

Peptide agonist design from GPCR sequences

Binbin Xia^{1,#}, Xingzheng Wang^{1,2,#}, Qiuxiang Tan^{2,3,#}, Gan Xi^{1,2,#}, Yixin Niu^{1,2}, Junhao Zhu¹,
Yanjie Li^{1,2}, Rui Ma^{1,2}, Qiang Zhao^{3,4,*}, Longchao Liu^{1,*}, Jun Wang^{1,2*}

1. CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China.
2. University of Chinese Academy of Sciences, Beijing 100049, China.
3. State Key Laboratory of Drug Research, State Key Laboratory of Chemical Biology, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China.
4. Zhongshan Institute for Drug Discovery, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Zhongshan, 528400, China.

These authors contributed equally to this work.

* Correspondence: junwang@im.ac.cn (JW), zhaoq@sim.ac.cn (QZ),
liulongchao@im.ac.cn (LCL)

Abstract

Simplifying protein agonists into peptides, or identifying alternative peptide agonists, could substantially enhance the pharmaceutical utility of many GPCR and is therefore highly desirable for biomedical applications. However, *de novo* peptide design for these receptors remains challenging, as agonism requires both binding to and activation of the target; such on-target activity is difficult to incorporate by current approaches that primarily aims at binding. Here we report that based on cross-attention mechanism, our new model GRADE accurately predicts agonist-GPCR pairing using only sequences, and designed peptide agonists across multiple targets with success rates of 20%-40%. Further with the CCR/CXCR family as a focal group, GRADE resulted in agonist with expanded agonistic spectrum than natural chemokines, as well as target-specific agonists. Structural determination revealed GRADE-designed agonists bind to CCR5 orthosteric pocket with novel, reversed polarity of natural chemokines. In a murine model of tumor immune clearance, selected agonists activated CCR5-expressing CD8⁺ T cells and anti-tumor immunity, in particular the CCR5-specific agonist removed cross-reactivity that is inherent to CCL5. With drastically reduced model complexity, GRADE offers an accurate and flexible method for GPCR peptide agonist design using only target sequences, and demonstrates the potential of cross-attention mechanism in future protein ligand design for expanded range of targets.

G-protein coupled receptors (GPCRs) constitute the largest family of cell surface membrane receptors and simultaneously the largest family of drug targets^{1,2}. Among GPCRs with known endogenous ligands, more than a hundred are activated by protein or peptide agonists, which bind to their targets and induce conformational changes in GPCRs, recruiting down-stream heterotrimeric G-proteins and arrestins to produce second messengers for signaling³. As a result of GPCRs' importance in biomedicine, there is continuing interest in modification or *de novo* design of novel peptide agonists, via mutations and modifications of natural ligands (e.g. GLP-1 biosimilars), and more recently structure-guided design of protein binders, exemplified by diffusion-based deep learning approaches^{4,5}. Agonist design however poses a unique challenge for the dual demand of binding and activation of target GPCR, and current structure-based approaches aimed at binding are not capable of defining on-target activity of the peptides, in particular agonism⁶. In addition, structure-based GPCR agonist design requires active-state structures, which are available for only a limited number of GPCRs (219 out of ca. 800)⁷, and as a result of low success rate high-throughput screening is still required to obtain binders from high-n number (thousands or more) of design drafts⁵.

Another important obstacle in GPCR agonist design is the potential cross-reactivity across multiple targets, which is difficult to incorporate prior to experimental validation^{6,8}. It is common for natural peptide agonist to be promiscuous among multiple GPCRs, due to similarities in the receptors' binding pocket structures^{9,10}. One example is the chemokine-receptor system, which includes CCRs and CXCRs and their ligands containing C-C and C-X-C motifs, respectively. Chemokine receptors constitute the largest and most complex GPCR subfamily and play central regulatory roles in the migration and function of diverse immune cells^{9,11,12}. Yet, chemokines and CCRs/CXCRs constitute a complex interaction network that prevents the direct application of many chemokines for therapy, for instance CCR1 can be activated by nine chemokines while CCL13 acts as agonist for up to five receptors. One focal CCR5 has been identified as a key regulator for CD8⁺ T cell activities and its activation leads to improved immune cell infiltration and cancer cell clearance¹³, yet its primary ligand CCL5 (RANTES) also activates CCR1 and CCR3, activation of which in eosinophils in turn suppresses immune clearance in the tumor micro-environment^{14,15}. Previous modification of CCL5 improved its selectivity for CCR5 over CCR1/3 (delta-EC₅₀ increased by 10-fold), but the modified ligand still activates CCR3 at micromolar concentrations¹⁶. The demand for managing ligand specificity also exceeds extant methods' ability, as they are not computationally effective or sufficiently accurate for considering multiple targets.

In parallel with structure-dependent binder design approaches, new methods that rely on language models and design binders from target sequences are emerging, hinting at the possibility of agonist generation directly from GPCR sequences, as an example of achieving on-target activity in the design process¹⁷. Here, we report that based on cross-attention mechanism from natural language processing (NLP), a peptide's on-target activity towards GPCR can be evaluated using only target sequences, and novel peptides with defined specificity can be designed from an augmented peptide library (Fig.1a). Despite being trained on a small

dataset and having substantially reduced model complexity, the resulted GRADE model outperforms existing models in predicting agonist-GPCR pairing. Using GRADE, we identified peptide agonists across multiple targets with experimental hit rates of 20%-40%. Focusing on the CCR/CXCR family, agonists with either broader-spectrum or more specific agonism can be discovered, which bind to CCR5 orthosteric pocket with reversed polarity compared to natural ligand CCL5. When administered at equal-mass doses, the selected agonists outperform CCL5 in a murine model of tumor immune clearance. Our approach thus provides a drastically simple, effective and flexible method for GPCR agonist design using only target sequences, and demonstrates the potential of cross-attention mechanism in function-aware, ligand-target interaction prediction and protein ligand design.

Results

Cross-attention mechanism predicts agonist-GPCR pairing

We started by asking whether agonist-GPCR pairing could be predicted directly from protein sequences without involving structural predictions or docking; computationally this requires separating relatively small number of peptide agonist-GPCR pairs from a large background of non-agonistic pairs. To address this problem, we developed GRADE (GPCR-targeted Ranking-guided Agonist Design Engine, Fig.1a-c), a sequence-based cross-attention model, taking encoded peptide sequences (by a short peptide-adapted masked protein language model) and GPCR sequences (encoded by ESM2¹⁸) as input. GRADE integrates the two representations through two parallel bidirectional cross-attention blocks and outputs an agonist confidence score (ACS) between 0 and 1.

The complete dataset contained 721 experimentally validated agonist pairs spanning 136 human GPCRs (114 class A, 18 class B and 4 class F receptors) and 380 unique peptide ligands; 609 positive pairs were assigned to training, 28 to validation and 84 to testing⁷. Negative pairs were generated by pairing peptides with non-cognate receptors at a ratio of approximately 1:70. A sum of 43,680 peptide-receptor pairs were used for training, 2,000 pairs were used for validation, and 6,000 peptide-receptor pairs (including 84 or 1.4% positive pairs) were held out as the test set (Extended Data Fig.1a,b; Supplementary Table 1). We used area under the precision-recall curve (AUPRC) as the primary metric for this imbalanced benchmark. Area under the receiver operating characteristic curve (AUROC) and Precision@K (the proportion of known agonist pairs among the K highest-scoring pairs) provided complementary measures of overall discrimination and top-rank enrichment, respectively.

Despite the small number of positive pairs available for training, we found that GRADE achieved an AUPRC of 0.7991 and an AUROC of 0.9788 on the test set (Fig.1d,e; Supplementary Table 2). In addition, Precision@10 and Precision@20 were both 1.00, and Precision@50 was 0.88. All top 20 predictions were positive pairs, demonstrating strong enrichment at the experimentally relevant top of the ranked list. We benchmarked GRADE against nine available sequence- and structure-based interaction predictors: the structure-based approaches comprised AlphaFold 3¹⁹, RoseTTAFold2-PPI²⁰ and SpatialPPIv2²¹ (Structure).

The sequence-based group comprised PLM-interact²², MINT²³, PepMLM¹⁷, two D-SCRIPT^{24,25} variants and SpatialPPIv2²¹ (Sequence). Each method was ranked by its output score on the same test pairs. AUROC, AUPRC and Precision@K were calculated from the fixed class labels and ranked scores, whereas MCC was calculated using a score threshold of 0.5; undefined scores were excluded before metric calculation. On the same test set, GRADE achieved the highest AUPRC, AUROC, Matthews' correlation coefficient (MCC) and Precision@K (Fig.1d,e, Extended Data Fig.1c-e; Supplementary Table 2); while AlphaFold 3 achieved AUROC of 0.8742 but with an AUPRC of only 0.1910, the remaining methods only achieved AUROC values between 0.4791 and 0.6203, and AUPRC values between 0.0157 and 0.0323. GRADE achieved high predictive performance for both Class A and Class B GPCRs (Extended Data Fig.1f), indicating that the enrichment effect was not driven by only one receptor class. Thus, in the case of peptide agonist-GPCR pairs with large data imbalance, GRADE achieved the highest precision of identifying positive pairs.

GRADE outperformed the other methods in predicting peptide agonist-GPCR pairs without requiring multiple-sequence alignment, sequence or template database searches, or structural input. GRADE contains only 3.3 million parameters (excluding encoders whose outputs can be precomputed), which is 112.1-fold smaller than AlphaFold 3 and also smaller than nine models included for comparison (Fig.1f; Supplementary Table 3). While three models are smaller than GRADE, the two D-SCRIPT variants and RoseTTAFold2-PPI, had AUPRC values no higher than 0.0202 and AUROC values no higher than 0.5382. Ablation test demonstrated that the cross-attention mechanism has the most impact on GRADE's performance, in terms of AUPRC, AUROC, MCC (Extended Data Fig.2a-c), Precision@10, Precision@20 and Precision@50 (Extended Data Fig.2d-f). In summary, GRADE's sequence-to-score workflow established a compact, sequence-only model for accurately predicting peptide agonism for GPCRs.

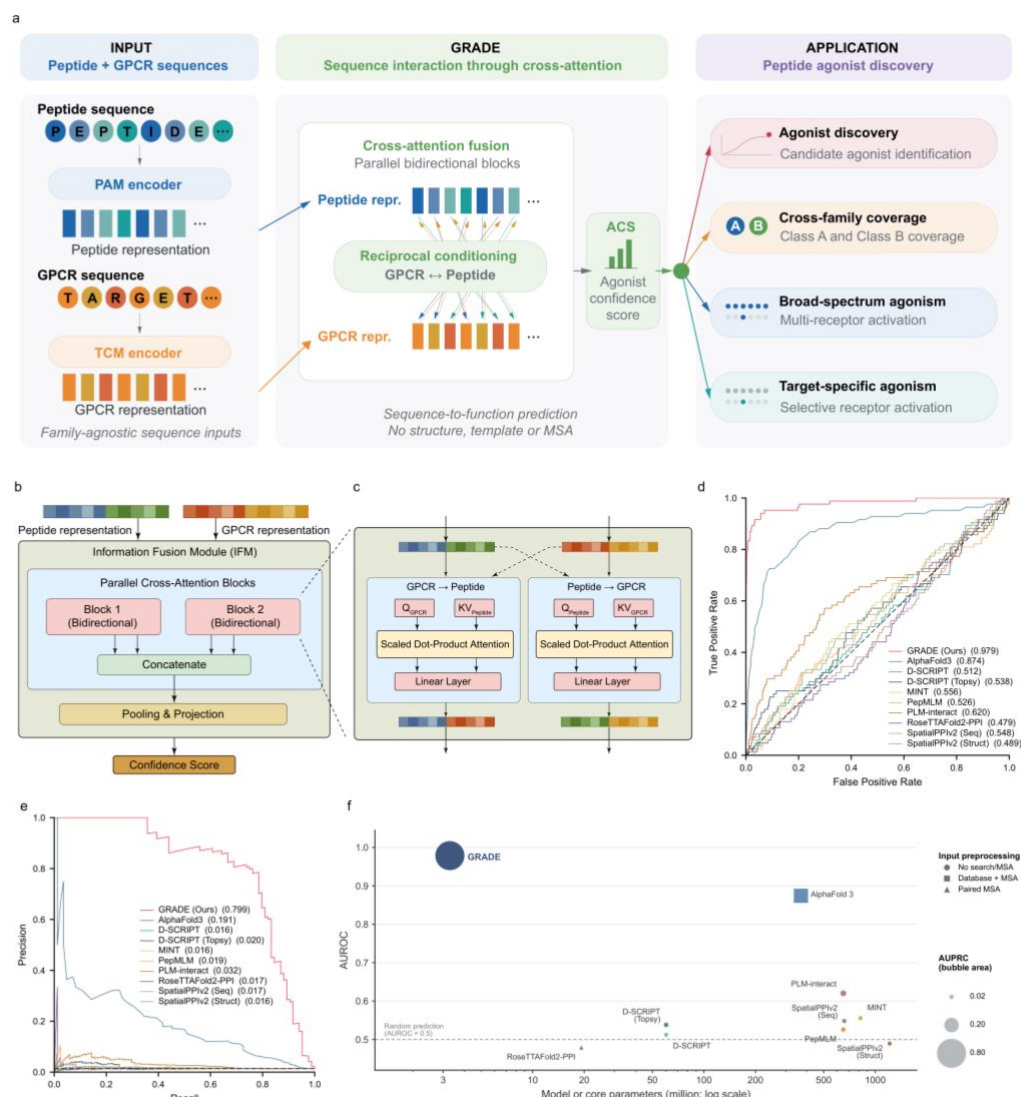


Fig.1 GRADE architecture and benchmark performance for peptide agonist-GPCR prediction.

a, Overview of the sequence-based GRADE framework. Peptide and GPCR sequences are encoded by the peptide-adapted module (PAM) and target-context module (TCM), respectively, and integrated through cross-attention to produce an agonist confidence score (ACS). The schematic summarizes agonist discovery, Class A and Class B coverage, broad-spectrum agonism and target-specific agonism; sequence letters are illustrative. **b**, Information fusion module (IFM). Two parallel bidirectional cross-attention blocks integrate peptide and GPCR representations; their outputs are concatenated, pooled and projected to an ACS. **c**, The two attention directions. GPCR-to-peptide attention uses the GPCR representation as queries (Q) and peptide representations as keys and values (KV); peptide-to-GPCR attention reverses these assignments. **d,e**, Receiver-operating-characteristic (d) and precision-recall (e) curves for GRADE and nine comparator methods on the held-out test set of 6,000 peptide-GPCR pairs, including 84 positive pairs. Parentheses report AUROC (d) and AUPRC, calculated as average precision (e). **f**, Predictive performance versus model or core parameter count. The horizontal axis shows parameter count in millions on a logarithmic scale, the vertical axis shows AUROC and bubble area is proportional to AUPRC. GRADE contains 3.29 million parameters in the IFM and classifier. Circles denote models requiring no database search or multiple-

sequence alignment (MSA), squares denote models requiring database search plus MSA, and triangles denote models requiring paired MSA. The dashed line denotes random prediction (AUROC = 0.5).

165 **GRADE captures interpretable ligand-GPCR interface features**

To determine whether GRADE predictions reflected biologically relevant features of peptide-receptor interaction, we employed the experimentally resolved CCL5-CCR5 paired structure (PDB 7F1R, Fig.2a) for a canonical interpretability test. From the cross-attention outputs of GRADE's information fusion module (IFM), we derived two residue-level attribution measures
170 for CCL5: agonism-margin attribution, which quantifies each residue's positive contribution to the agonist-minus-non-agonist logit margin using Gradient $\times$ Activation; and agonism-logit contribution, which decomposes the agonist-class logit into residue-level contributions by projecting ligand-query cross-attention outputs onto the agonist-class classifier weights.

We ranked all 68 CCL5 residues by positive agonism-margin attribution and assessed the
175 recovery of structure-defined CCR5 contacts among the top-ranked residues. Overall, the N terminus and 30s loop, which together form the principal CCR5-contacting surface, had significantly higher contributions than the remaining CCL5 regions (one-sided Mann-Whitney *U* tests, $P = 0.008$ and $P = 0.024$ for agonism-margin attribution and agonism-logit contribution, respectively; Fig.2b,c). Six of the top seven residues contacted CCR5 within 4.0 Å, exceeding
180 the random-selection reference (pointwise $P = 0.0007$; Fig.2d). Restricting the analysis to the 50 coordinate-resolved positions also recovered six contacts among the top seven residues ($P = 0.0047$).

We compared the prediction margin (the agonist-class logit minus the non-agonist-class logit) for canonical CCL5 paired with CCR5 with those for 1,000 composition-preserving shuffled
185 sequences and 1,000 uniformly random sequences of the same length. Canonical CCL5 had a prediction margin of 4.63, exceeding all control sequences and placing it 13.9 standard deviations above the mean of the shuffled controls (Fig.2e). Together, these analyses show that GRADE identifies structure-defined contact residues in the CCL5-CCR5 pair and distinguishes canonical CCL5 from shuffled and random-sequence controls.

190

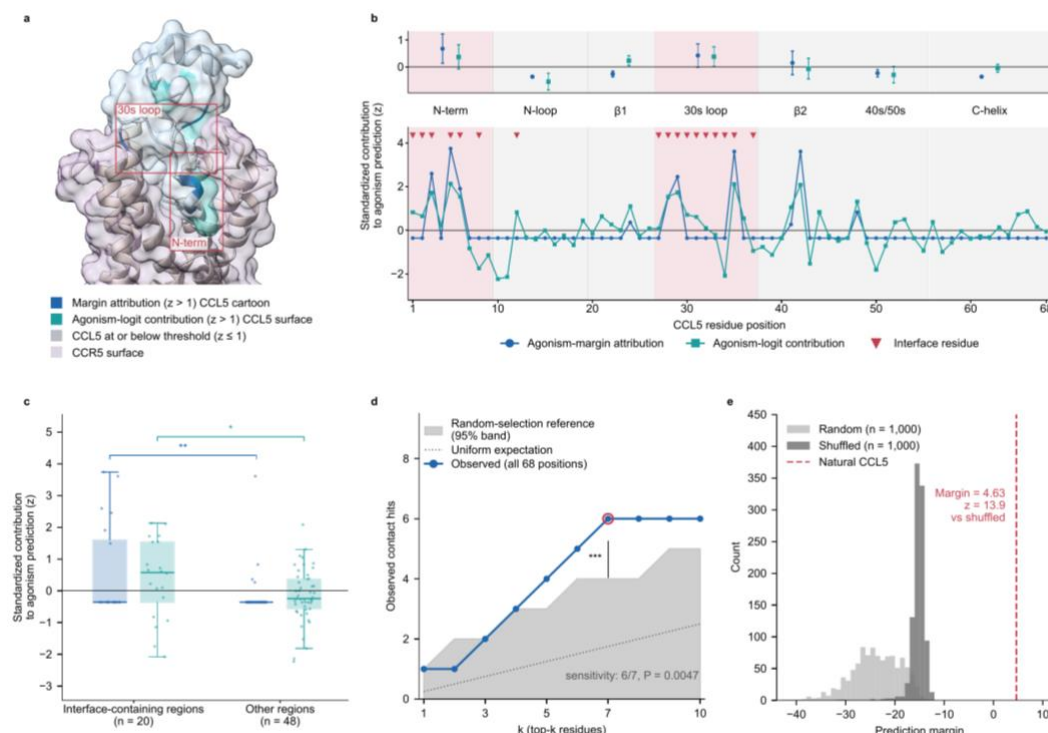


Fig.2 Ligand-side attribution of CCR5-CCL5 agonism prediction.

a, Canonical CCL5 residue attributions mapped onto the engineered CCR5-CCL5 complex (PDB 7F1R). Blue cartoon segments indicate positive agonism-margin attribution with $z > 1$; teal surface regions indicate agonism-logit contributions with $z > 1$; grey indicates $z \leq 1$; CCR5 is pale purple. Red boxes mark the interface-containing N terminus (residues 1-9) and 30s loop (27-37). **b**, Regional averages (top) and residue-level profiles (bottom) of agonism-margin attribution and agonism-logit contribution across all 68 canonical CCL5 residues. Both measures are standardized across the 68 residues using the sample standard deviation. Regional averages are shown as mean $\pm$ s.e.m. across residues within each region. Red triangles mark 17 interface residues whose minimum heavy-atom distance to CCR5 is ≤ 4.0 Å. **c**, Contributions in interface-containing regions ($n = 20$ residues) and other regions ($n = 48$). Boxes show medians and interquartile ranges; whiskers extend to the most extreme values within 1.5 times the interquartile range; points show individual residues. Brackets indicate one-sided Mann-Whitney U tests without multiple-comparison correction. **d**, Observed contacts recovered among the top-k residues ranked by positive margin attribution. The grey band shows the 2.5th-97.5th percentiles of 10,000 random selections from all 68 positions; the dotted line shows the expectation $k \times 17/68$. Six of the top seven residues are contacts (pointwise random-selection $P = 0.0007$; red circle). **e**, Prediction-margin distributions for 1,000 composition-preserving shuffled and 1,000 uniformly random sequences of CCL5 length. The dashed line marks canonical CCL5 (margin, 4.63; $z = 13.9$ relative to shuffled controls). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Agonist design from augmented peptide library

After validating GRADE as an accurate predictor for GPCR peptide agonists, we next constructed an augmented virtual peptide library for the design of new agonists. For this, we fine-tuned a causal protein language model on 7,617,212 natural peptide sequences and generated new sequences autoregressively (Fig.3a). After deduplication and removal of

sequences identical to natural entries, the generated peptides contained 9,479,470 unique sequences, yielding a final augmented peptide library (APL) of 17,096,682 non-redundant sequences (Supplementary Table 5). The peptides in the APL had a median length of 34 amino acids (range, 1-68). The NPL and GPL had broadly similar overall amino-acid compositions (Supplementary Table 7). The generated peptides had a mean log-transformed perplexity (logPPL) of 7.13, compared with 12.02 for natural peptides (Fig.3b), indicating greater sequence likelihood under the peptide distribution learned during fine-tuning; also CD-HIT²⁶ clustering at 50% sequence identity revealed the generated peptides had a cluster-dispersion ratio of 0.977, compared with 0.549 for the natural peptide library (Fig.3c; Supplementary Table 6). GPL combined increased sequence diversity with broad preservation of NPL-like properties, including disorder, solubility, isoelectric point, signal-peptide and transmembrane-region prevalence, and structural confidence (Fig.3d,e; Extended Data Fig.3a-f). UMAP²⁷ placed generated peptides partly within the natural peptide distribution and partly in distinct regions of sequence space (Fig.3f). Together, these results established the augmented peptide library as a substantially enlarged peptide space for GRADE design.

We first started with designing/screening peptide agonists for CCR5, a representative receptor of class A GPCR receptor and the chemokine receptor family. Using GRADE, we selected the five highest-ranked peptides from each of the NPL and GPL, finding that three are agonists (3/10) using a concentration threshold of 10 μ M (Supplementary Table 10). Dose-response measurements yielded an EC_{50} value of 128.86 nM for CCR5-n3, 19.83 nM for CCR5-n8 and 8.65 nM for CCR5-n9, compared with 88.05 nM for the natural ligand CCL5 (Fig.3g), and their normalized E_{max} values ranged from 68.3% to 71.5% of the CCL5 E_{max} ; though all new peptides are partial agonists, CCR5-n9 leads to a higher fitted response than CCL5 in the range of approximately 1-68 nM (CCR5-n9 versus CCL5: 20.1% versus 16.0% at 1 nM and 35.1% versus 30.9% at 10 nM, normalized to the net maximal CCL5 response; Fig.3g).

GRADE processed a substantially larger sequence-screening workload than four structure-based binder design workflows: BindCraft²⁸, PocketXMol²⁹, BoltzGen³⁰ and the RFdiffusion³¹-ProteinMPNN³²-AlphaFold 3¹⁹ pipeline (RF-MPNN-AF3) (Fig.3h,i). Against CCR5, we selected the top 10 design drafts from each method and tested them using the same reporter assay, finding that only BoltzGen yielded one agonist (1/10; BoltzGen-05 had an EC_{50} of 3.20 μ M and a normalized E_{max} of 34.8% of the CCL5 response) and none of the candidates from the other methods showed agonist activity (Fig.3g,j).

We further examined the agonist design capacity of GRADE across three additional GPCRs: the class A peptide receptor NK2R, the class B peptide receptor GIPR, and the orphan receptor GPR18 that was not included in the training set and has no peptide agonist reported. For each target, ten top-ranked peptide candidates were synthesized and tested. GRADE identified three agonists for NK2R (3/10), two for GIPR (2/10) and four for GPR18 (4/10) (Fig.3k). Several NK2R and GIPR agonists showed lower EC_{50} values or higher E_{max} values than their respective reference agonists. One NK2R agonist (NK2R-01) had an EC_{50} of 34.96 pM and an E_{max} of 71.4% relative to Neurokinin A, whereas NK2R-08 had an E_{max} of 138.9%. For GIPR, one peptide agonist had an EC_{50} of 128.30 nM and an E_{max} of 188.1% relative to tirzepatide. For the orphan receptor GPR18, GPR18-n9 had the lowest EC_{50} of 1.77 μ M, whereas the EC_{50} values of the other three peptides ranged from 6.72 to 11.36 μ M. These results show that GRADE

enabled peptide agonist discovery across GPCRs from different receptor classes, including receptors without known peptide agonists and absent from the training set (“zero-shot” cases), using only ten candidates per target.

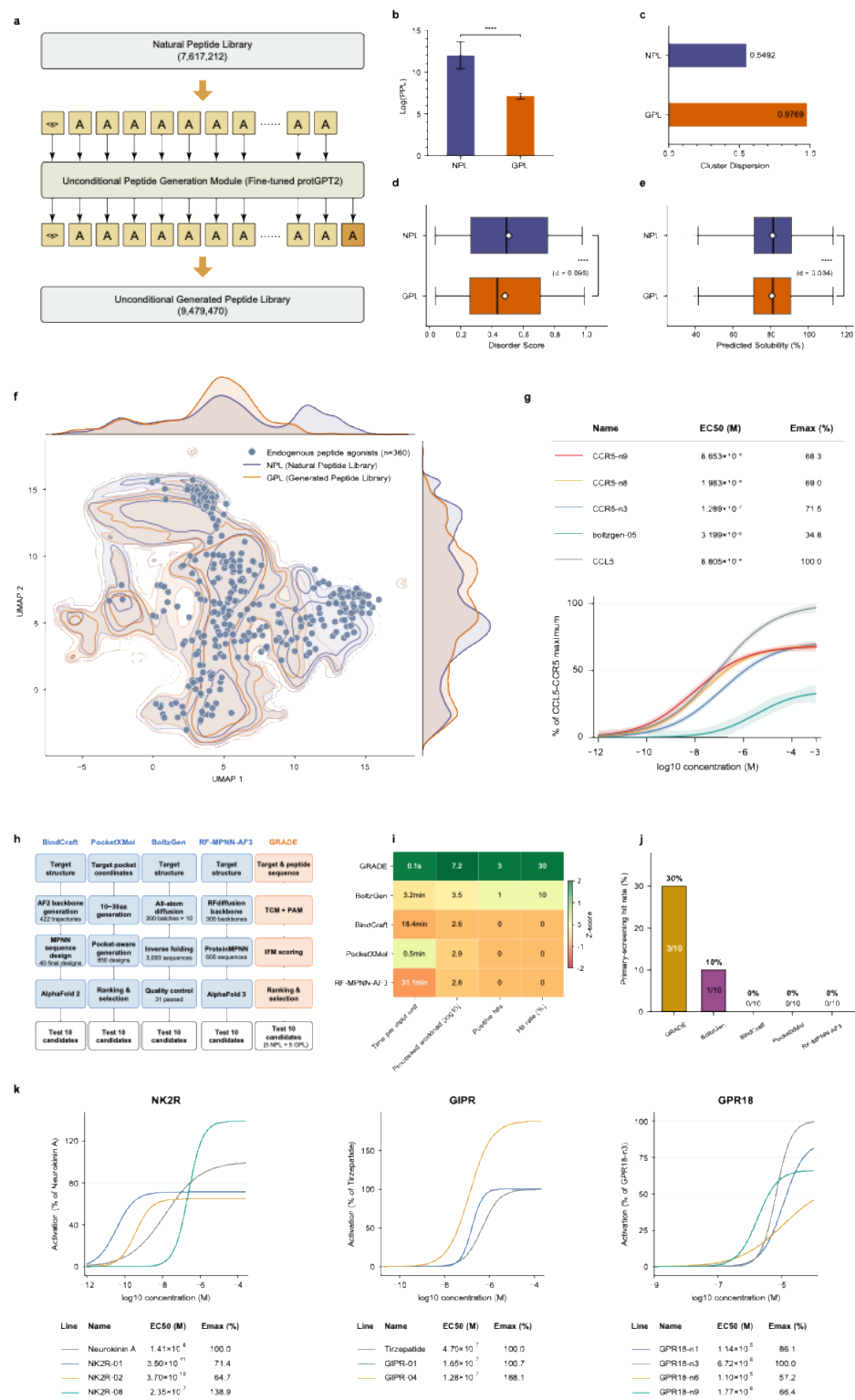


Fig.3 Generation and characterization of peptide libraries and experimental benchmarking of CCR5 agonist discovery.

a, Unconditional peptide generation using ProtGPT2 fine-tuned on the Natural Peptide Library (NPL; 7,617,212 sequences), yielding the Unconditional Generated Peptide Library (GPL; 9,479,470 non-redundant sequences after filtering). **b**, Mean log-transformed perplexity ($\text{Log(PPL)} \pm \text{s.d.}$ for NPL and GPL ($n = 7,617,212$ and $9,479,470$ sequences, respectively). NPL-GPL comparison uses a two-sided unpaired Welch's t-test, **** $P < 0.0001$. **c**, Cluster dispersion, defined as the number of CD-HIT clusters divided by sequence count at 50% sequence identity (NPL, 0.5492; GPL, 0.9769). **d,e**, Predicted intrinsic disorder (metapredict; **d**) and solubility (Protein-Sol; **e**). The solubility comparison in **e** is restricted to valid-output subsets (NPL, 5,551,089; GPL, 9,479,073 sequences). Boxes show medians and interquartile ranges, whiskers extend to the most extreme values within 1.5 times the interquartile range, and diamonds indicate means. Plots use 10,000 sequences per library; two-sided Mann-Whitney U tests and Cohen's d use separate 100,000-sequence samples per library. Cohen's d is calculated in the direction NPL minus GPL. **f**, UMAP projection of PAM sequence embeddings for NPL and GPL. Contours enclose 99.5%, 95%, 80% and 50% of within-library density; marginal distributions are normalized to unit area. Points show 360 unique endogenous peptide agonists in the same projection. **g**, CCR5 reporter dose-response curves for three GRADE-identified peptides (CCR5-n3, CCR5-n8 and CCR5-n9) and the BoltzGen candidate BoltzGen-05, with CCL5 as the reference agonist. The table reports EC_{50} and E_{max} values. Responses were baseline-corrected using each curve's fitted baseline and expressed as percentages of the fitted net maximal response to CCL5 at CCR5 (100%). E_{max} values were normalized to the same reference. Lines show fitted curves; shading indicates pointwise 95% confidence bands on the normalized scale. Each concentration was measured four times for CCL5 and three times for each candidate. **h**, Candidate-generation and selection workflows for GRADE and four structure-based design pipelines; ten candidates per method were tested. **i**, Computational and primary-screening performance of the five workflows, including execution time per workflow-specific input unit, processed workload ($\log_{10}$), hits and hit rate. Cell labels report metric values; colours show column-wise z-scores across workflows. Higher z-scores indicate shorter execution time ($-\log_{10}$ hours) or larger values of the other metrics. **j**, Primary-screening hit rates among ten tested candidates: GRADE, 3/10; BoltzGen, 1/10; and BindCraft, PocketXMol and RF-MPNN-AF3, 0/10 each. **k**, Dose-response curves of GRADE-identified peptides at NK2R, GIPR and GPR18. Activation is expressed relative to the fitted net maximal response of Neurokinin A at NK2R and tirzepatide at GIPR, respectively, and of GPR18-n3 at GPR18 (100%). Lines show fitted curves ($n = 3$). Horizontal dotted lines indicate 100% reference activation. Tables list fitted EC_{50} and E_{max} values.

GRADE customizes agonism spectrum of peptides

After establishing GRADE as an efficient tool for agonist design, we next explored its ability in customizing agonism spectrum of peptide agonists, achieving either broad-spectrum or highly specific agonists. For this task, we selected the human chemokine receptors (CCRs and CXCRs) as exemplary receptors, and used GRADE to score peptides from APL against 19 chemokine receptors. Using an ACS of 0.9 as a cut-off threshold, GRADE identified a total of 10 peptides that have potentially broader range of agonism (two peptides that are potential agonists against eight receptors, eight peptides against six receptors; Extended Data Fig.4a) than natural ligands (CCL13, five receptors; Fig.4a). These broad-spectrum candidates were shorter than the NPL background and had a higher mean cysteine content (Extended Data Fig.4b,c).

In the follow-up validation, a total of four peptides showed agonism spectra agreeing with
 305 GRADE prediction, corresponding to a success rate of 40%. In particular, a peptide CCRbn8-
 2 activated eight receptors and was therefore named octokine (Fig.4b); another three peptides
 activated six receptors. We focused on octokine and determined its EC₅₀ values against the eight
 receptors. These values ranged from 4.96 nM for CCR9 to 4.36 μM for CCR1, within the range
 reported for CCLs at their primary and cross-reactive targets (Fig.4c). Sequence-wise, the
 310 octokine has only a maximum gap-inclusive global sequence identity of 8.7% from the 44 core
 human chemokines (to CCL4; Fig.4b; Supplementary Table 9) and intriguingly has one C-C
 and two C-X-C motifs conjoined in a 6-AA window (CSCCLC), potentially explaining its broad
 activation range.

Promiscuity of natural chemokines is an important obstacle for its clinical applications, and we
 315 then examined whether GRADE can generate specific peptide agonist for CCRs/CXCRs as
 well. After filtering for an ACS > 0.9 at the intended receptor and ACSs < 0.6 at all other
 receptors, we ranked the candidates and retained up to ten per target from each library, yielding
 242 peptide candidates across 15 targets (Extended Data Fig.5a). The retained candidates
 showed a range of predicted selectivity ratios and receptor-dependent off-target ACSs
 320 (Extended Data Fig.5b,c,d,e). Within the CCR and CXCR families, each candidate scored
 highest at its intended receptor (Extended Data Fig.5f,g), and among 20 candidates predicted
 to be CCR5-specific lacked cysteine (Extended Data Fig.5h), we experimentally identified five
 agonists. No off-target activation was detected when these five peptides were tested against
 other CCRs/CXCRs (Fig.4d); they are hereafter referred to as CCR5-Unikines. Their EC₅₀
 325 values ranged from 1.60 μM to 6.40 μM, corresponding to 18.2- to 72.7-fold higher EC₅₀ values
 than the natural chemokine CCL5 (88.05 nM; Fig.4e). By contrast, CCR5-n9 also activated
 CCR3, with an EC₅₀ of 0.92 μM. The five CCR5-Unikines are also sequence-divergent from
 natural CCLs, with gap-inclusive global sequence identities to CCL5 of 6.4% to 19.2%, and
 comparing against the remaining 43 core human chemokines, the maximum global identity was
 330 19.7% (between CCR5-Unikine-g8 and CCL22), and the five experimentally validated CCR5-
 Unikines contain no cysteine, distinct from the conserved N-terminal cysteines of classical
 CCLs/CXCLs (Fig.4d).

We further validated seven target-specific peptides across six targets. All activated their
 GRADE-predicted target receptors at micromolar concentrations, whereas no activation of
 335 receptors cross-activated by the corresponding natural ligands was detected at concentrations
 up to 100 μM (Supplementary Table 14). In summary, by differentiating the potential of
 agonism towards either broader-spectrum or higher target specificity, GRADE achieved highly
 customized agonist design for this complex family of receptors.

33 residues from CCR5 C-terminus (F320-L352) to improve protein yield and homogeneity as previously reported. CCR5-n9 and CCR5-Unikine-g7 are abbreviated as n9 and g7, respectively.

360 The n9-CCR5-G₁₁ and g7-CCR5-G₁₁ complexes were determined by single-particle cryo-electron microscopy at nominal global resolution of 3.1 Å and 3.3 Å, respectively (Fig.5a,b, Extended Data Fig.6a-l and Supplementary Table 15). The resulting maps enabled modelling of the CCR5-G₁₁ protein complex together with the C-terminal segments and S30-S34 of n9 and S27-F36 of g7 (Extended Data Fig.7a,b).

365 We found that both the n9-CCR5 and g7-CCR5 complexes adopt closely related active receptor conformations, with a Cα r.m.s.d. of approximately 0.6 Å compared to the chemokine-bound CCR5 complex structure (Fig.5c,d). Both peptides bind in the canonical CRS2 pocket of CCR5, which is occupied by the N termini of endogenous chemokines in previously determined structures. However, in contrast to chemokines, they interact with CCR5 via their C termini
370 rather than their N termini. Moreover, despite occupying the same binding pocket, the two peptides adopt markedly distinct conformations: the C-terminal segment of CCR5-n9 adopts a compact U-shaped conformation, and is more deeply embedded within the central transmembrane pocket; while CCR5-Unikine-g7 follows an extended conformation that runs beneath ECL2, with its C-terminal segment penetrating into CRS2 (Fig.5e).

375 In terms of CCR5-n9, we found that the C-terminal segment comprising S30-S34 adopts a compact U-shaped conformation within CRS2 (Fig.5c). The S31-S32 region establishes a polar interaction network with residues including T82^{2.56}, L104^{3.28}, Y108^{3.32} and E283^{7.39}. K33 occupies the most deeply embedded region of the peptide. Its aliphatic portion is enclosed by F109^{3.33}, Y251^{6.51}, L255^{6.55} and M279^{7.35}, whereas its amino group forms a polar contact with
380 Q194^{5.38}. The terminal S34 subsequently folds back towards the extracellular side, completing the compact U-shaped configuration. In comparison, the CCR5-Unikine-g7 segment comprising S27-F36 extends from the extracellular vestibule into the CCR5 transmembrane pocket (Fig.5d). S27 and S28 occupy the membrane-distal end of the g7 segment, with their backbone approaching L174 in ECL2. Extensive interactions were observed between CCR5
385 and the C terminus of g7, with V32 packing against M279^{7.35}, and L33 and T34 contact T82^{2.56} and L104^{3.28}, respectively. I35 occupies a hydrophobic sub-pocket formed by T105^{3.29}, Y108^{3.32} and F109^{3.33}. At the base of the pocket, the terminal F36 provides the principal anchor through an edge-to-face aromatic interaction with F109^{3.33} and additional contacts with Q194^{5.38}, I198^{5.42}, Y251^{6.51} and L255^{6.55}.

390 Further superposition with MIP-1α- and CCL5-bound CCR5 reveals that both of our agonists occupy the canonical CRS2 pocket used by the N termini of endogenous chemokines (Fig.5f). For CCR5-n9, K33, S32 and S31 overlap approximately with residues S1, L2 and A3 of MIP-1α and with S1, P2 and Y3 of CCL5, respectively. These overlaps indicate that CCR5-n9 uses its C-terminal region to mimic the CCR5 occupancy of the chemokine N terminus, thereby
395 adopting reversed peptide-chain polarity. Its compact U-shaped conformation nevertheless reproduces the hook-like spatial arrangement of the deeply inserted chemokine residues, while the terminal S34 folds back towards the extracellular vestibule, such that its U-shaped conformation more closely recapitulates the deeply inserted chemokine N-terminal hook (Fig.5g). For CCR5-Unikine-g7, F36 and I35 occupy positions corresponding approximately
400 to S1 and L2 of MIP-1α and to S1 and P2 of RANTES, respectively. Thus, CCR5-Unikine-g7

reproduces the chemokine anchor position using its C terminus and engages CCR5 with reversed peptide-chain polarity (Fig.5h). Unlike full-length chemokines, however, neither CCR5-n9 nor CCR5-Unikine-g7 engages CRS1 or CRS1.5. The two peptides therefore act as compact, CRS2-directed C-terminal chemokine mimetics that stabilize active CCR5 using distinct binding ways: with a folded-back, distributed polar and hydrophobic interaction network for CCR5-n9, and an extended, terminal-aromatic anchoring mode for the CCR5-Unikine-g7.

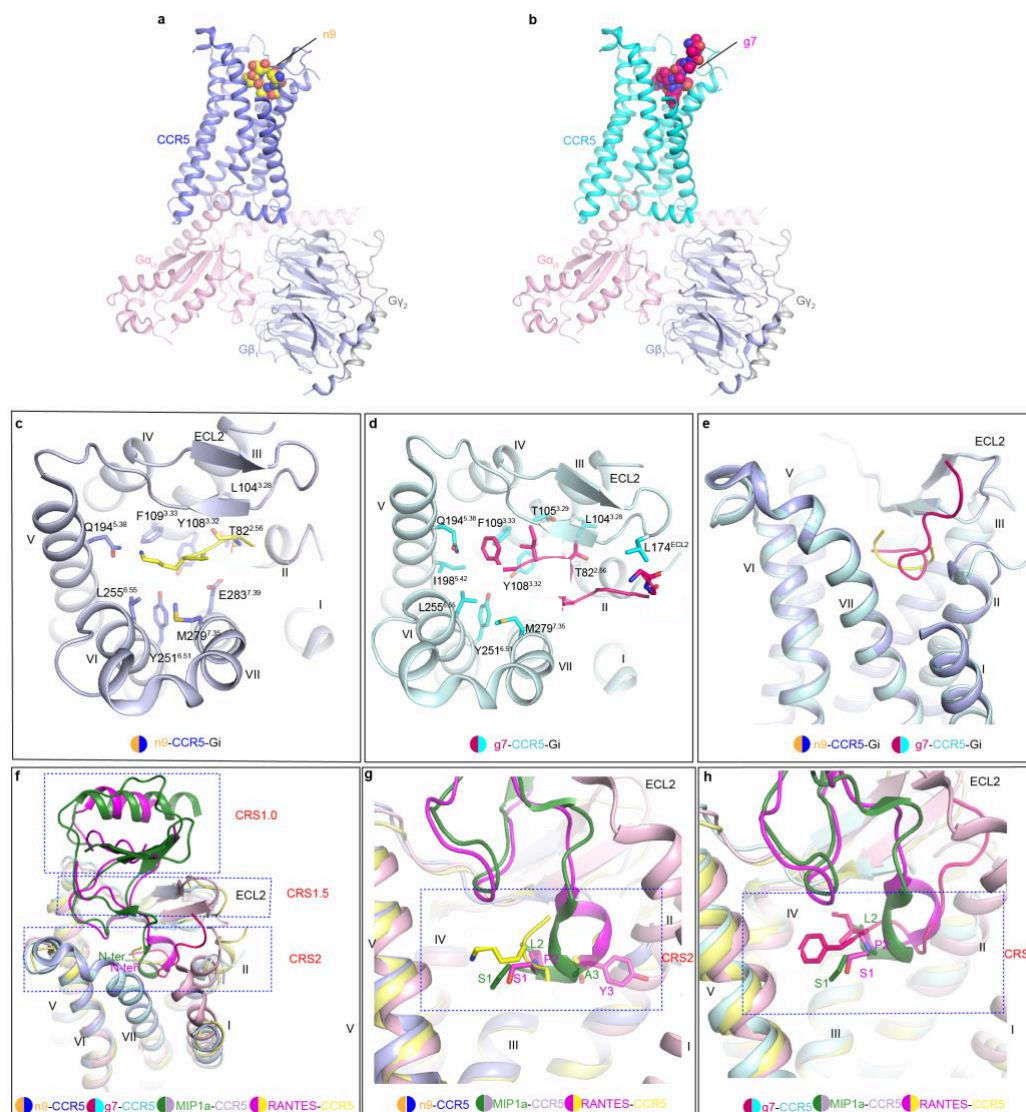


Fig.5 Cryo-EM structures of peptide-bound CCR5-Gi complexes and distinct recognition modes of n9 and g7.

n9 and g7 denote CCR5-n9 and CCR5-Unikine-g7, respectively. **a,b**, Overall structures of the n9-CCR5-Gi (a) and g7-CCR5-Gi (b) complexes, viewed parallel to the membrane. Peptides are shown as spheres to indicate their positions within the extracellular transmembrane cavity. **c,d**, Binding modes of the C-terminal segments of n9 (c) and g7 (d) in the CCR5 extracellular vestibule and transmembrane cavity, viewed from the extracellular side. CCR5 residues involved in peptide recognition are shown as sticks. **e**, Structural superposition of the n9-CCR5-Gi and g7-CCR5-Gi complexes, viewed parallel to the membrane. g7 follows an extended trajectory beneath ECL2, whereas n9 adopts a compact U-shaped conformation positioned more deeply and centrally within the transmembrane cavity.

f, Superposition of the n9-bound and g7-bound CCR5 structures with the MIP-1 α -bound and RANTES (CCL5)-bound CCR5 structures. Dashed boxes indicate chemokine recognition sites CRS1.0, CRS1.5 and CRS2. **g**, Close-up comparison of n9 with the chemokine N-terminal hooks in CRS2. n9 residues K33, S32 and S31 approximately overlap with the first three chemokine residues, illustrating a compact structural mimic of the deeply inserted chemokine N terminus. **h**, Close-up comparison of the deeply inserted g7 residues with the N termini of MIP-1 α and RANTES (CCL5) in CRS2. g7 residues I35 and F36 occupy positions corresponding approximately to the first two residues of the endogenous chemokines, illustrating the reversed peptide-chain polarity of g7 recognition.

GRADE-designed CCR5 agonists activate CD8⁺ T cells for effective tumor control

It has been established that CCR5 is highly expressed in CD8⁺ T cells, and the expression of primary ligand of CCR5 (CCL5) is positively correlated with the outcome of PD-1/cancer immune therapies. We thus examined the ability of our designed/identified CCR5 agonists in CD8⁺ T activation and the anti-tumor effect *in vivo*, using a murine xenograft tumor model using both wild-type mice, and transgenic mice expressing human CCR5 in which murine CCR5 is knocked-out. First, we established that CCR5-n9 and CCR5-Unikine-g7 activate both human and murine CCR5 (hCCR5 and mCCR5 respectively), and intriguingly both peptides have higher $E_{\max}$ than human CCL5 (hCCL5) towards mCCR5 (Fig.6a). CCR5-n9 and CCR5-Unikine-g7 suppressed MC38 tumor growth more effectively than hCCL5 in hCCR5-transgenic mice, while all three agonists significantly reduced tumor size following intratumoral treatment (Fig.6b,c). At this dosage, the molar amounts of CCR5-n9 and CCR5-Unikine-g7 were approximately 2.3 and 2.1 times that of hCCL5, respectively. Considering the differences in EC_{50} , CCR5-Unikine-g7 in particular is still comparatively under-dosed than hCCL5 for hCCR5. Consistent with the anti-tumor efficacy, intratumoral T cells had been rejunavated by CCR5-Unikine-g7 treatment as indicated by increased percentage of CD69⁺CD8⁺ T cells as well as the reduction of PD-1⁺TIM-3⁺ exhausted CD4⁺ T cells in hCCR5-transgenic mice (Fig.6d). CCR5-n9 and CCR5-Unikine-g7 also outperformed hCCL5 in wild-type C57BL/6J mice, with all three agonists significantly reducing MC38 tumor size following intratumoral treatment (Fig.6e,f).

In human, CCL5 activates multiple chemokine receptors besides CCR5, including CCR1, CCR3 and CCR4; CCR1⁺ and CCR4⁺ immune cells such as eosinophils and neutrophils have been implicated in immunosuppression of tumor microenvironment. In light of this, after establishing the efficacy of our CCR5 agonists in tumor immune clearance we next examined the specificity of these agonists. We first found that, against mouse CCRs, CCR5-n9 and CCR5-Unikine-g7 have similar cross-reactivities as in humans: hCCL5, and CCR5-n9 all activate non-CCR5 targets, in particular hCCL5 activates mCCR1, mCCR3, with EC_{50} values of 3.15 and 0.76 μ M, respectively, and CCR5-n9 activates mCCR1/3 with EC_{50} of 54.01 and 11.81 μ M, respectively; in contrast CCR5-Unikine-g7 only activates mCCR5 and does not activate the above-mentioned targets at up to 100 μ M concentration (Fig.6g). No mCCR4 activation was detected with hCCL5, CCR5-n9 or CCR5-Unikine-g7 under the tested conditions (Fig.6h). In an additional group using wild-type mice, tumors were harvested 24 h after the second treatment for immune profiling (Fig.6i). We found that while all treatment groups increased the frequencies of CD69⁺CD4⁺ and CD69⁺CD8⁺ T cells, among which the CCR5-n9 and CCR5-

Unikine-g7 had significantly higher frequencies in both groups compared to that of hCCL5; however CCR5-Unikine-g7 did not significantly increase the frequencies or absolute number of CCR1⁺ and CCR3⁺ cell populations, while hCCL5 and CCR5-n9 both led to significant increase in these two populations compared to CCR5-Unikine-g7 and blank control (Fig.6j,k).

We also investigated which CCR5-positive immune cell populations could be recruited by CCR5-n9 and CCR5-Unikine-g7. Although the relative frequencies of these cell populations were similar across groups, the absolute numbers of mCCR5⁺CD3⁺CD8⁺, mCCR5⁺CD3⁺CD8⁻, mCCR5⁺CD11b⁺ and mCCR5⁺CD11c⁺MHCII⁺ cells were significantly increased following CCR5-Unikine-g7 treatment (Fig.6l,m). These findings further demonstrate that CCR5-Unikine-g7 selectively recruits CCR5 positive immune cells. Together, these findings suggest that selective CCR5 agonists in particular preferentially promote the accumulation of CCR5-expressing immune cells, while avoiding the recruitment of non-CCR5 immune populations associated with the broader receptor activity of CCL5; this increased specificity ostensibly reduced the undesired cross-activation of immunosuppressive cells and enhanced antitumor immunity.

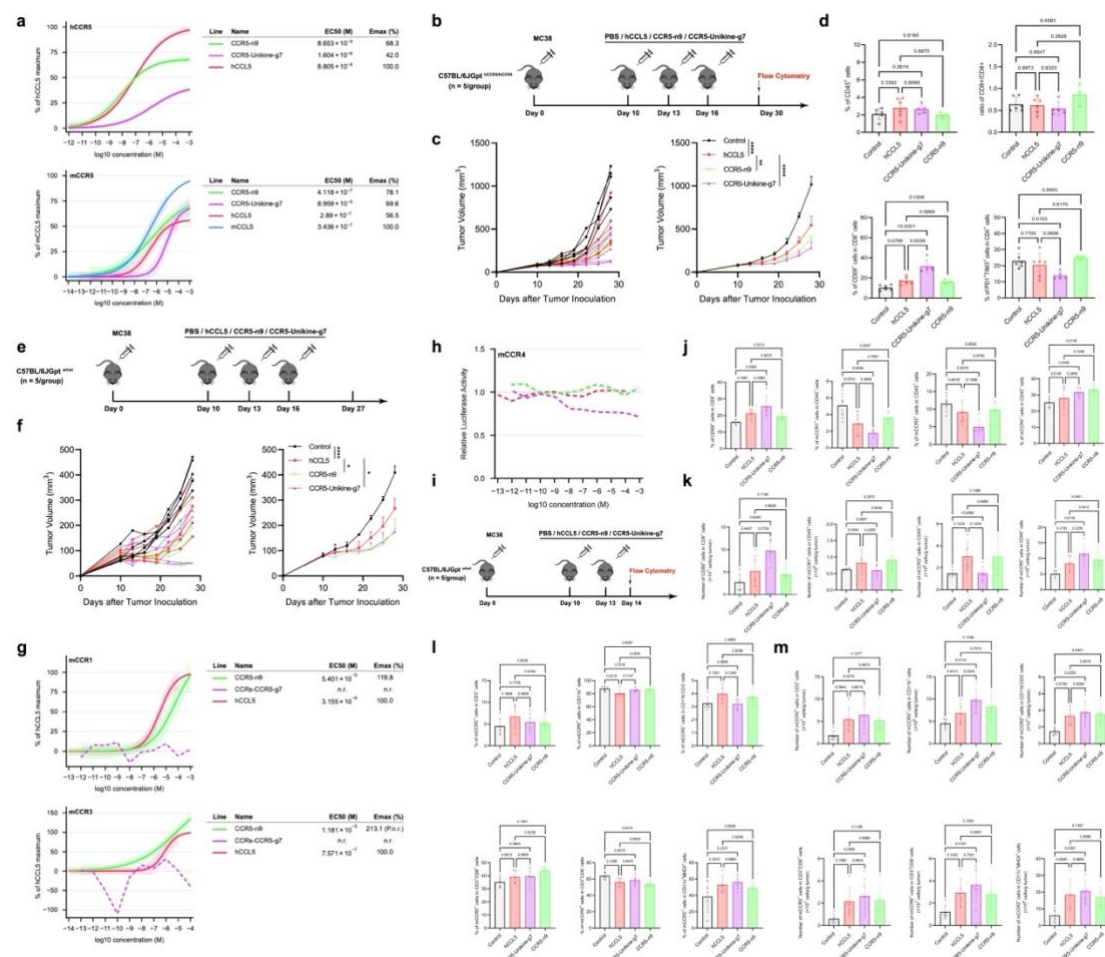


Fig.6 CCR5-n9 and CCR5-Unikine-g7 activate CD8⁺ T cells for effective tumor control.

a, Dose-response curves for activation of human and mouse CCR5, by CCR5-n9 and CCR5-Unikine-g7. EC₅₀ values were determined using the luciferase reporter assay described above. E_{max} values are expressed relative to the

maximal CCL5 response (100%). **b**, Treatment schedule for evaluating the antitumor activity of CCR5 agonists in hCCR5-transgenic mice. MC38 tumor-bearing mice received intratumoral injections of PBS, human CCL5 (hCCL5), CCR5-n9 or CCR5-Unikine-g7 on days 10, 13 and 16 after tumor inoculation (1 μ g per injection; n = 5 mice per group). **c**, Individual (left) and group (right) tumor growth curves for the hCCR5-transgenic mice described in **b**. Tumor volume was monitored every three days. **d**, Flow cytometric analysis of intratumoral T-cell phenotypes in hCCR5-transgenic mice on day 30 after tumor inoculation following CCR5 agonist treatment. Plots show the percentage of CD45⁺ cells, the CD8/CD4 T-cell ratio, CD8⁺CD69⁺ cells and PD-1⁺TIM-3⁺CD4 T cells (n = 3-6 mice per group). **e**, Treatment schedule for evaluating the antitumor activity of CCR5 agonists in wild-type C57BL/6J mice. MC38 tumor-bearing mice received intratumoral injections of PBS, hCCL5, CCR5-n9 or CCR5-Unikine-g7 on days 10, 13 and 16 after tumor inoculation (1 μ g per injection; n = 5 mice per group). **f**, Individual (left) and group (right) tumor growth curves for the wild-type C57BL/6J mice described in **e**. Tumor volume was monitored every three days. **g**, Dose-response curves for activation of mouse CCR1 and CCR3 by CCR5-n9 and CCR5-Unikine-g7. n.r., no detectable agonist response under the tested conditions; EC₅₀ and E_{max} were not determined. P.n.r., plateau not reached within the tested concentration range; the displayed E_{max} is an extrapolated fit estimate. **h**, Dose-response curves for activation of mouse CCR4 by CCR5-n9 and CCR5-Unikine-g7. No CCR4 activation was detected with hCCL5, CCR5-n9 or CCR5-Unikine-g7 under the tested conditions. **i**, Experimental schedule for flow cytometric analysis of tumor-infiltrating immune cells following CCR5 agonist treatment. C57BL/6J mice were inoculated subcutaneously with MC38 cells (1 $\times$ 10⁶ cells per mouse) and received intratumoral injections of PBS, hCCL5, CCR5-n9 or CCR5-Unikine-g7 on days 10 and 13 (1 μ g per injection). Tumors were collected on day 14 for flow cytometry (n = 5 mice per group). **j**, Frequencies of tumor-infiltrating CD8⁺CD69⁺ T cells, CD45⁺mCCR1⁺, CD45⁺mCCR3⁺ and CD45⁺mCCR5⁺ cells. **k**, Absolute numbers of the tumor-infiltrating immune cell populations listed in **j**. **l,m**, Frequencies (l) and absolute numbers (m) of intratumoral mCCR5⁺CD3⁺, mCCR5⁺CD11b⁺, mCCR5⁺CD11b⁻CD3⁻, mCCR5⁺CD3⁺CD8⁺, mCCR5⁺CD3⁺CD8⁻ and mCCR5⁺CD11c⁺MHCII⁺ cells. Data from two independent experiments are presented as mean $\pm$ s.e.m. Statistical significance is indicated by *P* values or the following symbols: ns, not significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

Discussion

In this work, we present GRADE, a lightweight, cross-attention-based language model, and demonstrate efficient and flexible design of peptide agonists for GPCRs. With the on-target activity considered in model construction, GRADE achieves agonist discovery within the first ten or fewer candidates (low-n) for multiple targets, and the agonism spectra of candidate peptides can be customized without incurring overwhelming computational cost, resulting in broader-spectrum agonists as well as highly specific agonists for the exemplary chemokine receptor family. Structurally, GRADE-designed agonists showed orthosteric interactions between agonists and the target receptor, yet with reversed binding polarity relative to endogenous ligands. Lastly in the murine tumor model, the new agonists both demonstrated improved activation of CD8⁺ T cells and stronger immune clearance of tumors, with the CCR5-specific agonist also free of undesired cross-activation of immune-suppressive cells.

Agonist-GPCR interaction in the orthosteric pockets is highly complex and often beyond the accuracy of structural prediction/docking approaches, making novel binder design a difficult task^{5,33}. More importantly, the on-target activities of binders are difficult to incorporate in structure-based approaches, exemplified here by the subtle properties in agonists that drive

GPCR activation beyond simple binding^{6,34}. Available on-target activity or “function-labeled” data are usually limited. For example, the 721 agonist-GPCR pairs used in this study are approximately two orders of magnitude fewer than the 10^5 protein complexes used to train structure-based approaches. This scarcity makes it unlikely that an agonism-oriented model can be constructed based on structure^{19,31}. Instead, relying on sequence information without structure input, we found that GRADE, based on cross-attention mechanism, nevertheless recaptures the implicit biophysical principles of GPCR agonism - that a peptide has the dual ability of binding to the target and contains driver AA(s) inducing target conformational changes³⁵. GRADE offers explainability comparable to that of structure prediction/docking-based approaches, as in the exemplary canonical agonist-GPCR pair GRADE captures the target-engaging interfaces of CCL5 down to residue level.

High success rates in protein binder design are desired to reduce the cost of experimental validation, and publicly available, open-source software have not reached on average double-digit range success rate; towards GPCR this rate is further lower^{5,28,30,36}. Here employing only sequences of GPCR targets and augmented peptide library, GRADE achieved the highest reported success rate in GPCR agonist design, reducing the experimental validation requirement to low-n (10 or fewer) range. Such findings highlight the value of small-scale but function-driven data (agonism), from which cross-attention can predict and design peptides/proteins with specified on-target activity. We demonstrated the ability of GRADE to design agonists for major classes of GPCRs with peptide ligands. In preliminary evaluations of targets without endogenous peptide ligands or orphan targets, GRADE also achieved comparable success rates in these “zero-shot” scenarios. It’s further conceivable that the principle of cross-attention could be transferred for constructing function-aware models for ligand discovery and design, in cases where function-labeled data are scarce: e.g. extended receptor families and peptide/protein ligands, enzymes and peptide inhibitors, antigens and neutralizing antibodies etc^{37–42}.

Activation of a range of GPCR targets is highly desired for clinical applications, including inciting anti-tumor immunity, modulating metabolism etc^{43–46}. CCR5 agonists are expected to elicit antitumor efficacy by increased T-cell infiltration and activation, offering an alternative to conventional pathway blockade such as PD-1/PD-L1^{47,48}. However, endogenous protein/peptide agonists for GPCRs, especially CCRs, commonly face the problem of target promiscuity, which poses an important obstacle for their application in therapeutics¹⁵. This is now overcome in our study as firstly GRADE consumes per design draft only fraction of computation resource of structure-based approaches, and secondly the experimental hit rates for agonist discovery across the tested targets range from 20% to 40%, enabling effective cross-examination on multiple targets. In the exemplary application of designing CCR/CXCR agonists, our new approach GRADE demonstrated its ability to control agonism spectrum of the peptide candidates, and discovered octokine with spectrum surpassing the broadest-acting chemokine, and target-specific CCR5-Unikines. Similarly, other GPCR families in need of more selective agonists for clinical applications, or in cases where synergy is desired among multiple targets, can benefit from customized agonist design using GRADE^{2,49–52}.

Commonly reported biases in extant protein design methods also included the ‘data leakage’ problems, that the designed drafts mimic existing proteins in the training datasets instead of

exploring novel protein spaces⁵³. In GRADE, we found that the validated peptides are distinct both in terms of sequences and in their binding modes. Overall, we found that the new agonists for multiple targets have <20% similarity to the known, endogenous ligands, for both class A and B GPCRs tested in this study; and selected agonists showed comparable potencies in the low concentration range to that of natural ligands, among which CCR5-n8 and CCR5-n9 exhibited lower EC₅₀ values than CCL5, suggesting that GRADE successfully predicts agonism for extremely divergent sequences. Even more strikingly, GRADE-generated CCR5-specific peptide agonists lacking cysteine residues activated CCR5, greatly expanding the agonist sequence space beyond known CCLs/CXCLs with N-terminal cysteines⁹. In the resolved structure of agonists-CCR5 pairs, GRADE-discovered agonists act as chemokine mimetics, but used the C-terminus instead of the N-terminus shared by all known chemokines in binding with the receptors; the ability of ‘inventing’ novel mode of binding and activation further supports that GRADE accurately models the interactions of completely new ligands and GPCRs without requiring structure prediction or docking^{9,54,55}.

In conclusion, GRADE built on cross-attention mechanism show-cases the ability to design protein binders with designed function beyond simple binding, and offers accurate, flexible approach for coming up with novel ligands for GPCRs of interest. Further expansion both within the GPCR protein families as well as in other families of interest will generate important peptide drug candidates, improve the efficacy of established targets as well as elucidate the potential of orphan targets for the development of therapies^{56,57}.

Code availability

The GRADE predictions reported in this study can be reproduced using the resources available in the GitHub repository (<https://github.com/devxia/GRADE>), which also provides environment setup instructions and a usage guide. Source code and model weight will be made public soon.

References

1. Zhang, M. *et al.* G protein-coupled receptors (GPCRs): advances in structures, mechanisms and drug discovery. *Signal Transduct. Target. Ther.* **9**, 88 (2024).
2. Lorente, J. S. *et al.* GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* **24**, 458–479 (2025).
3. Wu, F., Song, G., de Graaf, C. & Stevens, R. C. Structure and Function of Peptide-Binding G Protein-Coupled Receptors. *J. Mol. Biol.* **429**, 2726–2745 (2017).
4. Andersen, A., Lund, A., Knop, F. K. & Vilsbøll, T. Glucagon-like peptide 1 in health and disease. *Nat. Rev. Endocrinol.* **14**, 390–403 (2018).
5. Muratspahić, E. *et al.* De novo design of miniproteins targeting GPCRs. *Nature* **656**, 1044–1053 (2026).

6. Bender, A. *et al.* Artificial intelligence in drug discovery — what it is, where we stand and the path forward. *Nat. Rev. Drug Discov.* <https://doi.org/10.1038/s41573-026-01496-2> (2026) doi:10.1038/s41573-026-01496-2.
7. Herrera, L. P. T. *et al.* GPCRdb in 2025: adding odorant receptors, data mapper, structure similarity search and models of physiological ligand complexes. *Nucleic Acids Res.* **53**, D425–D435 (2025).
8. Winniffrith, A., Outeiral, C. & Hie, B. L. Generative artificial intelligence for de novo protein design. *Curr. Opin. Struct. Biol.* **86**, 102794 (2024).
9. Kleist, A. B. *et al.* Encoding and decoding selectivity and promiscuity in the human chemokine-GPCR interaction network. *Cell* **188**, 3603–3622.e27 (2025).
10. Hedderich, J. B. *et al.* The pocketome of G-protein-coupled receptors reveals previously untargated allosteric sites. *Nat. Commun.* **13**, 2567 (2022).
11. Bachelerie, F. *et al.* New nomenclature for atypical chemokine receptors. *Nat. Immunol.* **15**, 207–208 (2014).
12. Chen, K. *et al.* Chemokines in homeostasis and diseases. *Cell. Mol. Immunol.* **15**, 324–334 (2018).
13. Allen, S. J., Crown, S. E. & Handel, T. M. Chemokine: Receptor Structure, Interactions, and Antagonism. *Annu. Rev. Immunol.* **25**, 787–820 (2007).
14. Schall, T. J., Bacon, K., Toy, K. J. & Goeddel, D. V. Selective attraction of monocytes and T lymphocytes of the memory phenotype by cytokine RANTES. *Nature* **347**, 669–671 (1990).
15. Horuk, R. Chemokine receptor antagonists: overcoming developmental hurdles. *Nat. Rev. Drug Discov.* **8**, 23–33 (2009).
16. Dorgham, K. *et al.* Engineered CCR5 superagonist chemokine as adjuvant in anti-tumor DNA vaccination. *Vaccine* **26**, 3252–3260 (2008).
17. Chen, L. T. *et al.* Target sequence-conditioned design of peptide binders using masked language modeling. *Nat. Biotechnol.* **44**, 1002–1010 (2026).
18. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
19. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
20. Zhang, J. *et al.* Predicting protein-protein interactions in the human proteome. *Science* **390**, eadt1630 (2025).
21. Hu, W. & Ohue, M. SpatialPPIv2: Enhancing protein–protein interaction prediction through graph neural networks with protein language models. *Comput. Struct. Biotechnol. J.* **27**, 508–518 (2025).
22. Liu, D. *et al.* PLM-interact: extending protein language models to predict protein-protein

interactions. *Nat. Commun.* **16**, 9012 (2025).

23. Ullanat, V., Jing, B., Sledzieski, S. & Berger, B. Learning the language of protein-protein interactions. *Nat. Commun.* **17**, 1199 (2026).

640 24. Sledzieski, S., Singh, R., Cowen, L. & Berger, B. D-SCRIPT translates genome to phenome with sequence-based, structure-aware, genome-scale predictions of protein-protein interactions. *Cell Syst.* **12**, 969-982.e6 (2021).

25. Singh, R., Devkota, K., Sledzieski, S., Berger, B. & Cowen, L. Topsy-Turvy: integrating a global view into sequence-based PPI prediction. *Bioinformatics* **38**, i264–i272 (2022).

645 26. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

27. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

650 28. Pacesa, M. *et al.* One-shot design of functional protein binders with BindCraft. *Nature* **646**, 483–492 (2025).

29. Peng, X. *et al.* Unified modeling of 3D molecular generation via atomic interactions with PocketXMol. *Cell* **189**, 1904-1922.e28 (2026).

30. Stark, H. *et al.* BoltzGen: Toward Universal Binder Design. Preprint at <https://doi.org/10.1101/2025.11.20.689494> (2026).

655 31. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).

32. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).

660 33. Congreve, M., de Graaf, C., Swain, N. A. & Tate, C. G. Impact of GPCR Structures on Drug Discovery. *Cell* **181**, 81–91 (2020).

34. Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).

665 35. Ferguson, L. *et al.* DeorphaNN: Virtual screening of GPCR peptide agonists using AlphaFold-predicted active-state complexes and deep learning embeddings. *Mol. Cell* **86**, 3102-3117.e5 (2026).

36. Mille-Fragoso, L. S. *et al.* Efficient generation of epitope-targeted antibodies with Germinal. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-026-03187-0> (2026) doi:10.1038/s41587-026-03187-0.

670 37. Newstead, S. Future opportunities in solute carrier structural biology. *Nat. Struct. Mol. Biol.* **31**, 587–590 (2024).

38. De Bosscher, K., Desmet, S. J., Clarisse, D., Estébanez-Perpiña, E. & Brunsveld, L. Nuclear receptor crosstalk — defining the mechanisms for therapeutic innovation. *Nat. Rev.*

Endocrinol. **16**, 363–377 (2020).

39. Drew, D. & Boudker, O. Shared Molecular Mechanisms of Membrane Transporters. *Annu. Rev. Biochem.* **85**, 543–572 (2016).
40. Singh, K. *et al.* Recent Advances in Enzyme Inhibition: A Pharmacological Review. *Curr. Enzyme Inhib.* **20**, 2–19 (2024).
41. Patadiya, N., Panchal, N. & Vaghela, V. A REVIEW ON ENZYME INHIBITORS. *Int. Res. J. Pharm.* **12**, 60–66 (2021).
42. Hummer, A. M., Abanades, B. & Deane, C. M. Advances in computational structure-based antibody design. *Curr. Opin. Struct. Biol.* **74**, 102379 (2022).
43. Rozengurt, E., Sinnett-Smith, J. & Kisfalvi, K. Crosstalk between Insulin/Insulin-like Growth Factor-1 Receptors and G Protein-Coupled Receptor Signaling Systems: A Novel Target for the Antidiabetic Drug Metformin in Pancreatic Cancer. *Clin. Cancer Res.* **16**, 2505–2511 (2010).
44. Fu, Q., Shi, Q., West, T. M. & Xiang, Y. K. Cross-Talk Between Insulin Signaling and G Protein-Coupled Receptors. *J. Cardiovasc. Pharmacol.* **70**, 74–86 (2017).
45. Garg, P., Jallepalli, V. R. & Verma, S. Unravelling the CXCL12/CXCR4 Axis in breast cancer: Insights into metastasis, microenvironment interactions, and therapeutic opportunities. *Hum. Gene* **40**, 201272 (2024).
46. Ma, Y. *et al.* Structure-guided discovery of a single-domain antibody agonist against human apelin receptor. *Sci. Adv.* **6**, eaax7379 (2020).
47. Li, Y. *et al.* CCR5 expression defines the functional heterogeneity of tumor-infiltrating CD8⁺ T cells and predicts enhanced antitumor immunity and favorable prognosis in non-small cell lung cancer. *Clin. Transl. Oncol.* **28**, 2068–2085 (2026).
48. Woodward Davis, A. S. *et al.* The human tissue-resident CCR5⁺ T cell compartment maintains protective and functional properties during inflammation. *Sci. Transl. Med.* **11**, eaaw8718 (2019).
49. Powers, A. S. *et al.* Structural basis of efficacy-driven ligand selectivity at GPCRs. *Nat. Chem. Biol.* **19**, 805–814 (2023).
50. Huggins, D. J., Sherman, W. & Tidor, B. Rational Approaches to Improving Selectivity in Drug Design. *J. Med. Chem.* **55**, 1424–1444 (2012).
51. Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J. & Bork, P. Drug Target Identification Using Side-Effect Similarity. *Science* **321**, 263–266 (2008).
52. Santhanam, B., Sluter, M. & Babu, M. M. Exploring GPCR signaling pathway networks as cancer therapeutic targets. *Cell Genomics* **4**, 100560 (2024).
53. Mattsson, B. & Walters, W. P. Identifying and Addressing Systematic Data Leakage in Protein-Ligand Affinity Benchmarks. Preprint at <https://doi.org/10.64898/2026.06.29.735309> (2026).

- 710 54. Wu, B. *et al.* Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science* **330**, 1066–1071 (2010).
55. Barington, L. *et al.* Role of Conserved Disulfide Bridges and Aromatic Residues in Extracellular Loop 2 of Chemokine Receptor CCR8 for Chemokine and Small Molecule Binding. *J. Biol. Chem.* **291**, 16208–16220 (2016).
- 715 56. Smith, N. J. & Murray, F. Shifting our perspective on orphan G protein-coupled receptors. *Nat. Struct. Mol. Biol.* **31**, 582–583 (2024).
57. Meixner, E. *et al.* A substrate-based ontology for human solute carriers. *Mol. Syst. Biol.* **16**, e9652 (2020).
58. Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. in *Advances in Neural Information Processing Systems 32* (eds Wallach, H. *et al.*) 8024–8035 (Curran Associates, Inc., 2019).
- 720 59. Wolf, T. *et al.* Transformers: State-of-the-Art Natural Language Processing. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* 38–45 (Association for Computational Linguistics, 2020). doi:10.18653/v1/2020.emnlp-demos.6.
- 725 60. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
61. Ferruz, N., Schmidt, S. & Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* **13**, 4348 (2022).
62. Emenecker, R. J., Griffith, D. & Holehouse, A. S. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **120**, 4312–4319 (2021).
- 730 63. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R. & Warwicker, J. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33**, 3098–3100 (2017).
64. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
- 735 65. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
66. DiBerto, J. F., Smart, K., Olsen, R. H. J. & Roth, B. L. Agonist and antagonist TRUPATH assays for G protein-coupled receptors. *STAR Protoc.* **3**, 101259 (2022).
- 740 67. Olsen, R. H. J. *et al.* TRUPATH, an open-source biosensor platform for interrogating the GPCR transducerome. *Nat. Chem. Biol.* **16**, 841–849 (2020).
68. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
69. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
- 745 70. Chai Discovery. Chai-1: Decoding the molecular interactions of life. Preprint at

<https://doi.org/10.1101/2024.10.10.615955> (2024).

71. Zhang, H. *et al.* Structural basis for chemokine recognition and receptor activation of chemokine receptor CCR5. *Nat. Commun.* **12**, 4151 (2021).

72. Draper-Joyce, C. J. *et al.* Structure of the adenosine-bound human adenosine A1 receptor–
750 Gi complex. *Nature* **558**, 559–563 (2018).

73. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

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

755 75. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

76. 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).

77. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in
760 cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).

78. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

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

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

81. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).

82. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
770