

A unique peptide recognition mechanism by the human relaxin family peptide receptor 4 (RXFP4)

Yan Chen^{1†}, Qingtong Zhou^{1†}, Jiang Wang^{2,3,4†}, Youwei Xu⁵, Yun Wang⁶, Jiahui Yan^{5,7,8}, Yibing Wang², Qi Zhu⁶, Fenghui Zhao^{5,7}, Chenghao Li^{2,4}, Chuan-Wei Chen⁹, Xiaoqing Cai^{5,7}, Ross A.D. Bathgate^{10,11}, Chun Shen⁶, H. Eric Xu^{5,8,12}, Dehua Yang^{5,7,8,9,*}, Hong Liu^{2,4,*}, Ming-Wei Wang^{1,5,7,8,9,13,*}

¹Department of Pharmacology, School of Basic Medical Sciences, Fudan University, Shanghai 200032, China; ²State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; ³Lingang Laboratory, Shanghai 200031, China; ⁴School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, UCAS, Hangzhou 310024, China; ⁵The CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; ⁶Genova Biotech (Changzhou) Co., Ltd, Changzhou 213125, China; ⁷The National Center for Drug Screening, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; ⁸University of Chinese Academy of Sciences, Beijing 100049, China; ⁹Research Center for Deepsea Bioresources, Sanya, Hainan 572025, China; ¹⁰The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria 3052, Australia; ¹¹Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3052, Australia; ¹²School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China; ¹³Department of Chemistry, School of Science, The University of Tokyo, Tokyo 113-0033, Japan

†These authors contributed equally: Yan Chen, Qingtong Zhou, Jiang Wang

*e-mail: hliu@simm.ac.cn (H.L.); dhyang@simm.ac.cn (D.H.Y.); mwwang@simm.ac.cn (M.-W.W)

Abstract

Members of the insulin superfamily regulate a variety of biological processes through two types of target-specific but structurally conserved peptides, insulin/insulin-like growth factors and relaxin/insulin-like peptides. The latter bind to the human relaxin family peptide receptors (RXFPs), which are class A G protein-coupled receptors (GPCRs), to exert pleiotropic actions. Here, we report three cryo-electron microscopy structures of RXFP4–Gi protein complexes in the presence of the endogenous ligand insulin-like peptide 5 (INSL5) or one of the two small molecule agonists, compound 4 and DC591053, both were discovered through medicinal chemistry efforts. The B chain of INSL5 adopts a single α -helix that penetrates into the orthostatic pocket, while the A chain sits above the orthosteric pocket to interact with the extracellular surface of RXFP4, revealing a unique

peptide-binding mode previously unknown. Together with mutagenesis and functional analyses, the key determinants responsible for the peptidomimetic agonism and subtype selectivity were identified. DC591053 selectively mimicked the action of INSL5 at RXFP4 whereas compound 4 activated both RXFP3 and RXFP4. Comparison of peptide binding modes within the insulin superfamily displayed diverse interaction mechanisms distinct to each type of the peptides. Our findings not only provide valuable insights into ligand recognition and subtype selectivity among class A GPCRs, but also expand the knowledge of signaling mechanisms in the insulin superfamily.

Key words

Insulin superfamily, INSL5, peptidomimetics, RXFP4, selectivity

Introduction

The human relaxin family peptide (RXFP) receptors (RXFP1, RXFP2, RXFP3 and RXFP4) play physiological roles through peptide hormones relaxin, insulin-like peptide 3 (INSL3), relaxin-3, and insulin-like peptide 5 (INSL5), respectively¹. These peptides exert pleiotropic actions covering reproduction, cardiovascular adaptation, stress responses, metabolic control, colon motility and behavioral processes¹, thereby showing the therapeutic potential for a variety of disorders. Different from RXFP1 and RXFP2 that share a large extracellular domain containing 10 leucine-rich repeats (LRR) and a unique low-density lipoprotein class A module (LDL-A)²⁻⁵, RXFP3 and RXFP4 have distinct binding properties with a relatively short N-terminal tail rather than LRR. They possess 43% sequence identity and inhibit cAMP production via pertussis toxin-sensitive $\text{G}\alpha_{i/o}$ proteins⁶.

RXFP4, also known as GPCR142 or GPR100, is primarily distributed in peripheral tissues with the highest expression in the colorectum^{6,7}. Its endogenous ligand INSL5, secreted by the colonic L-cell, was originally identified as an incretin albeit with some controversies⁶⁻⁹. Their expression pattern together with impaired glucose and fat control shown in INSL5 or RXFP4 deficient mice indicate their involvement in energy metabolism^{6,7,10,11}. INSL5 has also been described as an orexigenic hormone¹² and RXFP4 was implicated in colon motility^{13,14}, colorectal cancer and nasopharyngeal carcinoma^{15,16}.

Despite these advances, difficulties in obtaining sufficient quantities of native INSL5 hampered our efforts in further exploring the biology of the peptide and its cognate receptor. Since relaxin-3 also binds to and activates RXFP4 *in vitro*⁴, it has been used as a surrogate ligand to study potential actions of INSL5 due to their shared tertiary structure closely related to insulin including two chains and three disulfide bonds¹⁷. In addition to peptidic analogues, small molecule modulators have been reported in recent years. Compound 4, an amidino hydrazone-based scaffold identified by Novartis, is an RXFP3/RXFP4 dual agonist¹⁸. However, selective RXFP4 agonists discovered via high-

throughput screening campaigns and follow-up structural modifications displayed deficiencies in solubility, potency and toxicity^{19,20}. This promoted us to develop a small molecule agonist (DC591053) with better affinity and selectivity for RXFP4 as presented in this study.

Here, we report three cryogenic electron microscopy (cryo-EM) structures of the human RXFP4–G_i complexes bound to INSL5, compound 4 and DC591053 with global resolutions of 3.19 Å, 3.03 Å and 2.75 Å, respectively. Together with mutagenesis and functional analyses, we describe a unique peptide-binding mechanism previously unseen in other class A G protein-coupled receptors (GPCRs) and provide useful information for structure-based design of RXFP4 agonists either as research probes or as drug candidates.

Results

Characterization of recombinant INSL5

The purity of recombinant INSL5 was over 90% by reverse phase high-performance liquid chromatography (RP-HPLC) and the molecular weight was determined to be 5,061.2 Da by mass spectrometry (MS), equivalent to that of native INSL5 peptide (5,061.8 Da; N-terminal Q of A chain not converted to pE) (**Supplementary Fig. 1**). As depicted in **Supplementary Fig. 1**, chymotrypsin cleavage resulted in 8 major peaks (labeled as ①-⑧) on RP-HPLC. The measured molecular masses of the individual peaks were identical to the theoretical values of the expected chymotrypsin-generated peptides, which allowed for 100% sequence coverage. The recombinant INSL5 peptide was subsequently verified for its bioactivity in Chinese hamster ovary (CHO-K1) cells stably transfected with RXFP4. As shown in **Supplementary Fig. 1**, it bound to RXFP4 with high affinity and was able to inhibit forskolin-induced cAMP responses ($pEC_{50} = 8.80 \pm 0.11$, $n = 3$; $pKi = 8.19 \pm 0.06$, $n = 3$, as measured in stably-transfected CHO-K1 cells) compared to the native INSL5 standard.

Characterization of DC591053

We screened our in-house tetrahydroisoquinoline library aimed at discovering novel RXFP4 agonists using cAMP accumulation assay. The lead compound, DC591053 ((*S*)-(7-ethoxy-6-methoxy-1-(2-(5-methoxy-1H-indol-3-yl)ethyl)-3,4-dihydroisoquinolin-2(1H)-yl)(morpholino)methanone), was identified and synthesized from the commercially available compound 4-hydroxy-3-methoxybenzaldehyde, followed by alkylation reaction, reduction, Wittig reaction, cyclization, asymmetric reduction reaction, and condensation reaction (**Supplementary Fig. 2a**). DC591053 is a white solid that characterized by ¹H, ¹³C NMR and high-resolution mass spectra (HRMS) and determined to be 96.9% pure by column chromatography analyses (¹H NMR

(500 MHz, DMSO-*d*₆) δ 10.58 (s, 1H), 7.21 (d, *J* = 8.7 Hz, 1H), 7.10 (d, *J* = 2.0 Hz, 1H), 6.93 (d, *J* = 2.3 Hz, 1H), 6.70 (dd, *J* = 8.7, 2.4 Hz, 1H), 6.64 (d, *J* = 5.6 Hz, 2H), 4.91 – 4.83 (m, 1H), 3.88 (q, *J* = 7.0 Hz, 2H), 3.74 (s, 3H), 3.69 (s, 3H), 3.57 (ddd, *J* = 9.1, 6.2, 2.8 Hz, 2H), 3.54 – 3.47 (m, 2H), 3.42 – 3.32 (m, 2H), 3.17 (ddd, *J* = 12.5, 6.1, 2.5 Hz, 2H), 3.03 (ddd, *J* = 12.8, 6.0, 2.6 Hz, 2H), 2.71 (ddd, *J* = 27.7, 13.5, 6.7 Hz, 3H), 2.64 – 2.56 (m, 1H), 2.07 (q, *J* = 12.7, 10.3 Hz, 2H), 1.26 (t, *J* = 7.0 Hz, 3H). ¹³C NMR (125 MHz, DMSO-*d*₆) δ 163.40, 152.87, 147.47, 146.11, 131.51, 130.04, 127.33, 125.40, 122.99, 113.58, 112.01, 111.98, 111.71, 110.87, 100.12, 65.84, 63.71, 55.40, 55.31, 53.95, 47.40, 36.55, 27.53, 21.84, 14.72. ESI-LRMS *m/z* 494.2 [M+H]⁺. ESI-HRMS *m/z* calcd for C₂₈H₃₆N₃O₅ [M+H]⁺ 494.2649, found 494.2650.) (**Supplementary Fig. 2b-d**). DC591053 demonstrated full agonism at RXFP4 both in competitive europium (Eu)-labelled R3/I5 binding and cAMP accumulation assays (*pEC*₅₀ = 7.24 ± 0.12, *n* = 3; *pKi* = 6.95 ± 0.14, *n* = 3, as measured in stably-transfected CHO-K1 cells). Importantly, DC591053 neither reacted with related RXFP3 nor with parental CHO-K1 cells (**Supplementary Fig. 2e-g**).

Overall structures

To prepare a high-quality human RXFP4–G_i complex, we added a haemagglutinin (HA) signal peptide to enhance receptor expression, followed by a 10× histidine tag as well as cytochrome b562RIL (BRIL) insertion at the N terminus, and applied the NanoBiT tethering strategy (**Supplementary Fig. 3a**)²¹⁻²³. The activity of the modified RXFP4 construct was confirmed by cAMP accumulation assay showing a response similar to that of the wild-type (WT). These complexes were then purified, resolved as monodispersed peaks on size-exclusion chromatography (SEC) and verified by SDS gel to ascertain all the expected components (**Supplementary Fig. 3b-d**). After sample preparation, cryo-EM data were collected, analyzed and 3-dimensional (3D) consensus density maps reconstructed (**Supplementary Fig. 4**) resulting in an overall resolution of 3.19 Å, 3.03 Å and 2.75 Å for the INSL5–RXFP4–G_i, compound 4–RXFP4–G_i and DC591053–RXFP4–G_i complexes, respectively (**Fig. 1, Supplementary Table 1**). These maps allowed us to build near-atomic level models for most regions of the complexes except for the flexible α-helical domain (AHD) of G_i, the N terminus (M1 to K34) and the intracellular loop (ICL) 1 between N66 to P72 of RXFP4 (**Fig. 1, Supplementary Fig. 5**). Because of the relatively high resolution of the three structures, the RXFP4-bound INSL5, compound 4 and DC591053 were well-defined in the EM density maps.

These structures share a similar conformation with root mean squared deviation (RMSD) of < 0.5 Å, including a hallmark outward movement of the intracellular half of transmembrane helix (TM) 6 relative to the X-ray structures of inactive β₂-adrenergic receptor or cholecystokinin A

receptor (CCK_{AR})²⁴⁻²⁶ (**Supplementary Fig. 6**) and a β -hairpin occurred in the second extracellular loop (ECL2) that is similar to the peptide-bound class A GPCR structures such as CCK_{AR}²⁵, cholecystokinin B receptor (CCK_{BR})²⁵, type 1 bradykinin receptor (B1R)²⁷, type 2 bradykinin receptor (B2R)²⁷ and C-C chemokine receptor type 1 (CCR1)²⁸. One significant difference is that INSL5 displayed a new binding mode to the cognate receptor (**Supplementary Fig. 7**): its C-terminal α -helix of the B chain penetrated into the transmembrane domain (TMD) core, such that the two terminus residues R23^B (B indicates that the residue belongs to the B chain of INSL5) and W24^B fully occupied the orthosteric pocket, while the A chain strengthened the binding by restraining the movement of the B chain through two inter-chain disulfide bonds (C8^A-C7^B and C21^A-C19^B) (**Fig. 2a**). Both compound 4 and DC591053 displayed a peptidomimetic feature by structurally and spatially mimicking the C-terminal tryptophan (W24^B) as a common chemotype; with ligand-specific recognition by TM5, TM7 and ECL2 to confer distinct subtype selectivity of RXFP4 over RXFP3 (**Fig. 3a, f, Supplementary Table 2**).

Peptide recognition

INSL5 anchored in the RXFP4 orthosteric binding pocket bordered by TMs 2-7 and ECLs 1-3, with its B chain inserting into the TMD bundle and contributing a majority of the receptor interaction sites, while the A chain docked above the orthosteric pocket and interacted with ECL2, ECL3 and solvent (**Fig. 2a-e**). Consistently, the interface area between RXFP4 and the B chain (1444 \AA^2) is significantly larger than that of the A chain (351 \AA^2).

The B chain of INSL5 exhibited a single amphipathic α -helix conformation²⁹ from E10^B to W24^B, with the C terminus W24^B being the deepest residue into the receptor core. The N-terminal residues (R5^B to L9^B) adopted a loop that is clasped by a short N-terminal α -helix of the A chain through one disulfide bond (C8^A-C7^B). W24^B contributed massive polar and nonpolar interactions to stabilize the peptide binding via both side chain indole and the carboxylic acid group. The former made a hydrogen bond with T121^{3,32} (superscripts denote Ballesteros-Weinstein numbering³⁰), as well as cation- π stacking with R208^{5,42} and π - π stacking with W97^{2,60}, F291^{7,35} and H299^{7,43}, while the latter pointed to TM5 with the formation of one hydrogen bond (via Q205^{5,39}) and one salt bridge (via R208^{5,42}) (**Fig. 2b**). These observations support the importance of a free carboxyl group in the B chain C terminus for high-affinity RXFP4 binding and signaling activity³¹, consistent with our mutagenesis studies showing that INSL5-induced cAMP responses were completely abolished in mutants T121^{3,32}A and R208^{5,42}A, profoundly reduced in mutant H299^{7,43}A (E_{\max} value by 70%) or markedly diminished in mutants W97^{2,60}A and Q205^{5,39}A by 20.4-fold and 5.3-fold, respectively (**Fig. 2f-g, Supplementary Table 4**). In addition, alanine replacement of W24^B and amidation of

the B chain C terminus significantly reduced INSL5-elicited agonistic activity as previously described³². Another important residue is R23^B whose side chain oriented towards TM2 and formed one salt bridge with E100^{2,63}. Mutation of E100^{2,63} to alanine (**Fig. 2f, Supplementary Table 4**) or arginine³³ both deprived the ability of INSL5 to activate RXFP4, in line with the reduced potencies reported for R23^BA³² or R23^BE³³. Interestingly, diverse peptide-receptor contacts were observed for the residue at 2.63 depending on physicochemical properties including positively charged [*e.g.*, R102^{2,63} in growth hormone secretagogue receptor (GHSR) and R84^{2,63} in formyl peptide receptor 1 (FPR1)] and negatively charged amino acids [*e.g.*, D93^{2,63} in C-X-C chemokine receptor type 4 (CXCR4)]. Besides the polar contacts, R13^B and S21^B made one salt bridge and one hydrogen bond with the side chain of D104^{2,67} and backbone oxygen of Q287^{7,31}, respectively, while Y17^B was stabilized by the π - π stacking from F105^{ECL1} (**Fig. 2c-d**). The hydrophobic residues in the B chain further strengthened INSL5 binding by hydrophobic contacts with both RXFP4 residues (F105^{ECL1}, V185^{ECL2}, C186^{ECL2}, V188^{ECL2}, L192^{45,51}, K273^{6,62}, W279^{ECL3} and Y284^{7,28}) and the A chain residues (L3^A and L17^A) via Y11^B, V15^B, I16^B, I18^B, C19^B and A20^B (**Fig. 2c-d**). Disruption of these hydrophobic contacts through mutants F105^{ECL1}A, K273^{6,62}A, W279^{ECL3}A and Y284^{7,28}A moderately decreased both potency and E_{max} of INSL5 (**Fig. 2f-g, Supplementary Table 4**), supported by weak or moderate decreases in binding affinity and potency when the B chain residues I12^B, V15^B, I16^B and I18^B were mutated to alanine³². Consistently, molecular dynamics (MD) simulations found that the C-terminal α -helix of the B chain could stably insert into the orthosteric pocket through its tip residues, evidenced by the interface area and representative minimum distances (R13^B–D104^{2,67}, R23^B–E100^{2,63} and W24^B–Q205^{5,39}/R208^{5,42}) (**Supplementary Fig. 8**).

Different from the binding mode of B chain that was largely buried by the TMD bundle, A chain solely interacted with several residues in ECL2 and ECL3 forming one salt bridge (via the side chain of K273^{6,62}), one hydrogen bond (via the side chain of R194^{ECL2}) and multiple hydrophobic contacts (via V185^{ECL2}, V277^{ECL3} and W279^{ECL3}) (**Fig. 2e**). As expected, alanine substitutions at K273^{6,62}, R194^{ECL2} and W279^{ECL3} modestly reduced INSL5 potency by 4.7-fold, 2.2-fold and 2.5-fold, respectively (**Fig. 2f-g, Supplementary Table 4**). Instead of direct interaction with RXFP4, A chain is likely to stabilize peptide binding by restraining the dynamics of INSL5 through three disulfide bonds and a hydrophobic patch (L3^A, L6^A, L17^A, L20^A, Y11^B, V15^B and I18^B), thereby maintaining the correct conformation of INSL5 for RXFP4 recognition and reducing the entropy cost during peptide binding. Functional and MD simulation studies are in agreement with this observation as deletion of the A chain completely abolished receptor binding and signaling activities of INSL5³⁴ (**Supplementary Fig. 9**), suggesting that B chain alone is not sufficient to sustain the α -helix conformation.

Receptor selectivity

Strong electron densities were observed for compound 4 from the orthosteric site of RXFP4 to ECL2, revealing a C-shaped conformation of compound 4, with the indole ring inserting deeply into the orthosteric binding pocket and its chlorobenzene moiety extending to the extracellular side (**Fig. 3a**). By displaying a conformation similar to the C-terminal residue W24^B of INSL5, the indole ring of compound 4 showed strong interactions with RXFP4 residues, forming two hydrogen bonds (via T295^{7.39} and H299^{7.43}), stacking contacts (via W97^{2.60}, R208^{5.42} and F291^{7.35}) and hydrophobic contacts (via L118^{3.29}, T121^{3.32} and V122^{3.33}). The central guanidine moiety was positively charged to mimic R23^B of INSL5 and made one salt bridge with the negatively charged side chain of E100^{2.63} as well as cation- π stacking interactions with F105^{ECL1}. The chlorobenzene group covered the orthostatic site and was close to ECL2 with the formation of multiple hydrogen bonds (via the backbone oxygen atom of L193^{45.52} and R194^{ECL2}) and hydrophobic contacts (via L192^{45.51} and L193^{45.52}) (**Fig. 3b**). Mutagenesis and structure-activity relationship (SAR) studies support these observations: mutants W97^{2.60}A, E100^{2.63}A, and H299^{7.43}A abolished cAMP responses, while T121^{3.32}A and R208^{5.42}A significantly impaired the potency of compound 4 by 20.1-fold and 6.6-fold, respectively (**Fig. 3c, Supplementary Table 4**); substitution of hydroxy by methoxyl at the indole 5 position or replacement of ethyl by a smaller methyl at the indole 7 position eliminated the hydrogen bonds with TM7 residues and weakened hydrophobic contacts with TM3 residues, respectively, thereby reducing the agonist potencies as reported previously¹⁸.

Since the sequence identity of the ligand-binding pocket between RXFP3 and RXFP4 is 86.36%, development of receptor subtype-selective ligands is significant challenge. Only six pocket residues are diversified: S159^{3.29}, S163^{3.33}, V249^{45.52}, H268^{5.39}, K271^{5.42} and V375^{7.39} for RXFP3, and L118^{3.29}, V122^{3.33}, L193^{45.52}, Q205^{5.39}, R208^{5.42} and T295^{7.39} for RXFP4. Compound 4 formed one hydrogen bond with the side chain of T295^{7.39} that is unlikely to occur in the equivalent position of RXFP3 (V375^{7.39}). However, two distinct amino acids in TM5 (Q205^{5.39}, R208^{5.42} for RXFP4 and H268^{5.39}, K271^{5.42} for RXFP3³⁵) were not contacted, which may limit the subtype selectivity. To overcome this hurdle, DC591053 was developed demonstrating a full agonism at RXFP4 ($pEC_{50} = 7.24 \pm 0.12$) without observable cross-reactivity with RXFP3 (**Fig. 3e**).

As shown in **Fig. 3f**, the indole ring of DC591053 occupied the orthosteric pocket in a similar manner as W24^B of INSL5 and compound 4. It also stabilized the RXFP4–G_i complex by stacking interactions with W97^{2.60}, R208^{5.42}, F291^{7.35} and H299^{7.43} as well as hydrophobic contacts with L118^{3.29}, T121^{3.32} and V122^{3.33} (**Fig. 3g**). Mutants W97^{2.60}A and T121^{3.32}A suppressed the ability of RXFP4 to inhibit cAMP production upon DC591053 stimulation (by 1.6-fold and 20.9-fold,

respectively), and H299^{7.43}A seriously affected the E_{max} value (reduced by 65%) (**Fig. 3h, Supplementary Table 4**). The methoxyl at the indole 5 position of DC591053 pointed towards TM7 with the formation of one hydrogen bond (via T295^{7.39}). Different from compound 4, the morpholine ring rendered DC591053 to form two moderate hydrogen bonds with Q205^{5.39} and R208^{5.42}, *i.e.*, an RXFP4-specific edge in the ligand-binding pocket, which may enhance the selectivity for RXFP4 (**Fig. 3f-g**). Consistently, R208^{5.42}A decreased the potency of DC591053 by 7.6-fold (**Fig. 3h, Supplementary Table 4**). Another notable difference is the replacement of guanidine moiety in compound 4 and R23^B of INSL5 by the urea group in DC591053, which is unlikely to make polar interactions with E100^{2.63}, in agreement with unchanged agonism of DC591053 at mutant E100^{2.63}A whose signaling is abolished for INSL5 and compound 4. To compensate for the contact gap caused by the above replacement, the tetrahydroisoquinoline moiety of DC591053 contributed multiple stacking interactions with F105^{ECL1}, R194^{ECL2} and F291^{7.35} and hydrophobic contacts with L190^{ECL2}, L192^{45.51} and P292^{7.36} (**Fig. 3g**), which are significantly stronger than that of compound 4. Removal of these contacts by mutants F105^{ECL1}A and R194^{ECL2}A reduced DC591053 potency by 4.9-fold and 8.1-fold, respectively (**Fig. 3h, Supplementary Table 4**). To further explore subtype selectivity, we performed an amino acid switch in equivalent positions between RXFP4 and RXFP3 around the ligand-binding pocket. Double mutant L118^{3.29}S/V122^{3.33}S in RXFP4 selectively affected the potency of DC591053 by 20.9-fold without notable influence on that of compound 4. As a comparison, S159^{3.29}L/S163^{3.33}V in RXFP3 reduced the potency of compound 4 by 26.9-fold. Similar phenomena were also observed in Q205^{5.39}H and R208^{5.42}K in RXFP4 (displayed more profound reduction for DC591053 than compound 4), while H268^{5.39}Q and K271^{5.42}R in RXFP3 exhibited dose-response features for compound 4 similar to the WT (**Supplementary Fig. 10b-c, Supplementary Table 4**). The results indicate that these sites may play important roles in subtype selectivity.

G_i coupling

G_i-coupling was almost identical among in the three complex structures (**Fig. 4a**), where G_i protein was anchored by the α 5 helix of G_i subunit, thereby fitting to the cytoplasmic cavity formed by TMs 2, 3 and 5-7 as well as ICLs 2 and 3, a phenomenon widely observed in other G_i-coupled structures such as GHSR³⁶, formyl peptide receptor 2 (FPR2)³⁷ and CCR1²⁸ (**Fig. 4a-b**). The hydrophobic patch at the C terminus of G_i, including I345^{G.H5.16} (superscripts refer to the common G α numbering system³⁸), L349^{G.H5.20}, C352^{G.H5.23}, L354^{G.H5.25} and F355^{G.5.26}, interacted with a series of surrounding hydrophobic residues in TMs 3, 5 and 6 by contributing massive hydrophobic contacts (via V142^{3.53}, V143^{3.54}, Y224^{5.58}, L227^{5.61}, F230^{5.64}, L231^{5.65}, V243^{6.32}, V244^{6.33}, V248^{6.37} and L251^{6.40}), three hydrogen bonds (R139^{3.50}-C352^{G.H5.23}, V142^{3.53}-N348^{G.H5.19}, and S247^{6.36}-L354^{G.H5.25}) and one salt

bridge (D240^{6,29}–K346^{G,H5,17}) (**Fig. 4c**). Unlike the short α -helix conformation that observed in FPR2, CCR1 and somatostatin receptor 2 (SSTR2), ICL2 of RXFP4 adopted a loop conformation and made one hydrogen bond (H152^{ICL2}–N348^{G,H5,19}) and multiple hydrophobic contacts via A147^{ICL2} and P149^{ICL2} with G_i (**Fig. 4d**). Consistent with the crucial role of ICL3 in signaling pathways of various GPCRs^{39–41}, three adjacent positively charged residues (R234^{ICL3}, R236^{ICL3} and R237^{ICL3}) and Q235^{ICL3} established a polar network through multiple salt bridges (via E309^{G,H4,26}, E319^{G,h4s6,12} and D342^{G,H5,13}) and several hydrogen bonds (via D338^{G,H5,9} and T341^{G,H5,12}) (**Fig. 4e**). Notably, one salt bridge between H8 and α 5 helix of G_i (E315^{8,49}–K350^{G,H5,21}) was found only in the cryo-EM structure of compound 4–RXFP4–G_i complex (**Fig. 4a**).

Class-wide comparison

Endogenous peptides mainly bind to class A and B1 GPCRs^{42,43}. Unlike its class B1 counterparts that have large extracellular domains, class A GPCRs usually adopt extended loop conformations during their insertion into the orthosteric pocket by the peptide N terminus [e.g., DAMGO⁴⁴, C-C chemokine ligand 15 (CCL15)²⁸, C-X-C motif chemokine ligand 8 (CXCL8)⁴⁵, A β 42⁴⁶, N-formyl humanin⁴⁶ and ghrelin³⁶], the peptide C terminus [e.g., angiotensin II^{47,48}, bradykinin²⁷, cholecystokinin-8 (CCK-8)⁴⁹, Des-Arg¹⁰-kallidin²⁷, gastrin-17²⁵, IMV449⁵⁰, neuromedin U⁵¹ and neuromedin S⁵¹] or the peptide middle region [e.g., α -melanocyte-stimulating hormone (α -MSH)⁵², arginine-vasopressin (AVP)⁵³ and somatostain-14⁵⁴], thereby achieving a significantly larger peptide-receptor interface area ($>1500 \text{ \AA}^2$) compared to that displayed by interaction with small molecules ($<1000 \text{ \AA}^2$) (**Fig. 5, Supplementary Fig. 7**). Of note, galanin, located far away from the receptor core^{55,56}, adopted an α -helical structure that sat flat on the top of the orthosteric pocket with formation of massive contacts with ECLs 1-3 and moderate interface area ($\sim 1600 \text{ \AA}^2$). Different from the above peptide-binding modes, INSL5 penetrates into the orthosteric pocket via its B chain C terminus by adopting a single α -helix conformation, which is distinct from all reported peptide-bound class A GPCRs but closer to those seen with class B1 structures bound by peptides, such as glucagon-like peptide-1 (GLP-1), glucose-dependent insulinotropic polypeptide (GIP) and glucagon whose N termini insert deeply into the TMD core. This organization resulted in a profound interface area (1761 \AA^2) for INSL5 and direct signal initiation via engagement of α -helix terminus W24^B. Obviously, this α -helix conformation was maintained by the three disulfide bonds, supported by the conserved three helical segments of INSL5 observed in solution-state NMR studies²⁹.

Mechanistic implication

Sharing the same structural scaffold (three α -helices constrained by one intra- and two inter-chain disulfide bonds) and the insulin signature (CC-3X-C-8X-C motif in the A chain), insulin, insulin-

like growth factors (IGFs) 1 and 2, relaxin 1-3 and INSL3-6 constitute the human insulin superfamily (**Fig. 6a**), an ancient family of functionally diverse proteins^{57,58}. While insulin and IGF-1 mainly bind to and activate cell surface tyrosine kinase receptors, *i.e.*, canonical insulin receptor (IR)/IGF-1 receptor (IGF-1R), and IGF-2 acts through the single-transmembrane glycoprotein IGF-2/manose-6-phosphate receptor (IGF-2R/M6PR); the actions of relaxin 1-3, INSL3 and INSL5 are mediated by respective GPCRs. The INSL5-bound RXFP4–G_i complex structure, together with abundant information on insulin and IGFs in the literature⁵⁹⁻⁶², provide an excellent opportunity to investigate the structural basis of the functional versatility with no cross-reactivity among members of this important peptide superfamily.

The peptide-binding pocket of RXFP4 is significantly different from that of the insulin and IGF-1 receptors. By arranging the residues at the extracellular halves of TMs 2-7, RXFP4 provides a typical class A GPCR pocket that is deeply buried and occluded for the penetration of the C-terminal α -helix of INSL5 B chain (α 1 in **Fig. 6b**). Meanwhile, the ECLs of RXFP4 interact with the C-terminal region of the second short α -helix of INSL5 A chain (α 3 in **Fig. 6b**). Such a binding mode suggests that the sequence and the length at the C-terminal ends of A and B chains are likely to play a key role in receptor activation and subtype selectivity. Consistently, the C-terminal truncation at the B chain of relaxin-2 greatly reduced agonist potency by ~100-fold compared to the native peptide⁶³. Such a truncation transformed relaxin-3 to an antagonist for RXFP3 and RXFP4⁶⁴. Because of the presence of additional residues at the C termini of both chains, insulin and IGFs produced massive sterically clashes with RXFP4 upon structure superimposition (**Supplementary Fig. S11a-b**), implying that they are unable to bind and activate RXFPs. As a comparison, the binding pockets of IR and IGF-1R are planar and largely solvent-exposed, where distinct segments of the conserved structural feature were used by insulin or IGF-1 for receptor recognition (**Fig. 6c, d**)⁶⁵. Specifically, both peptides utilized the hydrophobic residues at the two short α -helices (α 2 and α 3) as hydrophobic core to interact with the hydrophobic residues in IR and IGF-1R, whereas the extended C-terminal tail of insulin’s B chain sealed the cleft between the L1 domain and α -CT. Notably, IGF-1 further inserted into a groove formed by L1 and CR domains (CRDs) of IGF-1R via its extremely long C-domain loop. INSL5 that aligned to the insulin at site 1 eliminated interactions from the L1 domain- α -CT cleft and caused steric clashes with FnIII-1 and α -CT, respectively (**Supplementary Fig. 12a-b**). Similar phenomena were found when aligning INSL5 to IGF-1 bound by IGF-1R (**Supplementary Fig. 12c-d**). These observations reveal distinct ligand recognition mechanisms in the insulin superfamily and highlight that functional versatility is achieved by varying peptide sequence and ligand-binding pocket (**Fig. 6e**).

Discussion

As one of the most important peptide-binding receptor subfamilies, RXFPs are promising drug targets for multiple diseases. In this study, we present three G_i-bound RXFP4 structures in complex with its endogenous ligand INSL5, RXFP3/RXFP4 dual agonist compound 4 and RXFP4-specific agonist DC591053. Combined with mutagenesis, SAR analysis and MD simulations, mechanisms of INSL5 recognition, peptidomimetic agonism and subtype selectivity of RXFP4 were delineated, thereby expanding our understanding of the structural basis of functional versatility of the relaxin family peptide receptors.

The INSL5-bound RXFP4–G_i complex structure presents a unique peptide-binding mode previously unknown and helps us elucidate an additional mechanism of activation related to peptide-binding class A GPCRs. Unlike the loop or “lay-flat” α -helix conformations adopted by other reported class A GPCR bound peptides, the B chain of INSL5 exhibits a single α -helix conformation that penetrates into the orthosteric pocket, while the A chain, similar to the extracellular domain (ECD) of class B1 GPCR, sits above the orthosteric pocket to interact the extracellular half of B chain as well as the extracellular surface of RXFP4. Despite variable receptor interaction modes, both A chain and B chain are indispensable to the functionality of INSL5, indicating the essence of such a peptide architecture in executing its action. This phenomenon has not been reported previously among peptidic ligands for GPCRs, but is a common feature (three intra-peptide disulfide bonds) of the insulin superfamily members.

High resolution complex structures of compound 4- and DC591053-bound RXFP4 demonstrate both common and unique features of these two small molecule agonists in terms of peptidomimetic agonism and subtype selectivity. By structurally mimicking the C terminus residue W24^B, compound 4 and DC591053 occupy the bottom of the orthosteric pocket in a manner similar to INSL5 thereby displaying their peptidomimetic property. Meanwhile, the varying extents to which they contact RXFP4-specific residues form the foundation that governs receptor subtype selectivity, where DC591053 was discovered and validated as a RXFP4-specific agonist without observable cross-reactivity with RXFP3. Clearly, further structure-guided optimization of DC591053 towards better efficacy should be feasible with support of the near-atomic level structural information.

Members of the insulin superfamily mediate a diverse array of signaling pathways through one TM or seven TM receptors, representing an evolutionary lineage of functional versatility using a similar structural scaffold. To specifically activate corresponding receptors, two different and mutually exclusive peptide recognition modes (featured by α 1 helix of INSL5 that inserts deeply to a buried pocket of RXFP4 and α 2/ α 3 helices of insulin/ IGF-1 that closely covers the planar interface

of insulin receptor or IGF-1R) are employed, where variances in peptide sequence length and amino acid composition constitute the molecular basis of distinct functionalities. It appears that different regions of a peptide scaffold are able to interact with different types of receptors, conferring ligand specificity. In this manner, differences in signal transduction between IR/IGF-1R (via homo- or hetero-dimerization) and GPCRs (via individual conformational alterations) are preserved to maximize functional versatility with a conserved peptide scaffold, especially for signal initiation and propagation. Unlike insulin and IGF-1 that mainly change the relative subdomain orientations to trigger downstream signaling, INSL5, as shown by the cryo-EM structure reported here, deeply inserts into the orthosteric pocket of RXFP4 (particularly the terminal residues R23^B and W24^B of the B chain) and induces conformational rearrangements of the ligand-binding pocket that further propagate to the intracellular side and render the outward movement of the intracellular half of TM6 as well as the G protein coupling. This information will greatly expand our knowledge on the signaling mechanisms of the insulin superfamily and may advance the development of therapeutic agents for multiple diseases.

Methods

Construct

The full-length human RXFP4 was cloned into a modified pFastBac vector (Invitrogen) with HA signal peptide to enhance receptor expression, followed by a 10× histidine tag and BRIL insertion at the N terminus. LgBiT subunit (Promega) was fused at the C terminus of RXFP4 connected by a 15-amino acid polypeptide linker. A dominant-negative human G α_{i2} (DNGi2) were generated by introducing S47N, G203A, E245A and A326S substitutions in the G α subunit as previously described⁶⁶. The human G $\beta1$ with a C-terminal 15-amino acid polypeptide linker was followed by a HiBiT (peptide 86, Promega), and the scFv16 was modified with an N-terminal GP67 signaling peptide and a C-terminal 8× histidine tag. The engineered human G α_{i2} , G $\beta1$, bovine G $\gamma2$ and scFv16 were cloned into the pFastBac vector (Invitrogen), respectively.

Production of INSL5 peptide

Recombinant INSL5 was designed to be produced from a single-chain INSL5 precursor in which the B chain (24 residues) and the A chain (21 residues) were connected by a specific C-peptide with addition of a leader peptide at the N terminus. It was converted to two-chain human INSL5 by digesting with two proteinases after refolding (**Supplementary Fig. 1a**). Compared to the native hormone containing N-terminal pyroglutamate (pGlu, pE), the N-terminal glutamine (Gln, Q) of the recombinant INSL5 used in this study was not converted to pE (**Supplementary Fig. 1b**).

A gene encompassing the coding sequence of the INSL5 precursor (5' end with Nde I recognition sequence and start codon, 3' end with stop codon and Hind III recognition sequence) was designed and codon-optimized for high-level expression in *E. coli*. It was chemically synthesized and inserted into a pUC57 based vector (GenScript). The encoding DNA fragment of the INSL5 precursor was confirmed by DNA sequencing. The fragment of which was cleaved by Nde I and Hind III from the pUC57 plasmid and subsequently ligated into a pET vector that was pretreated with the same restriction enzymes using a T4-DNA polymerase. The expression construct was designated as pET-INSL5 plasmid and was transformed into competent *E. coli* cells derived from BL21 (DE3). After confirmation of the protein expression with IPTG induction, a single colony with higher level was selected, cultured and stored at -80°C for future fermentation.

The above cells were cultivated in LB medium (ThermoFisher Scientific) at 37°C and then inoculated for fermentation. At the end of fermentation, the biomass was harvested and the inclusion body was recovered for refolding. The refolded precursor was purified by chromatography and cleaved with proteinases to generate the two-chain INSL5 with three pairs of correct disulfide bonds. After chromatographic purification, the mature two-chain INSL5 was analyzed by non-reducing SDS-PAGE and RP-HPLC (**Supplementary Fig. 1**).

The primary structure was confirmed by peptide mapping and 2-dimensional (2D) liquid chromatograph (LC)-MS. The molecular weights of individual peptide peaks, generated by chymotrypsin, were measured by an online Orbitrap spectrometer (ThermoFisher Scientific) coupled to two sequential RP-HPLC procedures. The amino acid sequences of chymotrypsin-generated peptides were assigned by matching molecular weight measured with theoretical sequence of a peptide. The recombinant INSL5 peptide was subsequently verified for its bioactivity in CHO-K1 cells stably transfected with RXFP4 compared with an INSL5 standard (Phoenix Pharmaceuticals).

Synthesis of DC591053

The newly discovered RXFP4 agonist DC591053 was synthesized following procedures depicted in **Supplementary Fig. 2a**⁶⁷. Commercially available 4-hydroxy-3-methoxybenzaldehyde (**1-1**) was treated with iodine ethane to give **1-2**, which was refluxed in nitromethane to obtain **1-3** under the catalysis of ammonium acetate. Then compound **1-3** was reduced by LiAlH₄ to give key intermediate **1-4**. 5-Methoxy-1H-indole-3-carbaldehyde (**1-5**) was reacted with the Wittig reagent methyl 2-(triphenyl-λ5-phosphorylidene)acetate (**1-6**) to give the corresponding α,β-unsaturated ester **1-7**, which was converted to the saturated ester **1-8** by catalytic hydrogenation. Hydrolysis of compound **1-8** afforded the key intermediate acid **1-9**. Amide **1-10** was generated by a coupling reaction of intermediates 2-(4-ethoxy-3-methoxyphenyl)ethan-1-amine (**1-4**) and 3-(5-methoxy-

1H-indol-3-yl)propanoic acid (**1-9**). Then, amide **1-10** was treated with POCl_3 to afford the dihydroisoquinoline compound **1-11**. Asymmetric reduction with Noyori catalyst gave the *S*-isomer **1-12**, which was subjected to react with 4-morpholinocarbonyl chloride to provide the target product DC591053.

Preparation of scFv16

ScFv16 was expressed in High-FiveTM insect cells (ThermoFisher Scientific) as a secreted protein purified by Ni-sepharose chromatography column as described previously⁴⁴. The HiLoad 16/600 Superdex 75 column (GE Healthcare) was used to separate the monomeric fractions of scFv16 with running buffer containing 20 mM HEPES and 100 mM NaCl, pH 7.4. The purified scFv16 was flash frozen in liquid nitrogen with 10% glycerol and stored in -80°C until use.

Expression and purification of the RXFP4–G_i complex

Recombinant viruses of RXFP4, G α_{i2} , G $\beta1$ and G $\gamma2$ were generated using Bac-to-Bac baculovirus expression system (Invitrogen) in *Spodoptera frugiperda* (*Sf9*) insect cells (Invitrogen). P0 viral stock was produced by transfecting 5 μg recombinant bacmids into *Sf9* cells (2.5 mL, density of 1.5×10^6 cells per mL) for 96 h incubation and then used to produce high-titer P1 baculoviruses. High-FiveTM insect cells were grown to a density of 3.2×10^6 cells per mL and infected with RXFP4, G α_{i2} , G $\beta1$ and G $\gamma2$ P1 viral stocks at a ratio of 6: 1: 1: 1. The cells were cultured for 48 h at 27°C after infection and harvested by centrifugation at 2,000 rpm for 20 min.

The cell pellets were lysed in buffer [20 mM HEPES, 100 mM NaCl and 100 μM TCEP, pH 7.4, supplemented with 10% (v/v) glycerol and EDTA-free protease inhibitor mixture (Bimake)], and the membrane was collected at 30,000 rpm for 30 min followed by homogenization in the same buffer. The formation of RXFP4–G_i complex was initiated by addition of 10 mM MgCl₂, 1 mM MnCl₂, 5 mM CaCl₂, 25 mU/mL apyrase (NEB), 15 $\mu\text{g}/\text{mL}$ scFv16, ligands (20 μM INSL5, 50 μM Compound 4 or 50 μM DC591053), 100 μM TCEP and 100 U salt active nuclease (Sigma-Aldrich) supplemented with protease inhibitor cocktail for 1.5 h incubation at room temperature (RT). The membrane was then solubilized with 0.5% (w/v) lauryl maltose neopentyl glycol (LMNG, Anatrace) and 0.1% (w/v) cholesterol hemisuccinate (CHS, Anatrace) with additional protease inhibitor cocktail for 3 h at 4°C. The supernatant was isolated by centrifugation at 32,000 rpm for 1 h and incubated with Ni-NTA beads (GE Healthcare) for 1.5 h at 4 °C. The resin was collected and packed into a gravity flow column and washed with 10 column volumes of buffer A [20 mM HEPES, 100 mM NaCl, 5 mM MgCl₂, 1 mM MnCl₂ 100 μM TCEP, ligands (4 μM INSL5, 10 μM Compound 4 or 10 μM DC591053), 0.1% (w/v) LMNG, 0.02% (w/v) CHS and 30 mM imidazole, pH 7.4], followed by washing with 20 column volumes of buffer B [essentially the same as buffer A with

decreased concentrations of detergents 0.03% (w/v) LMNG, 0.01% (w/v) GDN and 0.008% (w/v) CHS containing 60 mM imidazole, pH 7.4]. The protein was eluted with five column volumes of buffer C (buffer B with 300 mM imidazole, pH 7.4). The complexes were then concentrated using a 100-kD Amicon Ultra centrifugal filter (Millipore) and subjected to Superdex 200 10/300 GL column (GE Healthcare) with running buffer containing 20 mM HEPES, 100 mM NaCl, 100 µM TCEP, ligands (4 µM INSL5, 10 µM Compound 4 or 10 µM DC591053), 0.00075% (w/v) LMNG, 0.00025% (w/v) GDN and 0.00025% (w/v) CHS, pH 7.4. The monomeric peak fractions were pooled and concentrated to 5-8 mg/mL.

Cryo-EM data acquisition

The purified complex samples (3 µL at 5-8 mg/mL) were applied to glow-discharged holey grids (Quantifoil R1.2/1.3, 300 mesh) and subsequently vitrified using a Vitrobot Mark IV (ThermoFisher) set at 100% humidity and 4°C. Cryo-EM images were acquired on a Titan Krios microscope (FEI) equipped with Gatan energy filter, K3 direct electron detector and ChimeraX EM3.7. The microscope was operated at 300 kV accelerating voltage, at a nominal magnification of 46,685 \times in counting mode, corresponding to a pixel size of 1.071 Å. Totally, 9,256 movies of the INSL5–RXFP4–G_i complexes, 9,562 movies of the compound 4–RXFP4–G_i complexes and 8,203 movies of the DC591053–RXFP4–G_i complexes were obtained, respectively, with a defocus range of -1.2 to -2.2 µm. An accumulated dose of 80 electrons per Å² was fractionated into a movie stack of 36 frames.

Cryo-EM data processing

Dose-fractionated image stacks were subjected to beam-induced motion correction using MotionCor2.1. A sum of all frames, filtered according to the exposure dose, in each image stack was used for further processing. Contrast transfer function parameters for each micrograph were determined by Gctf v1.06. Particle selection, 2D and 3D classifications were performed on a binned dataset with a pixel size of 2.142 Å using cryoSPARC v3.2.0 and RELION-3.1.1.

For the INSL5–RXFP4–G_i complex, auto-picking yielded 10,618,534 particle projections that were subjected to two rounds of reference-free 2D classification to discard false-positive particles or particles categorized in poorly defined classes, producing 3,267,126 particle projections for further processing. This subset of particle projections was subjected to a round of maximum-likelihood-based 3D classification with a pixel size of 2.142 Å, resulting in one well-defined subset with 2,201,257 projections. Further 3D classification with a mask on the receptor produced one good subset accounting for 524,035 particles, which were then subjected to 3D refinement and Bayesian polishing with a pixel size of 1.071 Å. After the last round of refinement, the final map

has an indicated global resolution of 3.19 Å at a Fourier shell correlation (FSC) of 0.143. Local resolution was determined using the Bsoft package (v2.0.3) with half maps as input maps.

For the compound 4–RXFP4–G_i complex, auto-picking yielded 4,796,219 particle projections that were subjected to two rounds of reference-free 2D classification to discard false-positive particles or particles categorized in poorly defined classes, producing 787,382 particle projections for further processing. This subset of particle projections was subjected to a round of maximum-likelihood-based 3D classification with a pixel size of 2.142 Å, resulting in one well-defined subset with 469,428 projections. Further 3D classification with a mask on the receptor produced one good subset accounting for 243,800 particles, which were then subjected to 3D refinement and Bayesian polishing with a pixel size of 1.071 Å. The map with an indicated global resolution of 3.03 Å at a FSC of 0.143 was generated from the final 3D refinement. Local resolution was determined using the Bsoft package (v2.0.3) with half maps as input maps.

For the DC591053–RXFP4–G_i complex, auto-picking yielded 8,996,005 particle projections that were subjected to two rounds of reference-free 2D classification to discard false-positive particles or particles categorized in poorly defined classes, producing 2,950,880 particle projections for further processing. This subset of particle projections was subjected to a round of maximum-likelihood-based 3D classification with a pixel size of 2.142 Å, resulting in one well-defined subset with 1,286,136 projections. Further 3D classification with a mask on the receptor produced one good subset accounting for 225,327 particles, which were then subjected to 3D refinement and Bayesian polishing with a pixel size of 1.071 Å. After the last round of refinement, the final map has an indicated global resolution of 2.75 Å at a FSC of 0.143. Local resolution was determined using the Bsoft package (v2.0.3) with half maps as input maps.

Model building and refinement

The cryo-EM structure of bradykinin–B2R complex (PDB code: 7F2O)²⁷ was used as the initial model of RXFP4 and scFv16, while the cryo-EM structure of A₁R–G_i complex (PDB code: 6D9H)⁶⁶ was used to generate the initial model of G proteins. For the structure of compound 4–RXFP4–G_i and DC591053–RXFP4–G_i complexes, the coordinates of INSL5–RXFP4–G_i complex were used as the starting point. Ligand coordinates and geometry restraints were generated using phenix.elbow in PHENIX v1.18⁶⁸. The model was docked into the electron microscopy density maps using UCSF Chimera v1.13.1⁶⁹, followed by iterative manual adjustment and rebuilding in COOT 0.9.4.1⁷⁰. Real space refinement was performed using PHENIX v1.18⁶⁸. The model statistics were validated using the module comprehensive validation (cryo-EM) in PHENIX v1.18⁶⁸. Structural figures were prepared in UCSF Chimera v1.13.1, UCSF ChimeraX v1.0 and PyMOL v.2.1 (<https://pymol.org/2/>). The final refinement statistics are provided in Table S1.

Molecular dynamics simulation

MD simulations were performed by Gromacs 2020.1. The INSL5–RXFP4 complexes were built based on the cryo-EM structure of the INSL5–RXFP4–G_i complex and prepared by the Protein Preparation Wizard (Schrodinger 2017-4) with the G protein and scFv16 removed. The receptor chain termini were capped with acetyl and methylamide. All titratable residues were left in their dominant state at pH 7.0. To build MD simulation systems, the complexes were embedded in a bilayer composed of 237 POPC lipids and solvated with 0.15 M NaCl in explicit TIP3P waters using CHARMM-GUI Membrane Builder v3.5⁷¹. The CHARMM36-CAMP force field⁷² was adopted for protein, peptides, lipids and salt ions. The Particle Mesh Ewald (PME) method was used to treat all electrostatic interactions beyond a cut-off of 10 Å and the bonds involving hydrogen atoms were constrained using LINCS algorithm⁷³. The complex system was first relaxed using the steepest descent energy minimization, followed by slow heating of the system to 310 K with restraints. The restraints were reduced gradually over 50 ns. Finally, restraint-free production run was carried out for each simulation, with a time step of 2 fs in the NPT ensemble at 310 K and 1 bar using the Nose-Hoover thermostat and the semi-isotropic Parrinello-Rahman barostat⁷⁴, respectively. The interface area was calculated by the program FreeSASA 2.0, using the Sharke-Rupley algorithm with a probe radius of 1.2 Å⁷⁵. Similar simulation procedure and analysis were adopted for the MD simulations of INSL5 and its B chain, which were placed in a cubic box that the boundary of the box was at least 15 Å to the solute.

Cell culture and transfection

CHO-K1 cells stably expressing human RXFP4 (hRXFP4-CHO) or RXFP3 (hRXFP3-CHO) were maintained in DMEM/F12 (Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS) and 2 mM L-glutamine. Human embryonic kidney 293T cells containing SV40 large T-antigen (HEK293T) were maintained in DMEM (Gibco) supplemented with 10% (v/v) FBS, 1 mM sodium pyruvate (Gibco), 100 units/mL penicillin and 100 µg/mL streptomycin at 37°C in 5% CO₂. For cAMP assays in mutants, HEK293T cells were seeded onto 6-well cell culture plates at a density of 7×10^5 cells per well. After overnight incubation, cells were transfected with pCMV6 vectors bearing WT or mutant receptors using Lipofectamine 3000 transfection reagent (Invitrogen). Following 24 h culturing, the transfected cells were ready for detection.

Eu-labelled binding assay

CHO-K1 cells stably transfected with RXFP3 or RXFP4 were plated onto pre-coated poly-L-lysine 96-well plates. The competitive binding assays were performed with 5 nM Eu-H3 B1-22R (RXFP3)

or Eu-R3/I5 (RXFP4) in the presence of increasing amounts of ligands as previously described^{19,76,77}. Time-resolved fluorescent measurements were carried out at an excitation wavelength of 340 nm and an emission wavelength of 614 nm on a BMG POLARstar plate reader (BMG Labtech, Melbourne, Australia). Binding was performed in at least three independent experiments with triplicate determinations within each assay. Data are presented as means \pm S.E.M. of specific binding and were fitted using a one-site binding curve in Prism software (GraphPad).

cAMP accumulation assay

Inhibition of forskolin-induced cAMP accumulation by INSL5, compound 4 and DC591053 was measured by a LANCE Ultra cAMP kit (PerkinElmer). Ligands were verified for their bioactivity in the beginning in hRXFP4-CHO, which were ready for use after 24 h culturing. For assaying mutants, HEK293T cells were used 24 h post transfection. Cells were digested with 0.02% (w/v) EDTA and seeded onto 384-well microtiter plates at a density of 8×10^5 cells/mL in cAMP stimulation buffer [HBSS supplemented with 5 mM HEPES, 0.1% (w/v) bovine serum albumin (BSA) and 0.5 mM 3-isobutyl-1-methylxanthine]. The cells were stimulated with different concentrations of ligands plus 1.5 μ M forskolin in RXFP4 and 4 μ M forskolin in RXFP3. After 40 min incubation at RT, the Eu-cAMP tracer and ULight-anti-cAMP working solution were then added to the plates separately to terminate the reaction followed by 60 min additional incubation. The time-resolved fluorescence resonance energy transfer (TR-FRET) signals were detected by an EnVision multilabel plate reader (PerkinElmer) with the emission window ratio of 665 nm over 620 nm under 320 nm excitation. Data were normalized to the maximal response of WT receptor.

Statistical analysis

All functional study data were analyzed using GraphPad Prism 8.3 (GraphPad Software) and presented as means \pm S.E.M. from at least three independent experiments. Dose-response curves were evaluated with a three-parameter logistic equation. The significance was determined with either a two-tailed Student's *t*-test or one-way ANOVA with Dunnett's multiple comparison test, and $P < 0.05$ was considered statistically significant.

Data availability

The raw data underlying Figs. 1e, 2f–g, 3c, 3e, 3h, 5b, Supplementary Figs. 1c, 1f–g, 2e–g, 3b–d, 4a–c, 10b–d and Supplementary Tables 3–4 are provided as a source data file. The atomic coordinates and electron microscopy maps have been deposited in the Protein Data Bank (PDB) under accession codes: 7YJ4 (INSL5–RXFP4–G_i complex), 7YK6 (compound 4–RXFP4–G_i complex) and 7YK7 (DC591053–RXFP4–G_i), as well as Electron Microscopy Data Bank (EMDB) under accession

codes: EMD-33871 (INSL5–RXFP4–G_i complex), EMD-33888 (compound 4–RXFP4–G_i complex) and EMD-33889 (DC591053–RXFP4–G_i complex), respectively. All relevant data are available from the authors and/or included in the manuscript or supplemental data. Source data are provided with this paper.

Acknowledgements

We are grateful to Jiao Yu, Tania Ferraro and Sharon Layfield for technical assistance. This work was supported by the National Natural Science Foundation of China 81872915 (M.-W.W.), 82073904 (M.-W.W.), 82121005 (D.Y.), 81973373 (D.Y.), 91953108 (H.L.) and 21704064 (Q.T.Z.); National Science & Technology Major Project of China–Key New Drug Creation and Manufacturing Program 2018ZX09735–001 (M.-W.W.) and 2018ZX09711002–002–005 (D.Y.); National Major Project of China Science and Technology Innovation 2030 for Brain Science and Brain-Inspired Technology 2021ZD0203400 (Q.T.Z.); the National Key Basic Research Program of China 2018YFA0507000 (M.-W.W.); Hainan Provincial Major Science and Technology Project ZDKJ2021028 (D.Y. and Q.T.Z.) and Shanghai Municipality Science and Technology Development Fund 21JC1401600 (D.Y.), the Victorian Government’s Operational Infrastructure Support Program (R.A.D.B.) and National Health and Medical Research Council of Australia Research Fellowship 1135837 (R.A.D.B.). The cryo-EM data were collected at Cryo-Electron Microscopy Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences.

Author contributions

Y.C. designed expression constructs, purified the receptor complexes, screened the specimen, prepared the final samples for cryo-EM data collection, conducted map calculation, built the models of the complexes, performed signaling experiments and participated in manuscript preparation; Q.T.Z. performed model building, structural analysis, MD simulations, figure preparation and participated in manuscript writing; J.W. synthesized DC591053 with the guidance of H.L; Y.W.X. performed structure refinement and model building under the supervision of H.E.X.; Y.W. and Q.Z. produced recombinant INSL5 under the supervision of C.S.; J.H.Y, F.H.Z., C.-W.C. and X.Q.C. took part in method development and functional experiments; Y.B.W and C.H.L assisted in the synthesis of DC591053; R.A.D.B. organized ligand binding assay; D.H.Y and M.-W.W. initiated the project, supervised the studies, analyzed the data, and wrote the manuscript with inputs from all co-authors.

Conflict of interest

The authors declare no conflict of interest.

References

- 1 Bathgate, R. A. *et al.* Relaxin family peptides and their receptors. *Physiol Rev* **93**, 405-480 (2013).
- 2 Halls, M. L. *et al.* Relaxin family peptide receptors RXFP1 and RXFP2 modulate cAMP signaling by distinct mechanisms. *Mol Pharmacol* **70**, 214-226 (2006).
- 3 Hsu, S. Y. *et al.* Activation of orphan receptors by the hormone relaxin. *Science* **295**, 671-674 (2002).
- 4 Kern, A. *et al.* The low-density lipoprotein class A module of the relaxin receptor (leucine-rich repeat containing G-protein coupled receptor 7): its role in signaling and trafficking to the cell membrane. *Endocrinology* **148**, 1181-1194 (2007).
- 5 Ng, A. *et al.* Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity. *Autophagy* **7**, 1082-1084 (2011).
- 6 Liu, C. *et al.* Identification of relaxin-3/INSL7 as a ligand for GPCR142. *J Biol Chem* **278**, 50765-50770 (2003).
- 7 Liu, C. *et al.* INSL5 is a high affinity specific agonist for GPCR142 (GPR100). *J Biol Chem* **280**, 292-300 (2005).
- 8 Luo, X. *et al.* The insulinotropic effect of insulin-like peptide 5 in vitro and in vivo. *Biochem J* **466**, 467-473 (2015).
- 9 Ang, S. Y. *et al.* Signal transduction pathways activated by insulin-like peptide 5 at the relaxin family peptide RXFP4 receptor. *Br J Pharmacol* **174**, 1077-1089 (2017).
- 10 Burnicka-Turek, O. *et al.* INSL5-deficient mice display an alteration in glucose homeostasis and an impaired fertility. *Endocrinology* **153**, 4655-4665 (2012).
- 11 Aparicio, S. *et al.* Use of GPR100 receptor in diabetes and obesity regulation. US Patent Appl 20080269118
- 12 Grosse, J. *et al.* Insulin-like peptide 5 is an orexigenic gastrointestinal hormone. *Proc Natl Acad Sci U S A* **111**, 11133-11138 (2014).
- 13 Diwakarla, S. *et al.* Colokinetic effect of an insulin-like peptide 5-related agonist of the RXFP4 receptor. *Neurogastroenterol Motil* **32**, e13796 (2020).
- 14 Pustovit, R. V. *et al.* A Novel Antagonist Peptide Reveals a Physiological Role of Insulin-Like Peptide 5 in Control of Colorectal Function. *ACS Pharmacol Transl Sci* **4**, 1665-1674 (2021).
- 15 Yang, X. *et al.* Identification and verification of HCAR3 and INSL5 as new potential therapeutic targets of colorectal cancer. *World J Surg Oncol* **19**, 248 (2021).
- 16 Li, S. B. *et al.* Autocrine INSL5 promotes tumor progression and glycolysis via activation of STAT5 signaling. *EMBO Mol Med* **12**, e12050 (2020).
- 17 Patil, N. A. *et al.* Relaxin family peptides: structure-activity relationship studies. *Br J Pharmacol* **174**, 950-961 (2017).
- 18 DeChristopher, B. *et al.* Discovery of a small molecule RXFP3/4 agonist that increases food intake in rats upon acute central administration. *Bioorg Med Chem Lett* **29**, 991-994 (2019).
- 19 Lin, G. Y. *et al.* High-throughput screening campaign identifies a small molecule agonist of the relaxin family peptide receptor 4. *Acta Pharmacol Sin* **41**, 1328-1336 (2020).
- 20 Lin, L. *et al.* Design, synthesis and pharmacological evaluation of tricyclic derivatives as selective RXFP4 agonists. *Bioorg Chem* **110**, 104782 (2021).
- 21 Xu, Y. *et al.* A distinctive ligand recognition mechanism by the human vasoactive intestinal polypeptide receptor 2. *Nat Commun* **13**, 2272 (2022).
- 22 Zhao, F. *et al.* Structural insights into multiplexed pharmacological actions of tirzepatide and peptide 20 at the GIP, GLP-1 or glucagon receptors. *Nat Commun* **13**, 1057 (2022).
- 23 Zhou, F. *et al.* Structural basis for activation of the growth hormone-releasing hormone receptor. *Nat Commun* **11**, 5205 (2020).
- 24 Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β 2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265 (2007).
- 25 Zhang, X. *et al.* Structures of the human cholecystokinin receptors bound to agonists and antagonists. *Nat Chem Biol* **17**, 1230-1237 (2021).
- 26 Zhou, Q. *et al.* Common activation mechanism of class A GPCRs. *Elife* **8**, e50279 (2019).
- 27 Yin, Y. L. *et al.* Molecular basis for kinin selectivity and activation of the human bradykinin receptors. *Nat Struct Mol Biol* **28**, 755-761 (2021).
- 28 Shao, Z. *et al.* Identification and mechanism of G protein-biased ligands for chemokine receptor CCR1. *Nat Chem Biol* **18**, 264-271 (2022).

29 Haugaard-Jonsson, L. M. *et al.* Structure of human insulin-like peptide 5 and characterization
of conserved hydrogen bonds and electrostatic interactions within the relaxin framework.
Biochem J **419**, 619-627 (2009).

30 Juan A. B. *et al.* Integrated Methods for the construction of three-dimensional models and
computational probing of structure-function relations in G protein-coupled receptors. *Methods
in Neurosciences* **25**, 366-428 (1995).

31 Patil, N. A. *et al.* The C-terminus of the B-chain of human insulin-like peptide 5 is critical for
cognate RXFP4 receptor activity. *Amino Acids* **48**, 987-992 (2016).

32 Hu, M. J. *et al.* Interaction mechanism of insulin-like peptide 5 with relaxin family peptide
receptor 4. *Arch Biochem Biophys* **619**, 27-34 (2017).

33 Wang, X. Y. *et al.* Identification of important residues of insulin-like peptide 5 and its receptor
RXFP4 for ligand-receptor interactions. *Arch Biochem Biophys* **558**, 127-132 (2014).

34 Belgı, A. *et al.* Minimum active structure of insulin-like peptide 5. *J Med Chem* **56**, 9509-9516
(2013).

35 Wong, L. L. L. *et al.* Distinct but overlapping binding sites of agonist and antagonist at the
relaxin family peptide 3 (RXFP3) receptor. *J Biol Chem* **293**, 15777-15789 (2018).

36 Liu, H. *et al.* Structural basis of human ghrelin receptor signaling by ghrelin and the synthetic
agonist ibutamoren. *Nat Commun* **12**, 6410 (2021).

37 Zhuang, Y. *et al.* Structure of formylpeptide receptor 2-G_i complex reveals insights into ligand
recognition and signaling. *Nat Commun* **11**, 885 (2020).

38 Flock, T. *et al.* Universal allosteric mechanism for G_α activation by GPCRs. *Nature* **524**, 173-
179 (2015).

39 Mozumder, S. *et al.* Comprehensive structural modeling and preparation of human 5-HT_{2A} G-
protein coupled receptor in functionally active form. *Biopolymers* **111**, e23329 (2020).

40 Blagotinsek Cokan, K. *et al.* Critical impact of different conserved endoplasmic retention motifs
and dopamine receptor interacting proteins (DRIPs) on intracellular localization and trafficking
of the D2 dopamine receptor (D2-R) isoforms. *Biomolecules* **10**, 1355 (2020).

41 Pydi, S. P. *et al.* The third intracellular loop plays a critical role in bitter taste receptor activation.
Biochim Biophys Acta **1838**, 231-236 (2014).

42 Davenport, A. P. *et al.* Advances in therapeutic peptides targeting G protein-coupled receptors.
Nat Rev Drug Discov **19**, 389-413, (2020).

43 Cong, Z. *et al.* Structural perspective of class B1 GPCR signaling. *Trends Pharmacol Sci* **43**,
321-334 (2022).

44 Koehl, A. *et al.* Structure of the micro-opioid receptor-G_i protein complex. *Nature* **558**, 547-552
(2018).

45 Liu, K. *et al.* Structural basis of CXC chemokine receptor 2 activation and signalling. *Nature*
585, 135-140 (2020).

46 Zhu, Y. *et al.* Structural basis of FPR2 in recognition of A β ₄₂ and neuroprotection by humanin.
Nat Commun **13**, 1775 (2022).

47 Asada, H. *et al.* The Crystal Structure of angiotensin II type 2 receptor with endogenous peptide
hormone. *Structure* **28**, 418-425 (2020).

48 Wingler, L. M. *et al.* Angiotensin and biased analogs induce structurally distinct active
conformations within a GPCR. *Science* **367**, 888-892 (2020).

49 Liu, Q. *et al.* Ligand recognition and G-protein coupling selectivity of cholecystokinin A
receptor. *Nat Chem Biol* **17**, 1238-1244 (2021).

50 Kato, H. E. *et al.* Conformational transitions of a neuropeptide receptor 1-G_{i1} complex. *Nature*
572, 80-85 (2019).

51 You, C. *et al.* Structural insights into the peptide selectivity and activation of human neuromedin
U receptors. *Nat Commun* **13**, 2045 (2022).

52 Ma, S. *et al.* Structural mechanism of calcium-mediated hormone recognition and G β
interaction by the human melanocortin-1 receptor. *Cell Res* **31**, 1061-1071 (2021).

53 Zhou, F. *et al.* Molecular basis of ligand recognition and activation of human V2 vasopressin
receptor. *Cell Res* **31**, 929-931 (2021).

54 Robertson, M. J. *et al.* Plasticity in ligand recognition at somatostatin receptors. *Nat Struct Mol
Biol* **29**, 210-217 (2022).

55 Duan, J. *et al.* Molecular basis for allosteric agonism and G protein subtype selectivity of
galanin receptors. *Nat Commun* **13**, 1364 (2022).

56 Jiang, W. *et al.* Structural insights into galanin receptor signaling. *Proc Natl Acad Sci U S A* **119**, e2121465119 (2022).

57 Lu, C. *et al.* New members of the insulin family: regulators of metabolism, growth and now ... reproduction. *Pediatr Res* **57**, 70R-73R (2005).

58 Shabanpoor, F. *et al.* The human insulin superfamily of polypeptide hormones. *Vitam Horm* **80**, 1-31 (2009).

59 Belfiore, A. *et al.* Insulin Receptor Isoforms in Physiology and Disease: An Updated View. *Endocr Rev* **38**, 379-431 (2017).

60 Li, J. *et al.* Synergistic activation of the insulin receptor via two distinct sites. *Nat Struct Mol Biol* **29**, 357-368 (2022).

61 Nielsen, J. *et al.* Structural Investigations of Full-Length Insulin Receptor Dynamics and Signalling. *J Mol Biol* **434**, 167458 (2022).

62 Xu, Y. *et al.* How IGF-II Binds to the Human Type 1 Insulin-like Growth Factor Receptor. *Structure* **28**, 786-798 (2020).

63 Hossain, M. A. *et al.* The minimal active structure of human relaxin-2. *J Biol Chem* **286**, 37555-37565 (2011).

64 Kuei, C. *et al.* R3(BΔ23 27)R/I5 chimeric peptide, a selective antagonist for GPCR135 and GPCR142 over relaxin receptor LGR7: in vitro and in vivo characterization. *J Biol Chem* **282**, 25425-25435 (2007).

65 Uchikawa, E. *et al.* Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor-ligand complex. *Elife* **8**, e48630 (2019).

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

67 Zhang, X. *et al.* Structure-aided identification and optimization of tetrahydro-isoquinolines as novel PDE4 inhibitors leading to discovery of an effective antipsoriasis agent. *J Med Chem* **62**, 5579-5593 (2019).

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

69 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

70 Emsley, P. *et al.* Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).

71 Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997-2004 (2014).

72 Guvench, O. *et al.* CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J Chem Theory Comput* **7**, 3162-3180 (2011).

73 Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* **4** (2008).

74 Aoki, K. M. *et al.* Constant-pressure molecular-dynamics simulations of the crystal-smectic transition in systems of soft parallel spherocylinders. *Phys Rev A* **46**, 6541-6549 (1992).

75 Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Res* **5**, 189 (2016).

76 Haugaard-Kedstrom, L. M. *et al.* Synthesis and pharmacological characterization of a europium-labelled single-chain antagonist for binding studies of the relaxin-3 receptor RXFP3. *Amino Acids* **47**, 1267-1271 (2015).

77 Lin, G. *et al.* High-Throughput Screening Campaign Identified a Potential Small Molecule RXFP3/4 Agonist. *Molecules* **26**, 7511 (2021).

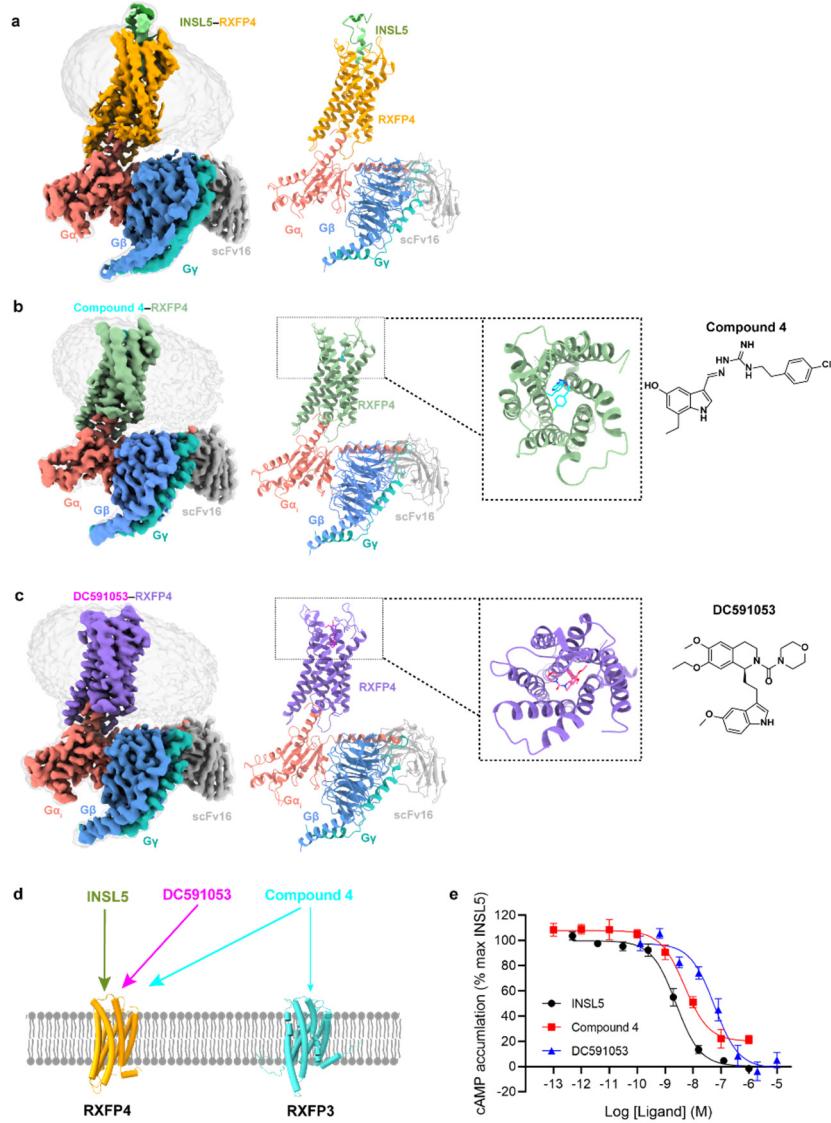


Figure 1. Cryo-EM structures of the RXFP4-G ι complexes. a-c, Cryo-EM density maps (left panel) and cartoon representation (middle) of the INSL5-RXFP4-G ι -scFv16 complex (a), compound 4-RXFP4-G ι -scFv16 complex (b) and DC591053-RXFP4-G ι -scFv16 complex (c). The chemical structures of compound 4 and DC591053 are shown in the right panel of (b) and (c). The A chain of INSL5 is shown in forest green and the B chain in light green, compound 4 in cyan and DC591053 in magenta. The corresponding RXFP4 is shown in orange, dark sea green and medium purple, respectively. G α_i in salmon, G β in cornflower blue, G γ in light sea green and scFv16 in dark gray. d, Activities of INSL5 and DC591053 in RXFP4 and cross-reactivity of compound 4 in RXFP3 and RXFP4. The thickness of lines indicates the strength of affinity. e, INSL5, compound 4 and DC591053 induced cAMP signaling in cells expressing wild-type RXFP4. Data shown are means \pm S.E.M. of at least three independent experiments performed in quadruplicate. Source data are provided as a source data file.

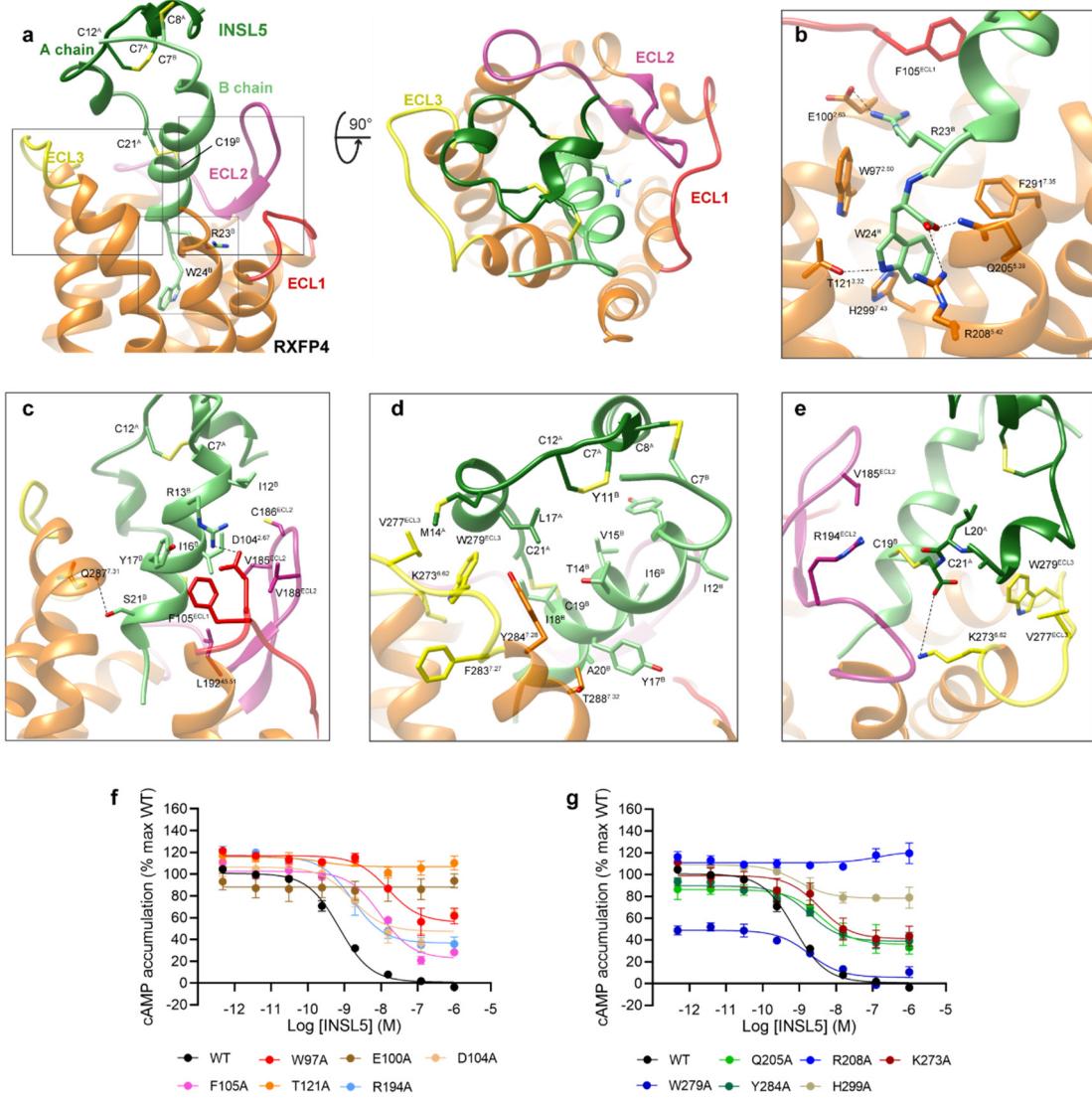


Figure 2. Molecular recognitions of INSL5 by RXFP4. **a**, Binding mode of INSL5 (green) with RXFP4 (orange), showing that the B chain of INSL5 (light green) penetrates into the TMD bundle using the C-terminal α -helix, whereas the A chain of INSL5 covers the orthosteric pocket mainly interacting with ECL2 and ECL3. **b**, Interactions of R23^B and W24^B with RXFP4 in the orthosteric pocket. **c-e**, Close-up views of the interactions between the B chain of INSL5 and the residues in ECLs of RXFP4. **f, g**, Effects of receptor mutations on INSL5-induced cAMP accumulation. Data are shown as means \pm S.E.M. of at least three independent experiments. Source data are provided as a source data file.

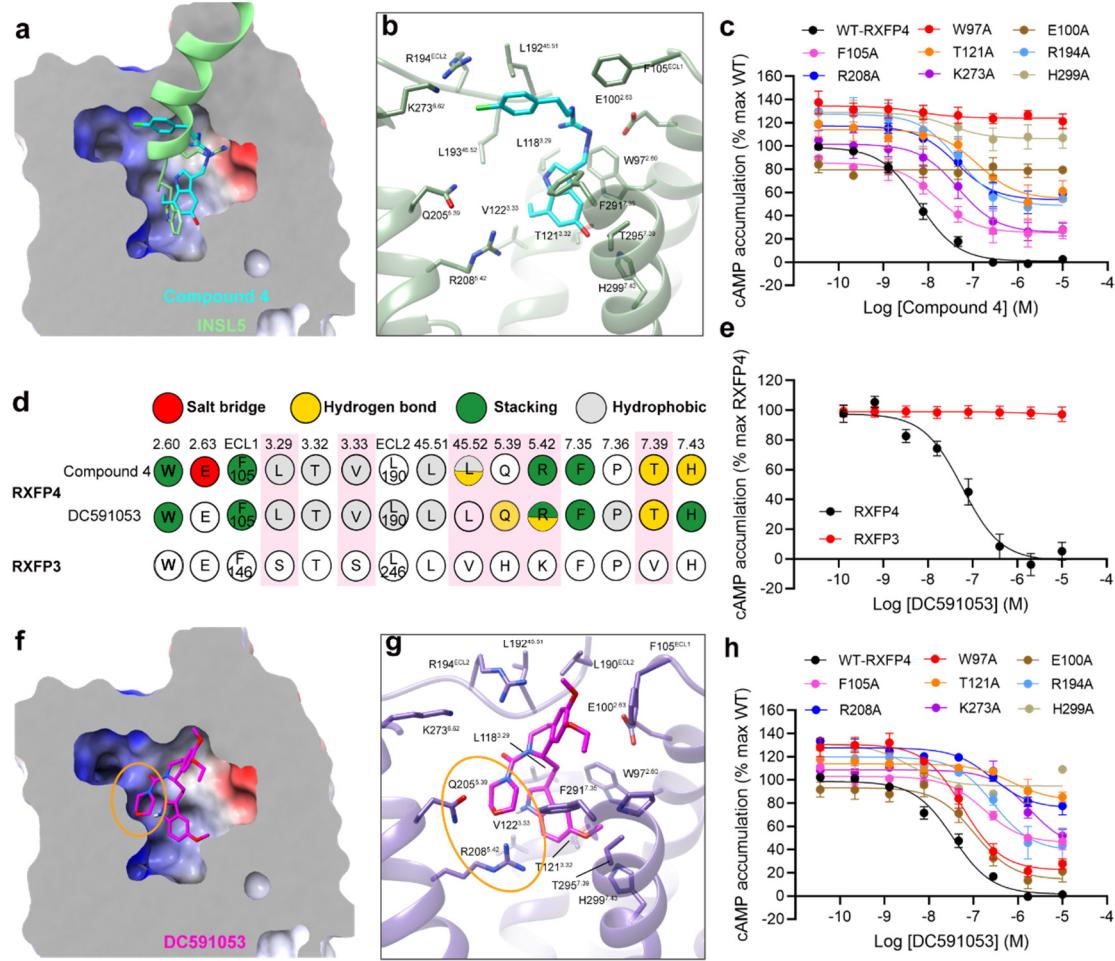


Figure 3. Peptidomimetic agonism and subtype selectivity demonstrated by compound 4 and DC591053. **a**, Cross-section of the compound 4 binding pocket in RXFP4. **b**, Detailed interactions of compound 4 with residues in RXFP4. **c**, Effects of receptor mutations on compound 4-induced cAMP accumulation. Data are shown as means \pm S.E.M. of at least three independent experiments. **d**, Schematic diagram of interactions between ligands and receptor. Amino acid residues of RXFP4 are colored red for salt bridge, yellow for hydrogen bond, green for π - π stacking and gray for hydrophobic interactions. Different residues around the binding pocket in RXFP3 and RXFP4 are highlighted in pink. **e**, Subtype selectivity of DC591053 at RXFP4 without observable cross-reactivity in RXFP3. **f**, Cross-section of the DC591053 binding pocket in RXFP4, with the RXFP4-specific edge in the ligand-binding pocket highlighted in orange circle. **g**, Detailed interactions of DC591053 with residues in RXFP4. **h**, Effects of receptor mutations on DC591053-induced cAMP accumulation. Data are shown as means \pm S.E.M. of at least three independent experiments. Source data are provided as a source data file.

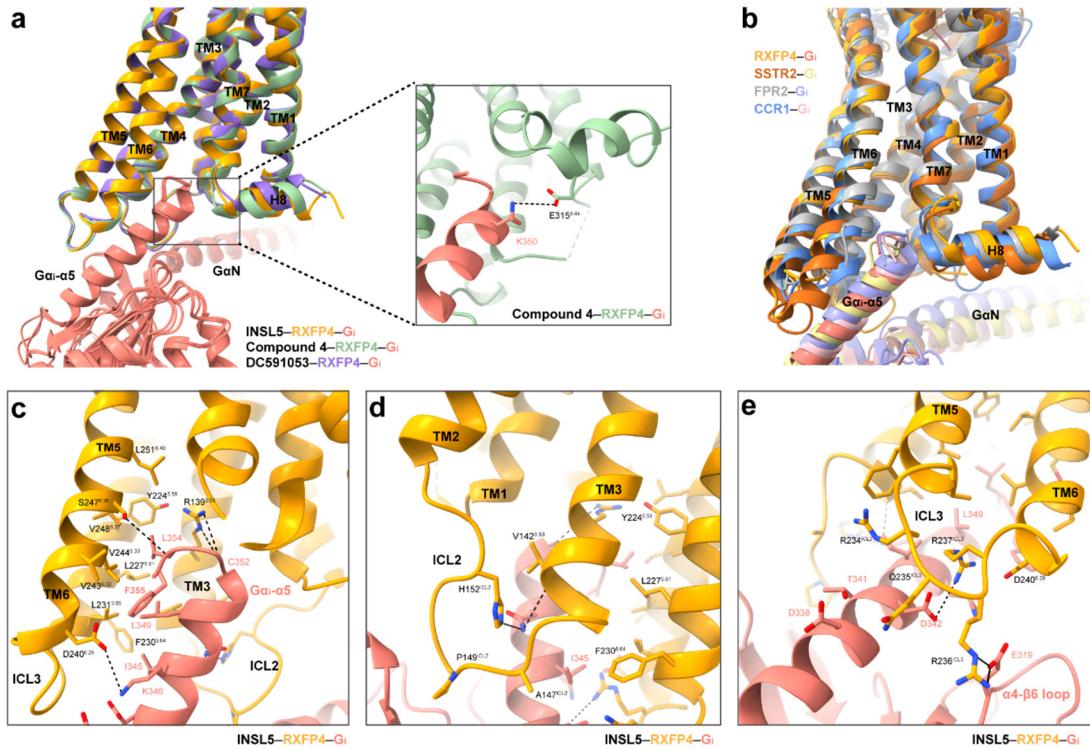


Figure 4. G protein coupling of RXFP4. **a**, G_i-coupling is almost identical among INSL5, compound 4 and DC591053-bound RXFP4 structures. The alignment is based on the receptor. One salt bridge between H8 and α5 helix of G_i was only found in the compound 4-bound RXFP4-G_i complex. **b**, Comparison of G protein coupling among INSL5-bound RXFP4, FPR2 (PDB code: 7WVV), CCR1 (PDB code: 7VL9) and SSTR2 (PDB code: 7T10). The receptors and G proteins were colored as the labeled. **c**, Interaction between RXFP4 (orange) and α5 helix of G_i (salmon) in the cavity of the cytoplasmic region. **d**, Interactions between ICL2 and G_i. **e**, Interactions between ICL3 and G_i. Polar interactions were shown as black dashed lines.

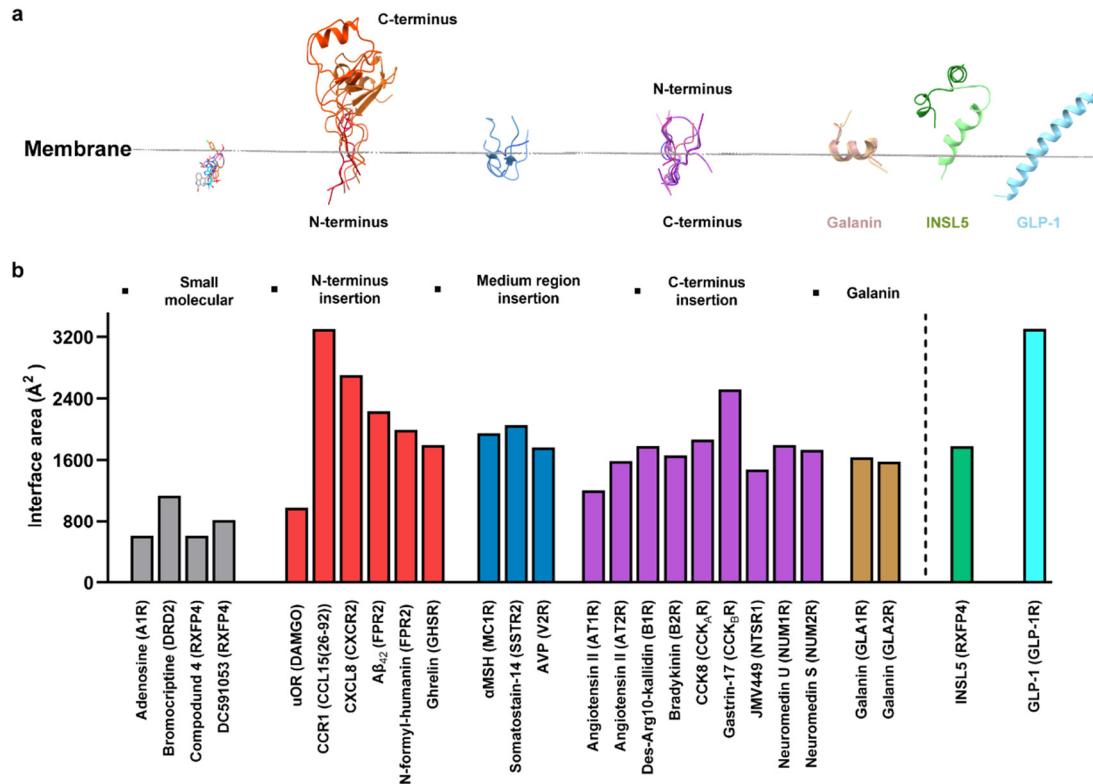


Figure 5. Comparison of ligand binding modes. **a**, Conformations of different ligands during their insertion into the orthosteric pocket. The posture of INSL5 embedded in the membrane was compared with small molecules, galanin, GLP-1 and other peptides with N terminus, medium region and C terminus insertion into the TMD core. Receptor structures were omitted for clarity. **b**, The interface area of peptide-receptor was measured with freeSASA. Small molecule in grey; N terminus insertion in red; medium region insertion in blue; C terminus insertion in purple; galanin in brown; INSL5 in green and GLP-1 in cyan.

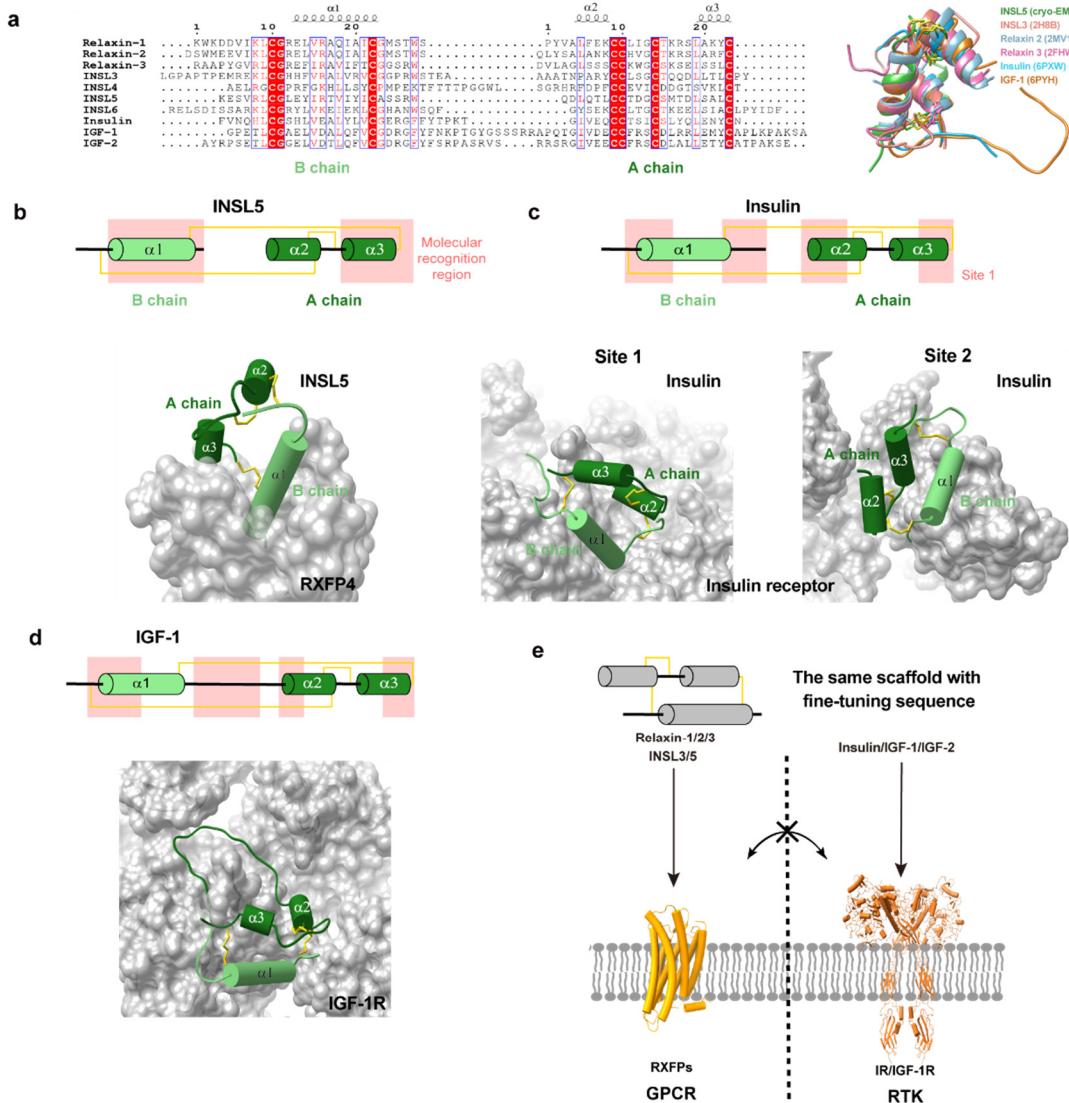


Figure 6. Ligand recognition in the insulin superfamily. **a**, Sequence alignment (left) and structure superimposition (right) of peptides in the insulin superfamily. **b-d**, Recognition of INSL5 (**b**), insulin (PDB code: 6PXW) (**c**) and IGF-1 (PDB code: 6PYH) (**d**) by cognate receptors, where the peptide and receptor are shown in cartoon and surface, respectively. The peptide segments responsible for recognition are highlighted in yellow-shaded region and the three disulfide bonds in each peptide are indicated by yellow solid lines. **e** Distinct ligand recognition modes in the insulin superfamily.