO-K1/PTH1R (DiscoverX, 93-0306C2) and CHO-K1/MRGPRX1 β-arrestin cells (DiscoverX, 93-0919C2) overexpressing the respective receptor and a split-β-galactosidase were cultured and plated according to the manufacturer's instructions. HEK293T (American Type Culture Collection (ATCC), CRL-3216) HEK293A (Thermo Fisher Scientific, R70507), SK-N-MC (ATCC, HTB-10) and HEK293/OXTR (Princess I. Imoukhuede laboratory) 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 (ATCC, CRL-1651) cells were cultured in DMEM containing 10% FBS only. CHO-K1/CXCR4 (GenScript, M00556) 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, 32180) and HeLa cells (Iain Cheeseman laboratory), 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. BHK-21 cells (ATCC, CCL-10) were stably transfected with a human GLP1R expression plasmid and a CRE-luciferase reporter plasmid. The GLP1R expression plasmid is based on pcDNA3.1(+), while the reporter plasmid used was pGL4.29[luc2P/CRE/Hygro] from Promega. A single cell clone, designated clone FCW467-12A/KZ10-1, was isolated for the assays. The growth media consisted of DMEM (Gibco, 61965026), 10% FBS (Gibco, 10100147), 1% penicillin-streptomycin (Gibco, 15140122), 500

μg/mL G418 (Gibco, 10131027), and 1% pyruvate (Gibco, 11360039). The cDNA coding regions of CXCR4, CXCR7 and CCR5 were cloned in the pcDNA3.1 vector containing an HA signal sequence and N-terminal FLAG tag from Genscript. The constructs used in the NanoBiT-based β-arrestin recruitment assays were cloned in the pCAGGS vector where the SmBiT tag was fused at the receptor C-terminus and LgBiT tag at the N-terminus of β-arrestin-1. The G protein subunit constructs used in the dissociation assays were generously gifted by Asuka Inoue. CHO-K1 cells (ATCC, CCL-61) were stably transfected with a pcDNA3.1(+) plasmid encoding human neurokinin-1-receptor (NK1R) and a pGL4.29[luc2P/CRE/Hygro] plasmid encoding a cAMP-response-element-luciferase (CRE-Luc). Transfected cells were selected under 600 μg/mL G418 (Gibco, 10131027) and 400 μg/mL hygromycin (Invitrogen, 10687010) and a single clone was isolated. Culture media consisted of DMEM/F-12 (Gibco, 21331-020), 10% FBS (Gibco, 10100147) and 1% penicillin-streptomycin (Gibco, 15140122). HTLA cells (HEK293 stably expressing a tetracycline transactivator (tTA)-dependent luciferase reporter and β-arrestin-2 fused to a TEV protease) (Bryan L. Roth laboratory) were maintained in DMEM supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin, 100 μg/mL streptomycin, 100 μg/mL hygromycin B, and 2 μg/mL puromycin at 37 °C in 5% CO_2. AMD3100 was obtained from Millipore Sigma (239820) and Dexamethasone Sodium Phosphate from Fresenius Kabi.

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)[44]. Briefly, 3×10^6 HEK293T 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 HEK293T cells, along with the pPACK packaging plasmid mix. GeneJuice (Sigma-Aldrich) 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 HEK293T cells cultured in a 24-well plate were transduced with GPCR lentivirus in the presence of 8 μg/mL polybrene (Sigma-Aldrich) in 500 L complete DMEM culture media. Immediately after transduction, HEK293T 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 obtained from Twist Bioscience. Combinatorial libraries were obtained 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 including the native library and different sorting pools were sequenced using Illumina NextSeq/MiSeq sequencing.

Yeast display

General yeast display methodologies were carried out with EBY100 yeast cells, as previously described[16,45]. Yeast clones for biofloating assay were grown in SDCAA medium at 30ºC while shaking at 200 rpm. Yeast cultures were induced in SGCAA 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 EBY100 strain cultures were grown in C-Trp-Ura media and induced in SGCAA. Cells were washed with PBSF (PBS with 1% BSA) and incubated with the Flag-tagged CXCR4 target (DIMA Biotech, SKU:FLP100074), OXTR (DIMA Biotech, SKU:FLP100429), CCR5 (SinoBiological, 13022-H91H-NA) or biotinylated GLP1R (SinoBiological 13944-H49H-B, GIPR (SinoBiological, 18774-H49H-B), GCGR (Acro Biosystems, GCR-H82E3), or PAC1R respectively. For the first round of sorting, cells were incubated with the FLAG-tagged CXCR4, CCR5 or OXTR nanodisc targets or biotinylated ECDs of GLP1R, GIPR, GCGR or PAC1R and labelled with corresponding antibodies simultaneously for 20 min whereas for the sorting rounds thereafter, cells were first pre-incubated with the target for 20 min and then labelled with corresponding antibodies for additional 20 min. Fluorescein isothiocyanate (FITC)-conjugated anti-c-Myc antibody (Immunology Consultants Laboratory, CMYC-45F, 1:50 dilution) was used to label miniproteins displayed on yeast cells. Phycoerythrin (PE)-conjugated anti-FLAG antibody (BioLegend, 637310, 1:15 dilution) was used to detect FLAG-tagged CXCR4, OXTR and CCR5 nanodisc targets while streptavidin-phycoerythrin (SAPE) (Thermo Fisher Scientific, S866, 1:15 dilution) was used to detect biotinylated ECDs of GLP1R, GCGR, GIPR and PAC1R. 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 and binding affinity (SC_{50}) estimated for GLP1R and PAC1R from next-generation sequencing (NGS) read counts[8]. All FACS data were analyzed in FlowJo v10.4.

For approaches involving NK1R, yeast EBY100 library cultures were grown in SDCAA media at 30°C while shaking at 200 rpm. Cultures were induced in SGCAA medium at 30°C while shaking at 200 rpm at an initial optical density of 0.5 (0.5×10^7 cells/mL). The following day, cells were washed with PBS + 0.5% BSA + 0.05% LMNG. For the first two rounds of sorting, cells were

incubated with the biotinylated, detergent-solubilized NK1R and labelled with corresponding staining reagents simultaneously for 30 min at room temperature whereas for the sorting rounds thereafter, cells were first pre-incubated with the target for 30 min and then labelled with corresponding antibodies for additional 30 min. Alexa Fluor 647-conjugated anti-c-Myc antibody (BioLegend, clone 9E10, 626810, 1:100 dilution) was used to label miniproteins displayed on yeast cells and streptavidin-Alexa Fluor 488 conjugate (Thermo Fisher Scientific, S32354, 1:200 dilution) was used to detect binding to biotinylated target. After three rounds of enrichment, affinity sorting using varying concentrations (100 pm - 1 μM) of the target was performed in a fourth round. The final sorting pools of the library were sequenced using Illumina NextSeq/MiSeq sequencing. All FACS experiments were performed using a SONY SH800.

Yeast SC_{50} estimation from FACS and Next Generation Sequencing

In order to quantify the level of binding for each of the designs, a previously described SC_{50} estimation was used[8]. The NGS data was used to quantify the composition of each sorting pool, noting what proportion of the entire pool each design represented. Using the ratio of the proportions between a child and parent pool along with the fraction of the cells collected on the FACS machine, the fraction of cells containing each design that survived each sort was calculated. For each design in each sort, the probability that the data could be generated from a binder with SC_{50} X was assessed for all X from fM to mM. These probabilities were multiplied together for all sorts and the interval where the probability of the SC_{50} was greater than 0.002 was reported to give a value resembling a 99.8th percent confidence interval. We also created the SC_{50}-RelativeEnrichment (SC_{50}RE). While the work from Cao and colleagues opted to eliminate designs carried by passenger plasmids with the use of minimum cell cutoffs, this method was proven to be not completely effective[8]. In fact, methods using enrichment (proportion_of_final_pool/proportion_of_expression_pool) do not suffer from passenger plasmids. We thus developed a method to include the power of enrichment to identify designs that were likely carried by passenger plasmids.

The SC_{50}RE represents the enrichment of a given design relative to the most enriched design with SC_{50} equal-to or worse-than the given design. In this way, if two designs claim to have the same SC_{50}, but one of them is 100x more enriched than the other, it is likely that the lesser-enriched design is from a passenger plasmid. In order to calculate the SC_{50}RE, one must choose which sorts to compare enrichment values from. We chose to use all non-avidity sorts where the sorting concentration is at least 10-fold greater than the predicted SC_{50}. These sorts were chosen because real binders should be saturating under these conditions and have robust enrichment (versus sorts with 100-folder concentration lower than the predicted SC_{50} where the predicted collection fraction is < 1% and the enrichment largely based on collection noise). For each design and sort that fit the previous criteria, the most-enriched design with SC_{50} equal-to or worse-than the current design was identified, and the ratios of their enrichment calculated (current_design_enrichment/most_enriched_enrichment). The lowest value obtained in any single sort was assigned as the current design's SC_{50}-RelativeEnrichment.

Mathematically, we can write the SC50-RelativeEnrichment for design i as

$$SC_{50}RE_i = min\{ E_{i,p} / E_p^{max} \; for \; p \; such \; that \; sort\_concentration_p > SC_{50,i} \times 10 \}$$

where for each pool p, $E_{i,p} = prop_{p,i} / prop_{expr.,i}$ is the enrichment compared to the expression pool and $E_p^{max} = max_j\{E_{p,j}\}$ is the largest enrichment of any design in pool p.

Designs with SC_{50}RE lower than 1/50 were suspected to be from passenger plasmids. Such a design would have an Enrichment 50x lower than another design with similar or worse SC_{50}.

Protein Purification of PAC1R ECD

The ECD of PAC1R, comprising residues 20-140 of isoform 1, was secreted from HEK293 cells as a fusion protein with an N-terminal His6-SUMOstar tag and a C-terminal (G4S)2-Avi-tag to facilitate purification and subsequent biotinylation. The culture supernatant was collected and subjected to affinity chromatography using a Ni Sepharose Excel column (Cytiva) to enrich the PAC1R protein. Following the affinity purification, the immobilized protein was biotinylated using a specific biotin ligase, and the N-terminal His6-SUMOstar tag was cleaved with SUMO protease to yield the native PAC1R extracellular domain. Subsequently, the protein was further purified by SEC using a Superdex S200 column (Cytiva) to separate the monomeric form of PAC1R from residual contaminants. The purified PAC1R ECD was concentrated to 1.6 mg/mL and stored in a buffer containing 20 mM HEPES, 150 mM NaCl, pH 7.4, at -80°C until required for downstream applications.

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 (PBS with 0.1% BSA) and then resuspended to a concentration of 2.5×10^6 cells/mL in PBSA containing Alexa Fluor 647-conjugated anti-c-Myc antibody (Cell Signaling Technology, clone 9B11, 2233) (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[18]. 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, Alexa 647-conjugated anti-c-Myc antibody (Cell Signaling Technology, clone 9B11, 2233, 1:100 dilution) was added to the mixture 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 SDCAA and grown for 1-2 days. The yeast were then induced in SGCAA 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 (approximately 600 yeast cells) on SDCAA plates and grown for 2 days. Individual clones were inoculated in 1 mL of liquid SDCAA media for 1-2 days, and subsequently induced for 1-2 days in 1 mL of SGCAA 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 Alexa 647-conjugated anti-c-Myc antibody (Cell Signaling Technology, clone 9B11, 2233, 1:100 dilution). 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:

For MRGPRX1 and PTH1R, GFP reporter vectors were generated by cloning full-length human MRGPRX1 (UniprotKB, Q96LB2) or a truncated sequence of PTH1R (UniProtKB, Q03431, amino acid residues 27-491 lacking the signal peptide and the C-terminal tail) into a lentiviral entry vector encoding an N-terminal mIGK 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-BsmBI-BFP-Myc-P2A-blast) using NEBridge Golden Gate Assembly Kit (BsmBI-v2) (New England Biolabs, E1602L). For PAC1R, a GFP reporter vector was generated by cloning a truncated sequence of PAC1R (UniprotKB, P41586, amino acid residues 21-468 lacking the signal peptide) into a lentiviral entry vector encoding an N-terminal HA signal peptide, FLAG tag, and GFP followed by a 2A peptide and blasticidin selection marker (pLenti/HA_SP-FLAG-eGFP-BsmBI-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-BsmBI) 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 ID: 2B98) and mCherry, followed by a BsmBI 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[46,47]. 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 min 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, PTH1R and PAC1R GFP reporters 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.[48]. 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[46], 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, PTH1R or PAC1R reporter and binder. A total of 9-10 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[46]. 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 GFP-GPCR with mCherry-binder. 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[49]. To screen for miniprotein antagonism, SK-N-MC cells (listed in the ICLAC register of misidentified cell lines) endogenously expressing CGRPR were used and a concentration-response curve of endogenous αCGRP was first generated. The G_s-mediated cAMP accumulation was measured in a final volume of 40 uL. The stimulation buffer containing 0.5 mM 3-isobutyl-1-methylxanine (IBMX, Sigma-Aldrich) was used for serial dilutions of tested ligands. Approximately 10 µL 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 αCGRP incubation for an additional 30 min under the same conditions. Concentration-response curves of miniprotein antagonists were either normalized to 1 µM of αCGRP (set to 100%) and fitted to three-parameter non-linear regression curves or normalized to maximal αCGRP response (set to 100%) and fitted

using Global Schild/Gaddum equation in GraphPad Prism v8 or v10. Screening for antagonism of CXCR4 binders was performed by measuring cAMP inhibition in cells stably expressing CXCR4. 3,500 cells per well in 10 µL were mixed with 5 µL of 4X CXCL12 and forskolin (5 µM final concentration), respectively. The reaction mixture was incubated for 30 min at 37°C and then 10 µL of cryptate-labeled cAMP and cAMP d2-labeled antibody, respectively, were added. CXCR4 antagonists were measured similarly to CGRPR antagonists. Binders were first pre-incubated (4X, 5 µL) with 5 µL of 3,500 cells for 30 min at 37°C followed by addition of an EC_{80} of CXCL12 (4X, 5 µL) and forskolin (4X, 5 µL). Concentration-response curves of miniprotein antagonists were either normalized to 1 µM of CXCL12 (set to 100%) and fitted to three-parameter non-linear regression curves or normalized to the maximal CXCL12 response (set to 100%) using Global Schild/Gaddum equation in GraphPad Prism v8. Intracellular cAMP levels were also measured using the GloSensor assay[50]. In brief, HEK293T cells were transfected with 3.5 µg of receptor (CXCR4 and CCR5) and 3.5 µg of F22 plasmid (Promega, E2301). After 14-16 h of transfection, cells were trypsinized and seeded in 96-well plates at a density of 200,000 cells per well in assay buffer (20 mM HEPES pH 7.4, 1X Hank's Balanced Salt Solution (HBSS) and 0.5 mg mL^{-1} D-luciferin (GoldBio, LUCNA-1G) and incubated for 1 h 30 min at 37 °C. The cells were further incubated with 1 µM concentrations of binder dCX1_001 for 30 min at room temperature. Basal luminescence was measured for 3 cycles in the FLUOstar Omega plate reader (BMG Labtech). After basal reading, 5 µM of forskolin was added to the cells and luminescence was measured for 8 cycles until the signal stabilized. This was followed by the addition of CXC12 or CCL5 at indicated concentrations. Luminescence was recorded for 11 cycles. Responses from cycles 5 to 10 were averaged, data normalized to the maximal CXCL12 or CCL5 response (set to 100%) and fitted to three-parameter non-linear regression curves GraphPad Prism v10.

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) and pcDNA3.1 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 maximal ratio and incubated for 15 min 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 25 mM HEPES, 0.1% w/v BSA (Sigma-Aldrich) and 0.5 mM IBMX, pH 7.4) and incubated for 30 min 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 HTRF kit (CisBio). The plates were read on a PHERAstar plate reader (BMG LABTECH). Values were converted to cAMP concentrations using a cAMP standard curve. Peptide agonist data were normalized to the vehicle control and the

maximal response of endogenous ligand for each receptor (set to 100%). Concentration-response curves of miniprotein antagonists were either normalized to 100 nM of endogenous peptides, (set to 100%), and fitted to three-parameter non-linear regression curves or normalized to the maximal peptide agonist response (set to 100%), and fitted using the Global Schild/Gaddum equation in GraphPad Prism v10. For site-directed mutagenesis, wild type CLR and N-terminally tagged CD33-FLAG RAMP1 in pcDNA3.1 were used as templates. Ala mutagenesis of selected residues either in RAMP1 or CLR were carried out using Quikchange™ site-directed mutagenesis using a single oligo targeted to the sense strand. Sequences were confirmed by whole plasmid sequencing, provided by Plasmidsaurus (https://plasmidsaurus.com). Mutant constructs (either mutant RAMP1 with WT CLR, or mutant CLR with WT RAMP1) were transiently transfected and the ability of dC2_049 and dC2_050 to inhibit αCGRP mediated cAMP responses were assessed as described above. In all experiments the WT CGRPR (WT CLR/WT RAMP1) were included as a control. Concentration-response curves were normalized to maximal αCGRP response (set to 100%) and fitted using Global Schild/Gaddum equation in GraphPad Prism v10.

HEK293A cells at 70-80% confluency were plated in 10 cm dishes and were transiently transfected with 5µg DNA plasmid per dish (GLP-1R with a 2x cMyc N-terminal epitope tag in pEF/V5/Dest, wild type GCGR pcDNA3.1 or GIPR in pcDNA3.1 using a DNA:PEI Max ratio of 1:6. After approx. 24 hours, cells were harvested and seeded into poly-D-lysine (Sigma-Aldrich) coated 96 well plates (Corning) at a density of 15,000 cells/well. Cells were incubated overnight at 37°C in 5% CO_2, before assay. On the day of experiment, 48 h post-transfection, cells were incubated in cAMP stimulation buffer (phenol red–free DMEM containing 25mM HEPES, 0.1% w/v bovine serum albumin (Sigma-Aldrich) and 0.5 mM IBMX, pH 7.4) for 30min at 37°C in 5% CO_2. To assess antagonistic activity of tested minibinders at GLP-1R, GCGR, and GIPR, varying concentrations of minibinders were added to the cells for 5 min, followed by subsequent stimulation of a single concentration of endogenous peptide (GLP-1R: 10 nM GLP-1 (7-36)NH_2, GCGR: 0.3 nM Glucagon (1-29), GIPR: 1 nM GIP(1-42)) for 30 min at 37°C. A concentration response curve of each endogenous ligand in the absence of minibinder was also included in each assay. After a 30 min incubation, the stimulation was terminated with addition of ice-cold ethanol. Measurements of cAMP concentrations were performed as described previously, with cAMP detected using the Cisbio HTRF kit (Cisbio) and read on PHERAstar plate reader (BMG LABTECH). Values were converted to cAMP concentrations using a cAMP standard curve. Data were normalized to the maximal response of endogenous ligands for each receptor (set to 100%) and fitted to three-parameter non-linear regression curves in GraphPad Prism v8. 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 (Corning, 3570). The cells were incubated overnight for 16 h at 37°C. 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 an EC_{80} concentration 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 solution or growth medium was added to each well. After a 30-min 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 h at 37°C with 5% CO_2. Following treatment, 40 µL of Bio-Glo™ Luciferase Assay Detection Solution (Promega, G7941) was added

to each well to initiate the luminescent reaction. Luminescence was then measured using a SpectraMax iD5 Multimode Plate Reader (Molecular Devices). Relative light unit (RLU) values were calculated by subtracting background luminescence (media alone plus Luciferase Assay Detection Solution) from each sample. Responses were normalized as (test sample − vehicle control) / (positive control ligand − vehicle control), with 1 μM αCGRP set to 100%, and fitted to three-parameter nonlinear regression curves in GraphPad Prism v8. To measure activity of designs targeting GLP1R, BHK21/hGLP1R/CRE-luciferase (clone FCW467-12A/KZ10-1) cells were seeded at a density of 5,000 cells per well in white, opaque 384-well plates (Revvity, 6007689) using growth media. The following day, growth medium was replaced with assay medium (DMEM, 10% FBS, 1% penicillin-streptomycin), into which miniproteins were diluted and added for equilibration. After a 30-min incubation at 37°C, 15 pM semaglutide (in-house produced) was added, followed by a 4 h incubation at 37°C. Subsequently, the cells were lysed, and a luciferase substrate was added using the Steady-Glo® kit (Promega, E2520). Luminescence was measured using the EnVision 2104 plate reader with ultrasensitive luminescence detection (Revvity). Concentration-response curves of miniprotein antagonists were normalized to maximal response of semaglutide (set to 100%) and fitted to three-parameter non-linear regression curves in GraphPad Prism v8.To assess pharmacological activity of miniproteins against NK1R, a CRE-Luc reporter gene assay was used. The assay utilizes luciferase expression under the control of a cAMP response element (CRE) to monitor NK1R signaling, as NK1R is known to signal through G_s, leading to an increase in cAMP production via adenylyl cyclase activation. CHO-K1/NK1R/CRE-Luc cells were seeded at a density of 10,000 cells per well in white, opaque 384-well plates (Greiner 781080) using culture media. The following day, miniproteins were diluted and added to the cells. For antagonists, cells were pre-incubated with miniproteins for 60 min before substance P (Tocris, 1156) at an EC_{80} concentration was added. For both agonists and antagonists, cells were stimulated with substance P or miniproteins for 3 h at 37°C and 5% CO_2 before a cell-lytic Steady-Glo® luciferase reagent (Promega, E2520) was added to the cells. After incubating for 20 min at room temperature, luminescence was measured using the EnVision 2104 plate reader with ultrasensitive luminescence detection (Revvity). To determine EC_{50} values of substance P and miniproteins, concentration-response curves were normalized to the maximal response of substance P (set to 100%) and fitted to three-parameter non-linear regression curves in GraphPad Prism v8. Data for miniprotein antagonists were normalized to the maximal response of substance P (set to 100%).

For calcium mobilization assays, CHO-K1/MRGPRX1 β-arrestin cells 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 min at 37°C. For agonism, cells were incubated with 10 μL of HBSS / 20 mM HEPES in the dark for 30 min at room temperature. The agonistic activity of ligands was measured on a FLIPR Tetra (MDS) or a FlexStation 3 (MDS). 10 μL of the sample (prepared at 4X concentration in HBSS / 20 mM HEPES) was added to the cells 5 sec before calcium mobilization was monitored for 2 min. The peak fluorescence intensity following ligand stimulation was divided by the average baseline signal per well and expressed as $\Delta F/F_0$ by subtracting 1. These values were normalized to the maximal response of BAM(8–22) (set to 100%) and fitted

using four-parameter nonlinear regression in GraphPad Prism v10. For antagonist measurements, after dye loading, 10 µL of the sample (prepared 3X) was added and cells were incubated for 30 min at room temperature. 10 µL of an EC_{80} of the agonist, prepared in HBSS / 20 mM HEPES, was added to the cells 5 sec before calcium mobilization was monitored for 2 min. The peak fluorescence intensity following ligand stimulation was divided by the average baseline signal per well and expressed as $\Delta F/F_0$ by subtracting 1. These values were normalized to the maximal response of BAM (8–22) (set to 100%) in GraphPad Prism v10.

Miniprotein activity against PTH1R was assessed by measuring β-arrestin-2 recruitment through complementation of split-β-galactosidase. CHO-K1 cells express an enzyme donor (ED) fragment of the β-galactosidase fused to the C-terminus of PTH1R while the enzyme acceptor (EA) fragment is fused to β-arrestin-2. In the event of activation, β-arrestin-2 is binding to the C-terminus of PTH1R leading to assembly of ED and EA forming an active β-galactosidase. After addition of substrate, chemiluminescence signal can be measured. For the assay, PathHunter® cells were seeded in culture media at a density of 5,000 cells per well in white, opaque 384-well plates (Greiner, 781080). The following day, cells were pre-treated with varying concentrations of miniproteins for 60 min at 37°C and 5% CO2. For stimulation, cells were treated with 400 or 200 pM human PTH (DiscoverX, 4011474) at 37C° for 90 min. Subsequently, PathHunter Detection Mix (DiscoverX, 93-0001) was added to lyse the cells and generate a chemiluminescence signal. Cells were incubated with PathHunter Detection Mix for 60 min before the signal was captured by using the Nexus EnVision plate reader (Revvity). Data were normalized to 200 nM or 400 nM of PTH1 and fitted to four-parameter non-linear regression curves in the GraphPad Prism v10.

Recruitment of β-arrestin-1 or β-arrestin-2 at chemokine receptors was assessed using a NanoBiT-based assay51,52. HEK-293T cells were transfected with 3.5 µg of C-terminal SmBiT-tagged CXCR4, CXCR7 or CCR5 together with N-terminal LgBiT-tagged β-arrestin-1 or β-arrestin-2. After 16–18 h, cells were harvested, resuspended in an assay buffer (5 mM HEPES, pH 7.4, 1× HBSS, 0.01% BSA, 10 µM coelenterazine (GoldBio, CZ05), and seeded into 96-well plates at 100,000 cells/well density. The plates were incubated at 37°C for 1.5 h before luminescence measurements. To measure ligand-induced responses, basal luminescence was recorded for three cycles using a FLUOstar Omega plate reader (BMG Labtech), followed by stimulation with the indicated ligand concentrations. Luminescence was then monitored over 11 cycles, and responses from cycles 5–10 were averaged. Initial screening of the CCR5 binders was carried out in both agonist and antagonist modes. For agonist mode of CCR5 binders, cells were treated with either 1 µM binder or CCL5 as control. Cells were pre-incubated with 1 µM binder for 30 min to assess β-arrestin-1 recruitment in antagonist mode, followed by stimulation with 100 nM of CCL5. Responses were baseline-corrected and normalized to the maximal CCL5 response (set to 100%). For CXCR4 selectivity measurements, 1 µM of dCX1_001 was pre-incubated for 30 min followed by varying concentrations of CXCL12 or CCL5. Responses were baseline-corrected and normalized as fold change with unstimulated condition set to 1. Data were fitted to three-parameter non-linear regression curves in Graphpad Prism v10. Concentration-response assay for binders (dCC1_005, dCC1_042) selected from the initial screening was carried out in the antagonist mode. Cells were pre-incubated with the indicated binder concentrations for 30 min, followed by CCL5 stimulation (EC_{80} = 0.316 µM). Data were normalized to the maximal CCL5 response (set to 100%) and fitted to three-parameter non-linear regression

curve in GraphPad Prism v10. Concentration-response assay for dCC1_002 selected from the initial screening was carried out in the agonist mode. Cells were incubated with the indicated binder concentrations for 30 min. Buffer-treated cells served as control. Data were normalized to the maximal CCL5 response (set to 100%) and fitted to three-parameter non-linear regression in Graphpad Prism v10.

β-arrestin-2 recruitment BRET1 assays at MRGPRX1 were performed as previously described, with minor modifications53. HEK293 cells were co-transfected with plasmids encoding RLuc8-tagged (C-terminal) MRGPRX1, GRK2, and mVenus–β-arrestin-2 at a 1:1:5 ratio. Culture conditions, plating density, incubation times, washing steps, and plate handling matched the BRET2 protocol. Coelenterazine-h was used as the luciferase substrate, and emissions were recorded at 475 nm (donor, RLuc8) and 535 nm (acceptor, mVenus) on a PHERAstar FSX plate reader. BRET1 ratios (acceptor/donor) were normalized to maximal BAM (8-22) response (set to 100%), and fitted to three-parameter non-linear regression curves in GraphPad Prism v10.

Ligand-induced Gα protein activation at chemokine receptors was measured using a NanoBiT-based G-protein dissociation assay[54]. In this system, LgBiT-tagged Gα subunit and SmBiT-tagged Gγ2 (harboring C68S mutation) subunit were co-expressed with the untagged Gβ1 subunit and N-terminal FLAG-tagged receptor (CXCR4 or CCR5). Upon ligand stimulation, the resulting decrease in luminescence was monitored. Briefly, HEK293T cells were transfected with a plasmid mixture consisting of 1 μg LgBiT-Gα, 4 μg Gβ1, 4 μg SmBiT-Gγ2 (C68S) and 1 μg CXCR4 or CCR5 (in pcDNA3.1). After 14–16 h of transfection, cells were trypsinized and resuspended in NanoBiT assay buffer (5 mM HEPES, pH 7.4, 1× HBSS, 0.01% BSA, 10 μM coelenterazine (GoldBio, CZ05), and seeded in 96-well plates (100,000 cells/well). After 1.5 h of incubation at 37 °C, basal luminescence was recorded for three cycles using a FLUOstar Omega plate reader (BMG Labtech). Cells were subsequently stimulated with indicated ligands, and luminescence signals were monitored for 11 cycles. The luminescence values at 10 min were basal-corrected and analysed using GraphPad Prism v10. For CXCR4 selectivity measurements, cells were pre-treated with 1 μM of dCX1_001 followed by stimulation with varying concentrations of CXCL12. Responses were baseline-correct, normalized to the unstimulated condition (set to 100%) and fitted to three-parameter non-linear regression curves in GraphPad Prism v10. For initial screening of CCR5 binders, Gα protein dissociation was measured in a single point in both agonist and antagonist modes. For antagonist mode, cells were pre-incubated with 1 μM binder for 30 min, followed by stimulation with 100 nM of CCL5. Responses were baseline-corrected and normalized to the unstimulated condition (set to 100%). For agonist mode, cells were treated with 1 μM binder, with CCL5 as a control, and the response was normalized to the unstimulated condition (set to 100%). For functional validation of the dCX1_001 structure, the predicted ligand–receptor binding mode was analyzed to delineate the interaction interface. This analysis indicated that residues interacting with dCX1_001 are distinct from those involved in canonical CXCL12–CXCR4 binding. Based on this prediction, eleven residues (E26A, C28A, D118A, E179A, D187A, Y190A, D193A, D262A, E268A, E277A, and H281A) were selected and subsequently mutated using a site-directed mutagenesis kit (New England Biolabs, E0554S). All constructs were verified by sequencing (Macrogen). Cells expressing CXCR4-WT or mutant receptors were pre-incubated with dCX1_001 for 30 min, while buffer-treated cells served as controls. Cells were then stimulated with CXCL12 at concentrations of 100 nM and 10 nM. Based on this initial screening,

seven mutants (E26A, E179A, Y190A, D193A, E268A, E277A, and H281A) that retained responsiveness to CXCL12 were selected for further characterization. These mutants were subjected to final validation experiments, in which cells were pre-incubated with increasing concentrations of dCX1_001 for 30 min, followed by stimulation with CXCL12 at EC_{80} concentrations. All responses were baseline-corrected, normalized to the unstimulated condition (set to 100%) and fitted to three-parameter non-linear regression curves in GraphPad Prism v10.

Gα protein dissociation BRET2 experiments at MRGPRX1 were performed as previously described, with minor adaptations[21,22]. HEK293T cells stably expressing human MRGPRX1 were generated by co-transfecting a PiggyBac transposase with a transposon plasmid encoding MRGPRX1 and a puromycin-resistance cassette flanked by inverted terminal repeats; cells with stable integrants were selected and maintained in DMEM containing 10% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin, and 2 µg/mL puromycin at 37 °C in 5% CO_2. For each experiment, 2–3 × 10^6 cells were plated per 6-cm dish and, 24 h later, transfected with TRUPATH Triple Gαq (Gαq-RLuc8, Gβ3, and Gγ9-GFP2 encoded on a single plasmid) at 1,000 ng per dish using TransIT-2020 (Mirus; 3 µL per µg DNA). After 24 h, cells were detached with 0.05% trypsin-EDTA, resuspended in DMEM + 1% dFBS, and seeded into poly-L-lysine–coated, white, clear-bottom 384-well plates at 1 × 10^4 cells per well. Twenty-four hours later, plates were backed with reflective white film, medium was removed, and wells were washed once with 20 µL phosphate-buffered saline containing PBS with 0.1% BSA. Serial dilutions of miniproteins (top concentration up to 30 µM, final) were prepared in PBS with 0.1% BSA. To each well, 20 µL miniprotein solution and 10 µL buffer containing coelenterazine-400a (5 µM, final) were added. After 15 min at 37 °C, emissions at 395 nm (donor, RLuc8) and 510 nm (acceptor, GFP2) were recorded on a PHERAstar FSX plate reader. BRET2 ratios were calculated as acceptor/donor (GFP2/RLuc8), normalized to the maximal response of BAM (8-22) (set to 100%) and fitted to three-parameter non-linear regression curves in GraphPad Prism v10.

Genome-wide profiling of off-target GPCR activation was performed with PRESTO-Tango, implemented with minor adjustments[23,55,56]. Poly-L-lysine–treated, white, clear-bottom 384-well plates were seeded with HTLA cells (10,000 cells in 40 µL DMEM + 1% dialyzed FBS). After approximately 6 h, cells were transfected with 20 ng plasmid DNA per well and allowed to express receptors overnight. On the following day, 10 µL of test compounds (miniproteins prepared in DMEM + 1% dFBS) was added to each well to achieve the screening concentration (3 µM). Plates were incubated overnight. For detection, medium was removed and 20 µL/well Bright-Glo diluted in assay buffer (HBSS, 20 mM HEPES, pH 7.4) was dispensed; after 20 min at room temperature in the dark, luminescence was recorded on a SpectraMax L microplate reader. Each plate included the dopamine D2 receptor stimulated with 100 nM quinpirole as positive control. Each receptor–compound condition was measured in quadruplicate at 0 and 3 µM miniprotein (final). Responses are reported as fold change relative to 0 µM..

The activity of binders targeting OXTR was evaluated in an IP1 accumulation assay (Cisbio). Stable OXTR-expressing cells (20,000 - 30,000 per well) were incubated overnight at 37°C in a 384 white well plate. The following day, media was removed and cells were equilibrated for 15 min at 37°C in 1X stimulation buffer. To assess agonism, binders were incubated for 30 min

followed by addition of d2-labeled IP1 and anti-IP1-cryptate prepared in a lysis detection buffer. To assess antagonism, binders were pre-incubated for 30 min at 37°C followed by addition of either an EC_{50}-EC_{80} of OXT to derive IC_{50} values or varying concentrations of OXT to derive *p*A2. After an incubation for 1 h at room temperature, the HTRF signal (ratio 665/620 nm) was measured on a Neo2 plate reader (Agilent). Concentration-response curves of miniprotein antagonists were either normalized to maximal OXT response (set to 100%) and fitted to three-parameter non-linear regression curves or normalized to the maximal OXT response (set to 100%) using Global Schild/Gaddum equation in GraphPad Prism v8.

Receptor surface expression assay

To assess cell surface expression of the CXCR4, CXCR7, and CCR5, whole cell surface ELISA was performed (**Supplementary Fig. 69**), as previously described[57]. Briefly, transfected cells were seeded at a density of 0.2 million cells per well 24 h post-transfection in 24-well plates (pre-coated with 0.01% poly-D-Lysine) and incubated for 24 h at 37 °C in a CO_2 incubator. After 24 h, the media was removed by aspiration and the cells were washed once with 400 µL 1X TBS. The cells were fixed by adding 300 µL 4% (w/v) PFA/paraformaldehyde, and were incubated for 20 min on ice. Excess PFA was removed by washing the cells thrice with 400 µL 1X TBS. The wells were blocked by incubating with 200 µL 1% (w/v) BSA prepared in 1X TBS for 1 h and 30 min. 200 µL HRP-conjugated anti-FLAG M2 antibody (Sigma-Aldrich, Clone M2 monoclonal, A8592, 1:10,000 dilution) was added to the wells and incubated for an additional 1 h and 30 min. To develop the signal, 200 µL of TMB/tetramethylbenzidine (Thermo Fisher Scientific, 34028) was added to the wells and incubated till the development of adequate color. The reaction was quenched by transferring 100 µL of the solution to a 96-well plate already containing 100 µL of 1 M H_2SO_4, and absorbance was measured at 450 nm using a multimode plate reader (Victor *X*4-Perkin-Elmer). To estimate the number of cells in each well, excess TMB was first removed by washing once with 400 µL 1X TBS, and the cells were incubated for 20 min in 0.2% (w/v) Janus Green B stain (Sigma-Aldrich, 201677). Excess stain was removed by repeated washing with Milli-Q water, and signal was developed by adding 800 µL 0.5 N HCl to each well. 200 µL of this colored solution was transferred to a 96-well plate and the absorbance was recorded at 595 nm. The signal was normalized by dividing the reading obtained at 450 nm with the reading obtained at 595 nm. Receptor surface expression was normalized to empty pcDNA vector (mock)-transfected cells (set to 1), and analyzed in the GraphPad Prism v10.

Surface plasmon resonance spectroscopy (SPR)

<u>Kinetic measurements for GLP1R binders</u>

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, BR100839) 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 for 120 sec and allowing them to dissociate for

300 sec. 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 (v3.2) or Biacore Insight Evaluation software (v5.0.18.22102). Data were analyzed in GraphPad Prism v10. 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. Sensorgrams were aligned, extracted using the Biacore Insight Evaluation Software (v5.0.18.22102) and analyzed in GraphPad Prism v10.

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 approximately 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 0.0128 nM, 0.064 nM, 0.32 nM, 1.6 nM, 8 nM, 40 nM, 200 nM, 1000 nM to 5000 nM 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[58]. 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, 94-001F) using baculovirus as reported previously[59].

The purification of CGRPR was conducted as described previously[59]. In brief, after removal of tags from CLR by addition of 3C protease (10 µg/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[60] 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 Titan Krios G4 microscope (Thermo Fisher Scientific) 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 Glacios microscope (Thermo Fisher Scientific), 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 (either v3.6.0.6389 or v3.9.1.8206 for the Krios or Glacios instruments, respectively).

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 for import into RELION 5.0[61]. Patch (4 x 4 patches) motion correction[62] was performed with MotionCor3. Contrast transfer function (CTF) estimation was performed with CTFFIND (version 4.1.14)[63]. 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 v4.6.0, 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 v4.6.0. 428,662 selected particles were refined using RELION 5.0 and subjected to particle polishing[64,65]. Using cryoSPARC v4.6.0, 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)[64,65]. 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, imported into RELION 5.0[61] and motion corrected using MotionCor3[62] (4 x 4 patch tracking) and had their CTF parameters estimated using CTFFIND 4.1.14[62]. Micrographs with robust CTF information beyond 5 Å were selected for further processing. Particles were picked using crYOLO (1.9.9[66]), 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[67], 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 v4.6.0, 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[68] followed by rounds of manual refinement in Coot v0.9.6[69] and real-space refinement in Phenix[70] v1.19.2 or v1.21.2.

Generation of MRGPRX1 constructs for cryo-EM

For the expression of MRGPRX1–Gα_q protein complex, the full-length DNA of human MRGPRX1 (UniProtKB: Q96LB2) was subcloned into a modified version of pFastBac1 (Invitrogen) baculovirus expression vector. Specifically, 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 Gα_q protein, the same mini-Gα_qiN 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–Gα_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, 103590166). 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, 94-001F) 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 PE-conjugated anti-baculovirus gp64 monoclonal antibody (Thermo Fisher Scientific, clone AcV1, 12-6991-82). 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 min 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) (Anatrace), 0.05% (w/v) cholesteryl hemisuccinate (CHS) (Anatrace), 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 min 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 SEC 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 (New England Biolabs) 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. 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[19,20]. In brief, the scFv16 gene was cloned into a modified pFastBac1 vector (Invitrogen), 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 (Sigma-Aldrich). Media chelating agents were quenched by the addition of 1 mM nickel chloride (Sigma-Aldrich) and 5 mM calcium chloride (Sigma-Aldrich) 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 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 sec to 6 sec. 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 (Thermo Fisher Scientific). 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 v4.2 using a multi-shot array as previously described. Data were collected at an exposure dose rate of approximately 15 electrons per pixel per sec as recorded from counting mode. Images were recorded for approximately 2.7 sec in 60 subframes to give a total exposure dose of approximately 50 electrons per $Å^2$. All subsequent classification and refinement steps were performed with cryoSPARC v4.6.0 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 (by Fourier shell correlation (FSC) using the 0.143-Å cutoff criterion) after local contrast transfer function (CTF) refinement and post-processing in cryoSPARC v4.6.0. Alternative post-sharpening was performed using deepEMhancer v0.17 and EMready v1.0.

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 ChimeraX v1.9 and Phenix v1.20.1. The models were manually adjusted in Coot v0.9.8.5 and then subjected to several rounds of real-space refinement refinement in Phenix v1.20.1. The model statistics were validated using Molprobity. Structure figures were prepared by either ChimeraX v1.9 or PyMOL v3.1.

Purification and cryo-EM analysis of CXCR4-binder complexes

For CXCR4-dCX1_001 structural studies, we first used the CXCR4-3 construct, which stabilizes the inactive receptor state as previously described[71,72]. A construct encoding an HA signal

sequence, FLAG tag, 10×His tag, and a TEV protease cleavage site at the N terminus of CXCR4-3 (HA-FLAG-His_{10}-TEV-CXCR4) was cloned into pFastBac1. CXCR4 was expressed in *Trichoplusia ni* High Five cells (Thermo Fisher Scientific, B85502) infected with high-titer recombinant baculoviruses and purified using established protocols[66]. The minibinder and maxibinder gene was synthesized (Integrated DNA Technologies) and cloned into the pET-28a(+) vector (Novagen) with an N-terminal TEV protease cleavage site, followed by superfolder GFP and a C-terminal His_{10} tag. Both binders were expressed in *E. coli* and purified by standard Ni^{2+}-affinity chromatography. The minibinder was first tested in a complex with CXCR4 by cryo-EM. However, due to its small size, the density corresponding to the binder could not be resolved for reliable particle alignment. Therefore, we proceeded with the maxibinder for subsequent structural studies. Initially, purified CXCR4 was incubated with a ten-fold molar excess of maxibinder. Complex formation was confirmed by analytical SEC and SDS-PAGE. Cryo-EM grids prepared with the concentrated complex on UltrAuFoil 1.2/1.3 grids showed extensive particle aggregation. Despite this, 19,620 movies were collected and processed in CryoSPARC v4.5.3, with 2D classification revealing a small subset of classes (~2%) showing binder engagement, but the particle number was insufficient for reliable 3D reconstruction. Given that the CXCR4-maxibinder complex was observed by SDS-PAGE, the complex was likely destabilized at the grid air–water interface. To improve complex stability, a fusion construct was generated by linking the maxibinder to the N-terminus of CXCR4 with a $(GSGSG)_4$ linker (maxibinder-linker-CXCR4). The fusion protein was successfully purified, as confirmed by SEC and SDS-PAGE. Cryo-EM grids prepared with the concentrated complex on UltrAuFoil 1.2/1.3 and 0.6/1 grids again exhibited severe aggregation. A smaller dataset (6,345 movies) was collected and processed, but 2D classification yielded only ~3% of particles with maxibinder engagement to CXCR4, which was still insufficient for 3D reconstruction. Reducing the sodium concentration decreased aggregation during SEC but had no effect on grid behaviour, where aggregation persisted. Although particle numbers increased slightly, they remained inadequate for structure determination. Attempts with the milder detergent GDN resulted in even greater aggregation during SEC, indicating it was unsuitable for CXCR4-maxibinder purification. Overall, SEC and SDS–PAGE confirmed complex formation, and multiple strategies including fusion constructs, varied detergents, altered sodium concentrations, and different grid types, were tested. Nonetheless, particles consistently aggregated on grids, and the proportion of well-behaved particles remained far below that required for 3D reconstruction, despite rare 2D classes showing binder engagement. Aggregation was not observed when just CXCR4 was put on a grid.

cDNA encoding full length wild-type CXCR4 was cloned in pVL1393 vector harboring N-terminus HA signal followed by FLAG-tag and M4 sequence and HRV-3C protease site. Sf9 cells (Expression Systems, 94-001F) were infected with CXCR4 expressing baculovirus at a density of 2.2 million cells per mL and were cultured for 68-72 h following which the cells were pellet down at 5,000 rpm for 15 min at 4℃ and flash frozen in liquid nitrogen and stored at -80℃ till further use. For receptor purification, CXCR4-expressing cell pellets were thawed and sequentially dounced homogenized in hypotonic (20 mM HEPES, pH7.4, 10 mM $MgCl_2$, 20 mM KCl, 1 mM PMSF and 2mM benzamidine) and hypertonic buffer (20 mM HEPES, pH7.4, 1M NaCl, 10 mM $MgCl_2$, 20 mM KCl, 1 mM PMSF and 2mM benzamidine) followed by centrifugation at 20,000 rpm for 20 min at 4℃. Subsequently, the pellets were resuspended in the solubilization buffer (20 mM HEPES, pH7.4, 450 mM NaCl, 1 mM PMSF and 2mM benzamidine) and the receptor was solubilized in 0.5% LMNG, 0.1% CHS and 2 mM iodoacetamide with constant tumbling for 2 h at 4℃. Following this, the lysate was diluted in a dilution buffer (20 mM HEPES, pH7.4, 2.5 mM

$CaCl_2$, 1 mM PMSF, 2 mM benzamidine), followed by spinning at 20,000 rpm for 20 min at 4℃. The lysate was then subjected to filtration through 0.45 µ bottle-top filter to clarify the supernatant. The lysate was then loaded on gravity flow columns containing M1-anti-FLAG beads, pre-equilibrated with low-salt buffer (20 mM HEPES, pH7.4, 150 mM NaCl, 2.5 mM CaCl2, 0.01% LMNG and 0.01% CHS) at a constant flow rate at 4℃. Subsequently, the non-specifically bound proteins were washed with a series of 3 washes of low-salt buffer alternated with 2 washes of high-salt buffer (20 mM HEPES, pH7.4, 350 mM NaCl, 2.5 mM CaCl2, 0.01% LMNG). The bound protein was competitively eluted with an elution buffer comprising 250 µg/mL FLAG-peptide in 20 mM HEPES, 150 mM NaCl, 2 mM EDTA and 0.01% LMNG. The eluted receptor was then subjected to cysteine blocking by incubating the eluate with 2 mM iodoacetamide at 10 min intervals and excess iodoacetamide was quenched with 2 mM L-cysteine. All the buffers were supplemented with 1 µM dCX1_001 and while 10 1 µM dCX1_001 was added in the elution buffer. dCX1_001-bound CXCR4 was then concentrated in a 100 KDa MWCO Cytiva concentrators at 3500 rpm at 4℃ to the SEC injectable volume and finally the complex was separated in a Superose 6 Increase™ (10/300GL), equilibrated with the SEC buffer (20 mM HEPES, pH7.4, 100 mM NaCl, 0.00075% LMNG, 0.0001% CHS, 0.00025% GDN). The complex was confirmed by running the SEC-peak fractions on 12% SDS-PAGE. The complex fractions were pooled together and concentrated in a 100 KDa MWCO cytiva concentrator to the final concentration of 17 mg/mL and flash frozen in liquid nitrogen and stored at 4℃ till further use.

For sample preparation, 3 µL of purified dCX1_001-CXCR4 complex was applied onto glow-discharged Quantifoil holey carbon grids (Au, R1.2/1.3) and vitrified in liquid ethane (-181°C) using a Vitrobot Mark IV (Thermo Fisher Scientific) at 4°C and 100% humidity. Cryo-EM movies were acquired on a Titan Krios microscope (Thermo Fisher Scientific) (300 kV) equipped with a Gatan K3 direct electron detector. Images were collected automatically using EPU software v3.8.1.7603 in counting mode at a nominal magnification of 165,000x (pixel size 0.53 Å) and defocus range of 0.8-1.8 µm, with a total dose of 75 $e^-/Å^{-2}$ fractionated across 74 frames.

Data processing was performed using cryoSPARC v4.5.3[67] unless otherwise stated. Dose-fractionated movie stacks were subjected to beam-induced motion correction with Patch motion correction (multi), followed by contrast transfer function (CTF) parameter estimation using Patch CTF estimation (multi). 29,827 dose-weighted, motion-corrected micrographs were selected for further processing. A total of 8,346,134 particles were picked using the blob-picker subprogram (size range: 120-260 Å) and subjected to multiple rounds of reference-free 2D classification to remove distorted particles and ice contaminants. This yielded 1,094,439 particle projections with clear secondary structure features, which were then used for ab-initio reconstruction into three classes. The particle stack from the 3D class showing features of the dCX1_001-CXCR4 complex was processed with the rebalance orientations subprogram, yielding 18,396 particles. Non-uniform refinement of this stack produced a map with a global resolution of 3.28 Å (0.143 FSC cut-off). Local resolution of the final map was estimated with the LocRes module with half-maps provided as input.

The starting coordinates of CXCR4 were obtained from the model of trimeric CXCR4 in complex with REGN7663 Fab (PDB: 8U4S), while the coordinates of the binder dCX1_001 were generated from AlphaFold using the primary sequence as input. These initial models were docked into the EM map using UCSF Chimera v1.15[73], followed by the “all-atom-refine” sub-module within Coot v0.9.6[74]. The resulting model was refined with phenix.real_space_refine combined with iterative manual adjustments in Coot v0.9.6. The final refined model showed decent statistics with more

than 97% of the residues lying within the most favored region of the ramachandran plot as assessed with Molprobity implemented in Phenix 1.20.1[75]. A complete data processing pipeline including all steps followed during data processing is provided as **Supplementary Figure 32**). All figures related to the cryo-EM structure were prepared with either UCSF Chimera v1.15 or ChimeraX v1.5[76].

Targeted mutagenesis

Design models of mM1_068, mM1_060, and mM1_034, together with cryo-EM structures of mM1_068 and mM1_060, were relaxed with Rosetta FastRelax and used as inputs for mutagenesis. Positions were selected for mutation if the residue at that site (i) showed a change in solvent-accessible surface area (SASA) upon binding or (ii) had SASA < 20 Å² in the unbound binder. All 19 amino acid substitutions were introduced at each site and evaluated with Rosetta *cartesian_ddg* to compute changes in binding free energy (ΔΔG, REU) and binder energy (REU). Substitutions were classified as favorable if they reduced ΔΔG by ≤ −1.0 REU without increasing monomer energy above +5 REU, or maintained ΔΔG ≤ 0.0 REU while lowering monomer energy below −5 REU. Favorable mutations were visually inspected and combined into single, double, and triple variants based on stabilizing biophysical interactions, yielding 152 variants for mM1_068, 78 for mM1_034, and 58 for mM1_060, for a total of 288 variants.

Pharmacokinetic study

Protein production

The unmodified miniprotein (mC2_022) and Fc-fused variant (mC2_022-Fc9) for the pharmacokinetic (PK) study were cloned, expressed and purified by Genscript. mC2_022 was expressed in CHO-S (Genscript) as a fusion protein with an N-terminal signal peptide (MGWSCIILFLVATATGVHS) followed by a 6xHis-tag, a TEV protease cleavage site (ENLYFQG), and the mC2_022 sequence, and purified by HisTrap FF Crude purification followed by TEV cleavage, and HisTrap FF Crude purification again to remove the tag. mC2_022-Fc9 was expressed in TurboCHO-Express 2.0 (Genscript) as a fusion protein with an N-terminal signal peptide (MARAWIFFLLCLAGRALA) followed by the mC2_022 sequence, a GS-linker (GGSGSGGSGSGSGGS), and the Fc9 domain[11], and purified by Protein A purification. Protein purity (>95% by coomassie; endotoxin ≤0.1 EU/mg) and identity (LC-MS) was confirmed by Genscript.

Ethical approval

All animal procedures were reviewed and approved by the Lundbeck veterinarians and were conducted in accordance with applicable institutional and national guidelines for the care and use of laboratory animals (PPL 2023-15-0201-01579). A formal sample size calculation was not performed. Sample sizes were selected based on prior experience with similar rodent PK studies, and published literature, while using the minimum number of animals required to achieve the scientific objective. Animals were not randomized but were allocated based on experimental logistics. Experimenters were blinded to the group allocation during experiments and analysis. The sex of the animals was female. The age of the animals at the start of the experiments was approximately 12 weeks. The species and strain used were C57BL/6jCR (Charles River).Study design

The pharmacokinetic study was conducted to compare systemic exposure profiles of unmodified and Fc-fused miniproteins following intravenous administration in mice. 16 C57BL/6 mice (8 per group) received a single IV dose of 3 mg/kg of either compound in sterile DPBS (pH 7.2), with a dose volume of 5 mL/kg.Animal handling

Mice were acclimated for ≥ 5 days prior to dosing and housed under standard conditions with enrichment. Body weights were recorded prior to dosing and weekly if study duration exceeded 7 days. All animals were sacrificed after final sampling.

Dosing and sample collection

Dosing solutions were prepared with 50% excess to ensure adequate volume and concentration (final: 0.25 mg/mL). Each mouse received 0.125 mL of dosing solution. Blood samples (approximately 50 µL) were collected at eight timepoints post-dose (5, 15, 30, 60, 180, 360, 480, and 1440 min), with each mouse contributing four samples. Plasma was separated by centrifugation (4°C, 2500 x g, 10 min) and stored at –80°C.

Bioanalysis

Plasma samples (≥ 25 µL) were analyzed for compound concentration using validated bioanalytical methods (UPLC-MS/MS). Sampling flexibility of ±10% was permitted. Endotoxin levels in dosing solutions were confirmed to be <1 EU/mg.

UPLC-MS/MS workflow

Plasma exposure of mC2_022 and mC2_022-Fc9 was measured by liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS). For both analytes, calibration standards were prepared in mouse plasma in the range 50 – 10000 ng/mL. Plasma samples and calibration standards were digested using a Waters digestion kit (ProteinWorks Auto-eXpress High 3 step protocol). After the addition of acid in the quench step, acetonitrile was added, and the samples were centrifuged. Hereafter, the supernatant was transferred to a new plate and chlorobutane was added to extract acetonitrile from the aqueous phase liquid-liquid extraction. The aqueous layer was transferred to a new plate. The aqueous extract was injected onto a Waters ACQUITY UPLC peptide CSH C18 column. Chromatography was conducted using a low gradient slope with 0.1% formic acid and acetonitrile + 0.1% formic acid as mobile phases. The UPLC system was connected to a XEVO TQ Absolute MS. For both analytes the tryptic peptide VLEVGDVGTQR (ion $586.8^{2+} \rightarrow 960.5^{+}$) was used as a quantifier ion. Two UPLC-MS/MS runs were conducted using 1 µL and 10 µL as injection volume to evaluate ion suppression. No suppression was observed. Pharmacokinetic analysis was performed by non-compartmental analysis of the composite mean plasma concentration time courses for each test compound (Phoenix v8.5).

In vivo **pharmacology of CXCR4 antagonist dCX1_001**

Ethical approval

All experiments involving animals were conducted in accordance with the institutional guidelines set forth by the University of Washington. The University of Washington receives accreditation from the Association for the Assessment and Accreditation of Laboratory Animal Care International (AALAC) and all live animal work conducted at the university is in accordance with the Office of Laboratory Animal Welfare (OLAW) Public Health Assurance (PHS) policy, USDA Animal Welfare Act and Regulations, the Guide for the Care and Use of Laboratory Animals and

the University of Washington's Institutional Animal Care and Use Committee (IACUC) policies. The studies were approved by the University of Washington IACUC (Protocol No. 3108-01). 6-8 weeks old C57BL/6-based transgenic mice that contained the human CD46 genomic locus and provide CD46 expression at a level and in a pattern similar to humans ($hCD46^{+/+}$ mice) were described earlier[77]. No statistical methods were used to predetermine sample size. Sample sizes were chosen based on prior experience with similar in vivo experiments and published studies in the field. The expected effect sizes are large with relatively low variability, allowing detection of biologically meaningful differences with modest group sizes. Sample sizes were minimized in accordance with ethical guidelines for animal use. Animals were randomly assigned to experimental groups by a technician in a blinded manner. No stratification or blocking was used. Investigators were not blinded during experimental procedures. For the mouse study, age matched male and female mice were included and sourced from the same breeding colony.

HSPCs mobilization in mouse

HSPCs were mobilized in mice by one subcutaneous (s.c.) injections AMD3100 (5 mg/kg) or one s.c. injection of dCX1_001 (5 mg/kg). For HDAd injection, in addition, animals received dexamethasone (10 mg/kg, intraperitoneal (i.p.)) 16 and 2 hr before virus injection to blunt innate toxicity associated with intravenous HDAd injection. 45 min after AMD3100 or dCX1_001, animals were intravenously (i.v.) injected with virus vectors through the retro-orbital plexus (4×10^{10} viral particles per mouse). A second HDAd injection was given 30 min later. Peripheral blood was collected under sterile conditions at the indicated time intervals to assess mobilization efficiency. At 72 h post-injection, mice were euthanized by i.p. administration of an overdose of tribromoethanol, and the maximum volume of blood was collected via cardiac puncture.

In vitro colony forming potential

For each time point, 0.1 mL of blood was lysed with 1× RBC lysis buffer for 15 min at room temperature. Lysis was quenched by adding PBS without calcium and magnesium, followed by centrifugation at 500 × g. The cell pellet was resuspended in 3 mL of MethoCult GF M3434 medium (STEMCELL Technologies, 03434) and plated into 6-well plates without introducing air bubbles. Plates were incubated for 14 days, after which colonies were scored.

Hematological analysis

50 µL of whole blood in the EDTA tube were subjected to complete blood count analysis using Hemavet 950FS auto blood analyzer (Drew Scientific).

In vivo HSPCs transduction with HDAd vector

Helper-dependent adenoviral vectors expressing GFP under EF1a promoter (HDAds) were prepared as described previously[78]. Mice were pretreated with dexamethasone 16 h and 2 h prior to HDAd administration. Mobilization was induced using AMD3100 or dCX1_001 as described above, and HDAds were administered i.v. at 30 min and 1 h after mobilization. Two doses were given, each containing 4×10^{10} viral particles. Mice were euthanized 72 h later to analyze the transduction. Spleen and bone marrow were collected for flow cytometry analysis as described below.

Cytokine Bead Array

Proinflammatory cytokine release was analyzed in mouse plasma using BD Cytometric Bead Array (CBA) Mouse Inflammation Kit (BD Biosciences, 23-12720) according to the manufacturer's instructions. Briefly, plasma samples collected at indicated time points post injection and stored in -80'C until the analysis. Cytokine standards were incubated with a mixture of fluorescence-coded capture beads specific for individual cytokines, followed by incubation with a PE-conjugated detection antibody at room temperature. After incubation, samples were acquired on a BD FACSymphony A3, and at least 300 events per bead population were collected. Cytokine concentrations were determined by comparison to standard curves generated for each analyte using CBA analysis software.Tissue processing

Blood: Fifty to one hundred microliters of whole blood was lysed with 1× RBC lysis buffer for 15 min at room temperature. Lysis was quenched by adding PBS without calcium and magnesium, followed by centrifugation at 500 × g. The cell pellet was resuspended in 100 µL of FACS buffer (Thermo Fisher Scientific, 00-4222-26).

Spleen: Single-cell suspensions were prepared by gently dissociating spleens through a 70 µm cell strainer. Red blood cells were lysed with 1× RBC lysis buffer as described above.

Bone marrow: Hind limbs were harvested and washed with PBS. Bone marrow cells were flushed from the bones using an insulin syringe filled with PBS, followed by red blood cell lysis with 1× RBC lysis buffer as described above.

Flow cytometry for LSK and lineage cells

LSK cells: 5 million cells were stained with a 10 µL of biotin-conjugated lineage (Lin) antibody cocktail (Miltenyi Biotec, 130-092-613) at 4°C for 30 min and washed with a FACS buffer. Subsequently, cells were stained with streptavidin-conjugated APC (Thermo Fisher Scientific, S868, 1:100 dilution), BV711-conjugated anti-c-Kit (BioLegend, clone 2B8, 105835, 1:50 dilution), and PE-Cy7-conjugated anti-Sca1 (Thermo Fisher Scientific, clone D7, 25-5981-82, 1:100 dilution) for 30 min at room temperature.

Lineage analysis: 1 million cells were stained with following antibodies at room temperature for 30 min: BV421-conjugated anti-CD45 (Thermo Fisher Scientific, clone 30-F11, 404-0451-82, 1:200 dilution), APC-conjugated anti-CD3 (BioLegend, clone 17A2, 100236, 1:50 dilution), PE-Cy7-conjugated CD19 (Thermo Fisher Scientific, eBio 1D3, 25-0193-82, 1:100 dilution) and BV711-conjugated Gr-1 (BioLegend, RB6-8C5, 108443, 1:50 dilution). Unbound antibodies were removed by washing with a FACS buffer, and samples were analyzed immediately using a FACSymphony A3 flow cytometer (BD Biosciences).

Vector copy number analysis

Vector copy number analysis was done as described earlier[79]. In brief, total genomic DNA was isolated using Purelink Genomic DNA isolation kit as per the manufacturer's instructions (Thermo Fisher Scientific, K18200). 9.6 ng of DNA was used for vector copy number analysis using primers binding to GFP (FWD: TCGTGACCACCCTGACCTAC, REV: GGTCTTGTAGTTGCCGTCGT) qPCR was performed using the Power SYBR™ Green PCR Master mix. (Thermo Fisher Scientific, 4367659. Each reaction was run in triplicates. Serial dilutions of purified HDAd-GFP viral DNA were used as a standard curve. VCN is calculated as A total of 9.6 ng DNA (9,600 pg/6 pg/cell = approximately 1,600 cells) was used for a 10 µL reaction.

Statistical analysis

In vivo pharmacological data of dCX1_001 were analyzed using two-sided two-way ANOVA or two-sided Kruskal Wallis, with Sidak's and Dunn's multiple comparisons tests, (GraphPad Prism v8), respectively. *In vitro* pharmacological data of dCX1_001 mutants were analyzed using a two-sided unpaired t test with Welch's correction (GraphPad Prism v10). OPS-RD data were analyzed using a one-sided Kolmogorov-Smirnov test, with *p* values adjusted using the Benjamini-Hochberg procedure (SciPy 1.0).

Data availability

Cryo-EM maps and models have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) under the following accession codes: mM1_068/MRGPRX1, EMD-70205 and 9O7N; mM1_060/MRGPRX1, EMD-70230 and 9O8L; dC2_049/GCRPR, EMD-48385 and 9MM5; dC2_050/CGRPR, EMD-48424 and 9MNI; dCX1_001/CXCR4, EMD-68747 and 22XC. MetaGen-derived scaffolds are available at https://files.ipd.uw.edu/pub/metagen_scaffolds/8920_scaffolds.tar.gz. OPS-RD data are available at https://files.ipd.uw.edu/pub/GPCRs/OPS_RD/MRGPRX1.zip, https://files.ipd.uw.edu/pub/GPCRs/OPS_RD/PAC1R.zip and https://files.ipd.uw.edu/pub/GPCRs/OPS_RD/PTH1R.zip. The NGS data have been deposited in the NCBI database under the accession code PRJNA1451936 with raw sequencing reads available in the Sequence Read Archive. Source data are provided with this publication.

Code availability

The Rosetta macromolecular modeling suite (https://www.rosettacommons.org) is freely available to academic and non-commercial users. Commercial licenses for the suite are available via the University of Washington Technology Transfer Office. Code for using RFdiffusion v1.0 is available at https://github.com/RosettaCommons/RFdiffusion#binder-design. Scripts for ProteinMPNN, ProteinMPNN-FastRelax and AF2 with initial guess are available at https://github.com/dauparas/ProteinMPNN and https://github.com/nrbennet/dl_binder_design, respectively. Code for running partial and iterative partial RFdiffusion v1.0 are available at https://github.com/RosettaCommons/RFdiffusion#partial-diffusion and https://github.com/davidekim/ppi_iterative_opt, respectively. Code used for estimating yeast SC_{50} values from FACS and NGS data can be found at https://files.ipd.uw.edu/pub//estimate_affinity/estimate_affinity_from_ngs_sc50er.py. MotionCor3 for motion correction and the EPU_group_AFIS.py script are available at https://github.com/czimaginginstitute/MotionCor3 and (https://github.com/DustinMorado/EPU_group_AFIS), respectively.

Methods references

39. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

40. Qin, L. *et al.* Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **347**, 1117–1122 (2015).

41. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).

42. Hsia, Y. *et al.* Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021).

43. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).

44. 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).

45. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).

46. Feldman, D. *et al.* Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **17**, 476–512 (2022).

47. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17 (2019).

48. Feldman, D. *et al.* Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **17**, 476–512 (2022).

49. Muratspahić, E. *et al.* Design and structural validation of peptide–drug conjugate ligands of the kappa-opioid receptor. *Nat. Commun.* **14**, 8064 (2023).

50. Kumar, B. A., Kumari, P., Sona, C. & Yadav, P. N. GloSensor assay for discovery of GPCR-selective ligands. in *Methods in Cell Biology* vol. 142 27–50 (Elsevier, 2017).

51. Maharana, J. *et al.* Structural snapshots uncover a key phosphorylation motif in GPCRs driving β-arrestin activation. *Mol. Cell* **83**, 2091-2107.e7 (2023).

52. Dwivedi-Agnihotri, H. *et al.* An intrabody sensor to monitor conformational activation of β-arrestins. in *Methods in Cell Biology* vol. 169 267–278 (Elsevier, 2022).

53. Wang, Y. *et al.* Structures of the entire human opioid receptor family. *Cell* **186**, 413-427.e17 (2023).

54. Inoue, A. *et al.* Illuminating G-Protein-Coupling Selectivity of GPCRs. *Cell* **177**, 1933-1947.e25 (2019).

55. Fink, E. A. *et al.* Structure-based discovery of nonopioid analgesics acting through the α_{2A}-adrenergic receptor. *Science* **377**, eabn7065 (2022).

56. Singh, I. *et al.* Structure-based discovery of conformationally selective inhibitors of the serotonin transporter. *Cell* **186**, 2160-2175.e17 (2023).

57. Measuring surface expression and endocytosis of GPCRs using whole-cell ELISA. in *Methods in Cell Biology* vol. 149 131–140 (Elsevier, 2019).

58. Liang, Y.-L. *et al.* Cryo-EM structure of the active, Gs-protein complexed, human CGRP receptor. *Nature* **561**, 492–497 (2018).

59. Josephs, T. M. *et al.* Structure and dynamics of the CGRP receptor in apo and peptide-bound forms. *Science* **372**, eabf7258 (2021).

60. 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).

61. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

62. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

63. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

64. 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).

65. 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).

66. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).

67. 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).

68. 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).

69. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

70. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 531–544 (2018).

71. Wu, B. *et al.* Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science* **330**, 1066–1071 (2010).

72. Qin, L. *et al.* Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **347**, 1117–1122 (2015).

73. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

74. Emsley, P. & Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

75. Adams, P. D. *et al. PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

76. Pettersen, E. F. *et al.* UCSF ChimeraX : Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

77. Kemper, C. *et al.* Membrane cofactor protein (MCP; CD46) expression in transgenic mice. *Clin. Exp. Immunol.* **124**, 180–189 (2002).

78. Palmer, D. & Ng, P. Improved system for helper-dependent adenoviral vector production. *Mol. Ther.* **8**, 846–852 (2003).

79. Richter, M. *et al.* In vivo transduction of primitive mobilized hematopoietic stem cells after intravenous injection of integrating adenovirus vectors. *Blood* **128**, 2206–2217 (2016).

Acknowledgements

We acknowledge the use of BioRender for preparation of Fig. 1a-b,e (Norn C. (2026), https://BioRender.com/u6mjw6t, https://BioRender.com/w5y0f4s, and https://BioRender.com/7qlmy52) and Fig. 5c (Muratspahić, E. (2026), https://BioRender.com/i2nj5sj). We thank Luki Goldschmidt and Kandise VanWormer, respectively, for maintaining computational and wet laboratory resources at the Institute for Protein Design. We also thank Brian Trippe for assistance with the analysis of yeast display data. We thank Nilanjana Banerjee and Divyanshu Tiwari for expressing and purifying CXCR4 and Gokul Nair for assistance with pharmacological assays. We thank Ulla Wahlers for supporting protein production. We thank Marina Mohr for supporting operations at Skape Bio. We thank Iain Cheeseman and Princess I. Imoukhuede for kindly providing HeLa and HEK293/OXTR cells, respectively. We thank Anders Rudebeck for supporting computational resources at the BioInnovation Institute. We thank Mette M. Rosenkilde for discussions concerning pharmacology. We acknowledge Angeli Tongson, Jason Walters, Gordon Leung, Quishi Wang and Paolo Gonzales for supporting pharmacological characterization of MRGPRX1 designs and Esperanza Rivera de Torre for assisting in circular dichroism studies. We thank Rie Christensen for excellent technical in vivo support at Lundbeck. We also acknowledge our colleagues at the NNRCC. A.K.S. is a recipient of the Sonu Agrawal Memorial Chair from IIT Kanpur.

Funding

E.M. discloses support for the research of this work from Austrian Science Fund (FWF), Erwin Schrödinger program (J-4663). C.N., D.F., A.M.B., P.R.S., F.D. and E.P.T.H. disclose support for the research of this work from BioInnovation Institute Foundation (BII22SG1021010, BII24SG1021475, BII24SG1022030). C.N. discloses support for the research of this work from Novo Nordisk Foundation (NNF18OC0030446). D.B. and G.R.L disclose support for the research of this work from Howard Hughes Medical Institute (GR020267). E.M. and D.B. disclose support for the research of this work from Novo Nordisk (GR018355). D.B. discloses support for the research of this work from Defense Threat Reduction Agency (HDTRA1-21-1-0038), Microsoft (GF117374, Microsoft Protein Prediction Research), The Audacious Project at the Institute for Protein Design, (PG117878, PG117879, PG117866), The Nordstrom Barrier Institute for Protein Design Directors Fund (GF124659), The Open Philanthropy Project Universal Flu Vaccine Fund (GF129461). D.B., S.V.T. and G.R.L. disclose funding from The Open Philanthropy Project Improving Protein Design Fund (GF129460). D.B. and T.S. disclose funding from The Wu Tsai Protein Innovation Fund (GF151772). A.M. discloses funding support from Cancer Research Grand Challenge provided by Cancer Research UK (GR050755). D.B. discloses funding from Department of the Defense and Defense Threat Reduction Agency (HDTRA1-21-1-0007, GR013444). D.B. and J.Z.Z. disclose funding from National Institutes of Health's National Cancer Institute (R01CA240339, GR009231, K99-CA293001), and National Institutes of Health's National Institute on Aging (R01AG063845, GR009173). D.B. discloses support for the research of this work from the U.S. Department of Energy, Office of Science, including resources of the National Energy Research Scientific Computing Center (BER-ERCAP0022018). B.E.K. and B.L.R. disclose support for the research of this work from National Institutes of Health NIDA (R01DA055656) and the NIMH Psychoactive Drug Screening Program. P.M.S. and D.W. disclose

support for the research of this work from National Health and Medical Research Council of Australia Investigator (2025694, 2026300). J.B.S. discloses funding for the research of this work from NSF (2143160), Department of Defense (W81XWH-21-1-0891), NIH NIDCR (R21DE031436) and CureSearch for Children's Cancer award. The research on chemokine receptors and de novo protein design in A.K.S.'s laboratory is supported by a Senior Fellowship of the DBT Wellcome Trust India Alliance (IA/S/20/1/504916), ANRF (ANRF/ARG/2025/007393/LS), Department of Biotechnology (BT/PR53019/MED/30/2528/20) and Lady Tata Memorial Trust.

Author contributions

E.M., D.F., K.D., C.G.T., C.N. and D.B. conceptualized the study. D.E.K., T.S., M.Ba., I.S. and 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., J.E.S., C.N. and M.J. designed binders. A.N. and M.J. performed meta-analysis of designed binders. D.F. and C.N. conceived the OPS-RD HTS assay. D.F., A.M.B., P.R.S. and E.P.T.H. developed cell lines and performed the OPS-RD assay. E.M., X.Q., A.M.B., D.E.K., T.S., L.M., S.V.T., P.P., Y.R. and C.N. expressed and purified binders. E.M., X.Q., F.D., P.K., P.N.H.T., L.O., W.C., N.L., MBr., Y.W., L.A., J.H.V., S.M., K.S., A.D. and S.S. pharmacologically characterized binders. J.F. and L.H. performed biofloating assay. E.M., X.Q., J.Z.Z., A.M., L.T., G.R.L., I.G., D.K.V., L.D. and A.B. performed yeast display experiments. B.C. improved the yeast SC_{50} estimation method from FACS and NGS. J.C., B.P.C., M.J.B. and P.N.H.T. determined cryo-EM structure of CGRPR binders, performed the CGRPR mutagenesis and generated associated figures. Q.C., R.B. and M.G. determined cryo-EM structure of CXCR4 binders. B.L.R. and B.E.K. determined cryo-EM structure of MRGPRX1 binders. J.S., T.D. and X.Y. expressed and purified PAC1R and NK1R. E.P.T.H., L.S., J.J.W., J.F.B., C.B. and A.A.A. designed constructs with improved PK properties and conducted in vivo studies. K.V.K. performed in vivo studies with dCX1_001 antagonist of CXCR4. E.M. and C.N. drafted the work and prepared figures. E.M., D.F., D.E.K., X.Q., A.M.B., P.R.S., J.H.V., E.P.T.H., M.J., B.P.C., P.K., Q.C., J.F., S.M., A.D., S.S., K.V.K., N.L., M.Br., Y.W., K.D., L.D., A.B., J.G.E., J.B.S., A.L., A.K.S., P.M.S., B.L.R., B.E.K., D.W., C.G.T., C.N. and D.B. reviewed, edited and revised the work. E.M., P.H., K.D., J.G.E., J.B.S., A.L., A.K.S., P.M.S., B.E.K., B.L.R., D.W., C.G.T., C.N. and D.B. provided research funding and supervision.

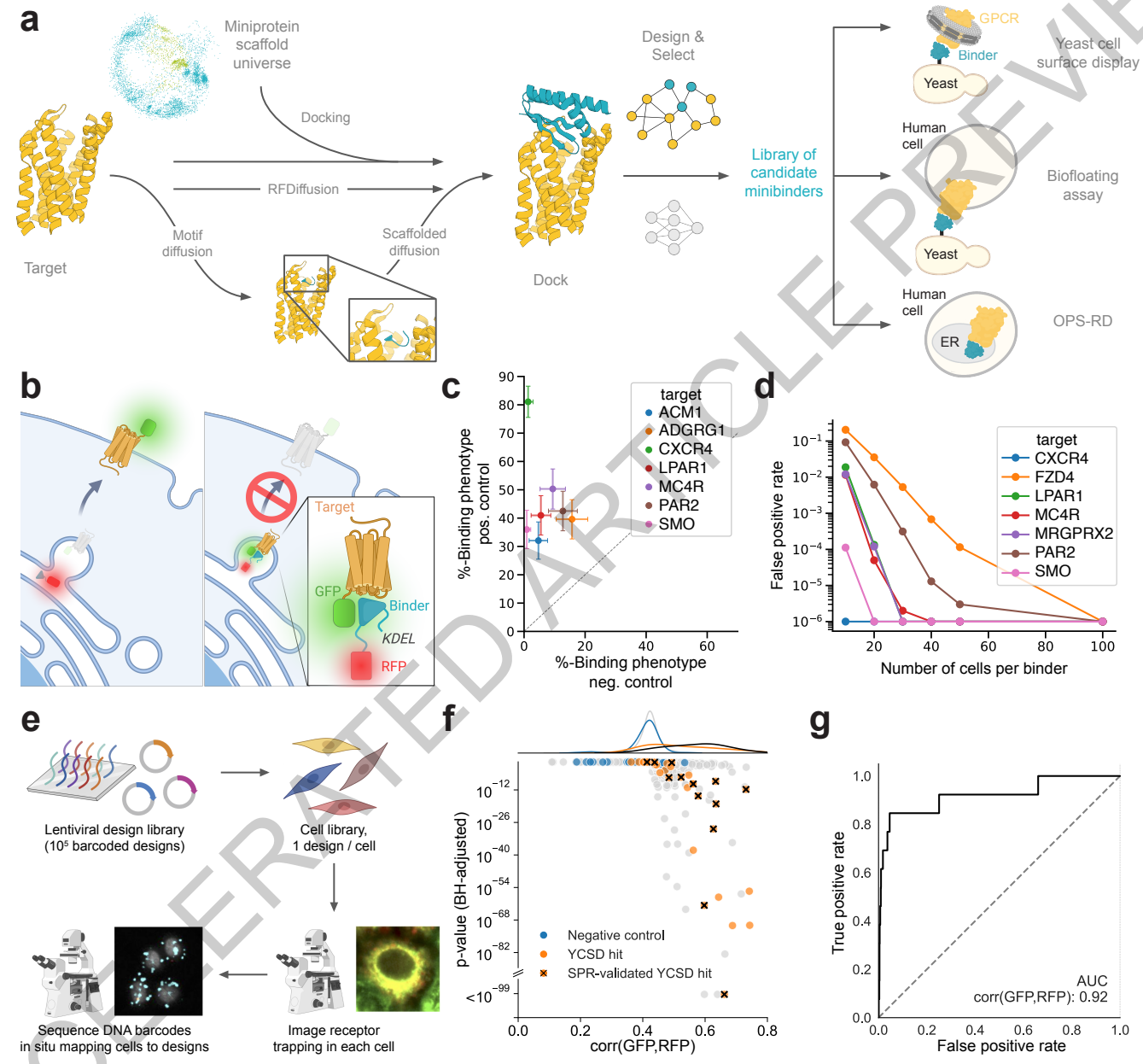
Competing interests

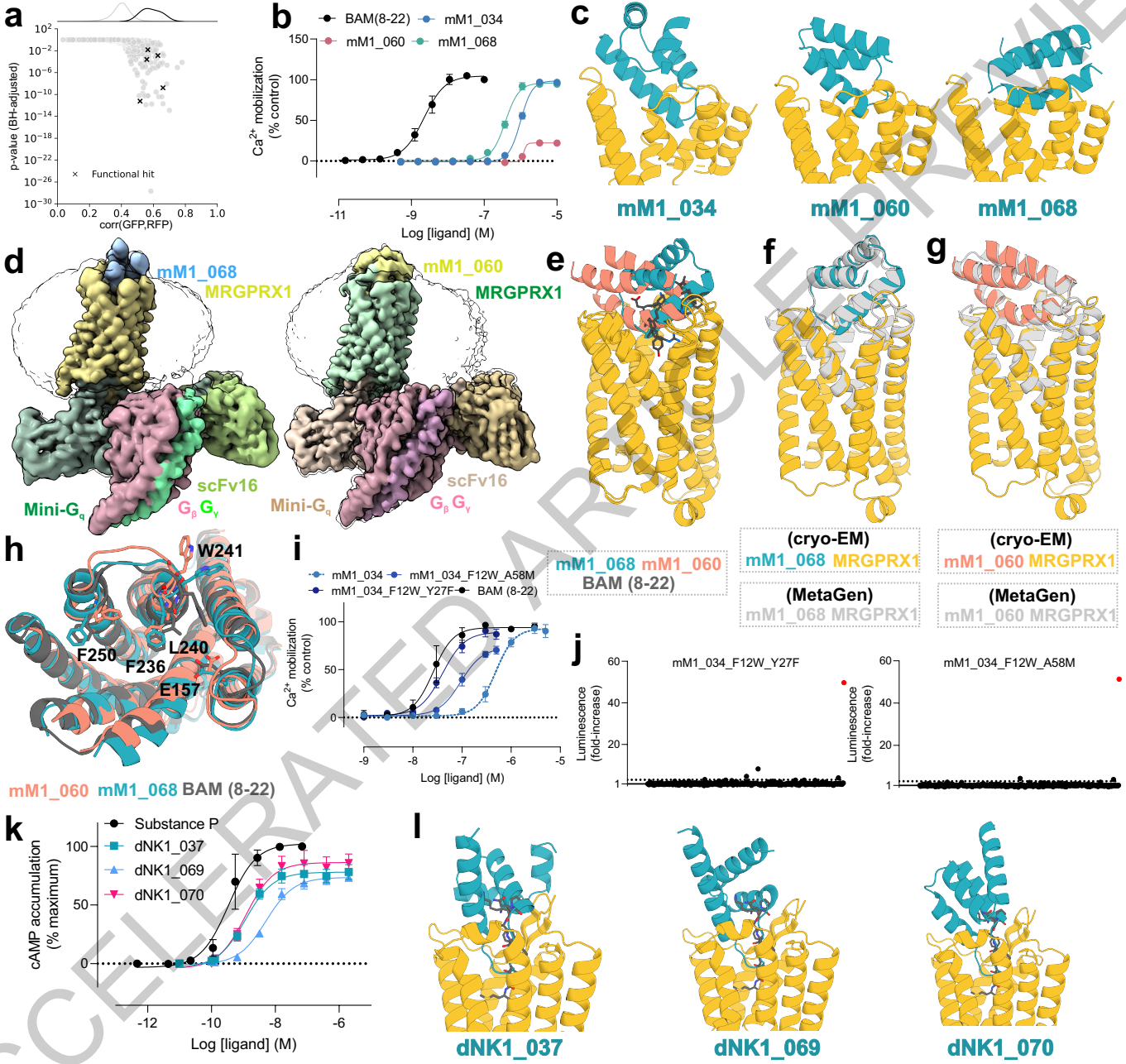
E.M., D.F., D.E.K., X.Q., A.M.B., P.R.S., J.H.V., E.P.T.H., M.J., F.D., J.F., K.D., N.L., M.Br., Y.W., L.J.S., J.B.S., C.N. and D.B. are listed as inventors or major contributors on records of innovation and associated provisional patent applications relating to discoveries described in this manuscript, at the University of Washington and Skape Bio. The Baker lab has received sponsored research funding from Novo Nordisk in support of the GLP1R, PAC1R, NK1R, and PTH1R research described in this manuscript. A.B., L.D., N.L, M.Br, J.S., K.D. and Y.W. were employees of Novo Nordisk at the time of the work, and some hold minor shareholdings as part of their employment. E.M., D.F., D.E.K., A.M.B., P.R.S., L.J.S., C.N., E.P.T.H., M.J, J.H.V., A.N., C.G.T., B.L.R. and D.B. are shareholders of Skape Bio ApS. Skape Bio is seeking to commercialize aspects of the research described in this manuscript. B.L.R. is a member of the S.A.B. of ImprintBio, XyloBio and Epiodyne. All other authors declare no competing interests.

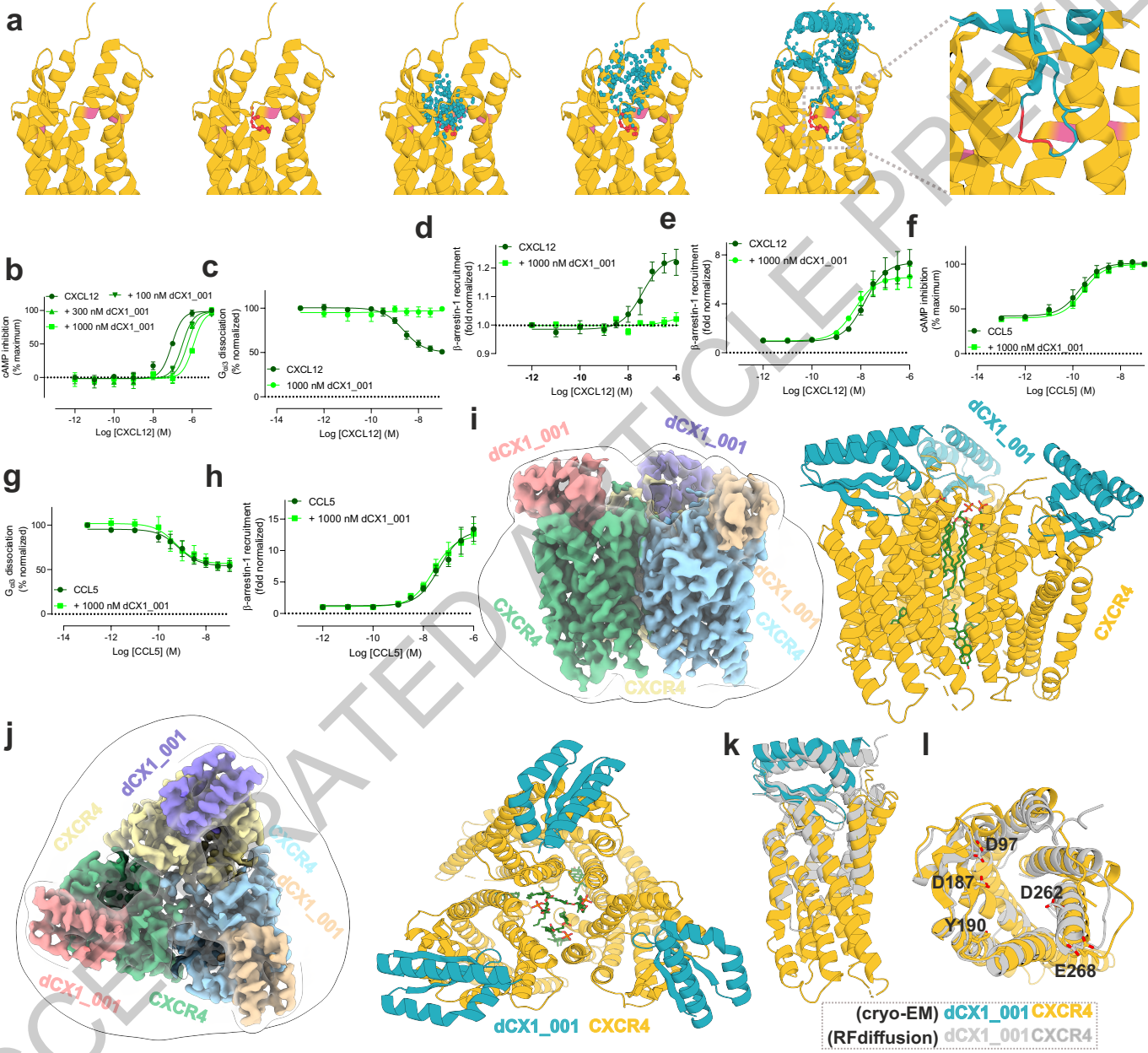
Extended Data Table 1. Cryo-EM data collection, refinement, and validation statistics for MRGPRX1-mM1_068 and MRGPRX1-mM1_060. The table summarizes data acquisition parameters, resolution estimates, and model quality statistics for the final refined MRGPRX1-miniprotein complexes.

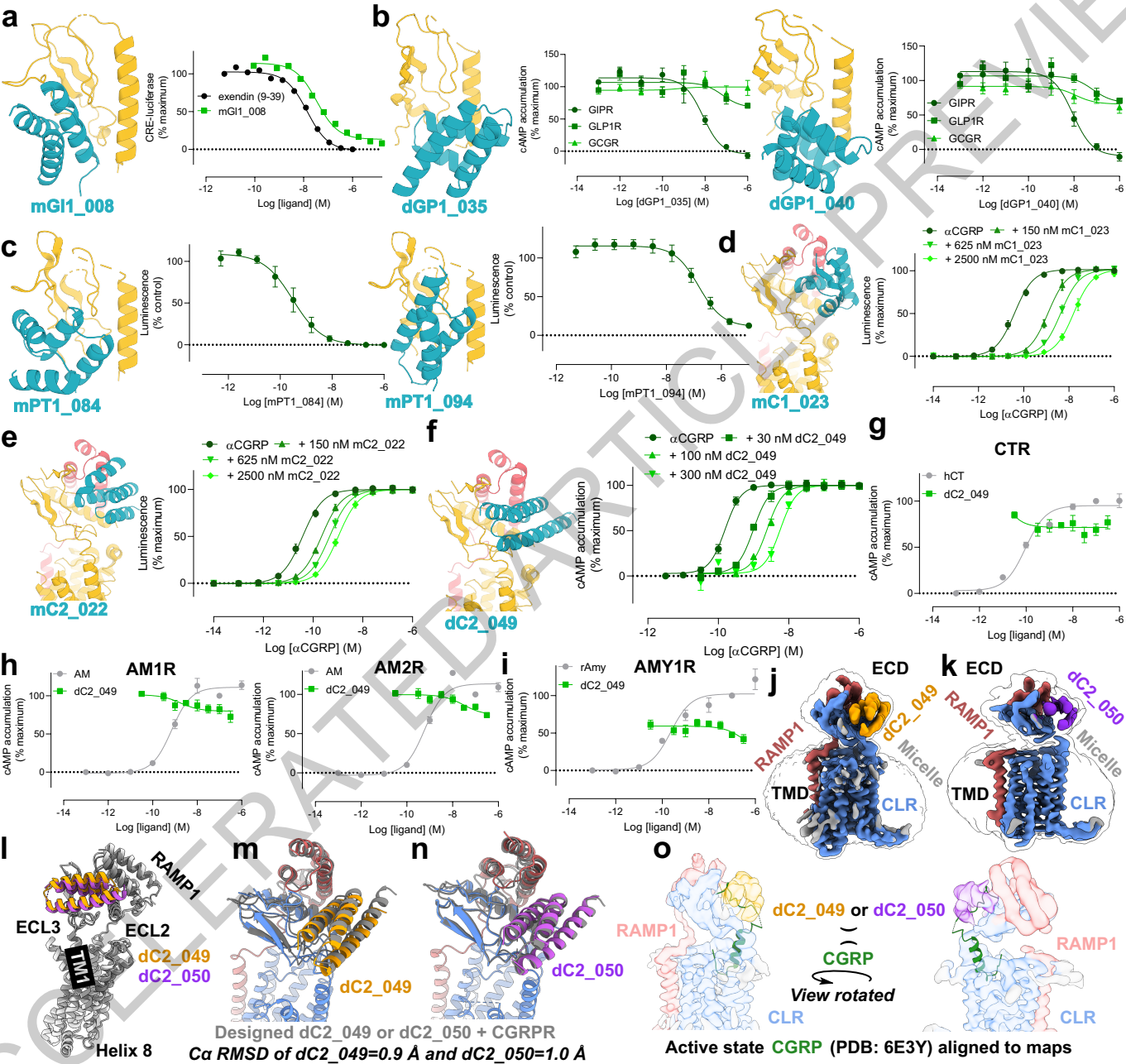
Extended Data Table 2. Cryo-EM data collection, refinement, and validation statistics for CXCR4-dCX1_001. The table reports data collection parameters, map resolution metrics, and model validation statistics for the final refined CXCR4-miniprotein complex.

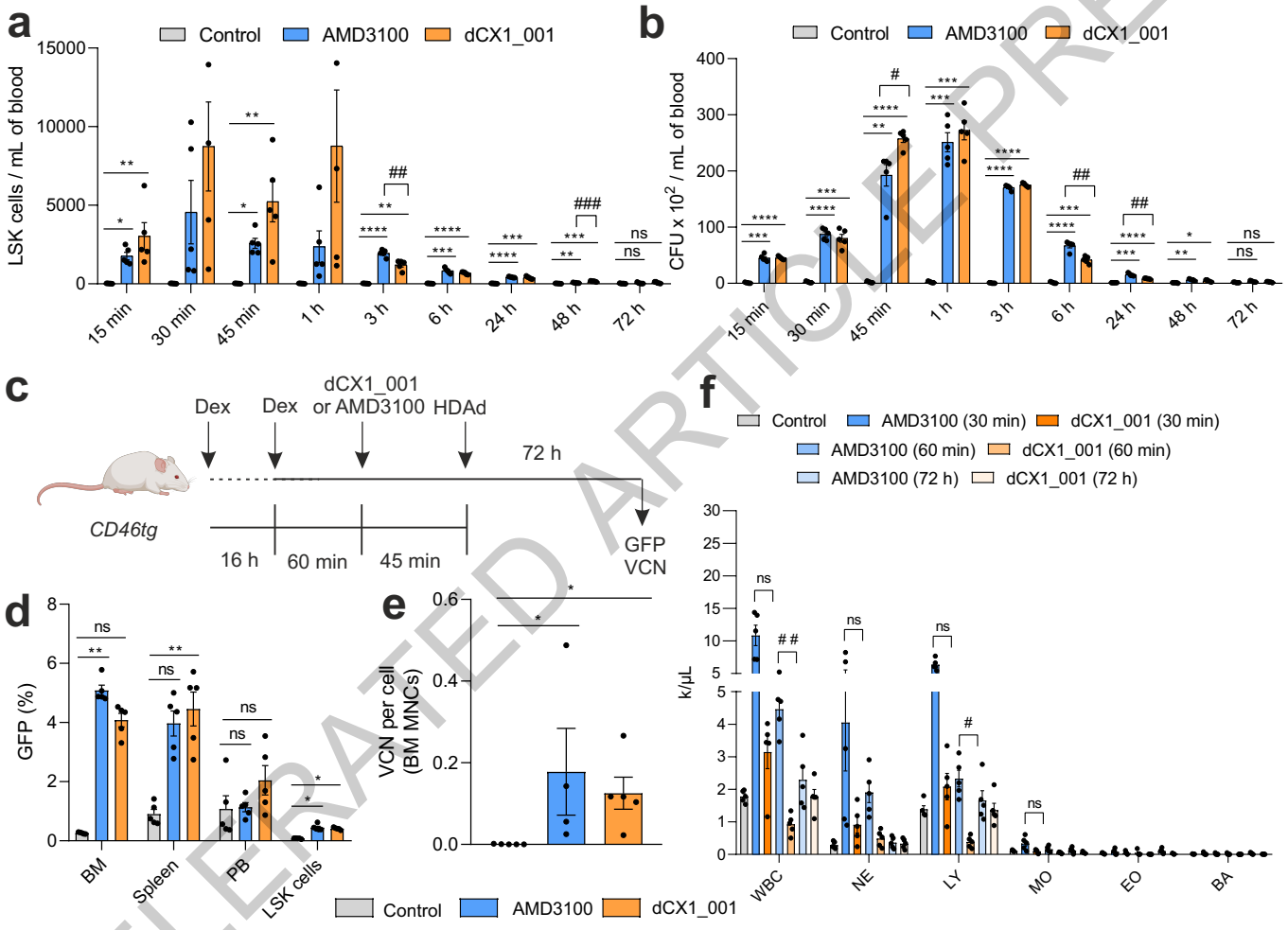
Extended Data Table 3. Cryo-EM data collection, refinement, and validation statistics for CGRPR-dC2_049 and CGRPR-dC2_050. The table provides data acquisition and refinement parameters together with resolution and model quality metrics for the final refined CGRPR-miniprotein complexes.











	MRGPRX1 mM1_068 EMD-70205 PDB: 9O7N	MRGPRX1 mM1_060 EMD-70230 PDB: 9O8L
Data collection and processing		
Magnification	45,000	45,000
Voltage (kV)	200	200
Electron exposure (e–/Å^2)	54.3	54.0
Defocus range (μm)	-1.0 to -1.6	-1.1 to -1.5
Pixel size (Å)	0.876	0.876
Symmetry imposed	C1	C1
Initial particle images (no.)	706,976	412,046
Final particle images (no.)	237,264	201,129
Map resolution (Å)	3.29	3.13
FSC threshold	0.143	0.143
Refinement		
Initial model used (PDB code)	8DWC	8DWC
Model composition		
Non-hydrogen atoms	8810	8700
Protein residues	1135	1138
Ligands	1	1
B factors (Å^2)		
Protein	99.3	83.5
Ligand	159.5	120.6
R.m.s. deviations		
Bond lengths (Å)	0.006	0.003
Bond angles (°)	1.230	0.578
Validation		
MolProbity score	2.00	1.63
Clashscore	24.29	9.38
Poor rotamers (%)	0.0	0.00
Ramachandran plot		
Favored (%)	97.4	97.3
Allowed (%)	2.6	2.7
Disallowed (%)	0.0	0.00

Extended Data Table 1

	CXCR4 dCX1_001 EMD-68747 PDB: 22XC
Data collection and processing	
Magnification	165,000x
Voltage (kV)	300
Electron exposure (e–/Å^2)	75
Defocus range (μm)	-0.8 to -1.8
Pixel size (Å)	0.53
Symmetry imposed	C3
Initial particle images (no.)	8,346,134
Final particle images (no.)	18,396
Map resolution (Å)	3.28
FSC threshold	0.143
Refinement	
Initial model used (PDB code)	8U4S
Model resolution (Å)	3.8
FSC threshold	0.5
Model composition	
Non-hydrogen atoms	8,623
Protein residues	1040
Ligands	D21: 3, CLR: 3
B factors (Å^2)	
Protein	73.75
Ligand	72.84
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.567
Validation	
MolProbity score	1.47
Clashscore	6.46
Poor rotamers (%)	0
Ramachandran plot	
Favored (%)	97.44
Allowed (%)	2.56
Disallowed (%)	0

Extended Data Table 2

	CGRPR dC2_049 EMD-48385 PDB: 9MM5	CGRPR dC2_050 EMD-48424 PDB: 9MNI
Data collection and processing		
Magnification	165,000	120,000
Voltage (kV)	300 kV	200 kV
Electron exposure (e–/Å^2)	50	50
Defocus range (μm)	-0.6 to -1.2	-0.6 to -1.4
Pixel size (Å)	0.75	0.86
Symmetry imposed	C1	C1
Initial particle images (no.)	3,942,967	4,782,642
Final particle images (no.)	285,458	482,106
Map resolution (Å)	3.18	4.06
FSC threshold	0.143	0.143
Refinement		
Initial model used (PDB code)	*Ab initio*	*Ab initio*
Model composition		
Non-hydrogen atoms	3946	3604
Protein residues	534	516
Ligands	0	0
R.m.s. deviations		
Bond lengths (Å)	0.003	0.004
Bond angles (°)	0.714	0.871
Validation		
MolProbity score	1.52	1.62
Clashscore	5.18	6.87
Poor rotamers (%)	0	0
Ramachandran plot		
Favored (%)	96.36	96.39
Allowed (%)	3.64	3.61
Disallowed (%)	0	0

Extended Data Table 3

nature portfolio

Corresponding author(s): Edin Muratspahic, Christoffer Norn

Last updated by author(s): Apr 15, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The Rosetta macromolecular modeling suite (https://rosettacommons.org/) is freely available to academic and non-commercial users. Commercial licenses for the suite are available through the University of Washington Technology Transfer Office. RFdiffusion v1.0, ProteinMPNN, AlphaFold2, partial and iterative partial RFdiffusion v1.0, and Rosetta were used for designing and filtering designs. Cryo-EM data were collected on Titan Krios G4, Glacios, Titan Krios, and Talos Arctica microscopes (Thermo Fisher Scientific), using automated acquisition in EPU v3.6.0.6389, v3.9.1.8206 and v3.8.1.7603, and SerialEM v4.2.
Data analysis	Data were analyzed using Python v3.1, GraphPad Prism v8 and v10, NumPy v1.26.4 and pandas v2.2.2 and visualized using seaborn v0.13.2 and matplotlib v3.9.2. Structures were visualized and images were generated using PyMOL v3.1, UCSF Chimera v1.15 and ChimeraX v1.5 and v1.9. FACS data were analyzed using FlowJo v10.4. Cryo-EM data were analyzed using cryoSPARC v4.5.3 and v4.6.0, Coot v0.9.6 and v0.9.8.5 and Phenix v1.19.2, v1.21.2 and v1.20.1. Pharmacokinetic (PK) data were analyzed using Phoenix v8.5.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Cryo-EM maps and models have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) under the following accession codes: mM1_068/MRGPRX1, EMD-70205 and 9O7N; mM1_060/MRGPRX1, EMD-70230 and 9O8L; dC2_049/GCRPR, EMD-48385 and 9MM5; dC2_050/CGRPR, EMD-48424 and 9MNI; dCX1_001/CXCR4, EMD-68747 and 22XC. MetaGen-derived scaffolds are available at https://files.ipd.uw.edu/pub/metagen_scaffolds/8920_scaffolds.tar.gz. OPS-RD data are available at https://files.ipd.uw.edu/pub/GPCRs/OPS_RD/MRGPRX1.zip, https://files.ipd.uw.edu/pub/GPCRs/OPS_RD/PAC1R.zip and https://files.ipd.uw.edu/pub/GPCRs/OPS_RD/PTH1R.zip. The NGS data have been deposited in the NCBI database under the accession code PRJNA1451936 with raw sequencing reads available in the Sequence Read Archive. Source data are provided with this publication.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No research involving human participants was conducted in this study. Not applicable.
Reporting on race, ethnicity, or other socially relevant groupings	No research involving human participants was conducted in this study. Not applicable.
Population characteristics	No research involving human participants was conducted in this study. Not applicable.
Recruitment	No research involving human participants was conducted in this study. Not applicable.
Ethics oversight	No research involving human participants was conducted in this study. Not applicable.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For cryo-EM studies, the number of micrographs collected was determined by available microscope time and the requirement to obtain sufficient particle numbers for high-resolution 3D reconstruction. For in vitro pharmacological and biophysical binding assays, and OPS-RD, no statistical methods were used to predetermine sample size. Sample sizes varied depending on the assay, with the majority of experiments performed with at least three biological replicates or independent experiments, while a few preliminary assays were performed with fewer replicates. Exact numbers of biological replicates and independent experiments are reported in the figure legends and Supplementary Tables. For in vivo mouse experiments, no statistical methods were used to predetermine sample size. Sample sizes were chosen based on prior experience with similar experiments and established standards in the field and were sufficient to reproduce effects while using the minimum number of animals required to achieve the scientific objective. Exact numbers of mice used for in vivo experiments are provided in the figure legends.
Data exclusions	A small number of data points were excluded from some concentration–response curves due to technical artefacts at specific concentrations. No formal exclusion criteria were pre-established. Excluded data points are indicated as blanks in the Source data files.
Replication	The majority of experiments were performed with at least three biological replicates or independent experiments, while a few preliminary assays were performed with fewer replicates. Exact numbers of biological replicates and independent experiments are reported in the figure legends and Supplementary Tables. All experiments that were independently repeated, including in vitro pharmacological and biophysical binding assays, yielded consistent results. For in vivo mouse experiments, each condition was tested in a single experiment using the indicated number of animals, and results were consistent across animals within the experiment.

Randomization

Cryo-EM studies did not involve allocation of samples to experimental groups; therefore, randomization was not applicable. For in vitro pharmacological and biophysical binding assays, no randomization was performed because no group allocation was used. A subset of miniproteins was independently reproduced in in vitro pharmacological assays in a second laboratory, yielding consistent results.

For in vivo mouse experiments, two independent studies were performed: a pharmacokinetic (PK) study targeting CGRPR and an HSPC mobilization study targeting CXCR4. In the PK study, animals were allocated based on experimental logistics. In the HSPC study, animals were randomly assigned to experimental groups in a blinded manner by a technician.

Blinding

For cryo-EM and biophysical binding studies, blinding was not performed because data acquisition and analysis were based on automated, instrument-derived readouts without subjective assessment. For in vitro pharmacological assays, blinding was not performed because experiments did not involve group allocation.

In mouse experiments, blinding was not performed during experimental procedures in the HSPC mobilization study (CXCR4). In the pharmacokinetic (PK) study (CGRPR), experimenters were blinded to group allocation during experiments and analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Anti-c-Myc antibodies (Immunology Consultants Laboratory, CMYC-45F; BioLegend, clone 9E10, 626810; and Cell Signaling Technology, clone 9B11, 2233; conjugated to FITC or Alexa Fluor 647), PE-conjugated anti-FLAG antibody (BioLegend, 637310), HRP-conjugated anti-FLAG M2 antibody (Sigma-Aldrich, A8592), PE-conjugated anti-baculovirus gp64 antibody (Thermo Fisher Scientific, clone AcV1, 12-6991-82), anti-human IgG (Fc) monoclonal antibody (Cytiva, Human Antibody Capture Kit, BR100839), biotin-conjugated lineage (Lin) antibody cocktail (Miltenyi Biotec, 130-092-613), BV711-conjugated anti-mouse c-Kit antibody (BioLegend, clone 2B8, 105835), BV421-conjugated anti-mouse CD45 antibody (Thermo Fisher Scientific, clone 30-F11, 404-0451-82), APC-conjugated anti-mouse CD3 antibody (BioLegend, clone 17A2, 100236), PE-Cy7-conjugated anti-mouse CD19 antibody (Thermo Fisher Scientific, eBio1D3, 25-0193-82), BV711-conjugated anti-mouse Gr-1 antibody (BioLegend, clone RB6-8C5, 108443)

Validation

Anti-c-Myc antibodies recognize the EQKLISEEDL epitope tag and are reported by the manufacturers to be validated for immunocytochemistry, ELISA, and immunoelectrophoresis, and suitable for Western blot, ELISA, and immunofluorescence applications. These reagents were used for detection of epitope-tagged miniproteins displayed on yeast cells. The PE-conjugated anti-FLAG antibody detects FLAG-tagged proteins and is validated for flow cytometry and intracellular flow cytometry. This reagent was used for detection of FLAG-tagged GPCR nanodisc targets. The HRP-conjugated anti-FLAG M2 antibody is validated by the manufacturer for Western blot, ELISA, and immunoprecipitation. It was used to confirm cell-surface expression of FLAG-tagged CXCR4, CXCR7 and CCR5 receptors by ELISA.
The PE-conjugated anti-baculovirus gp64 antibody is validated by the manufacturer for flow cytometry–based detection of infected insect cells and viral titration. In this study, it was used to determine baculovirus titers by flow cytometric analysis of gp64-positive Sf9 cells. An anti-human IgG (Fc) monoclonal antibody is validated by the manufacturer for capture of human or humanized IgG in immunoassay and biosensor-based interaction formats, including ELISA and surface plasmon resonance applications. In this study, it was used to capture GLP1R–Fc on the sensor chip surface for subsequent SPR measurement of miniprotein binding.
A biotin-conjugated lineage (Lin) antibody cocktail is validated by the manufacturer for flow cytometric analysis of mouse hematopoietic lineage populations. The BV711-conjugated anti-mouse c-Kit antibody and BV421-conjugated anti-mouse CD45 antibody are validated for flow cytometric detection of mouse hematopoietic stem and immune cell populations. The APC-conjugated anti-mouse CD3 antibody, PE-Cy7-conjugated anti-mouse CD19 antibody, and BV711-conjugated anti-mouse Gr-1 antibody are validated by the manufacturers for flow cytometric detection of mouse T cells, B cells, and myeloid cells, respectively. These antibodies were used for flow cytometric assessment of the in vivo efficacy of the dCX1_001 antagonist in mice.

No additional antibody validation was performed beyond that provided by the manufacturer.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293T, SK-N-MC, COS-7, BHK-21, and CHO-K1 cells were obtained from American Type Culture Collection (ATCC). HEK293A cells were obtained from Thermo Fisher Scientific. CHO-K1/CXCR4 and CHO-K1/Cre-Luc/CGRPR stable cell lines, as well as CHO-S and TurboCHO-Express 2.0 cells, were obtained from GenScript. Lenti-X 293T cells were obtained from Takara, and HeLa cells were obtained from the laboratory of Iain Cheeseman. CHO-K1/PTH1R and CHO-K1/MRGPRX1 cells were obtained from DiscoverX. HTLA and HEK293/OXTR cells were obtained from the laboratories of Bryan L. Roth and Princess I. Imoukhuede, respectively. Spodoptera frugiperda (Sf9) and Trichoplusia ni cells were obtained from Expression Systems. Trichoplusia ni High Five cells were obtained from Thermo Fisher Scientific.
Authentication	Cell lines were not authenticated, but were routinely monitored for normal morphology and growth characteristics.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination, but were routinely monitored for contamination.
Commonly misidentified lines (See ICLAC register)	SK-N-MC cells (listed in the ICLAC register of misidentified cell lines) were used for CGRPR binder screening due to their endogenous expression of CGRPR. Key antagonists identified in this system were further pharmacologically validated in COS-7 cells transiently expressing human CGRPR.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	For the pharmacokinetic (PK) study (CGRPR), C57BL/6J mice (approximately 12 weeks of age) were used. For the HSPC mobilization study (CXCR4), C57BL/6-based transgenic mice aged 6–8 weeks were used.
Wild animals	This study did not involve wild animals.
Reporting on sex	For the pharmacokinetic (PK) study (CGRPR), female mice were used. For the HSPC mobilization study (CXCR4), male and female mice were used.
Field-collected samples	This study did not involve animals collected from the field.
Ethics oversight	For the pharmacokinetic (PK) study (CGRPR), all animal procedures were reviewed and approved by the Lundbeck veterinarians and conducted in accordance with applicable institutional and national guidelines for the care and use of laboratory animals. For the HSPC mobilization study (CXCR4), all animal experiments were reviewed and approved by the University of Washington Institutional Animal Care and Use Committee (IACUC) and conducted in accordance with institutional guidelines and relevant national regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	No research involving plants was conducted in this study. Not applicable.
Novel plant genotypes	No research involving plants was conducted in this study. Not applicable.
Authentication	No research involving plants was conducted in this study. Not applicable.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Sample preparation was performed as described in the manuscript.
Instrument	Flow cytometry was performed using a Sony SH800, a BD FACSymphony A3, or a CytoFLEX flow cytometer.
Software	Sorting was performed as described in the manuscript.
Cell population abundance	The percentage of sequences that match the designed binder in each NGS sample was analyzed.
Gating strategy	The FSC/SSC gate was used to select single living yeast cells. The positive gate for flow cytometry was used to select cells which express both the binder protein and show binding signal above the negative sort. For peripheral blood mononuclear cells (PBMNCs), cells were first gated based on FSC and SSC (P1), followed by selection of $CD45^+$ cells. Within this population, $CD3^+$ T cells, $CD19^+$ B cells, and $Gr\text{-}1^+$ cells were identified using fluorescence gates.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.