Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Structural basis for chemotype diversity in small molecule GLP-1 receptor agonist drug discovery

Fan Wu
Structure Therapeutics Inc.
Danfeng Song
Structure Therapeutics Inc.
Wei Huang
Structure Therapeutics Inc.
Haizhen Zhang
Structure Therapeutics Inc.
Chunliang Lu
Structure Therapeutics Inc.
Qinghua Meng
Structure Therapeutics Inc.
Hui Lei
Structure Therapeutics Inc.
Xichen Lin
Structure Therapeutics Inc.
Ting Mao
Structure Therapeutics Inc.
Xianqiang Song
Structure Therapeutics Inc.
Raymond C. Stevens
Structure Therapeutics Inc.
Chris de Graaf
Structure Therapeutics Inc.
Yingli Ma
yingli.ma@structuretx.com

Structure Therapeutics Inc.

Article

Keywords:

Posted Date: October 12th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7597307/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The glucagon-like peptide-1 receptor (GLP-1R) is an established therapeutic target for treating obesity and related diseases with several approved injectable peptide agonists on the market. Small molecule GLP-1R agonists, which offer great potential in accessibility, patient compliance, and maintenance compared to their peptide counterparts, are now progressing into late-stage clinical trials. Four different GLP-1 small molecule agonist chemotypes are currently known and here we compare their diverse binding modes and receptor conformations and introduce a novel Class B GPCR binding pocket nomenclature. To further understand the binding pocket and structure-based drug discovery opportunities, five novel high resolution cryo-EM structures of GLP-1R bound to small molecule agonists with different biased signaling properties were determined (aleniglipron, lotiglipron, compound 73, compound 3b, and compound 355), together with a comparative structural analysis of other known GLP-1 small molecule structures. We demonstrate how complementary lipophilic hotspot and water network analyses of the multiple GLP-1R structures provide new insights into GLP-1R agonist structure-activity relationships (SAR) and receptor activation mechanism, enabling new GLP-1 small molecule drug design strategies.

INTRODUCTION

The glucagon-like peptide-1 receptor (GLP-1R) belongs to the class B GPCR subfamily, whose members are essential components in human physiological processes. GLP-1R serves as a valuable drug target for many diseases including obesity, diabetes, metabolic syndrome, osteoporosis, migraine, depression, and anxiety[1,2]. The glucagon-like peptide (GLP)-1 is a gastrointestinal peptide or neuroendocrine hormone that acts on GLP-1R to: 1) enhance glucose-dependent insulin secretion; 2) inhibit glucagon secretion; and 3) slow gastric emptying, reducing food intake. GLP-1 and the GLP-1R pathway is an important mechanism for obesity and related diseases[3–5]. Several approved GLP-1R targeting peptide therapeutics such as exenatide, liraglutide, lixisenatide, semaglutide, and tirzepatide have provided solid clinical validation for obesity and type 2 diabetes mellitus (T2DM) treatment and have over time circumvented issues inherent to peptide therapeutics (Fig. 1a). While successive generations of these peptides have improved half-lives, reducing dosing schedules from twice daily to once weekly, with the latest formulations for liraglutide and semaglutide having even enabled limited oral dosing, issues still exist for these peptides in terms of convenience as well as the scalability required to serve the global population with obesity and T2DM that numbers in the 100s of millions[6].

A drug-like, oral small molecule agonist of GLP-1R has long been considered a “holy-grail” of next generation obesity and T2DM therapeutics, offering advantages over peptide therapeutics with superior properties such as good oral absorption, dosing flexibility, and potentially better timing with naturally occurring incretin cycles[8,9]. However, the identification of such small molecules to replace peptides, particularly for class B GPCRs has been challenging. Peptide activation of class B GPCR requires a characteristic two-site binding model starting with the C-terminus of the peptide hormone initiating recognition with the extracellular domain (ECD). This in turn allows the peptide’s N-terminus to engage

deeply within the binding pocket at the core of the transmembrane domain (TMD), triggering conformational change proximal to the intracellular region and activating the downstream signaling pathway[10–14]. The interaction network between the peptide and its cognitive GPCR is extensive and difficult to achieve by a small molecule drug (Fig. 1b). Thus, despite decades of searching, few small molecule compounds have shown suitable efficacy on GLP-1R.

Despite these challenges, several compounds with reported activity on GLP-1R have been developed. These small molecules include: TT-OAD2[15], Boc5[16–18], WB4-24[18], CHU-128[19,20], orforglipron[21], danuglipron[22], lotiglipron[23], compound 3b[24], compound 73[25–27], compound 355[28], aleniglipron[29], LSN3160440[30], LSN3318839[31], BETP[32], and cpd-2[33]. Of these, LSN3160440 and LSN3318839 act as PPI stabilizers at the interface of GLP-1R and peptide agonists, whereas cpd-2 and BETP are covalent modulators for GLP-1R. These PPI stabilizers and allosteric modulators are not covered in the current work. The remainder of these small molecules are non-peptide orthosteric agonists that can fully activate GLP-1R. Of these, orforglipron completed phase III clinical trial with promising results, while the development of danuglipron was discontinued after Phase II clinical trial due to safety findings. The rest are either in early clinical development or they did not advance to clinical stage (Table 1). These setbacks are largely associated with large molecular weight, lack of drug-like properties, and lower intrinsic activity relative to endogenous peptide counterparts. Consequently, there remains a strong need for novel, improved, and drug-like GLP-1R agonists that preserve the therapeutic benefits of peptides while addressing their inherent disadvantages.

Structure-based drug discovery is a powerful approach for optimizing and designing new drugs as it can reveal the interaction mode between ligand and receptor with high resolution molecular insight. Here, we introduce a novel Class B GPCR binding pocket nomenclature to enable the comparison of the diverse binding modes and receptor conformations of eleven small molecule GLP-1R agonists, covering four different small molecule chemotypes targeting the orthosteric intrahelical GLP-1R TMD. We present a comparative structural analysis of cryo-EM structures of GLP-1R bound to diverse small molecule agonist chemotypes, including five novel, previously unpublished, high resolution cryo-EM structures (lotiglipron, compound 3b, compound 355, compound 73, aleniglipron). Combined with computational lipophilic hotspot and water network analyses, this research provides new detailed insights into the GLP-1R small molecule agonist structure-activity relationship (SAR), receptor activation mechanisms, and small-molecule drug design strategies and serves as a framework to inform future molecular design and optimization efforts.

RESULTS

GLP-1R small molecule agonists belong to four chemotype classes

We have grouped the reported eleven GLP-1R small molecule agonists into four chemotypes based on their chemical structure: TT-OAD2 defines chemotype 1, Boc5 defines chemotype 2, danuglipron as a representative compound of chemotype 3, and orforglipron and aleniglipron represent chemotype 4. The

physical-chemical property, chemotype class, clinical development stage, and available structural information for these small molecule agonists are summarized in Table 1.

Table 1. Structure determination of representative GLP-1R small molecules.

Chemotypes	Compound	Clinical stage	Resolution (Å)	PDB code	Heavy atoms	Number of pockets	Water molecules
Chemotype 1	TT-OAD2	Terminated phase I	3.0	6ORV	61	4	-
Chemotype 2	Boc5	-	2.6	7X8R	76	5	-
	WB4-24	-	3.1	7X8S	72	5	-
Chemotype 3	danuglipron	Terminated Phase II	2.5 2.8 3.2	6X1A 7LCL 7LCK	41	4	23
	lotiglipron	Terminated Phase I	2.3	9VC2	41	4	26
	compound 355	-	3.0	9VC4	43	4	14
	compound 3b	-	3.0	9VC5	44	4	10
Chemotype 4	orforglipron	Phase III	3.1	6XOX	65	6	-
	CHU-128	-	2.1	6X19	65	6	30
	aleniglipron	Phase II	2.5	9VC1	66	6	45
	compound 73	-	2.4	9VC3	65	6	38

GLP-1R peptide and orthosteric small molecule agonists binding pocket nomenclature

We introduce a novel Class B GPCR binding pocket nomenclature to enable the comparison of the diverse binding modes and receptor conformations of peptide and small molecule chemotypes targeting the orthosteric intrahelical GLP-1R TMD, as presented in Figure 2-3. This pocket nomenclature system assigns seven different pocket regions, following the backbone trace of the peptide agonist (tirzepatide) from the top of the orthosteric intrahelical binding site targeted by peptide agonists (corresponding to

position Asp^{15}) towards the deep TMD binding pocket (targeted by the N-terminus of peptide agonists), as shown in Figure 2b, c. Each of the seven binding pocket regions are further differentiated into subpockets by annotating the pocket number with ‘a’, ‘b’, or ‘c’. For example, the residue sidechain of tirzepatide covers seven different subpockets, including: subpockets 1a (Aib^{13}) and 1b (Asp^{15}) between the ECD, Extracellular Loop 1(ECL1) and the top of TM1/TM2, subpocket 2a (Tyr^{10}) between the top of TM1/TM2, subpocket 3a (Phe^{6}) between the top of TM1/TM7, subpocket 6a (Aib^{2}) between TM5/TM6, subpocket 6b (Asp^{3}) between TM6/TM7 towards the bottom of the intrahelical binding site, and subpocket 7a (Tyr^{1}) between TM3/TM5 (Figure 2d, e). This nomenclature, combined with the generic GPCR residue numbering system[7] and the common alpha helical peptide ligand framework allows translation of GLP-1R ligand binding analysis to other Class B GPCRs.

GLP-1R orthosteric small molecule agonists binding pocket

In the past six years, experimental cryo-EM structures of six unique GLP-1R small molecule agonist complexes have been reported. When combined with the five novel structures we determined in this work, structural information is now available for eleven reported GLP-1R small molecule agonists. We conducted structural analysis of the binding pockets of all four chemotype classes, comparing the small molecule agonists binding results with the seven subpockets associated with tirzepatide binding, and present the peptide-shared and small molecule-specific pockets below (Figure 2).

TT-OAD2 of chemotype 1 has a distinct binding mode and does not bind to any of the peptide interacting pockets: subpocket 2c between TM1/TM2; subpockets 4a/4b between TM3/TM4; and subpocket 4c that protrudes between ECL1, TM3, and ECL2, placing its 3,4-dichloro-benzyl moiety into the extrahelical membrane environment thereby explaining the largely lipophilicity driven SAR of this chemotype[15] (Figure 2b, 3b). The three other GLP-1R small molecule agonist chemotypes target at least one of the seven subpockets identified in tirzepatide binding (Figure 3a). Subpocket 1a (Aib^{13}) accommodates the central scaffolds of both chemotypes 3 and 4, albeit in a different conformation for chemotype 3 (Figure 3d, e) that is associated with an inward movement of ECL1 (Figure 4c). Subpocket 2a (Tyr^{10}) accommodates one of the furan rings of Boc-5 and the 6-membered ring of the indazole of orforglipron (Figure 3c, e). Subpocket 3a (Phe^{6}) is specifically targeted by the substituted phenyl ring of orforglipron (Figure 3e). Subpockets 4a and 4b are simultaneously targeted by Boc-5 (-((tert-butoxycarbonyl)amino)phenyl) and orforglipron (tetrahydropyran-indole) in chemotypes 2 and 4, whereas danuglipron (oxetane) of chemotype 3 only binds subpocket 4a (Figure 3a, c-e). Subpocket 6a is specifically bound by one of the furan rings of Boc-5 (Figure 3c), while subpockets 6b and 7a are not targeted by any of the small molecule agonists.

Together the four small molecule chemotypes target twelve different binding subpockets. Only subpocket 4a is targeted by all four chemotypes, whereas subpocket 4b is targeted by chemotypes 1, 2, and 4 and subpockets 1a and 2a accommodate 2 of the 4 chemotypes (Figure 3a). The eight other

pockets are small molecule chemotype-specific (subpockets 1b, 3a, 3b, 4c, 5a, 6a), reflecting the diversity of small molecule agonist binding modes (Figure 3).

Diverse orthosteric small molecule agonist binding modes cover unique combinations of GLP-1R subpockets and conformations

Overall, representative compounds of all four small molecule agonist chemotypes demonstrated orthosteric binding through extensive interactions with ECD, TM1/TM2/TM3, ECLs and TM7 (Figure 4). Among the four chemotypes, chemotypes 2 and 4 share a similar binding mode that makes close contact with TM1, TM2-ECL1-TM3, and TM7 (Figure 4b, d). Chemotype 1 (TT-OAD2) overlaps only partially with Boc5 and orforglipron in chemotypes 2 and 4. The 3,4-dichloro-benzyl ring moiety in TT-OAD2 interacts with TM3 and ECL2 in a similar fashion as the 4-((tert-butoxycarbonyl)amino)phenyl and Tetrahydropyran-indole groups in Boc5 and orforglipron, whereas the 3,4-dichloro-benzyl ring in TT-OAD2 protrudes beyond the orthosteric pocket, embedding between TM2 and TM3 into the extrahelical membrane environment (Figure 4a,b,d).

Danuglipron of chemotype 3 targets a unique GLP-1R conformation associated with significant inward movement of ECL1 (ECL1-in) and TM7/ECL3 (TM7-in) compared to the orforglipron-, Boc5-, and TT-OAD2-bound structures (Figure 4). While all four chemotypes share conserved interactions with several residues in TM2 (K197$^{2.67b}$, L201$^{2.71b}$) and TM3 (F230$^{3.33b}$) (Figure 5f), danuglipron forms unique interactions with several residues in ECL1 (L217^{ECL1}, Q221^{ECL1}) via its fluorobenzyl-oxy moiety (Figure 5d). In addition, the danuglipron chemotype-induced TM7-in conformation forms a unique subpocket 5a between ECL2/TM3 (T298$^{ECL245.52b}$, F230$^{3.33b}$ and M233$^{3.36b}$) and the top of TM7/ECL3 (R380$^{7.35b}$, F381$^{7.36b}$ and L384$^{7.39b}$), accommodating the benzimidazole carboxylic acid group of danuglipron (Figure 3d, Figure 5d, f). Similar interactions with residues L384$^{7.39b}$ and L388$^{7.43b}$ in TM7 were observed in the peptide binding mode as well (Figure 5f). The ECL3 in the TT-OAD2, Boc5, and orforglipron structures exhibits an open conformation and forms almost no interactions with the ligands (Figure 4, 5). This suggests that ECL3 is crucial for danuglipron binding but may not be involved in the binding of the other small molecule ligands we studied. Therefore, for small molecule agonists, achieving a few key interactions with the receptor is sufficient for ligand binding and receptor activation.

The top of TM7/ECL3 in the TT-OAD2, Boc5, and orforglipron structures exhibits an open conformation, previously reported as critical to efficient β-arrestin recruitment and biased signaling. Consistent to this conformational feature, TT-OAD2, Boc5, and orforglipron are biased GLP-1R agonist. In contrast, danuglipron, associated with a closed TM7/ECL3 conformation, is a non-biased agonist[21].

In general, all four small molecule chemotypes bind in the orthosteric pocket, sharing conserved interaction sites that are important for inducing GPCR conformational change and Gs binding site exposure. These four small molecule chemotypes also demonstrate considerably different binding mode features. As TT-OAD2 and Boc5 are not in active development due to lack of drug-like properties, and since danuglipron and orforglipron have reached clinical trials, we present below a thorough comparison

of compounds within and between chemotypes 3 and 4 to uncover insights into the specific binding mode of these two chemotype classes to help guide the design of next generation GLP-1R small molecule agonist drugs.

Druggability analysis of small molecule chemotypes

Following the binding pocket analysis, we performed complementary computational druggability analyses of the eight different representative molecules of chemotypes 3 and 4, covering the chemical space of current GLP-1R small molecule agonist clinical agents. We used GRID Molecular Interaction Field (GRID-MIF) analysis to describe the shape of the binding pocket surface and to identify lipophilic hotspots[34,35]. Combined with WaterFLAP water network analysis, applied previously for the analysis and structure-based drug design (SBDD) of different Class A GPCRs[36–38], we show how the displacement of high-energy (relative to bulk solvent, “unhappy”) water molecules from lipophilic hotspots and the stabilization of water molecule networks drives small molecule agonist binding. Here, we selected danuglipron and aleniglipron representative analogues as representative GLP-1R agonists of chemotypes 3 and 4 to further explain SAR[22,29] and illustrate the interplay between lipophilic subpocket engagement and water-network stabilization.

Chemotype 3

The integrated comparative structural and computational analyses provide several new insights into GLP-1R- small molecule agonist interactions, as demonstrated by the new lotiglipron, compound 3b, and compound 355 structures, which were poorly resolved, not observed, or inconsistent in the previously reported danuglipron-bound GLP-1R structures[20] (Figure 6a-d).

Here we observe that R380$^{7.35b}$ does not form an H-bond interaction with the carboxylic acid moiety of danuglipron, but it is observed in the lotiglipron, compound 3b, and compound 355 structures, explaining the importance of this acidic moiety in SAR[8]. Comparative structural and computational analysis of danuglipron, lotiglipron, compound 3b, and compound 355 reveals water-mediated H-bond networks furthermore showing how the agonist’s carboxylic acid moiety plays a role in stabilizing water networks in lipophilic pocket 5a via w1 (lotiglipron), w2 (danuglipron, lotiglipron, compound 3b, compound 355), and w3 (lotiglipron, compound 355). This provides another role for the carboxylic acid moiety beyond direct ionic/H-bond interactions with the receptor. While the nitrile moiety does not form an H-bond with Q221^{ECL1} in danuglipron structure, the new lotiglipron, compound 3b, and compound 355 structures show polar interactions between F/Cl groups and the NH sidechain amide moiety of Q221^{ECL1} (compound 3b, compound 355) and/or Q37 (lotiglipron, compound 3b, compound 355) in subpocket 1b. These interaction networks explain the role of these F/Cl moieties in GLP-1R binding beyond lipophilic interactions, consistent with SAR studies of danuglipron, which demonstrate that small polar substituents at the 4-position (e.g., chloro, fluoro, cyano) of CN,F-substituted benzyl group confer the highest potency. Removing the 4-position substituent of pyridine (compound 3-2) in danuglipron analogue compound 3-1 results in a greater than 10-fold reduction in potency[22] (Figure 7a). Comparison

of danuglipron, lotiglipron, compound 3b, and compound 355 structures provide insights into the role of conformational design for optimizing and chemical variation of this chemotype. In danuglipron, the pyridine-methoxy dihedral conformation (to avoid repulsion of the N and O lone pairs) provide a conformational lock, placing the CN,F-substituted benzyl moiety in subpocket 1b (Figure 6a). In lotiglipron and compound 3b, the rigidified dioxy-ring directs the pyridine moiety into subpocket 1b and locks the pyridine N into a favorable location between the lipophilic hotpots (Figure 6b-c). In compound 355, the conformational preference of the CH2-O linker with respect to the pyrimidine ring (avoid O-N repulsion) locks its bioactive conformation towards subpocket 1b (Figure 6d). In all cases above, polar moieties (O, N) are placed at specific locations between lipophilic hotspots in subpockets 1a and 1b. $Q221^{ECL1}$ forms an H-bond with the oxetane oxygen atom in the lipophilic hotspot of subpocket 4a[8,23] (Figure 6a-d). The SAR data of danuglipron analogue compound 3-4 confirm its importance, as deletion of the oxetane (compound 3-5) or replacement with methyl group (compound 3-1) results in close to 10-, and 100-fold potency losses, respectively (Figure 7a).

Previous studies have identified $W33^{ECD}$ as a key residue in danuglipron-induced receptor activation and that it contributes significantly to the binding and activity of danuglipron, consistent with mutation studies[8]. Comparative structural analyses suggest that W33 is critical for placing the trimethylene oxide moiety of danuglipron within subpocket 4a and for stabilizing the conformation of the pyridine ring in subpocket 1b. Our comparative analysis shows the importance of using multiple high-resolution structures in SAR interpretation and SBDD targeting class B GPCRs. Together, these observations indicates some key design principles for chemotype 3: (i) small polar substituents at the pyridine 4-position enhance potency via halogen-mediated polar interactions; occupying hydrophobic volume (subpocket 1b) is beneficial. (ii) incorporation of oxetane substituents can balance potency and physicochemical properties by leveraging lipophilic hotspot H-bonding; and (iii) carboxylic acids positioned to stabilize structured waters, particularly in pocket 5a, provide an additional mechanism for potency gains beyond direct receptor contacts.

Chemotype 4

As above, the integrated comparative structural and computational analysis again provides several new insights into GLP-1R-small molecule agonist interactions, this time demonstrated in the new compound 73 and aleniglipron structures, which were poorly resolved, not observed, or inconsistent in the previously reported orforglipron and CHU-128 bound GLP-1R structures[20,21] (Figure 8).

We observe that the indazole moieties of orforglipron and CHU-128 show different orientations (flipped 180°) in subpockets 2a and 2b (Figure 8a, b). The indazole moiety of the new compound 73 structure adopts a binding mode that is consistent with the CHU-128 orientation (but not with the previously reported orforglipron binding mode). The confirmed indazole binding mode is consistent in the druggability analysis (showing the polar N is placed between the lipophilic hotspots). The THP/morpholine rings of orforglipron and CHU-128 show different orientations (flipped 180°) resulting in different locations of the gem-dimethyl group in subpocket 4b. The THP ring and gem-dimethyl moiety of

the new compound 73 structure adopt a binding mode that is consistent with orforglipron (but not with the previously reported CHU-128 binding mode) (Figure 8a-c). The confirmed location of the gem-dimethyl group is in line with reported SAR[39] and consistent with our structure-based druggability analysis combining GRID and WaterFLAP analyses, placing the moiety in the lipophilic hotspots close to high energy apo water molecules.

The methylated cyclopropyl moiety of CHU-128 is solved in an incorrect stereochemical geometry[20], while the same moiety is consistently solved in the correct stereochemical geometry of the orforglipron and the new aleniglipron structure (Figure 8a,b,d). K197$^{2.67b}$ does not form an H-bond interaction with the oxadiazol-5-one group of the CHU-128 structure, although this H-bond is observed in the orforglipron, compound 73 and aleniglipron structures, explaining the importance of this acidic isostere for GLP-1R potency[40]. In SAR studies of aleniglipron, substituting oxadiazol-5-one group with methyl group (compound 4-1) or carboxyl group (compound 4-2) decreases activity by approximately 10- and 3000-fold, respectively, highlighting the importance of critical polar interactions [29](Figure 7b). Similarly, Y220^{ECL1} does not form an H-bond interaction with the tetrahydropyran group of orforglipron, but forms H-bond in the CHU-128, compound 73, and aleniglipron structures, explaining the importance of this accepting oxygen in SAR. Substitution of this oxygen with fluorine (compound 4-3) in aleniglipron leads to a two-fold loss in potency, consistent with the contribution of this moiety to favorable desolvation.

In addition to resolving the apparent discrepancies above between the earlier CHU-128 and orforglipron structures, the new compound 73 and aleniglipron structures provide additional insights into SAR and SBDD strategies. The new compound 73 structure shows how one of the meta-substituents of the phenyl ring of orforglipron chemotype is of particular importance[26], displacing unfavorable water molecules from the lipophilic hotspot in subpocket 3a with its para-fluor and single meta-cyclo-propyl moieties (Figure 8b). It's also consistent with the SAR data of aleniglipron. Deletion of the fluoro and methyl substituents (compound 4-5) in phenyl ring of aleniglipron analogue compound 4-4 diminishes potency approximately 15-fold, while substitution with isopropyl (compound 4-6) produces more dramatic losses of close to 100-fold. Computational water network analysis of the new aleniglipron and compound 73 structures shows how the central amide carbonyl O plays a role in stabilizing the otherwise unhappy water w1 via a bridging H-bond network with T298^{ECL2} in subpocket 4a (Figure 8c, d). The new aleniglipron structure shows how the introduction of the phosphonyl O provides an opportunity to form a water mediated H-bond interaction with K202^{ECL1} in subpocket 2b, while forming an intramolecular H-bond with the amine NH (Figure 8d). The removal of the entire phosphonyl oxygen group (compound 4-7) reduces activity by approximately 200-fold, featuring crucial polar contacts (phosphonyl-O networks), as well as the diethylphosphoryl-phenyl also helps fill hydrophobic volume adjacent to 2a/2b (Figure 7b, 8d). In addition, water network analysis reveals that the central amide carbonyl oxygen stabilizes an otherwise high-energy water molecule (w1) through a bridging H-bond interaction with T298^{ECL2} in subpocket 4a, further contributing to ligand efficacy.

Similar to its role in danuglipron binding, W33^{ECD}A mutation completely abolishing orforglipron-induced receptor activity. The structural analysis of chemotype 4 ligand-bound complexes reveal that W33^{ECD} forms extensive hydrophobic interactions with the ligands and places the indole and THP moieties in the lipophilic hotspots of subpocket 4a and subpocket 4b, respectively (Figure 8f). These findings confirm the essential role of W33^{ECD} in stabilizing chemotype 4 binding and activity, which is consistent with in vitro potency data[21].

Together, these observations establish some key design principles for chemotype 4 ligands: (i) retain a 4-fluoro-3,5-xylyl–like vector into subpocket 3a to occupy the lipophilic hotspot and displace high-energy waters (avoiding i-Pr/dimethyl swaps that erode potency); (ii) preserve a diethyl phosphoryl-phenyl handle in 2a with an intact phosphonyl O to enable a water-bridged contact to K202^{ECL1} and intramolecular preorganization—loss of which drives 10–300× potency penalties; and (iii) engage subpocket 4b with a THP or functionally equivalent lipophile while positioning the central amide carbonyl to stabilize a minimal, productive water network via T298^{ECL2} (4a)—thus coupling water displacement with selective water stabilization for maximal binding gain.

The structural and computational analyses of chemotype 3 and chemotype 4 ligands highlight common principles that can inform the design of next-generation GLP-1R agonists. Beyond scaffold differences, successful ligands consistently (i) occupy key lipophilic hotspots to maximize hydrophobic complementarity, (ii) displace high-energy waters to relieve entropic penalties, and (iii) incorporate polar or acidic substituents that stabilize structured water networks or form productive receptor contacts. These insights provide a transferable framework for scaffold hopping and the rational design of novel GLP-1R agonists with improved potency and drug-like properties.

DISCUSSION AND CONCLUSIONS

We have introduced a novel Class B GPCR pocket nomenclature to enable the comparison of the diverse binding modes and receptor conformations of different small molecule chemotypes targeting the orthosteric intrahelical GLP-1R TMD. This common framework is in line with earlier GPCR subfamily-specific druggability assessment and binding pocket naming[36], recent comparative binding pocket nomenclatures proposed for allosteric GPCR binding pockets[41], and is enabled by generic GPCR residue numbering system[7].With the increasing structural insights into Class B GPCR-ligand interactions, this system provides a potential basis for structural chemogenomics-based binding site analysis and ligand design approaches, as described for Class A GPCR subfamilies[42–44] and other protein families with more extensive structural information, like kinases[45].

According to comparative analysis of chemotype 3 and 4 binding pockets, we revealed shared interactions between the chemotype 3 and 4 ligands and the GLP-1R TMD, including TM1, TM2-ECL1-TM3, and ECL2, specifically involving residues L141$^{1.36b}$, K197$^{2.67b}$, L201$^{2.71b}$, F230$^{3.33b}$, M233$^{3.36b}$, and T298$^{45.52b}$. These conserved interactions overlap with part of the GLP-1 peptide binding interaction

network, indicating that the TM1, TM2-ECL1-TM3, and ECL2 regions constitute a critical part of the orthosteric pocket for stable ligand binding and receptor activation.

Outside of the conserved pattern, chemotype 3 ligands interact more specifically with TM7 (R380$^{7.35b}$, F381$^{7.36b}$, and L384$^{7.39b}$) while chemotype 4 ligands primarily interact with residues in TM1 (P137$^{1.32b}$, E138$^{1.33b}$, L144$^{1.39b}$, Y145$^{1.40b}$, and Y148$^{1.43b}$). These structural insights suggest a combination design strategy of next-generation GLP-1R agonist molecules beyond the four chemotypes discussed in this study. Our structural analysis clearly shows the chemotype 3 and 4 ligands do not utilize TM5 and chemotype 4 ligands do not utilize ECL3 in their binding, which is vastly different from what is observed in GLP-1 peptide binding mode. This suggests another design direction for novel small molecules: to design chemical moieties that can mimic the extensive interactions in TM7, TM5, and ECL3 to potentially enhance small molecule binding affinity and activity.

Identification of small molecule compounds with drug-like properties and comparable potency has been particularly challenging for class B GPCRs. The challenges include: the difficulty for small molecules to mimic the entirety of a peptide's interaction network; understanding which part of this interaction network is crucial and essential for binding and activity; and the difficulty for small molecules to engage ECD, which has been well-established as crucial for peptide recognition and receptor activation. The structure analyses described in this work show that ligands from all four chemotypes make only limited interaction, if any, with ECD, clearly indicating a different role of ECD in peptide recognition and binding from its role in small molecule binding.

The conformational changes observed in structures of chemotype 3 and 4 ligands and the GLP-1 peptide also strongly demonstrate that class B GPCRs are intrinsically dynamic and adopt different helical bundles and ECL conformations upon ligand binding. To design novel scaffolds of small molecule drugs for this type of dynamic receptors, protein co-folding prediction algorithms like AlphaFOLD3[46], combined with molecular dynamic simulation and experimental cryo-EM data, will build a stronger structural understanding of the orthosteric pocket, supporting design ideas proactively.

The current work demonstrates the value of using multiple structures and combining structural and computational druggability analyses to provide a comprehensive and higher resolution view of protein-ligand interactions to guide SBDD. The five novel structures presented in this work provide several new detailed insights into GLP-1R-small molecule agonist interactions that were poorly resolved, not observed, or inconsistent in the previously reported danuglipron-bound and orforglipron/CHU-128-GLP-1R structures[20,21]. The computational analyses provide orthogonal information on the importance of targeting lipophilic hotspots to displace high energy, unfavorable water molecules as drivers of Class B GPCR druggability and ligand binding. Furthermore, the structural set presented here highlights the key role of stabilizing water mediated interaction networks as additional key determinants of Class B GPCR-ligand binding.

In summary, we have grouped the recently reported small molecule GLP-1R agonists into four chemotype classes and examined their binding mode to deepen our understanding of molecule design. Our detailed analysis of commonality and uniqueness across these chemotype classes, comparing to GLP-1R peptide agonists suggests promising combination design strategy to optimize current chemotype ligands as well as to identify novel scaffolds and chemotypes.

EXPERIMENTAL SECTION

Expression and purification of small molecule agonist-bound GLP-1R-Gs-Nb35 complex

Human GLP-1R (Arg24 - Ser463) fragments fused with N-terminal HA-FLAG- tags and C-terminal TEV-10× His tags were synthesized and inserted into a modified pFastBac-1 vector (BamHI/HindIII sites) under the polh promoter (Genewiz). A dominant-negative Gαs (DNGαs) was generated as previously described[23] by introducing eight mutations: S54N, G226A, E268A, N271K, K274D, R280K, T284D, and I285T. Human $G\beta_1$ and $G\gamma_2$ were inserted into a pFastDual plasmid and were prepared by insertion of His_6-tagged human $G\beta_1$ fragment (BamHI/HindIII site) under the polh promoter and of $G\gamma_2$ fragment (XhoI/SphI site) under the p10 promoter, respectively (Genewiz).

Human GLP-1R, human DNGαs, and His_6-tagged human $G\beta_1$ and $G\gamma_2$ were expressed in sf9 insect cells using baculovirus and then infected with three separate baculoviruses at a ratio of 2:1:1 for GLP-1R, DNGαs, and $G\beta_1\gamma_2$. The cells were infected at a density of $2x10^6$ cells per ml and collected 48 h after infection and cell pellets were stored at − 80°C.

Lotiglipron, compound 3b, compound 355, compound 73, and aleniglipron were synthesized according to reported procedures[23,24,26,28,29]. All compounds are > 95% pure by HPLC.

The cell pellet was thawed, homogenized, and lysed in a buffer containing 20 mM HEPES pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$, and complete Protease Inhibitor Cocktail tablet (Roche), and the membranes were collected by centrifugation at 40,000 *g*. This step was repeated one additional time before the GLP-1R-Gs complex was formed by the addition of 50 μM ligands, 10 μg/mL Nb35[47,48], and 25 mU/mL apyrase to the homogenized membrane suspension. The suspension was incubated for 1 h at room temperature before the membranes were isolated by centrifugation at 40,000 *g* for 30 min. The membrane-bound complex was solubilized by 0.5% (w/v) lauryl maltose neopentyl glycol (LMNG) supplemented with 0.05% (w/v) cholesteryl hemisuccinate (CHS) for 2 h at 4°C in the presence of 50 μM ligands, 10 μg/mL Nb35, and 25 mU/mL apyrase. Insoluble material was removed by centrifugation at 40,000 *g* for 30 min, and the supernatant was incubated with TALON resin (Clontech) and 20 mM imidazole overnight at 4°C. The resin was washed by 20 column volumes of wash buffer A [20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 0.01% (w/v) LMNG, 0.001% (w/v) CHS, 10 μM ligands, and 30 mM imidazole] and 10 column volumes of wash buffer B [20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 0.01% (w/v) LMNG, 0.001% (w/v) CHS, 10 μM ligands, and 50 mM imidazole]. The complex was eluted by 5 column volumes

of elution buffer [20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 0.01% (w/v) LMNG, 0.001% (w/v) CHS, 30 μM ligands, and 300 mM imidazole]. The eluted protein was concentrated using an Amicon Ultra Centrifugal Filter (MWCO 100 kDa). The complex was subjected to size-exclusion chromatography (SEC) on a Superdex 200 Increase 10/300 column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 1 μM ligands, 0.00075% (w/v) LMNG and 0.000075% (w/v) CHS. The eluted peak fraction containing the small molecule agonist-GLP-1R-Gs-Nb35 was diluted to 10 mg/mL and used for single particle EM samples preparation and data collection.

Cryo-EM sample preparation and data collection

A 3.0 μL aliquot of the concentrated sample was applied to glow-discharged gold grids (Ultrafoil R1.2/1.3, Au 300 mesh, Quantifoil GmbH) and was vitrified using a Vitrobot Mark IV (Thermo Fisher Scientific) under 100% humidity at 4°C. For aleniglipron-GLP-1R-Gs, lotiglipron-GLP-1R-Gs, and compound 73-GLP-1R-Gs, data collection was carried out on a Titan Krios G3i transmission electron microscope (Thermo Fisher Scientific) operating at 300 kV and equipped with a Falcon 4i direct electron detector and a Selectris energy filter. Automated data acquisition was performed using the EPU software (Thermo Fisher Scientific). Movies were recorded at a nominal magnification of 130,000× in counting mode, yielding a calibrated pixel size of 0.932 Å. Each movie comprises 34 subframes with a total dose of 50 e − per $Å^2$. A total of 6557, 4266 and 7038 movies were collected for aleniglipron-GLP-1R-Gs, lotiglipron-GLP-1R-Gs, and compound 73-GLP-1R-Gs, respectively. For compound 355-GLP-1R-Gs and compound 3b-GLP-1R-Gs, data collection was carried out on a Titan Krios G4 transmission electron microscope (Thermo Fisher Scientific) operating at 300 kV and equipped with a Falcon 4 direct electron detector. Automated data acquisition was performed using the EPU software (Thermo Fisher Scientific). Movies were recorded at a nominal magnification of 96,000× in counting mode, yielding a calibrated pixel size of 0.81 Å. Each movie comprises 32 subframes with a total dose of 50 e − per $Å^2$. A total of 6489 and 7605 movies were collected for compound 355-GLP-1R-Gs and compound 3b-GLP-1R-Gs, respectively.

Cryo-EM data processing

For the aleniglipron-GLP-1R-Gs, lotiglipron-GLP-1R-Gs, compound 73-GLP-1R-Gs, compound 3b-GLP-1R-Gs, and compound 355-GLP-1R-Gs, initial data processing was performed in cryoSPARC 4.5[49] using standard procedures, including patch motion correction, CTF estimation, and template-based particle picking. After extraction, particles were subjected to 2 rounds of 2D classification to remove false positives and noise. The best classes were used for heterogeneous refinement, followed by a resolution-gradient refinement using three low-pass filtered references (at 5, 15, and 30 Å) to further enrich for high-quality particles. To recover additional good particles potentially discarded in earlier steps, a seed-facilitated 3D classification approach was employed. Particles from the first round of 2D classification were randomly divided into subgroups and combined with a previously selected high-quality particle set

used as seeds. Each combined group underwent heterogeneous refinement and resolution-gradient refinement, leading to an expanded particle dataset. To further clean up the particle dataset, ab initio models were generated and used for an additional round of heterogeneous refinement, followed by non-uniform refinement. To enhance the resolution of specific structural features, local refinement was subsequently performed using a soft mask encompassing the transmembrane domain (TMD) and extracellular domain (ECD). Finally, for the aleniglipron-GLP-1R-Gs, lotiglipron-GLP-1R-Gs, and compound 73-GLP-1R-Gs, reference-based motion correction was performed, which contributed to further improvement of the map quality.

Model building and refinement

The cryo-EM structure of CHU-128-GLP-1R-Gs (PDB: 6X18) was used as the initial model for the aleniglipron- and compound 73-GLP-1R-Gs complexes, while PF-06882961-GLP-1R-Gs (PDB: 6X1A) served as the template for the lotiglipron-, compound 355, and compound 3b-GLP-1R-Gs complex. Initial models were fitted into the EM density maps using UCSF Chimera 1.16[50], followed by iterative manual adjustments in WinCoot 0.9.8[51] and automated refinement in Phenix 1.21[52]. Final models were validated using the comprehensive validation (cryo-EM) module in Phenix 1.21.

Protein Structure preparation for computational analysis

Protein preparation was performed using the Schrödinger Drug Discovery Suite 2024-4 (https://www.schrodinger.com). The Protein Preparation Wizard was used to process the cryo-electron microscopy (cryo-EM) structure, with all water molecules retained throughout the procedure. Hydrogen atoms were added and minimized, while the coordinates of heavy atoms were left unaltered. Protonation and tautomeric states of hetero groups were assigned using Epik at a target pH of 7.4 ± 1.0 to approximate physiological conditions. All other settings were maintained at their default values.

GRID Molecular Interaction Field (GRID-MIF) analysis and lipophilic hotspot calculation

Druggability assessment of the pseudo-apo GLP-1R binding pockets are performed using the GRID molecular interaction field (GRID-MIF) analysis of energetically favorable regions for ligand interactions[34,35]. GRID maps were calculated, contoured (transparent solid), and colored in the following manner: the C1 lipophilic probe in yellow (at − 3.0 kcal/mol) and the C3 methyl group probe in gray (at -0.001 kcal/mol) defined the pocket surface in terms of how close a ligand carbon atom can reside.

WaterFLAP water network analysis

WaterFLAP water networks[36–38] and relative energetics were calculated on the pseudo-apo binding site (with the ligand removed) and in the presence of the small molecule agonist (complex). WaterFLAP water is shown as spheres and color-coded by relative energetic scoring as follows: red (very unfavorable) when predicted free energy (ΔG) is higher than 2.5 kcal/mol; yellow (unfavorable) when ΔG is between 1.0 and 2.5 kcal/mol; cyan if ΔG is between − 1.0 and − 2.0 kcal/mol; and blue when ΔG < − 2.0 kcal/mol. All WaterFLAP free energy estimations are relative to bulk solvent.

Generic GPCR Residue Numbering

The generic GPCR residue numbering system[7] used throughout this paper includes two numbers (*X*.*N*). The first (1–7) denotes the transmembrane helix domain (TMD) and the second indicates the residue position relative to the most conserved amino acid in the helix (which is assigned the number 50), with a 'b' added at the end to denote that this is Class B GPCR residue numbering framework. Conserved residue positions in extracellular loop 2 (ECL2, between TM4 and TM5) is defined as $C^{45.50b}$. For example, 2.67b indicates the residue 17 positions after the most conserved amino acid in class B GPCR TM2 ($H^{2.50b}$). If an amino acid is followed by its residue number, the generic GPCR residue numbering is included as a superscript[7] (e.g., $K197^{2.67b}$ in GLP-1R).

Declarations

Acknowledgments:

We thank Gye Won Han for her assistence in structure refinement and model QC, Lier Liu for his help with Cryo-EM sample preparation and Zhuming Zhang for comments and feedbacks on the manuscript. The EM data were collected at the Instrumental Analysis Center, Jiaotong University, Shanghai and with the support of facility managers.

Author Contributions:

F.W. and D.S. optimized constructs, expressed, and purified proteins, determined the structures, and wrote the initial manuscript; X.S. mentored the cryo-EM work and contributed to the preparation of the manuscript. W.H. performed computational structural analyses and contributed to the preparation of the manuscript. H.Z. Q. M. and C.L. contributed to chemotype SAR analysis. X.L., H.L., Q.M., H.Z., and C.L. contributed to GLP-1R small molecule agonist design, identification, and synthesis. T.M. provided in vitro characterization insight of all molecules discussed in this work. R.C.S mentored the project. C.d.G and Y.M. proposed the project and finalized the manuscript. All authors were involved in discussions and provided comments on the manuscript.

Competing interests:

The authors declare no competing financial interests.

Data availability:

Atomic coordinates and structure factors for lotiglipron-, compound 355-, compound 3b-, aleniglipron-, and compound 73 - GLP-1R structure have been deposited in the Protein Data Bank (PDB) with identification codes: 9VC2, 9VC4, 9VC5, 9VC1, and 9VC3, as well as Electron Microscopy Data Bank (EMDB) under accession codes: EMD-64940, EMD-64942, EMD-64943, EMD-64939, and EMD-64941. All relevant data are available from the authors and are included in the manuscript or Supplementary Information. Correspondence and requests for materials should be addressed to chris.degraaf@structuretx.com and yingli.ma@structuretx.com.

References

1. Graaf, C. de *et al.* Glucagon-Like Peptide-1 and Its Class B G Protein–Coupled Receptors: A Long March to Therapeutic Successes. *Pharmacol Rev* **68**, 954–1013 (2016).
2. Donnelly, D. The structure and function of the glucagon-like peptide-1 receptor and its ligands: GLP-1R structure-function. *British Journal of Pharmacology* **166**, 27–41 (2012).
3. Alavi, S. E. & Seyed Ebrahim Alavi. GLP-1 peptide analogs for targeting pancreatic beta cells. *Drug Discovery Today* **26**, 1936–1943 (2021).
4. Nauck, M. A. & Meier, J. J. Incretin hormones: Their role in health and disease. *Diabetes Obesity Metabolism* **20**, 5–21 (2018).
5. Knudsen, L. B. & Lau, J. The Discovery and Development of Liraglutide and Semaglutide. *Front. Endocrinol.* **10**, 155 (2019).
6. Huang, X., Wu, Y., Ni, Y., Xu, H. & He, Y. Global, regional, and national burden of type 2 diabetes mellitus caused by high BMI from 1990 to 2021, and forecasts to 2045: analysis from the global burden of disease study 2021. *Front. Public Health* **13**, 1515797 (2025).
7. Isberg, V. *et al.* Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends in Pharmacological Sciences* **36**, 22–31 (2015).
8. Griffith, D. A. *et al.* A Small-Molecule Oral Agonist of the Human Glucagon-like Peptide-1 Receptor. *J. Med. Chem.* **65**, 8208–8226 (2022).
9. Brameld, K. A., Kuhn, B., Reuter, D. C. & Stahl, M. Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused Analysis. *J. Chem. Inf. Model.* **48**, 1–24 (2008).
10. Runge, S., Wulff, B. S., Madsen, K., Bräuner-Osborne, H. & Knudsen, L. B. Different domains of the glucagon and glucagon-like peptide-1 receptors provide the critical determinants of ligand selectivity. *British J Pharmacology* **138**, 787–794 (2003).
11. Zhao, L.-H. *et al.* Differential Requirement of the Extracellular Domain in Activation of Class B G Protein-coupled Receptors. *Journal of Biological Chemistry* **291**, 15119–15130 (2016).
12. Hollenstein, K. *et al.* Insights into the structure of class B GPCRs. *Trends in Pharmacological Sciences* **35**, 12–22 (2014).

13. De Graaf, C. *et al.* Extending the Structural View of Class B GPCRs. *Trends in Biochemical Sciences* **42**, 946–960 (2017).
14. Zhou, Q., Zhao, F., Zhang, Y., Yang, D. & Wang, M.-W. Structural pharmacology and mechanisms of GLP-1R signaling. *Trends in Pharmacological Sciences* **46**, 422–436 (2025).
15. Zhao, P. *et al.* Activation of the GLP-1 receptor by a non-peptidic agonist. *Nature* **577**, 432–436 (2020).
16. He, M. *et al.* A continued saga of Boc5, the first non-peptidic glucagon-like peptide-1 receptor agonist with in vivo activities. *Acta Pharmacol Sin* **33**, 148–154 (2012).
17. Chen, D. *et al.* A nonpeptidic agonist of glucagon-like peptide 1 receptors with efficacy in diabetic *db* / *db* mice. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 943–948 (2007).
18. Cong, Z. *et al.* Structural basis of peptidomimetic agonism revealed by small-molecule GLP-1R agonists Boc5 and WB4-24. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2200155119 (2022).
19. Griffith, D. A. *et al. A Small-Molecule Oral Agonist of the Human Glucagon-like Peptide-1 Receptor*. http://biorxiv.org/lookup/doi/10.1101/2020.09.29.319483 (2020) doi:10.1101/2020.09.29.319483.
20. Zhang, X. *et al.* Differential GLP-1R Binding and Activation by Peptide and Non-peptide Agonists. *Molecular Cell* **80**, 485-500.e7 (2020).
21. Kawai, T. *et al.* Structural basis for GLP-1 receptor activation by LY3502970, an orally active nonpeptide agonist. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 29959–29967 (2020).
22. WO2018109607A1 - Glp-1 receptor agonists and uses thereof - Google Patents. https://patents.google.com/patent/WO2018109607A1/en?oq=WO2018109607A1.
23. WO2019239319A1 - Glp-1 receptor agonists and uses thereof - Google Patents. https://patents.google.com/patent/WO2019239319A1/en?oq=WO2019239319A1.
24. WO2023057427A1 - Certain 2,5-diazabicyclo[4.2.0]octanes as glp-1 receptor modulators - Google Patents. https://patents.google.com/patent/WO2023057427A1/en.
25. Zhang, H. *et al.* Binding sites and design strategies for small molecule GLP-1R agonists. *European Journal of Medicinal Chemistry* **275**, 116632 (2024).
26. WO2022017338A1 - Tetrahydropyrazolo-pyrazinyl-dihydroimidazolone or tetrahydropyrazolo-pyridinyl-dihydroimidazolone compounds and methods of using same - Google Patents. https://patents.google.com/patent/WO2022017338A1/en?oq=WO2022017338.
27. SID 507528867 - PubChem. https://pubchem.ncbi.nlm.nih.gov/substance/507528867.
28. US20250092025A1 - Gpcr receptor agonists, pharmaceutical compositions comprising the same, and methods for their use - Google Patents. https://patents.google.com/patent/US20250092025A1/en.
29. WO2021155841A1 - Heterocyclic glp-1 agonists - Google Patents. https://patents.google.com/patent/WO2021155841A1/en?oq=WO2021155841A1.
30. Bueno, A. B. *et al.* Structural insights into probe-dependent positive allosterism of the GLP-1 receptor. *Nat Chem Biol* **16**, 1105–1110 (2020).

31. Li, J. *et al.* Molecular basis of enhanced GLP-1 signaling mediated by GLP-1(9–36) in conjunction with LSN3318839. *Acta Pharmaceutica Sinica B* **14**, 5069–5073 (2024).
32. Eng, H. *et al.* Demonstration of the Innate Electrophilicity of 4-(3-(Benzyloxy)phenyl)-2-(ethylsulfinyl)-6-(trifluoromethyl)pyrimidine (BETP), a Small-Molecule Positive Allosteric Modulator of the Glucagon-Like Peptide-1 Receptor. *Drug Metabolism and Disposition* **41**, 1470–1479 (2013).
33. Cong, Z. *et al.* Molecular insights into ago-allosteric modulation of the human glucagon-like peptide-1 receptor. *Nat Commun* **12**, 3763 (2021).
34. Wade, R. C., Clark, K. J. & Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **36**, 140–147 (1993).
35. Sciabola, S. *et al.* High-Throughput Virtual Screening of Proteins Using GRID Molecular Interaction Fields. *J. Chem. Inf. Model.* **50**, 155–169 (2010).
36. Rappas, M. *et al.* Comparison of Orexin 1 and Orexin 2 Ligand Binding Modes Using X-ray Crystallography and Computational Analysis. *J. Med. Chem.* **63**, 1528–1543 (2020).
37. Mason, J. S., Bortolato, A., Congreve, M. & Marshall, F. H. New insights from structural biology into the druggability of G protein-coupled receptors. *Trends in Pharmacological Sciences* **33**, 249–260 (2012).
38. Baroni, M., Cruciani, G., Sciabola, S., Perruccio, F. & Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands And Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **47**, 279–294 (2007).
39. AU2022205222B2 - Pyrazolopyridine Derivative Having GLP-1 Receptor Agonist Effect - Google Patents. https://patents.google.com/patent/AU2022205222B2/en?q=(Pyrazolopyridine+derivative+glp-1+receptor+agonist+effect)&oq=Pyrazolopyridine+derivative+having+glp-1+receptor+agonist+effect.
40. US12187724B2 - Pyrazolopyridine derivative having GLP-1 receptor agonist effect - Google Patents. https://patents.google.com/patent/US12187724B2.
41. Peter, S. *et al.* Comparative Study of Allosteric GPCR Binding Sites and Their Ligandability Potential. *J. Chem. Inf. Model.* **64**, 8176–8192 (2024).
42. Vass, M. *et al.* Chemical Diversity in the G Protein-Coupled Receptor Superfamily. *Trends in Pharmacological Sciences* **39**, 494–512 (2018).
43. Vass, M. *et al.* Aminergic GPCR–Ligand Interactions: A Chemical and Structural Map of Receptor Mutation Data. *J. Med. Chem.* **62**, 3784–3839 (2019).
44. Arimont, M. *et al.* Structural Analysis of Chemokine Receptor–Ligand Interactions. *J. Med. Chem.* **60**, 4735–4779 (2017).
45. Van Linden, O. P. J., Kooistra, A. J., Leurs, R., De Esch, I. J. P. & De Graaf, C. KLIFS: A Knowledge-Based Structural Database To Navigate Kinase–Ligand Interaction Space. *J. Med. Chem.* **57**, 249–277 (2014).

46. Callaway, E. Major AlphaFold upgrade offers boost for drug discovery. *Nature* **629**, 509–510 (2024).
47. Rasmussen, S. G. F. *et al.* Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
48. WO2012175643A2 - Binding domains directed against gpcr:g protein complexes and uses derived thereof - Google Patents. https://patents.google.com/patent/WO2012175643A2/en?oq=WO2012175643A2.
49. 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).
50. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
52. Adams, P. D. *et al. PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).

Figures

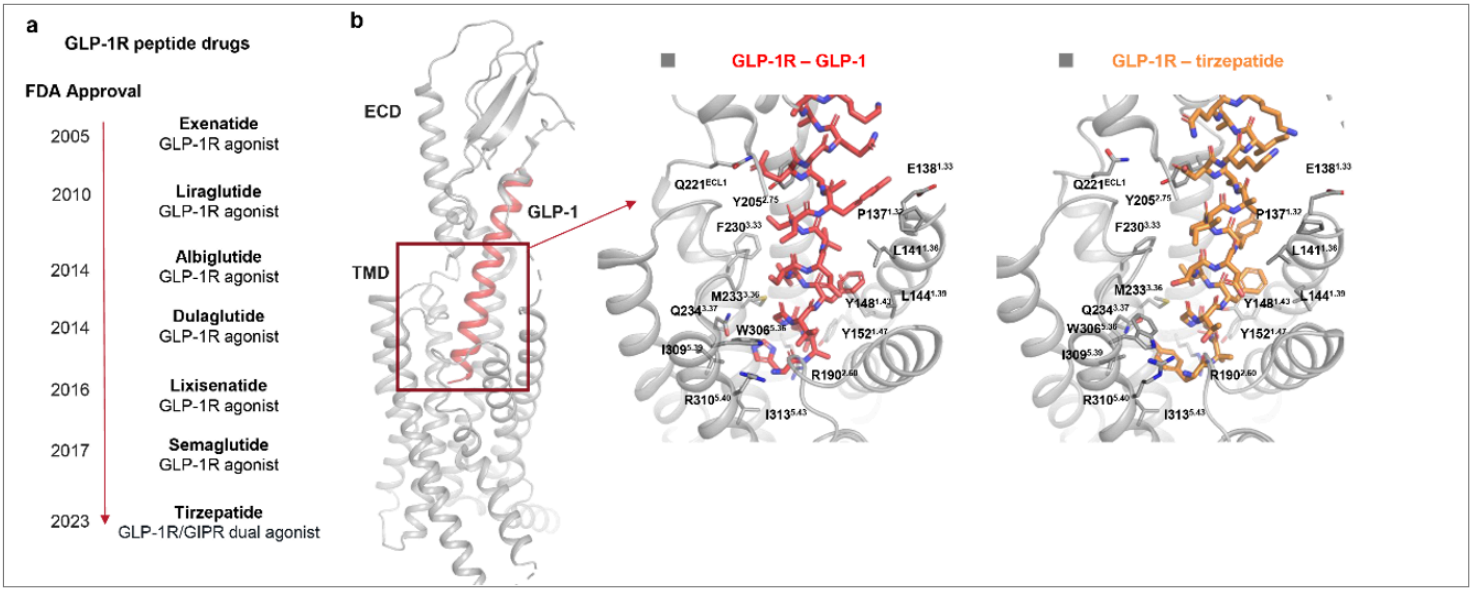

Figure 1

GLP-1R peptide agonists and their binding modes. (a) History of approved GLP-1R targeting peptide therapeutics. (b) Overall binding mode of GLP-1 on GLP-1R (Left); Interaction bindings sites of GLP-1 (center) and tirzeaptide (Right) on GLP-1R Transmembrane domain (TMD). GLP-1R is shown as grey cartoons, GLP-1 peptide is displayed as red cartoon or sticks, tirzepatide is shown as yellow sticks. GLP-1R residues involved in the interactions are shown as grey sticks and labeled with generic GPCR residue numbering[7] in superscript.

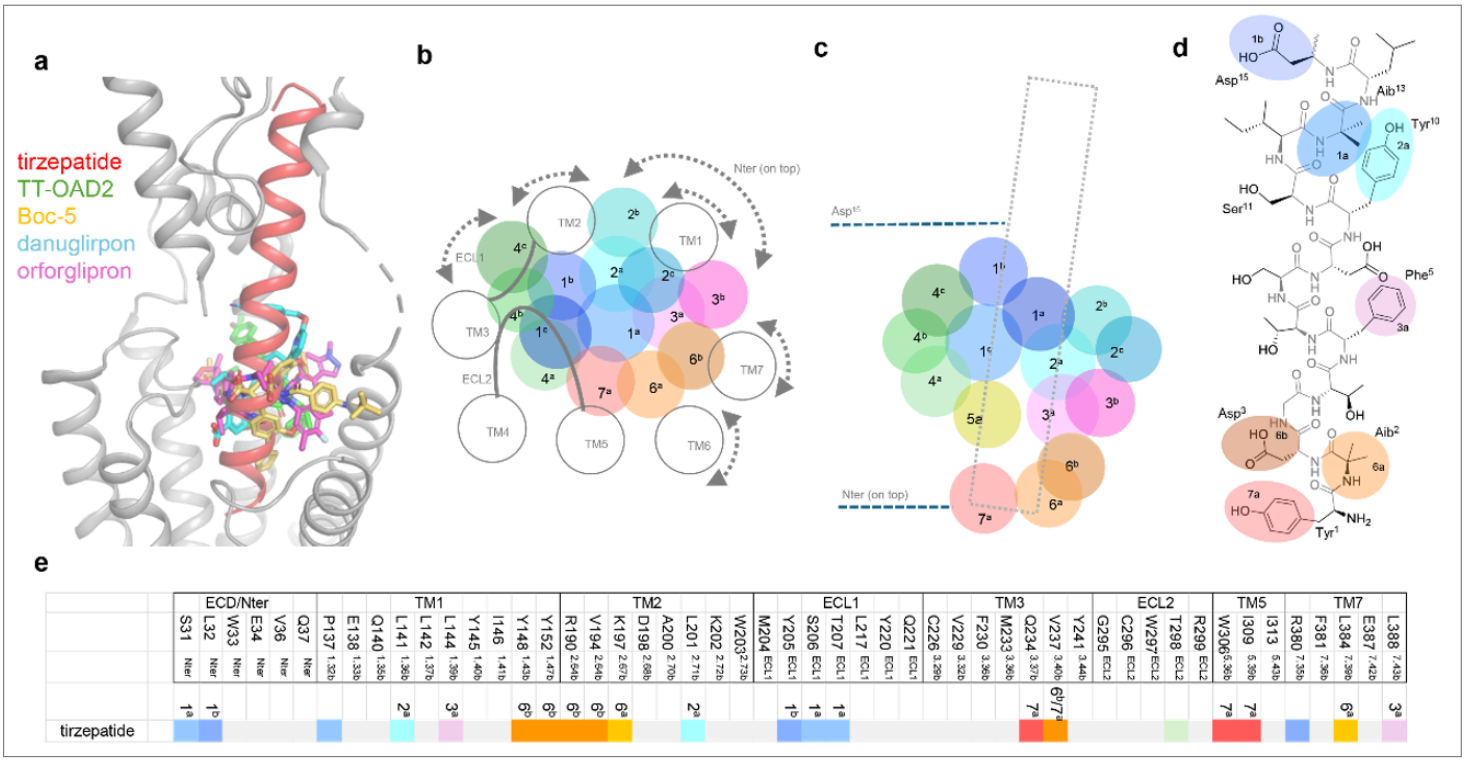

Figure 2

GLP-1R agonist and binding pocket nomenclature. (a) Overlay of tirzepatide peptide (PDB:7RGP) to four small molecule orthosteric agonists: TT-AOD2 (PDB:6ORV), Boc-5 (PDB:7X8R), orforglipron (PDB:6XOX), and danuglipron (PDB:6X1A). (b-c) Schematic depiction top-down (b) and side view (c) of the 7 different pocket regions covered by peptide and small molecule agonists targeting the orthosteric TM binding pocket of GLP-1R. (d) Schematic 2D map of the 7 different pocket regions with respect to the peptide chemical structures. (e) Interaction fingerprints (IFPs) of tirzepatide peptide agonist with different binding subpockets and binding site residues as defined in panel b. In addition to the Uniprot residue number, the generic GPCR residue numbering[7] is provided in a separate column.

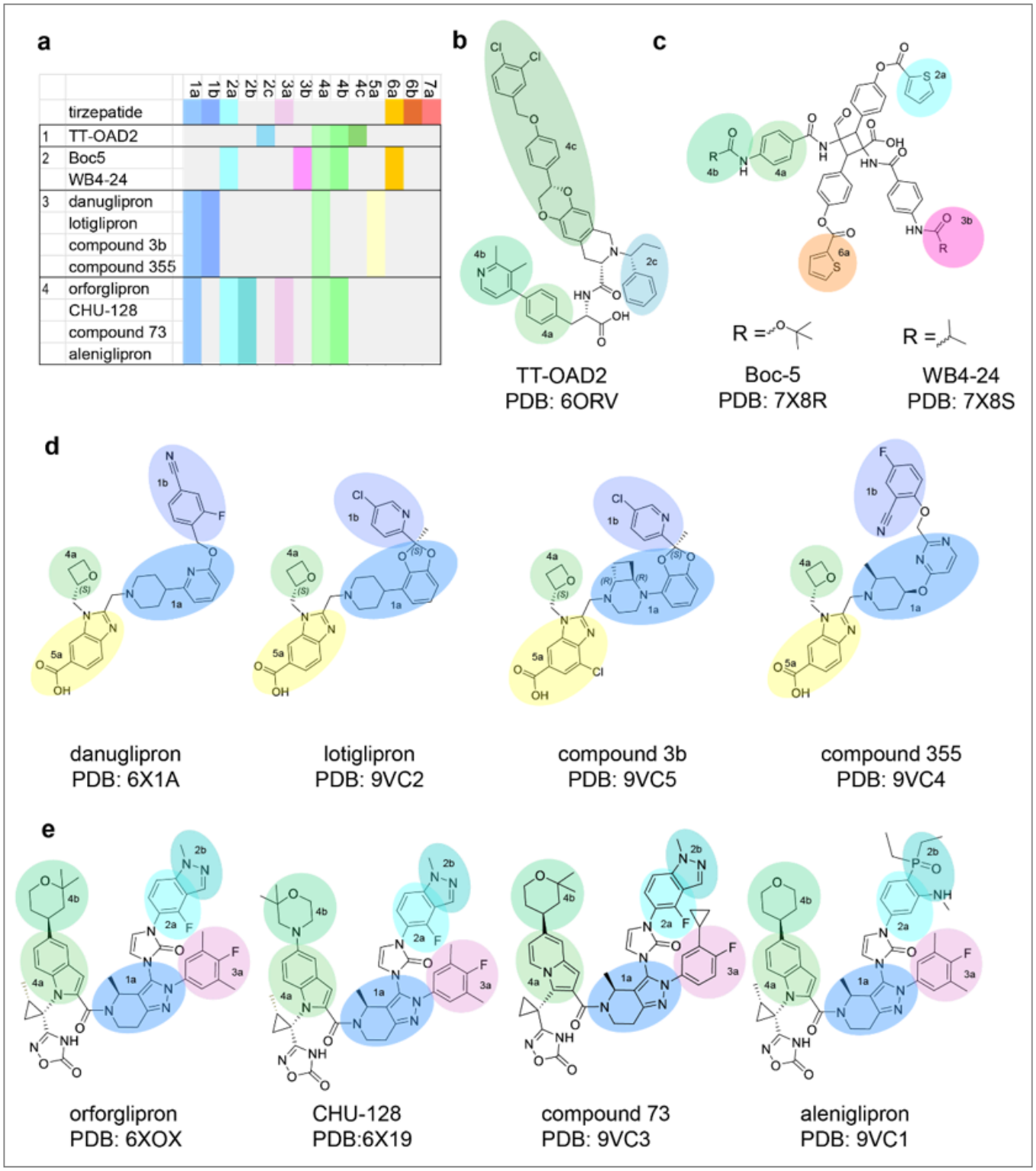

Figure 3

GLP-1R orthosteric small molecule agonists versus peptide binding pocket nomenclature. (a) Interactions of peptide and small molecule agonists with different binding pockets. (b-e) Schematic 2D maps of the pockets defined in panel a to the chemical structures of small molecule agonistsTT-OAD2 (PDB: 6ORV) (b); Boc5 (PDB: 7X8R) and WB4-24 (PDB:7X8S) (c); danuglipron (PDB: 6X1A), lotiglipron

(PDB: 9VC2), compound 3b (PDB: 9VC5), and compound 355 (PDB: 9VC4) (d); and orforglipron (PDB: 6XOX), CHU-128 (PDB: 6X19), compound 73 (PDB: 9VC3), and aleniglipron (PDB: 9VC1) (e).

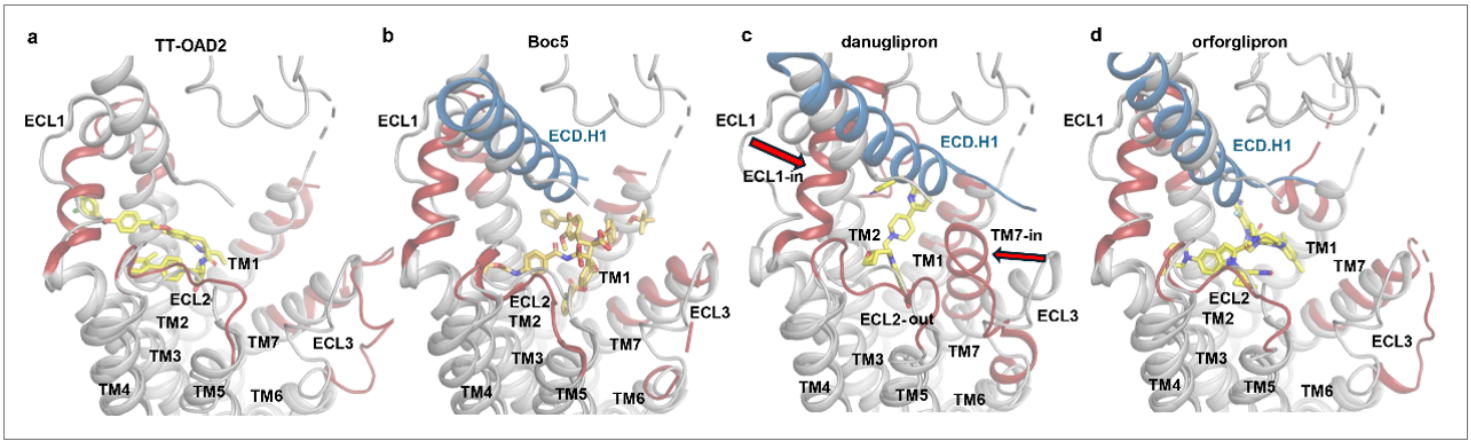

Figure 4

Conformational diversity induced by GLP-1R small molecules. (a-d) Conformational changes induced by small molecule agonists TT-OAD2, Boc5, orforglipron, and danuglipron compared to endogenous peptide GLP-1. The structures are displayed as grey cartoon. ECD.H1 and ECL1-ECL3 in small molecule agonists-bound GLP-1R structures are colored in blue and red, respectively. Small molecule agonists TT-OAD2, Boc5, danuglipron, and orforglipron are shown as yellow sticks.

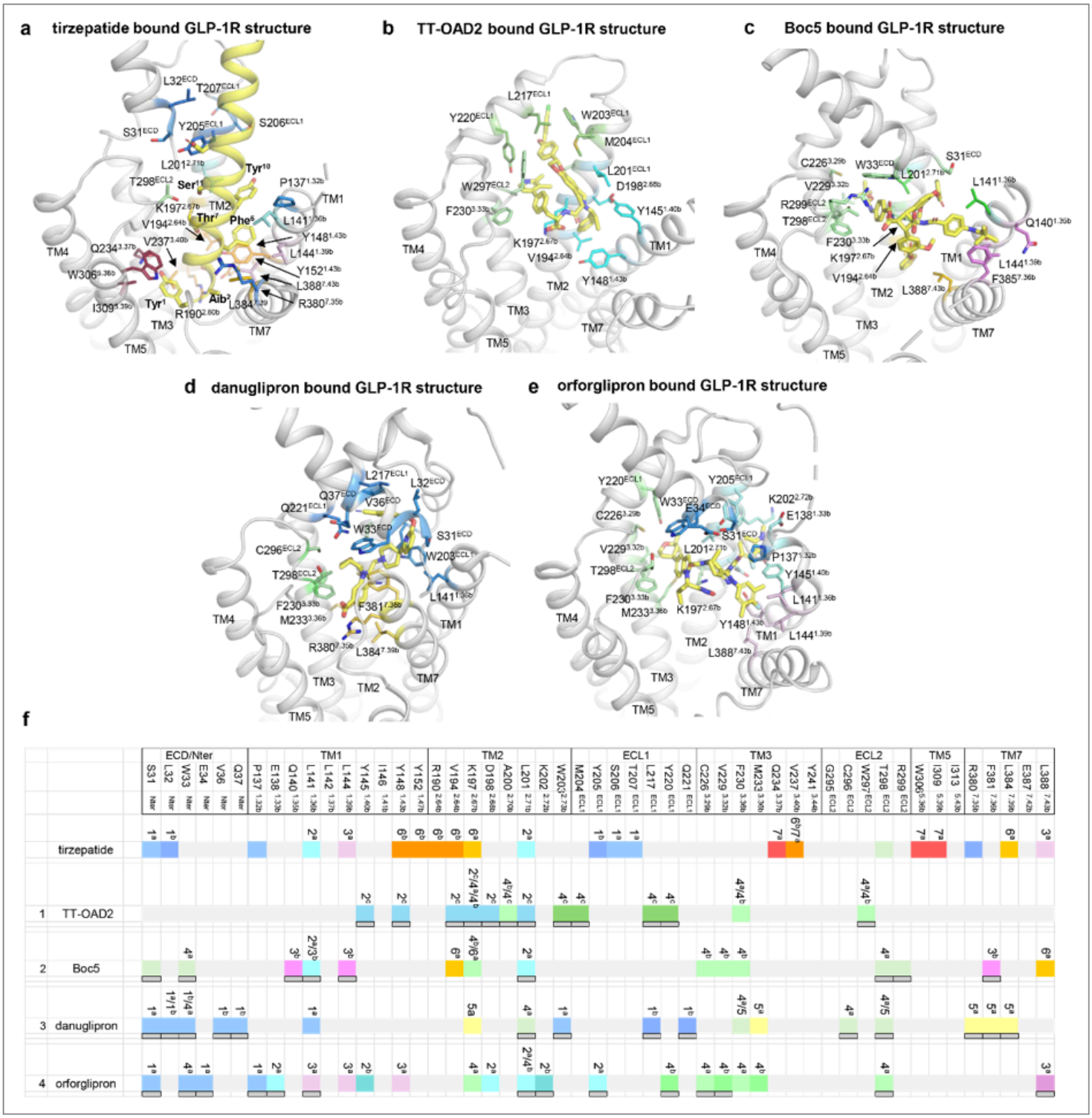

Figure 5

GLP-1R small molecule agonists binding pockets and interaction fingerprints (IFPs). (a-e) GLP-1R small molecule agonists binding pockets. The structures are shown as grey cartoons and ligands as yellow sticks. Residues involved in the interactions are shown as sticks and labeled with generic GPCR residue numbering in superscript[7] . (f) IFPs of the GLP-1R binding site residues with different small molecule agonists shown in panels a-e.

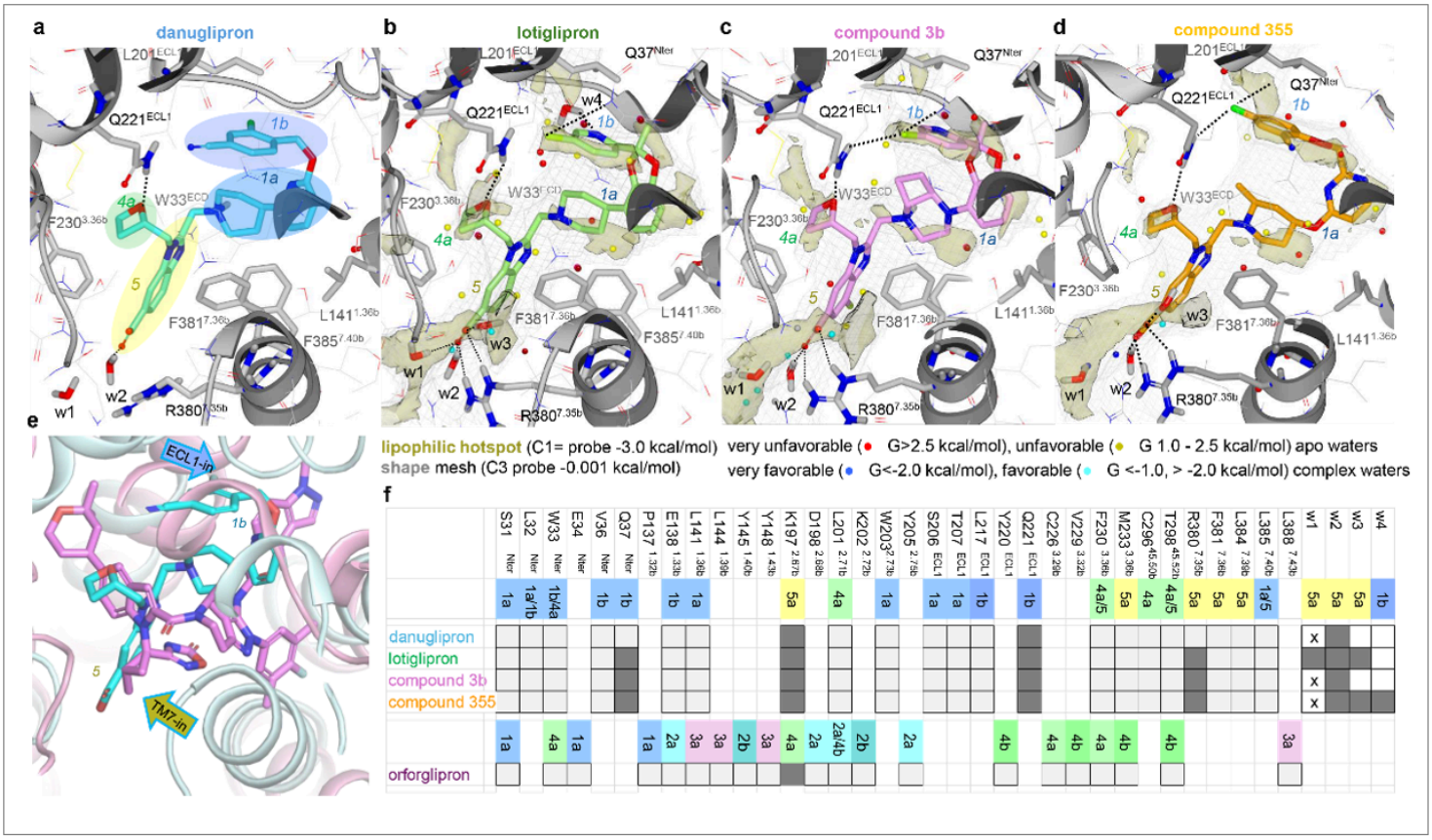

Figure 6

Comparative druggability analysis of GLP-1R small molecule agonists of danuglipron chemotype. (a-d) GLP-1R small molecule agonists binding modes with schematic depiction of binding pockets as defined in Figure 3. Key residues for receptor-ligand interactions described in the text are shown as sticks and H-bond interactions are depicted as dotted black lines. Binding site shape is shown as grey mesh and lipophilic hotspots are colored transparent solid yellow. Energetically very unfavorable (△G>2.5 kcal/mol, red) (red), unfavorable (yellow) (△G ³1.0, £2.5 kcal/mol, yellow) apo water, and very favorable (blue) (△G<-2.0 kcal/mol, blue) complex water molecule locations, as defined by WaterFLAP are shown as colored dots. (e) Comparison of orforglipron (magenta) and danuglipron (cyan) binding modes and GLP-1R conformations, with key movements described in the text indicated with colored arrows. (f) IFPs of GLP-1R binding site residues with different small molecule agonists shown in panels A-D. Dark grey boxes indicate residues that are involved in H-bond interactions while light grey boxes indicate non-polar interactions. The presence of water molecule position w1, stabilized by agonist binding, is indicated with an ‘x’.

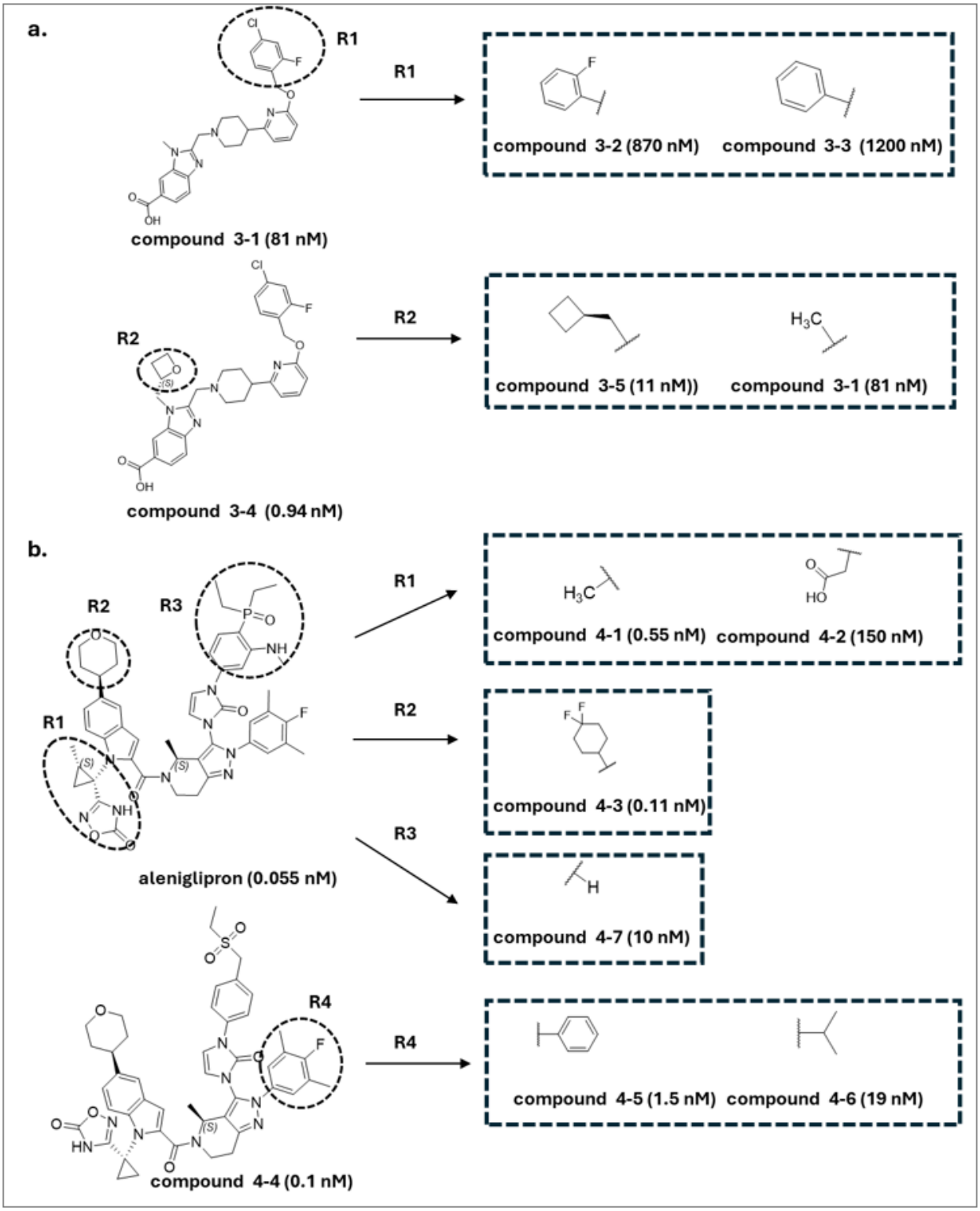

Figure 7

SAR of danuglipron and aleniglipron representative analogues. (a) Chemical structure and assay data of danuglipron analogues were obtained from the published patent WO2018109607A1[22]. (b) Chemical structure and assay data of aleniglipron and its analogues were obtained from the published patent WO2021155841A1[29].

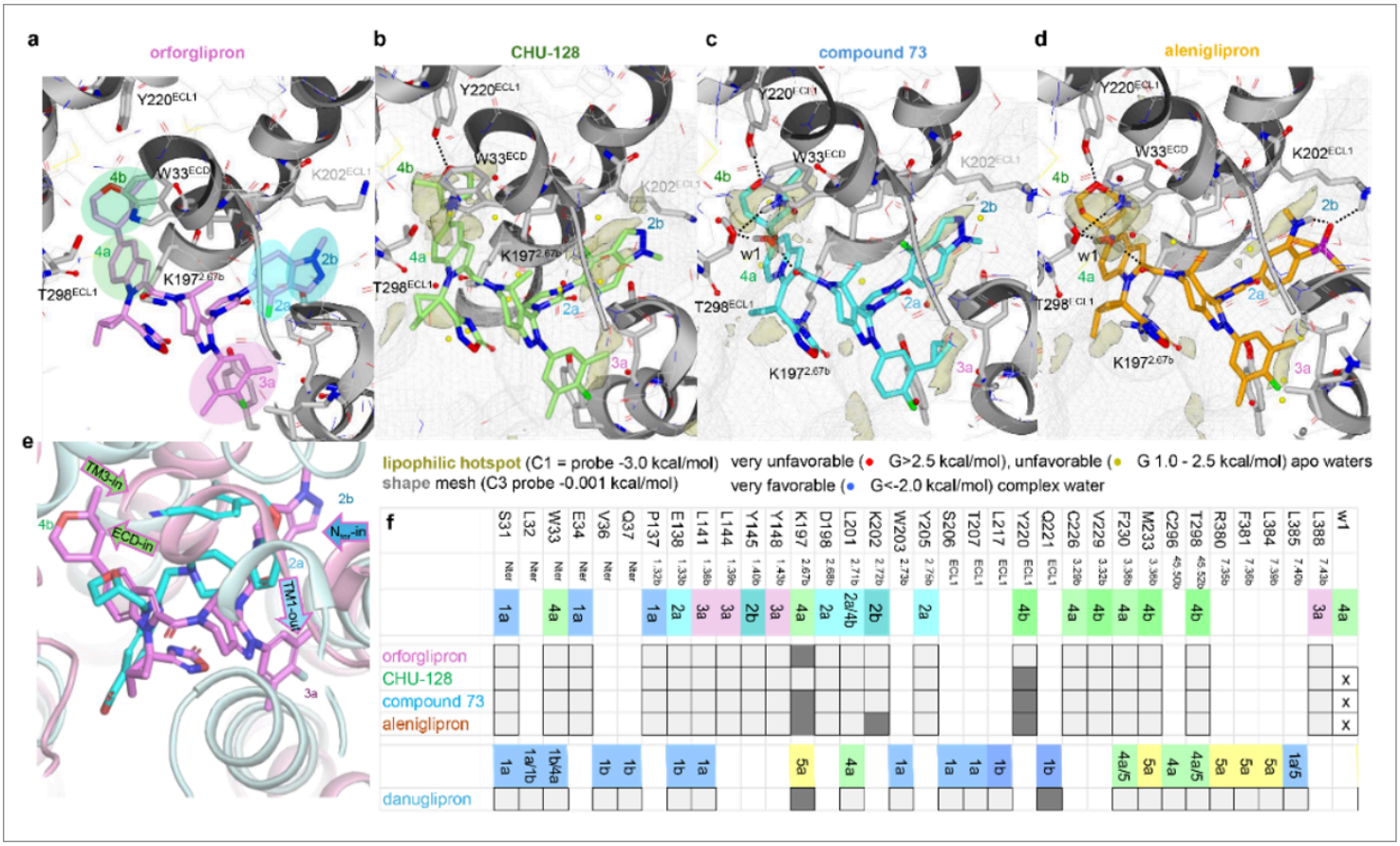

Figure 8

Comparative druggability analysis of GLP-1R small molecule agonists of orforglipron chemotype. (a-d) GLP-1R small molecule agonists binding modes. Figure legend as in Figure 6. Favorable (△G £-1.0, ³-2.0 kcal/mol) complex water molecules as defined by WaterFLAP are colored cyan. (e) Comparison of orforglipron (magenta) and danuglipron (cyan) binding modes and GLP-1R conformations, with key movements described in the text indicated by colored arrows. (f) IFPs of GLP-1R binding site residues with different small molecule agonists shown in panels A-D. Figure legend as in Figure 6.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GLP1RSMASupplementaryfigures.pdf