



# Structural basis of peptidomimetic agonism revealed by small-molecule GLP-1R agonists Boc5 and WB4-24

Zhaotong Cong<sup>a,1</sup>, Qingtong Zhou<sup>a,1</sup>, Yang Li<sup>a</sup>, Li-Nan Chen<sup>b,c</sup>, Zi-Chen Zhang<sup>d</sup>, Anyi Liang<sup>e</sup>, Qing Liuf<sup>b</sup>, Xiaoyan Wu<sup>f</sup>, Antao Dai<sup>f</sup>, Tian Xia<sup>e</sup>, Wei Wu<sup>d</sup>, Yan Zhang<sup>b,c</sup>, Dehua Yang<sup>f,g,h,i,2</sup>, and Ming-Wei Wang<sup>a,f,g,h,i,2</sup>

Edited by Robert Lefkowitz, HHMI, Durham, NC; received January 5, 2022; accepted March 28, 2022

Glucagon-like peptide-1 receptor (GLP-1R) agonists are effective in treating type 2 diabetes and obesity with proven cardiovascular benefits. However, most of these agonists are peptides and require subcutaneous injection except for orally available semaglutide. Boc5 was identified as the first orthosteric nonpeptidic agonist of GLP-1R that mimics a broad spectrum of bioactivities of GLP-1 *in vitro* and *in vivo*. Here, we report the cryoelectron microscopy structures of Boc5 and its analog WB4-24 in complex with the human GLP-1R and G<sub>s</sub> protein. Bound to the extracellular domain, extracellular loop 2, and transmembrane (TM) helices 1, 2, 3, and 7, one arm of both compounds was inserted deeply into the bottom of the orthosteric binding pocket that is usually accessible by peptidic agonists, thereby partially overlapping with the residues A8 to D15 in GLP-1. The other three arms, meanwhile, extended to the TM1-TM7, TM1-TM2, and TM2-TM3 clefts, showing an interaction feature substantially similar to the previously known small-molecule agonist LY3502970. Such a unique binding mode creates a distinct conformation that confers both peptidomimetic agonism and biased signaling induced by nonpeptidic modulators at GLP-1R. Further, the conformational difference between Boc5 and WB4-24, two closed related compounds, provides a structural framework for fine-tuning of pharmacological efficacy in the development of future small-molecule therapeutics targeting GLP-1R.

glucagon-like peptide-1 receptor | type 2 diabetes mellitus | cryoelectron microscopy | peptidomimetic agonism

It is estimated that diabetes affects the health of more than 10% adults worldwide, with ~90% of those suffering from type 2 diabetes mellitus (T2DM) (1). Glucagon-like peptide-1 receptor (GLP-1R) agonists are effective in treating T2DM and obesity, with salient benefits for cardiovascular diseases (2–4). Despite the therapeutic success, this class of medications is suboptimal due to the requirement of subcutaneous injection. With the established therapeutic effects that derive mainly from the potentiation of glucose-dependent insulin secretion and the suppression of appetite (5), considerable efforts have been made in developing orally available GLP-1R agonists. Semaglutide (Rybelsus) is the first and only noninjectable formulation in the clinic despite its low bioavailability and frequent gastrointestinal complaints, such as nausea and vomiting (6, 7). To circumvent these problems, the search for nonpeptidic agonists suitable for oral administration has never stopped.

A number of small-molecule GLP-1R agonists in different chemotypes have been documented, including quinoxalines, sulfonylthiophenes, pyrimidines, phenylalanine derivatives, substituted cyclobutanes, and azoanthracene and oxadiazanthracene derivatives (8, 9). A few compounds presented in the patent literature have progressed to clinical trials (10). These compounds were reported to mimic GLP-1 to induce insulin release, body weight decrease, and cardiovascular disease risk reduction, with variable pharmacokinetic profiles (9). Recently, insulinotropic activity upon oral intake of LY3502970 or PF-06882961 was demonstrated in cynomolgus monkeys (11). A similar result was obtained in mice bearing humanized GLP-1R as well after oral dosing of TTP273 (12).

Boc5, a substituted cyclobutane, was identified as the first orthosteric nonpeptidic agonist at GLP-1R (13). Our previous studies show that it is of peptidomimetic nature with dual efficacies in mouse models of T2DM and obesity (14). Pharmacologically, Boc5 is capable of eliciting cyclic adenosine monophosphate (cAMP) accumulation to the same level achieved by GLP-1 and stimulating insulin secretion both *in vitro* and *in vivo* (13). Chronic oral and intraperitoneal administration of Boc5 in *db/db* mice reduced hemoglobin A1c (HbA1c), blood glucose, and weight gain (13–15). WB4-24, a closely related analog of Boc5, exhibited more potent *in vitro* and *in vivo* bioactivities (15, 16). While Boc5 and WB4-24 are not “druggable” due to their poor solubility

## Significance

Glucagon-like peptide-1 receptor (GLP-1R) agonists are efficacious in the treatment of type 2 diabetes and obesity. While most clinically used agents require subcutaneous injection, Boc5, as the first orthosteric nonpeptidic agonist of GLP-1R, suffers from poor oral bioavailability that hinders its therapeutic development. The cryoelectron microscopy structures of Boc5 and its closely related analog WB4-24 presented here reveal a binding pocket located deeper in the transmembrane domain for nonpeptidic GLP-1R agonists. Molecular interaction with this site may facilitate a broad spectrum of *in vivo* agonistic activities, in addition to that with the upper helical bundles presumably responsible for biased signaling. These findings deepen our understanding of peptidomimetic agonism at GLP-1R and may help design better drug leads against this important target.

Author contributions: M.-W.W. designed research; Z.C., Q.Z., Y.L., Z.-C.Z., Q.L., X.W., and A.D. performed research; T.X., W.W., and Y.Z. contributed new reagents/analytic tools; Z.C., Q.Z., Y.L., L.-N.C., Z.-C.Z., A.L., Q.L., T.X., W.W., Y.Z., and D.Y. analyzed data; and Z.C., Q.Z., D.Y., and M.-W.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>Z.C. and Q.Z. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: dhyang@simm.ac.cn or mwwang@simm.ac.cn.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2200155119/-/DCSupplemental>.

Published May 13, 2022.

and low oral bioavailability (15), they may represent a unique pharmacophore for optimization. Nonetheless, they could certainly serve as a useful tool to probe the molecular mechanism by which nonpeptidic ligands recognize and activate GLP-1R. Here we report the cryoelectron microscopy (cryo-EM) structures of the GLP-1R–G<sub>s</sub> protein complex bound to Boc5 or WB4-24 to provide valuable insights into the structural basis of their binding modes and signaling properties.

## Results

**Structure Determination.** To increase the binding of Boc5 and WB4-24 to the GLP-1R–G<sub>s</sub> complex during purification, the solubility of the compounds was improved by approximately 1,000 times using a surfactant (poloxamer 188 [P188]) that had no significant impact on receptor activation (*SI Appendix, Fig. S1A*). The complexes were further stabilized by Nb35 and Nano-BiT tethering strategy (17, 18), purified using size-exclusion chromatography, and verified by SDS gel (*SI Appendix, Fig. S1 E and F*). The cryo-EM was used to determine the structures of the complexes bound to Boc5 and WB4-24 at a global nominal resolution of 2.61 Å and 3.09 Å, respectively (Fig. 1 and *SI Appendix, Figs. S2 and S3 and Table S1*), with well-defined density for each transmembrane (TM) helix, the compounds, and a majority of the loop regions (*SI Appendix, Fig. S4*). The cryo-EM maps allowed us to build an unambiguous model for most regions of the complexes except for the flexible  $\alpha$ -helical domain of G $\alpha_s$  and the stalk between TM1 and the extracellular domain (ECD), which were poorly resolved in most cryo-EM G protein-coupled receptor (GPCR)–G<sub>s</sub> complex structures. Extracellular loop (ECL) 3 between V370 and T377 and intracellular loop (ICL) 3 between N338 and D344 were also poorly resolved and not modeled for both structures.

**Receptor Activation.** In line with our previous studies (16), Boc5 and WB4-24 displayed a full agonism as GLP-1 at eliciting G<sub>s</sub>-mediated cAMP responses with half maximal effective concentration values of  $45 \pm 1.12$  nM and  $23 \pm 1.13$  nM, respectively (*SI Appendix, Fig. S1 C and Table S2*). No detectable  $\beta$ -arrestin recruitment was observed for both Boc5 and WB4-24 (*SI Appendix, Fig. S1 D and Table S2*), implying that their glycemic and weight loss effects might be achieved by avoiding  $\beta$ -arrestin-mediated receptor desensitization (19).

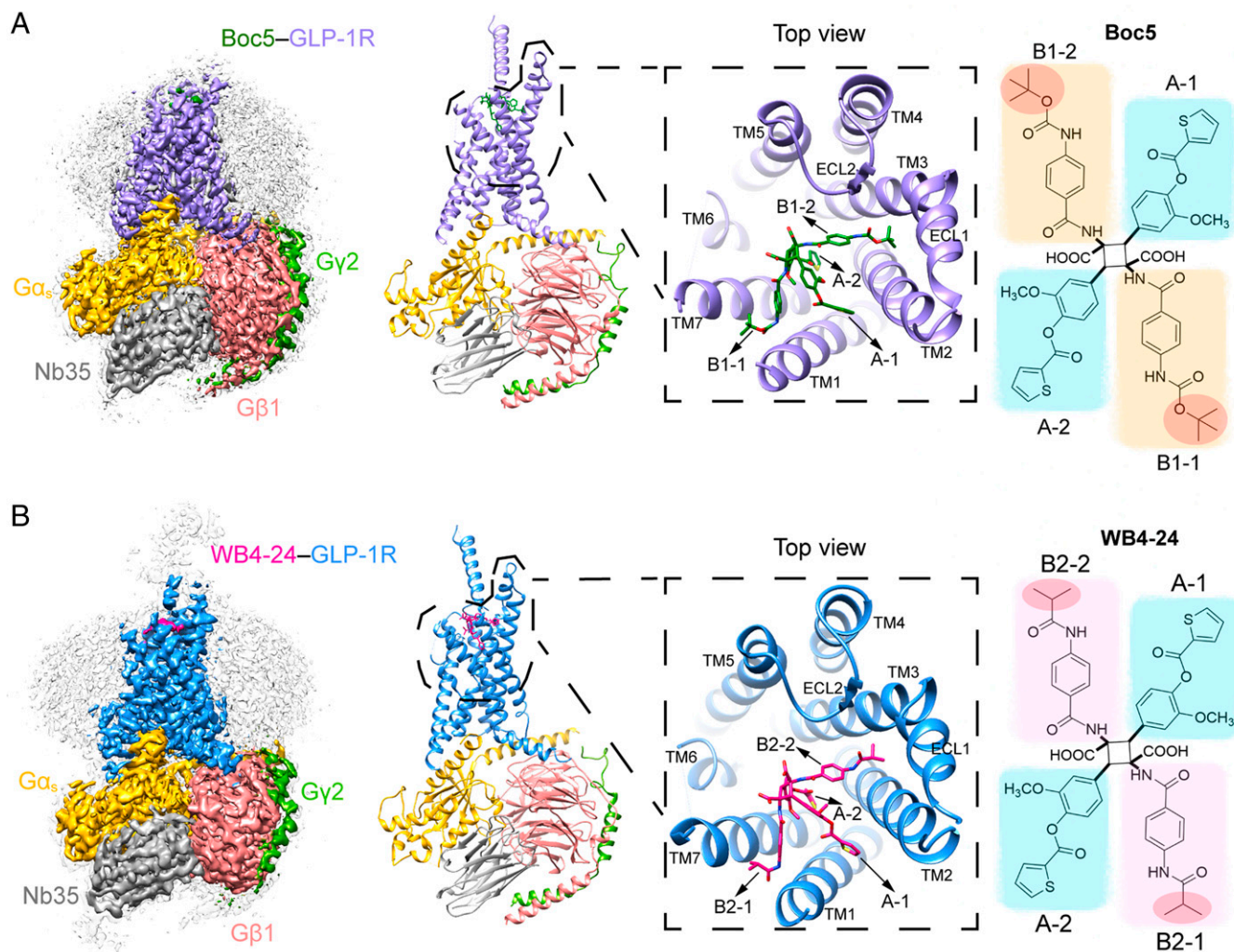
Compared to the inactive GLP-1R structure (20), WB4-24 binding and G protein coupling helped the receptor undergo significant conformational changes, including the ECD movement toward ECL2, the inward movement of the extracellular tip of TM1, the outward movement of the extracellular tip of TM2 and the extracellular half of TM7, and the subsequent outward movement of the intracellular part of TM6 (Fig. 2A). While similar conformation in the intracellular side was observed in all agonist-bound GLP-1R structures, the extracellular half that agonists directly interact with displayed ligand-dependent structural features. Compared to the peptidic agonists including GLP-1, the N-terminal  $\alpha$ -helix of ECD folded down toward the transmembrane domain (TMD) to stabilize the binding of both WB4-24 and Boc5, whereas the ECL1 reorganized its conformation to accommodate the insertion of the ligands into the TM2–TM3 cleft (Fig. 2A). Consistently, truncation of the N-terminal  $\alpha$ -helix or the entire ECD of GLP-1R completely abolished specific binding and cAMP signaling of Boc5 and WB4-24 (*SI Appendix, Fig. S5B*).

WB4-24 anchored in the orthosteric pocket by extensively interacting with residues in TMs 1 to 3, 7, ECL2, and the

N-terminal  $\alpha$ -helix through its four arms (labeled in Fig. 1), while the residues in TMs 4 to 6, ECL1, and ECL3 provided negligible direct contacts (Fig. 2B). Specifically, the 3-methoxy-4-((thiophene-2-carbonyl)oxy)phenyl group of the compounds (arm A-2, blue shading in Fig. 1) deeply penetrated into the bottom of the orthosteric pocket and partially overlapped with the positions occupied by peptide residues A8<sup>GLP-1</sup> to D15<sup>GLP-1</sup> but approached TM1 in a manner closer than that of GLP-1 (Fig. 2C). At the base of the binding pocket of GLP-1-bound GLP-1R, A8<sup>GLP-1</sup>, E9<sup>GLP-1</sup>, and F12<sup>GLP-1</sup> engaged in an extensive network with TM1 and TM7, including hydrogen bonds with Y152<sup>1.47b</sup> [class B1 GPCR numbering in superscript (21)] and R190<sup>2.60b</sup> and hydrophobic interactions with L141<sup>1.36b</sup>, L144<sup>1.39b</sup>, and Y148<sup>1.43b</sup> in TM1 and E387<sup>7.42b</sup> and L388<sup>7.43b</sup> in TM7 (22). A majority of these GLP-1 interacting residues at the binding pocket base also made contact with the thiophene moiety of WB4-24, including L141<sup>1.36b</sup>, L144<sup>1.39b</sup>, Y148<sup>1.43b</sup>, Y152<sup>1.47b</sup>, L384<sup>7.39b</sup>, and L388<sup>7.43b</sup>, despite their distinct chemotypes and binding poses between WB4-24 and GLP-1. Other notable interactions with arm A-2 were the hydrophobic contacts with V194<sup>2.64b</sup> in TM2, which receives the support of our mutagenesis study where V194A dramatically reduced the efficacy of cAMP responses induced by either Boc5 or WB4-24, with a stronger influence seen with WB4-24-treated cells (Fig. 2B and *SI Appendix, Fig. S6 and Table S3*). The N-terminal residues H7<sup>GLP-1</sup> and A8<sup>GLP-1</sup> that mainly interact with TM3 and TM5 are crucial for receptor activation, as the truncation of these two residues, i.e., GLP-1[9-36]NH<sub>2</sub>, lowers both the affinity and the efficacy of GLP-1 (23). Limited by the molecular size, both WB4-24 and Boc5 failed to directly interact with the residues in TM3 and TM5 like GLP-1, but they converged to induce a similar conformation of the central polar network at the bottom of the orthosteric pocket. These features may contribute to the peptidomimetic properties of the two ligands.

The other 3-methoxy-4-((thiophene-2-carbonyl)oxy)phenyl group (arm A-1, blue shading in Fig. 1B) and the 4-((*tert*-butoxycarbonyl)amino)phenyl group (arm B2-1, pink shading in Fig. 1B) of WB4-24 extended to the TM1–TM2 and TM2–TM3 clefts, respectively. In our molecular dynamics simulation experiments, these two arms steadily interacted with the two clefts (*SI Appendix, Fig. S7*) also occupied by other small-molecule agonists, such as LY3502970 (Fig. 2D) and TT-OAD2 (Fig. 2E), where a similar V-shaped orientation within this region was commonly adopted. The 2,2-dimethyl-tetrahydropyran and 4-fluoro-1-methyl-indazole arms of LY3502970 interacted with multiple residues including W33<sup>ECD</sup>, L141<sup>1.36b</sup>, K197<sup>2.67b</sup>, L201<sup>2.71b</sup>, F230<sup>3.33b</sup>, and T298<sup>ECL2</sup> (Fig. 2D) in a manner similar to WB4-24. However, the 4-((*tert*-butoxycarbonyl)amino)phenyl group (arm B2-2) of WB4-24 is bulkier than the 3,5-dimethyl-4-fluorophenyl ring of LY3502970, thereby conferring distinct interactions with the residues in TM1 and TM7 including Q140<sup>1.35b</sup>, F381<sup>7.36b</sup>, and F385<sup>7.40b</sup> (Fig. 2B). Together, these specific conformations of the four arms of WB4-24 establish a unique binding mode for this class of molecules.

**Improved Bioactivity.** A series of Boc5 analogs were designed to improve bioactivities by varying the length and steric hindrance of R1 groups (red shading in Fig. 1). When the *tert*-butoxycarbonyl of Boc5 (red shading in Fig. 1A) was changed to smaller moieties, the potency of ligand binding and receptor activation were improved, with the isobutyryl (red shading in Fig. 1B) as the optimum (16). Quantitative analysis of both Boc5- and WB4-24-bound GLP-1R structures may provide



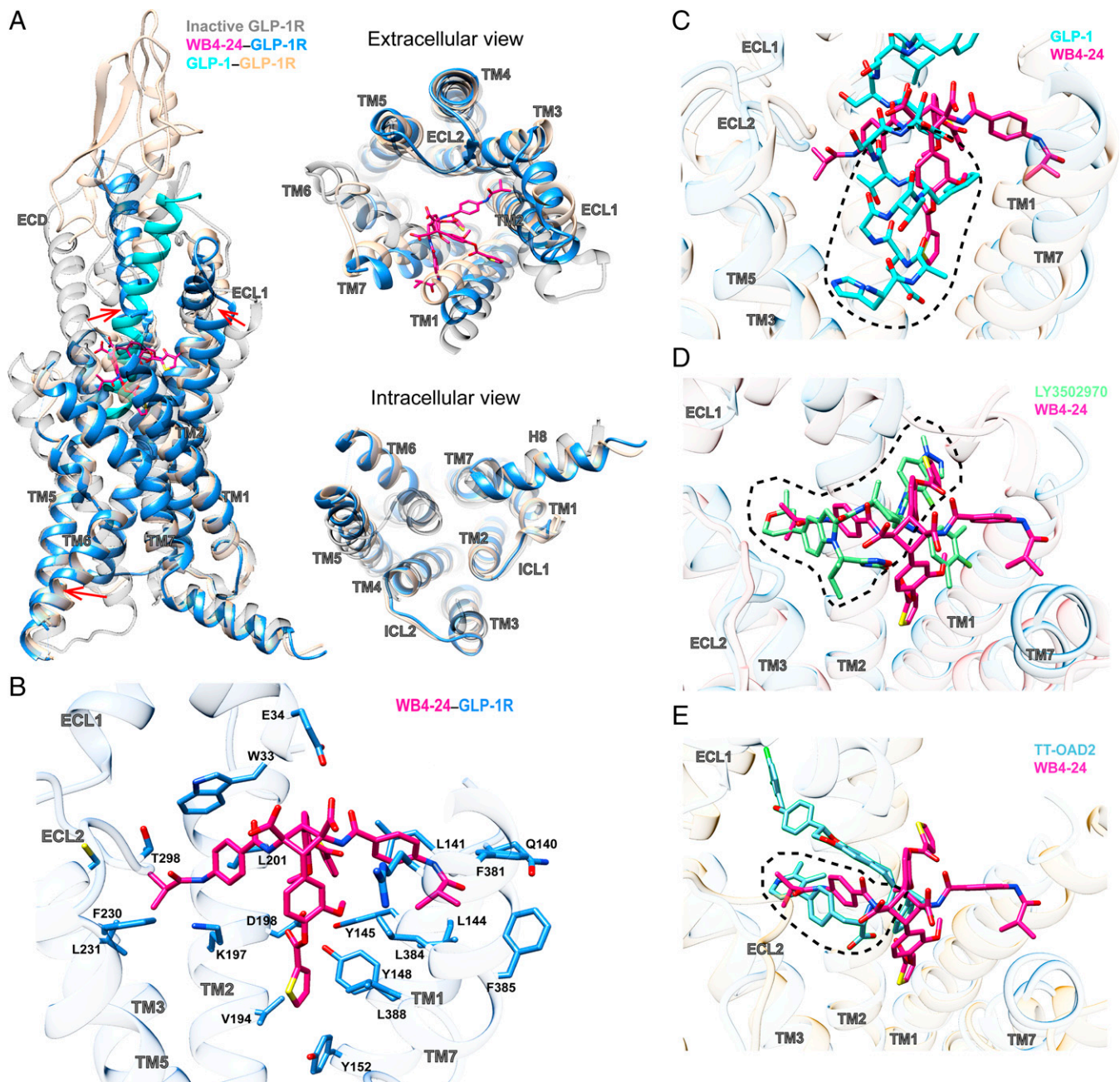
**Fig. 1.** Overall cryo-EM structures of the Boc5- and WB4-24-bound GLP-1R- $G_s$  complexes. (A) The cryo-EM map with a disk-shaped micelle (Left) and cartoon representation (Middle) of the Boc5-bound complex. The top view shows the binding pose of Boc5. (Right) chemical structure of Boc5 (molecular weight: 1,018) with two 3-methoxy-4-((thiophene-2-carbonyl)oxy)phenyl groups (blue shade) named A-1 and A-2, respectively, and two 4-((tert-butoxycarbonyl)amino)phenyl groups (orange shade) named B1-1 and B1-2, respectively. (B) The cryo-EM map with a disk-shaped micelle (Left) and cartoon representation (Middle) of the WB4-24-bound complex. The top view shows the binding pose of WB4-24. (Right) chemical structure of WB4-24 (molecular weight: 1,074) with two 3-methoxy-4-((thiophene-2-carbonyl)oxy)phenyl groups (blue shade) that are the same as A-1 and A-2 of Boc5, and two 4-((tert-butoxycarbonyl)amino)phenyl groups (pink shade) named B2-1 and B2-2, respectively. Moieties that are different between Boc5 and WB4-24 are highlighted in red circles, Boc5-bound GLP-1R in purple, WB4-24-bound GLP-1R in dodger blue, Ras-like domain of  $G\alpha_s$  in gold,  $G\beta$  subunit in deep pink,  $G\gamma$  subunit in green, Nb35 in gray, Boc5 in dark green, and WB4-24 in magenta.

mechanistic insights into the increased bioactivity of WB4-24 relative to Boc5.

Structural superimposition revealed that the two molecules adopt a similar binding pose (Fig. 3A), but notable differences in ligand recognition and associated TM bundle conformation were observed (Fig. 3 B and C). Due to a larger steric hindrance of R1 groups, Boc5 pushed the extracellular tips of TMs 2, 3, and 7 to move outward by 1.6 Å (measured by C $\alpha$  carbon of Y205<sup>2,75b</sup>), 1.2 Å (measured by C $\alpha$  carbon of S225<sup>3,28b</sup>), and 2.0 Å (measured by C $\alpha$  carbon of R380<sup>7,35b</sup>), respectively, compared with WB4-24 (Fig. 3 B and C). Given the fundamental role of TM6-ECL3-TM7 in GLP-1R-mediated signaling (12, 24), these subtle differences between Boc5 and WB4-24 in TM7 conformation may alter their potencies. Notably, R380<sup>7,35b</sup> in TM7 formed a salt bridge with the carboxylic group of WB4-24, in a manner similar to D15<sup>GLP-1</sup> in GLP-1 (Fig. 3D). Previous structure-activity relationship (SAR) studies demonstrated that the two carboxylic groups of Boc5 are essential for its bioactivity, whose modification by either amide or ester led to a total loss of activity (16). The

importance of R380<sup>7,35b</sup> was confirmed by alanine mutagenesis that almost abolished the activity of WB4-24 but had a weaker impact on Boc5-induced cAMP production (Fig. 3E and *SI Appendix*, Table S3). Besides the conformational changes at the TM, sidechain orientation of several residues including L144<sup>1,38b</sup> and K197<sup>2,67b</sup> were distinct between Boc5 and WB4-24. Their alanine mutation had a slightly reduced effect on Boc5 than WB4-24-elicited cAMP signaling (Fig. 3E and *SI Appendix*, Table S3). Collectively, these results provide a quantitative measure of pharmacology between two closely related molecules at a near-atomic level.

**Peptidomimetic Site.** Structural comparison of the GLP-1R structures bound by various small-molecule agonists and GLP-1 suggests that Boc5 and WB4-24 mimic the peptidic binding to an extent previously unseen. Fig. 4A shows that both compounds penetrated into the GLP-1R TMD pocket deeper than all other nonpeptidic ligands reported previously (by 2~4 Å as measured vertically). Such a binding mode enabled them to directly interact with the bottom of the binding pocket only

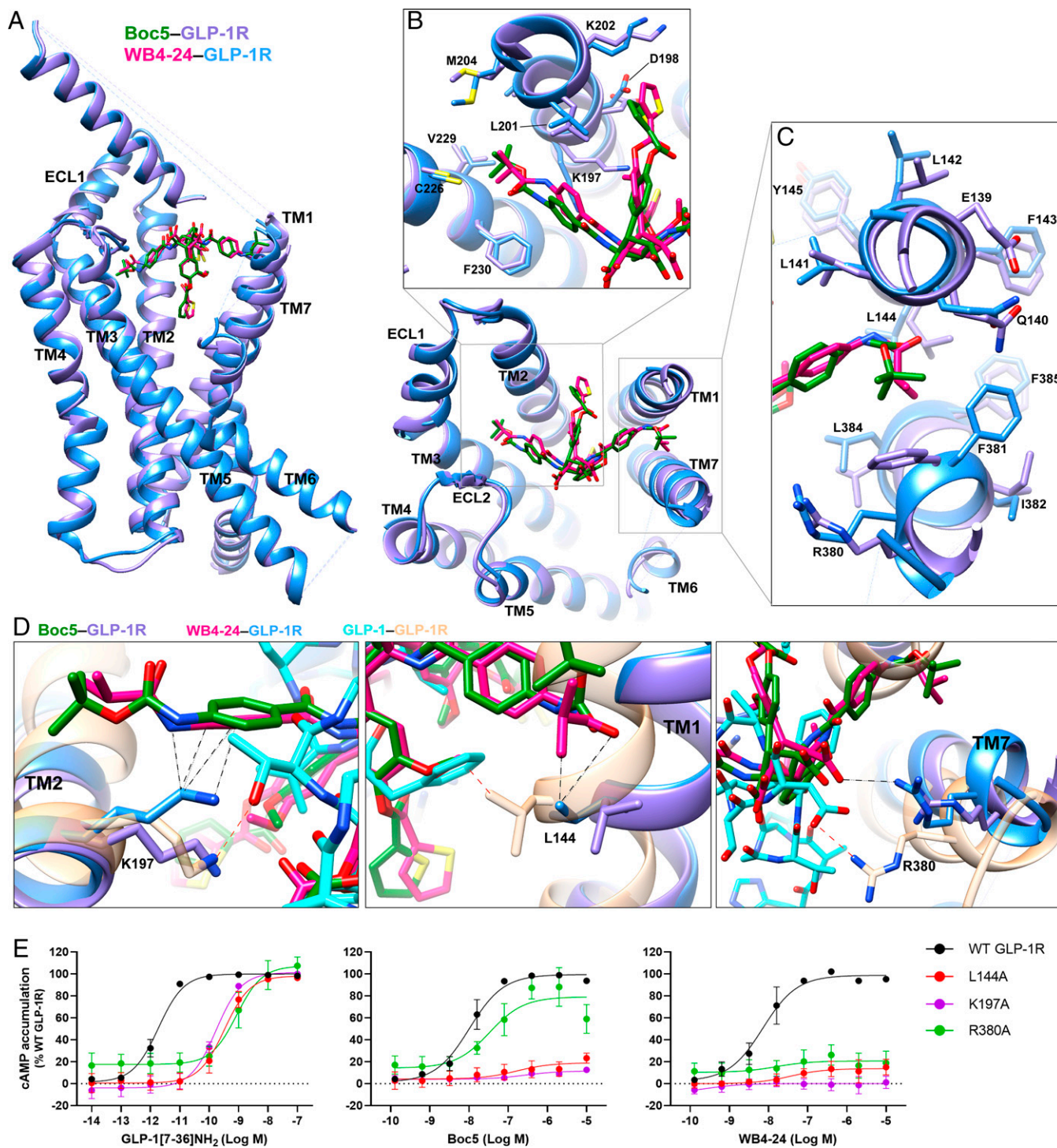


**Fig. 2.** Structural analysis of the WB4-24-bound GLP-1R. (A) Structural comparison of the WB4-24-bound GLP-1R with inactive and GLP-1-bound GLP-1R complexes (PDB code: 6X18 and 6LN2). Arrows show the transition from the inactive conformation to the WB4-24-induced active state. (B) Interactions of WB4-24 within the TM binding cavity. WB4-24 and its interaction residues are shown by sticks. (C) Overlay of the WB4-24- and GLP-1-bound receptors reveals that WB4-24 has a partial overlap with the N terminus of GLP-1 in the TM binding cavity. (D) Superimposition of the WB4-24- and LY3502970-bound receptors reveals a similar binding pose and a large degree of overlap in the GLP-1R TM cavity. (E) Superimposition of the WB4-24- and TT-OAD2-bound receptors reveals a partial overlap in the TM2-TM3 cleft.

accessible by peptidic agonists (Fig. 4 B–D). The interface area between Boc5 and the pocket was consistently  $360 \text{ \AA}^2$ , almost half of GLP-1 ( $737 \text{ \AA}^2$ ), significantly larger than other small molecules ( $\sim 250 \text{ \AA}^2$  for LY3502970, CHU-128, and TT-OAD2;  $90 \text{ \AA}^2$  for PF-06882961; and  $40 \text{ \AA}^2$  for RGT1383). The latter two compounds additionally occupied the cleft between ECL1 and the N-terminal  $\alpha$ -helix of ECD, thereby stabilizing the complex by forming multiple strong contacts with ECL1 (Fig. 4B). As a comparison, ECL1 contributed negligible contacts in the cases of Boc5 and WB4-24. Besides this pocket, the TM1-TM7 cleft was also occupied by Boc5 or WB4-24 in a manner more efficient than that of all the other small molecules, supported by the agonist-

TM1/7 interface area calculated for Boc5 ( $680 \text{ \AA}^2$ ) and WB4-24 ( $749 \text{ \AA}^2$ ), respectively, close to that of GLP-1 ( $890 \text{ \AA}^2$ ) and remarkably larger than other nonpeptidic ligands, especially TT-OAD2 (Fig. 4D). These two sites bound by Boc5 and WB4-24 offer valuable clues for the design of peptidomimetic/small-molecule GLP-1R agonists through optimization of the existing ligands and the preparation of agonists capable of binding to both peptidomimetic sites or penetrating into the deep orthosteric pocket and TM1-TM7 cleft. Structure-based virtual screening to identify scaffolds is another option.

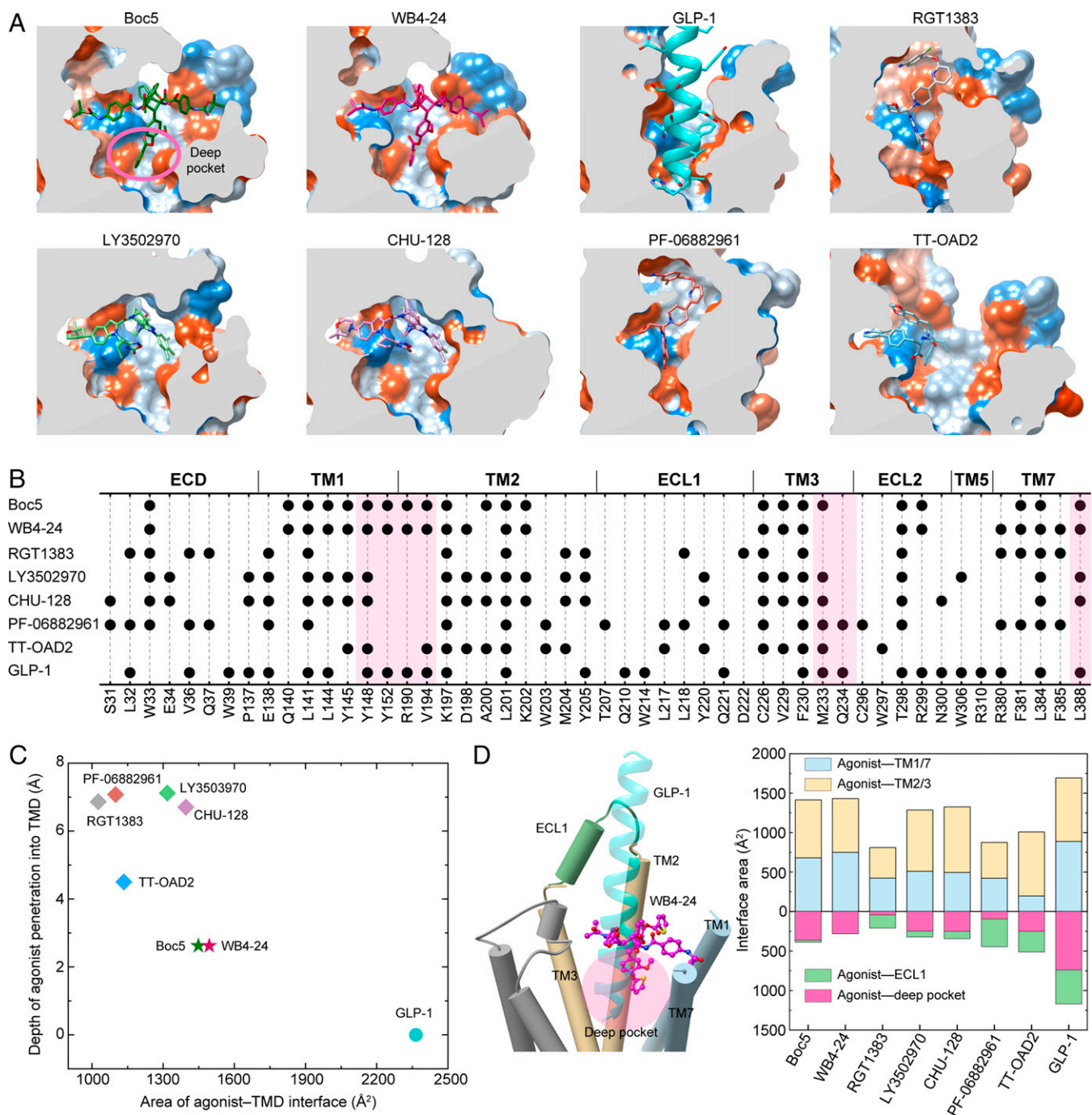
Examination of various molecular recognition patterns of GLP-1R by nonpeptidic agonists reveals a common region (the



**Fig. 3.** Comparison of GLP-1R conformation and ligand binding pocket stabilized by Boc5 and WB4-24. (A) Overlay of Boc5- and WB4-24-bound GLP-1R structures. (B and C) Close-up of the TM bundles viewed from the extracellular side reveals conformational differences in TM2/TM3 (B) and TM1/TM7 (C) backbones when stabilized by different arms of Boc5 and WB4-24. (D) WB4-24 makes more extensive contacts with TM bundles than Boc5 by interacting with residues L144, K197, and R380, which contribute to the agonism of GLP-1. Coloring for A–D denotes the segments as highlighted (GLP-1 in cyan, GLP-1-bound GLP-1R in sandy brown, Boc5 in dark green, Boc5-bound GLP-1R in purple, WB4-24 in magenta, WB4-24-bound GLP-1R in dodger blue). The interactions of the residues with WB4-24 and GLP-1 are indicated by black and red dashed lines, respectively. (E) Mutagenesis analysis of residues L144, K197, and R380 for GLP-1, Boc5, and WB4-24 show the requirement of these contacts for receptor activation. Data are shown as means  $\pm$  SEM from at least three independent experiments.

extracellular tips of TM2 and TM3) that constitutes an anchoring platform for agonist binding. Specifically, the sidechains of K197<sup>2,67b</sup> and F230<sup>3,33b</sup> pointed to the TMD core and uplifted a small molecule, allowing the latter to be further covered by the folded-down N-terminal  $\alpha$ -helix of ECD, particularly W33<sup>ECD</sup> (Fig. 4B). The TM1-TM2 cleft to which the A-1 branch of Boc5

or WB4-24 inserted was also touched by CHU-128, LY3502970, and TT-OAD2 but not RGT1383 or PF-06882961. It was reported that lipidated K10 in peptide 20 and positive allosteric modulator LSN3160440 bound exactly to this cleft via the formation of more interactions with receptors (25), indicating a potential region for lead optimization. Similar phenomena were



**Fig. 4.** Peptidomimetic binding mode of Boc5 and WB4-24. (A) Structural comparison of the binding conformations of peptidic and nonpeptidic agonists of GLP-1R. The receptor is shown in surface representation and colored from dodger blue for the most hydrophilic region to white, and to orange-red for the most hydrophobic region. Small-molecule agonists and GLP-1 are shown as sticks and ribbon, respectively. (B) Schematic diagram of the interactions between peptidic and nonpeptidic agonists and GLP-1R. (C) Scatterplot of agonist binding in GLP-1R. All structures were superimposed on the GLP-1-bound GLP-1R coordinates from the OPM database (PDB: 6X18) using the TMD C $\alpha$  atoms. The x and y axis represent into the depth of the agonist that penetrated the GLP-1R TMD pocket and the buried surface area between the agonist and TMD, respectively. The former is defined as the minimum vertical distance between the heavy atoms of a certain agonist and the carboxyl oxygen atom of E9<sup>P</sup> in GLP-1 (the deepest atom of GLP-1 that inserted into the TMD), while the latter was calculated using freeSASA. (D) Interface comparison across peptidic and nonpeptidic agonists of GLP-1R. Different regions of the GLP-1R TMD pocket including TM1/7, TM2/3, ECL1 (residues S206 to D222), and the deep pocket (defined as the collection of the residues Y148<sup>1,43b</sup>, F195<sup>2,65b</sup>, M233<sup>3,36b</sup>, V281<sup>4,57b</sup>, R310<sup>5,40b</sup>, F367<sup>6,56b</sup>, and L388<sup>7,43b</sup>-W420<sup>8,61b</sup>) were subjected to the interface area calculation.

also observed in the TM2-TM3 cleft, to which all other nonpeptidic ligands (TT-OAD2 in particular) except RGT1383 and PF-06882961 penetrated using one of the arms. Despite a deficiency in making contacts with the above two regions (TM1-TM2 and TM2-TM3 clefts), RGT1383 and PF-06882961 successfully triggered an inward movement of ECL3 and the extracellular end of TM7, consequently creating a tightly packed orthosteric pocket to accommodate agonist binding (Fig. 4 A and

B). These analyses highlight the diversity in signaling initiation by nonpeptidic ligands.

## Discussion

Boc5 was identified as the first nonpeptidic GLP-1R agonist with a broad spectrum of in vitro and in vivo bioactivities as GLP-1 (13–15). To explore the SAR of this compound, more

than 50 derivatives with various modifications on the cyclobutane core and four arms (16) were synthesized and evaluated. Among them, WB4-24 stood out as the most potent analog of Boc5 (15, 16, 26). However, suffering from low oral bioavailability due to metabolism by esterase in the gut as well as the limitation of photosynthesis (15), this class of GLP-1R agonists is precluded from further development. Meanwhile, a handful of nonpeptidic ligands with distinct pharmacological properties were identified in the past decades including TT-OAD2 (12), LY3502970 (27), CHU-128 (22), PF-06882961 (22), and RGT1383 (28). While a few have entered clinical trials, none was regulatory-approved for clinical use.

The cryo-EM structures of GLP-1R in complex with Boc5 or WB4-24 presented here expanded our knowledge of ligand recognition and receptor activation from three aspects. First, the unique binding mode of Boc5 and WB4-24 whose one arm (A-2) inserted deeply into the bottom of the orthosteric pocket usually accessible only by peptidic agonists reveals a previously unknown interaction site for small-molecule agonists. Second, quantitative analysis between Boc5- and WB4-24-bound GLP-1R structures, together with the mutagenesis data, provide mechanistic insights into the improved bioactivity of WB4-24 over Boc5 at the near-atomic level. Third, our systematic review of all the reported GLP-1R structures bound by nonpeptidic ligands (Fig. 5) uncovers both common and unique features of ligand recognition distinctively related to the pharmacological property of one another. The extracellular tips of TM2 and TM3 are one common interacting region that makes extensive contacts with all bound ligands regardless of peptidic or nonpeptidic nature. Meanwhile, individual ligands may also have their specific interaction sites: Peptides including GLP-1 have rich contacts with the bottom of the orthosteric pocket (the deep pocket) and the ECL1 but not the TM1-TM2 or TM2-TM3 cleft; TT-OAD2 steadily inserts into the TM2-TM3 cleft while CHU-128 and LY3502970 additionally occupy the TM1-TM2 cleft; both PF-06882961 and RGT1383 alternatively chose the ECL1 and fold the ECL3 and the extracellular tip of TM7 inward as interaction sites. Surprisingly, Boc5 and WB4-24, having four long extended arms, simultaneously occupy the deep pocket, TM1-TM2, TM2-TM3, and TM1-TM7 clefts, allowing them to achieve peptidomimetic agonism and consequently show GLP-1-like in vivo activities such as insulin release, appetite suppression, HbA1c reduction, and weight loss (*SI Appendix, Table S5*). Unlike LY3502970 (27), CHU-128 (22), PF-06882961 (22), or RGT1383 (28), which failed to elicit agonist activity due to the mutation of W33 conserved in human and primate GLP-1Rs, W33<sup>ECL3</sup> only reduced specific binding and cAMP signaling of Boc5 and WB4-24, indicating a differentiated receptor interaction mode (*SI Appendix, Fig. S8*).

These analyses provide a molecular basis of pharmacological consequences rendered by GLP-1R agonists of different chemical natures. It is known that TM7 conformation is important for peptidomimetic properties of small-molecule agonists (27). In the PF-06882961-bound structure, TM7 moved inward dramatically, forming a similar polar interaction network observed in the GLP-1-bound structure. Thus, PF-06882961 was able to better mimic the pharmacological effects of GLP-1 than its counterparts reported in the literature (*SI Appendix, Table S5*). Such an inward movement was followed by TM6 to form an extensive polar interaction network involving R380<sup>7.35b</sup>, thereby constituting the conformation of TM6-ECL3-TM7. In addition, the contraction of TM7 induced by PF-06882961

moved F385<sup>7.40b</sup> to the TM1-TM7 cleft to stabilize overall side-chain rearrangements among key residues located in TMs 1 and 2 and mediated by residues Y148<sup>1.43b</sup>, K197<sup>2.67b</sup>, and D198<sup>2.68b</sup> (22). Intriguingly, in the Boc5- and WB4-24-bound structures, the interactions with TM1, TM2, and TM7 were mainly made by the A-1 and A-2 arms, supporting an essential role played by thiophene groups in receptor activation. Removal of them in A-2 would impair the interaction with TM7, resulting an inactive metabolite M-2 (15). Moreover, compared to full agonists, ligands that exhibit biased agonism toward G protein-dependent cAMP signaling are associated with a distinct conformation of TM6-ECL3-TM7, such as LY3502970 (27), TT-OAD2 (12), and CHU-128 (22). Although the interaction with TM7 in the WB4-24-bound GLP-1R was similar to that in GLP-1- or PF-06882961-bound structures, its contact with ECL3 was not observed, which may account for the biased agonism of Boc5 and WB4-24. Indeed, alanine mutations of D372<sup>ECL3</sup> and E373<sup>ECL3</sup> in ECL3 that eliminate the interaction with GLP-1 resulted in biased signaling toward cAMP accumulation (*SI Appendix, Fig. S9*). Overall, these results imply that stabilizing the conformation of the TM6-ECL3-TM7 region would be required for unbiased as well as peptidomimetic agonism at GLP-1R. Clearly, the newly identified binding pocket of Boc5 and WB4-24 as well as their interaction mode with the receptor provide valuable clues for developing the next generation of nonpeptidic modulators with better therapeutic profiles.

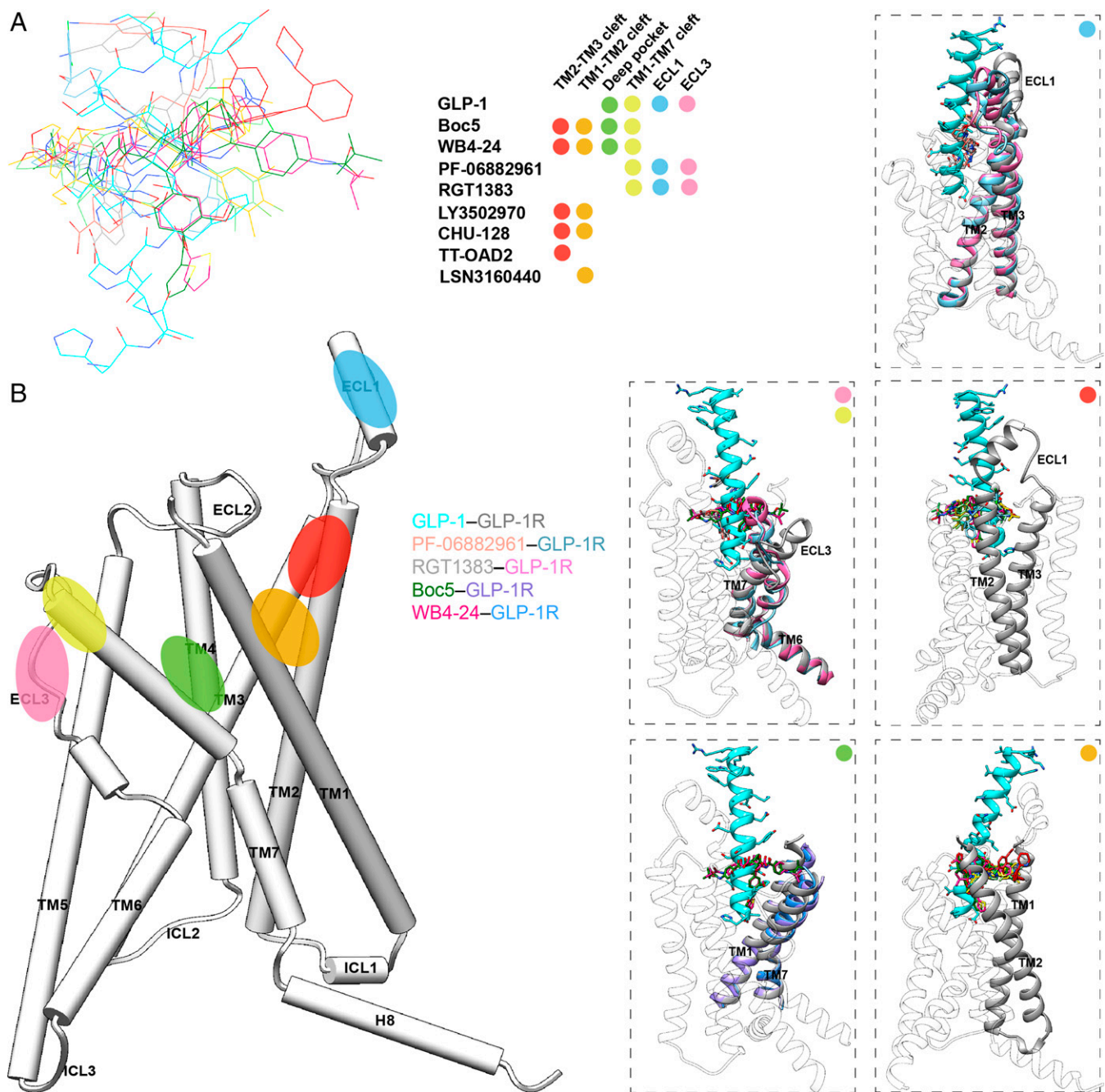
## Methods

**Cell Culture.** *Spodoptera frugiperda* (Sf9) insect cells (Expression Systems) were grown in ESF 921 serum-free medium (Expression Systems) at 27 °C and 120 rpm. HEK293T cells (American Type Culture Collection) were cultured in DMEM (Life Technologies) supplemented with 10% FBS (Gibco) and maintained in a humidified chamber with 5% CO<sub>2</sub> at 37 °C.

**Construct Preparation.** The human GLP-1R was modified with its native signal sequence (M1-P23) replaced by the hemagglutinin signal peptide to facilitate receptor expression. To obtain a GLP-1R-G<sub>s</sub> complex with good homogeneity and stability, we used the NanoBIT tethering strategy, in which the C terminus of rat Gβ1 was linked to the HiBiT subunit with a 15-amino acid polypeptide (GSSGGGSGGGGSSG) linker and the C terminus of GLP-1R was directly attached to the LgBiT subunit followed by a tobacco etch virus (TEV) protease cleavage site and a double maltose binding protein tag. An engineered G<sub>s</sub> construct (G112) was designed based on mini-G<sub>s</sub> as previously described (29, 30) and was used to purify the Boc5 or WB4-24-bound GLP-1R-G<sub>s</sub> complex. The constructs were cloned into both the pcDNA3.1 and pFastBac vectors for functional assays in mammalian cells and protein expression in insect cells, respectively. All modifications of the receptor had no effect on ligand binding and receptor activation (*SI Appendix, Fig. S1B and Table S3*). Other constructs including the wild-type and GLP-1R mutants were cloned into the pcDNA3.1 vector for cAMP accumulation and whole-cell binding assays.

**Solubility Analysis.** The excess amounts of Boc5 and WB4-24 were dissolved by shaking at 37 °C and 180 rpm for 72 h in normal saline, 5 mM HEPES, 20% PEG400, or 2% P188, respectively. After centrifugation at 13,000 rpm for 10 min, the supernatant was extracted and diluted to an appropriate concentration with methanol. The saturated solubility of the compounds was determined using high-performance liquid chromatography and estimated by fitting the data to the standard concentration curves.

**Formation and Purification of Complex.** The Bac-to-Bac Baculovirus Expression System (Invitrogen) was used to generate high-titer recombinant baculovirus for GLP-1R-LgBiT-2MBP, G112, Gβ1-HiBiT, and Gy2. P0 viral stock was produced by transfecting 5 μg recombinant bacmids into Sf9 cells (2.5 mL, density of 1 million cells per mL) for 96 h incubation and then used to produce P1 and P2 baculoviruses. GLP-1R-LgBiT-2MBP, DNGαs, Gβ1-HiBiT, and Gy2 were coexpressed at a multiplicity of infection ratio of 1:1:1:1 by infecting Sf9 cells at



**Fig. 5.** Systematic comparison of ligand recognition patterns of different nonpeptidic GLP-1R agonists. (A) Binding poses of the reported nonpeptidic agonists, GLP-1 (residues H7-L20 are shown), and a positive allosteric modulator, LSN3160440. Each ligand is displayed with a different color (GLP-1, cyan; PF-06882961, salmon; RGT1383, gray; Boc5, dark green; WB4-24, magenta; LY3502970, light green; CHU-128, yellow; TT-OAD2, sky blue; and LSN3160440, orange red). The binding sites of these ligands are located mainly in six regions represented by color dots, including the clefts of TM1-TM2 (orange), TM2-TM3 (red), and TM1-TM7 (yellow); ECLs 1 (blue) and 3 (pink); and the deep orthosteric pocket (green). (B) Positions of the binding sites identified among published GLP-1R structures. The dashed boxes show the binding modes of the agonists in specific regions. Colors are consistent with binding sites in panel (A).

a density of 3 million cells per mL with the P2 baculovirus (viral titers > 90%). Culture was harvested by centrifugation for 48 h post infection and cell pellets were stored at  $-80^{\circ}\text{C}$  until use.

To prepare the Boc5 or WB4-24 solution, the compound was first dissolved in DMSO with a final concentration of 100 mM and then distributed by 2% P188 solution (1:10, vol/vol) under ultrasonic condition at room temperature (RT) until a clear solution was obtained. The cell pellets were thawed and lysed in a buffer containing 20 mM Hepes, pH 7.5, 100 mM NaCl, 10% (vol/vol) glycerol, 10 mM  $\text{MgCl}_2$ , 1 mM  $\text{MnCl}_2$ , and 100  $\mu\text{M}$  TCEP supplemented with a EDTA-free protease inhibitor mixture (Bimake) by dounce homogenization. The complex formation was initiated by the addition of the Boc5 or WB4-24 solution (1:100, vol/vol), 10  $\mu\text{g}/\text{mL}$  Nb35, and 25  $\text{mU}/\text{mL}$  apyrase (New England Bio-Labs). After

1.5 h incubation at RT, the membrane was solubilized in the buffer above supplemented with 0.5% (wt/vol) LMNG (Anatrace) and 0.1% (wt/vol) CHS (Anatrace) for 2 h at  $4^{\circ}\text{C}$ . The supernatant was isolated by centrifugation at 65,000 g for 30 min and incubated with amylose resin (New England Bio-Labs) for 2 h at  $4^{\circ}\text{C}$ . The resin was then collected by centrifugation at 500 g for 10 min and washed in a gravity flow column (Sangon Biotech) with five column volumes of buffer containing 20 mM Hepes (pH 7.5), 100 mM NaCl, 10% (vol/vol) glycerol, 5 mM  $\text{MgCl}_2$ , 1 mM  $\text{MnCl}_2$ , 25  $\mu\text{M}$  TCEP, and 0.1% (wt/vol) LMNG-0.02% (wt/vol) CHS supplemented with 10  $\mu\text{M}$  Boc5 or WB4-24, followed by washing with 15 column volumes of buffer containing 20 mM Hepes (pH 7.5), 100 mM NaCl, 10% (vol/vol) glycerol, 5 mM  $\text{MgCl}_2$ , 1 mM  $\text{MnCl}_2$ , 25  $\mu\text{M}$  TCEP, and 0.03% (wt/vol) LMNG-0.01% (wt/vol) GDN (Anatrace)-0.008% (wt/vol) CHS supplemented



with 10  $\mu$ M Boc5 or WB4-24. The protein was then incubated overnight with TEV protease (customer-made) on the column to remove the C-terminal 2MBP tag in the buffer above at 4 °C. The flow-through was collected the next day and concentrated with a 100 kDa molecular weight cutoff concentrator (Millipore). The concentrated product was loaded onto a Superdex 200 increase 10/300 GL column (GE Healthcare) with a running buffer containing 20 mM Hepes (pH 7.5), 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 100  $\mu$ M TCEP, and 0.00075% LMNG-0.00025% GDN-0.0002% (wt/vol) CHS supplemented with 2  $\mu$ M Boc5 or WB4-24. The fractions for a monomeric complex were collected and concentrated to 15 to 20 mg/mL for cryo-EM examination.

**Expression and Purification of Nb35.** Nb35 with a C-terminal 6  $\times$  His-tag was expressed in the periplasm of *Escherichia coli* BL21 (DE3), extracted and purified by nickel affinity chromatography as previously described (31). The HiLoad 16/600 Superdex 75 column (GE Healthcare) was used to separate the monomeric fractions of Nb35 with a running buffer containing 20 mM Hepes, pH 7.5, and 100 mM NaCl. The purified Nb35 was flash-frozen in 30% (vol/vol) glycerol by liquid nitrogen and stored at  $-80$  °C until use.

**Cryo-EM Data Acquisition.** The concentrated sample (3.5  $\mu$ L) was applied to glow-discharged carbon grids (Quantifoil R1.2/1.3, 300 mesh), and subsequently vitrified using a Vitrobot Mark IV (ThermoFisher Scientific) set at 100% humidity and 4 °C. Cryo-EM images were collected on a Titan Krios microscope (FEI) equipped with a Gatan energy filter and a K3 direct electron detector. The microscope was operated at a 300 kV accelerating voltage, at a nominal magnification of 46,685 $\times$  in counting mode, corresponding to a pixel size of 1.071 Å. The total exposure time was set to 7.2 s with intermediate frames recorded every 0.2 s, resulting in an accumulated dose of 80 electrons per Å<sup>2</sup> with a defocus range of  $-1.2$  to  $-2.2$   $\mu$ m. In total, 10,654 images for the Boc5-GLP-1R-G<sub>s</sub> complex and 10,148 images for the WB4-24-GLP-1R-G<sub>s</sub> complex were collected.

**Image Processing.** Dose-fractionated image stacks were subjected to beam-induced motion correction using MotionCor2.1 (32). 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 (33). Particle selection, 2-dimensional classifications, ab initio model generation, and homogeneous refinement were performed on cryoSPARC.

For the Boc5-GLP-1R-G<sub>s</sub> complex, template-picking yielded 6,785,990 particles that were subjected to three-dimensional (3D) classifications with a mask on the receptor to discard false-positive particles or particles categorized in poorly defined classes, producing particle projections for further processing. This subset of particles was subjected to further 3D autorefinement with a mask on the complex. A selected subset containing 610,843 projections was subsequently subjected to 3D refinement and Bayesian polishing with a pixel size of 1.045. After the last round of refinement, the final map had an indicated global resolution of 2.61 Å at a Fourier shell correlation (FSC) of 0.143. Local resolution was determined using the Bsoft package with half-maps as input maps (34).

For the WB4-24-GLP-1R-G<sub>s</sub> complex, a total of 8,405,468 particles were picked using a template-picker. The 3D classifications were performed with a mask on the receptor to discard false-positive particles or particles categorized in poorly defined classes, producing 3,866,515 particle projections for further processing. This subset of particles was subjected to further 3D autorefinement with a mask on the complex and were subsequently subjected to a round of 3D classifications with a mask on the ECD. A data set of 747,282 particles was subjected to 3D refinement, yielding a final map with a global nominal resolution at 3.09 Å by the 0.143 criteria of the gold-standard FSC. Half-reconstructions were used to determine the local resolution of each map.

**Model-Building and Refinement.** The structures of the CHU-128-GLP-1R-G<sub>s</sub> (22) (Protein Data Bank [PDB]: 6X19) was used as an initial template for model-building of the Boc5- and WB4-24-bound complexes. Ligand coordinates and geometry restraints were generated using phenix.elbow. Models were docked into the electron microscopy density map using UCSF Chimera. This starting model was then subjected to iterative rounds of manual adjustment and automated refinement in Coot (35) and Phenix (36), respectively. The final refinement statistics were validated using the module comprehensive validation (cryo-EM) in Phenix. Structural figures were prepared in Chimera, Chimera X,

and PyMOL (<https://pymol.org/2/>). The final refinement statistics are provided in *SI Appendix, Table S1*.

**cAMP Accumulation Assay.** Peptide- or small-molecule-stimulated cAMP accumulation was measured by a LANCE Ultra cAMP kit (PerkinElmer). Briefly, after 24 h transfection with various constructs, HEK293T cells were digested by 0.2% (wt/vol) EDTA and washed once with PBS. Cells were then resuspended with stimulation buffer (HBSS supplemented with 5 mM Hepes, 0.5 mM IBMX, and 0.1% (wt/vol) BSA, pH 7.4) to a density of 0.6 million cells per mL and added to 384-well white plates (3,000 cells per well). Different concentrations of ligand in stimulation buffer were added, and the stimulation lasted for 40 min at RT. The reaction was stopped by adding 5  $\mu$ L Eu-cAMP tracer and ULIGHT-anti-cAMP. After 1 h incubation at RT, the plate was read by an EnVision plate reader (PerkinElmer) to measure time resolved-fluorescence resonance energy transfer signals (excitation: 320 nm, emission: 615 nm and 665 nm). A cAMP standard curve was used to convert the fluorescence resonance energy transfer ratios (665/615 nm) to cAMP concentrations.

**Whole-Cell Binding Assay.** HEK293T cells were seeded into 96-well plates (PerkinElmer) coated with poly-D-lysine hydrobromide (Sigma-Aldrich) at a density of 30,000 cells per well and incubated overnight. After 24 h transfection, cells were washed twice and incubated with blocking buffer (DMEM supplemented with 33 mM Hepes, and 0.1% (wt/vol) BSA, pH 7.4) for 2 h at 37 °C. Then, radiolabeled <sup>125</sup>I-GLP-1 (30 pM, PerkinElmer) and increasing concentrations of unlabeled ligand were added and competitively reacted with the cells in binding buffer (PBS supplemented with 10% [wt/vol] BSA, pH 7.4) at RT for 3 h. After that, cells were washed with ice-cold PBS and lysed by 50  $\mu$ L lysis buffer (PBS supplemented with 20 mM Tris-HCl and 1% [vol/vol] Triton X-100, pH 7.4). Finally, 150  $\mu$ L of scintillation mixture (OptiPhase SuperMix, PerkinElmer) was employed and radioactivity (counts per minute) was determined by a scintillation counter (MicroBeta2 plate counter, PerkinElmer).

**$\beta$ -arrestin 1/2 Recruitment Assay.** HEK293T cells were seeded at a density of 30,000 cells per well into 96-well culture plates pretreated with poly-D-lysine hydrobromide. After incubation for 24 h to reach 80% confluence, the cells were transiently transfected with HA-GLP-1R-Rluc8 and  $\beta$ -arrestin 1/2-Venus at a 1:9 mass ratio using lipofectamine 3000 reagent (Invitrogen) and cultured for another 24 h. Thereafter, cells were washed once and incubated for 30 min at 37 °C with HBSS buffer (pH 7.4) supplemented with 0.1% BSA and 10 mM Hepes. Five micromolars of coelenterazine h (Yeasen Biotech) was then added and incubated for 5 min in the dark. The bioluminescence resonance energy transfer (BRET) signals were detected with an EnVision plate reader by calculating the ratio of emission at 535 nm over emission at 470 nm. A 1.5 min baseline of BRET measurement was taken before the addition of the ligand and BRET signal was measured at 10 s intervals for a further 9 min. After removing baseline and background readings by subtracting the average values of the baseline measurement and the average values of vehicle-treated samples, respectively, the area-under-the-curve (AUC) across the time-course response curve was determined. Concentration-response curves were plotted using the total AUC during the time of measurement post ligand addition.

**Molecular Dynamics Simulation.** Molecular dynamics simulation studies were performed using Gromacs 2020.1. The Boc5-GLP-1R and WB4-24-GLP-1R complexes were built based on the cryo-EM structures and prepared by Protein Preparation Wizard (Schrodinger 2017-4) with the G protein and Nb35 nanobody removed. The missing ECD and loops in the GLP-1R were generated by molecular superposition, using UCSF Chimera, of the corresponding regions in the CHU-128-GLP-1R-G<sub>s</sub> complex (22) (PDB: 6X19). The receptor was capped with acetyl and methylamide. All titratable residues were left in their dominant state at pH 7.0. The ligand-bound GLP-1R complexes were embedded in a bilayer composed of 211 POPC lipids and solvated with 0.15 M NaCl in explicit TIP3P waters using CHARMM-GUI Membrane Builder v3.5 (37). The CHARMM36-CAMP force field (38) was adopted for protein, lipids, and salt ions. Boc5 and WB4-24 were modeled with the CHARMM CGenFF small-molecule force field (39), program version 2.5. The Particle Mesh Ewald method was used to treat all electrostatic interactions beyond a cutoff of 12 Å, and the bonds involving hydrogen atoms were constrained using the LINCS algorithm (40). The complex system was first relaxed using the steepest descent energy minimization, followed by slow

heating of the system to 310 K with restraints. Finally, 1,000 ns production simulations without restraints were carried out, with a time step of 2 fs in the isothermal-isobaric ensemble at 310 K and 1 bar using the v-rescale thermostat and the semi-isotropic Parrinello-Rahman barostat (41), respectively.

**Statistical Analysis.** All functional study data were analyzed using Prism 7 (GraphPad) and presented as means  $\pm$  SEM from at least three independent experiments. Concentration–response curves were evaluated with a three-parameter logistic equation. The significance was determined with either a two-tailed Student's *t* test or a one-way ANOVA, and *P* < 0.05 was considered statistically significant.

**Data Availability.** The atomic coordinates and the electron microscopy maps have been deposited in the PDB under accession codes 7X8R (Boc5–GLP-1R–G<sub>s</sub> complex) and 7X8S (WB4-24–GLP-1R–G<sub>s</sub> complex) and in the Electron Microscopy Data Bank under accession codes EMD-33057 (Boc5–GLP-1R–G<sub>s</sub> complex) and EMD-33058 (WB4-24–GLP-1R–G<sub>s</sub> complex), respectively. All relevant data are available from the authors and/or included in the manuscript or *SI Appendix*.

**ACKNOWLEDGMENTS.** We are indebted to H. Eric Xu, Honglei Ma, Peng-Fei Lan, and Ming Lei for valuable discussions. This work was partially supported by the National Natural Science Foundation of China 81872915 (M.-W.W.), 82073904 (M.-W.W.), 82121005 (D.Y.), 81973373 (D.Y.), and 21704064 (Q.Z.);

the 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.); the National Science and Technology Major Project of China–Innovation 2030 for Brain Science and Brain-Inspired Technology 2021ZD0203400 (Q.Z.); the National Key Basic Research Program of China 2018YFA0507000 (M.-W.W.); Novo Nordisk-CAS Research Fund grant NNCAS-2017–1-CC (D.Y.); and the SA-SIBS Scholarship Program (D.Y.). The cryo-EM data were collected at the Cryo-Electron Microscopy Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences.

Author affiliations: <sup>a</sup>Department of Pharmacology, School of Basic Medical Sciences, Fudan University, Shanghai 200032, China; <sup>b</sup>Department of Biophysics, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China; <sup>c</sup>Department of Pathology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China; <sup>d</sup>School of Pharmacy, Fudan University, Shanghai 201203, China; <sup>e</sup>School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China; <sup>f</sup>The National Center for Drug Screening, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; <sup>g</sup>School of Graduate Studies, University of Chinese Academy of Sciences, Beijing 100049, China; <sup>h</sup>The CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; <sup>i</sup>Department of Bioactivity Screening, Research Center for Deepsea Bioresources, Sanya, 572025, China; and <sup>j</sup>Department of Chemistry, School of Science, The University of Tokyo, Tokyo 113-0033, Japan.

1. J. Reed, S. Bain, V. Kanamarlapudi, A review of current trends with type 2 diabetes epidemiology, aetiology, pathogenesis, treatments and future perspectives. *Diabetes Metab. Syndr. Obes.* **14**, 3567–3602 (2021).
2. F. K. Saraiva, A. C. Sposito, Cardiovascular effects of glucagon-like peptide 1 (GLP-1) receptor agonists. *Cardiovasc. Diabetol.* **13**, 142 (2014).
3. D. J. Drucker, Mechanisms of action and therapeutic application of glucagon-like peptide-1. *Cell Metab.* **27**, 740–756 (2018).
4. Y. M. Cho, C. E. Merchant, T. J. Kieffer, Targeting the glucagon receptor family for diabetes and obesity therapy. *Pharmacol. Ther.* **135**, 247–278 (2012).
5. S. L. Samson, A. Garber, GLP-1R agonist therapy for diabetes: Benefits and potential risks. *Curr. Opin. Endocrinol. Diabetes Obes.* **20**, 87–97 (2013).
6. T. K. Thethi, R. Pratley, J. J. Meier, Efficacy, safety and cardiovascular outcomes of once-daily oral semaglutide in patients with type 2 diabetes: The PIONEER programme. *Diabetes Obes. Metab.* **22**, 1263–1277 (2020).
7. E. E. Wright Jr., V. R. Aroda, Clinical review of the efficacy and safety of oral semaglutide in patients with type 2 diabetes considered for injectable GLP-1 receptor agonist therapy or currently on insulin therapy. *Postgrad. Med.* **132**, 26–36 (2020).
8. D. Wootten *et al.*, Differential activation and modulation of the glucagon-like peptide-1 receptor by small molecule ligands. *Mol. Pharmacol.* **83**, 822–834 (2013).
9. H. J. Choe, Y. M. Cho, Peptidyl and non-peptidyl oral glucagon-like peptide-1 receptor agonists. *Endocrinol. Metab. (Seoul)* **36**, 22–29 (2021).
10. C. Liu, Y. Zou, H. Qian, GLP-1R agonists for the treatment of obesity: A patent review (2015–present). *Expert Opin. Ther. Pat.* **30**, 781–794 (2020).
11. A. Saxena *et al.*, 353-OR: Oral small molecule GLP-1R agonist PF-06882961 robustly reduces plasma glucose and body weight after 28 days in adults with T2DM. *Diabetes* **69**, 353-OR (2020).
12. P. Zhao *et al.*, Activation of the GLP-1 receptor by a non-peptidic agonist. *Nature* **577**, 432–436 (2020).
13. D. Chen *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).
14. H. Su *et al.*, Boc5, a non-peptidic glucagon-like peptide-1 receptor agonist, invokes sustained glycemic control and weight loss in diabetic mice. *PLoS One* **3**, e2892 (2008).
15. M. He *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).
16. Q. Liu *et al.*, Cyclobutane derivatives as novel nonpeptidic small molecule agonists of glucagon-like peptide-1 receptor. *J. Med. Chem.* **55**, 250–267 (2012).
17. J. Duan *et al.*, Cryo-EM structure of an activated VIP1 receptor-G protein complex revealed by a NanoBiT tethering strategy. *Nat. Commun.* **11**, 4121 (2020).
18. F. Zhou *et al.*, Structural basis for activation of the growth hormone-releasing hormone receptor. *Nat. Commun.* **11**, 5205 (2020).
19. M. Lucey *et al.*, Disconnect between signalling potency and in vivo efficacy of pharmacokinetically optimised biased glucagon-like peptide-1 receptor agonists. *Mol. Metab.* **37**, 100991 (2020).
20. F. Wu *et al.*, Full-length human GLP-1 receptor structure without orthosteric ligands. *Nat. Commun.* **11**, 1272 (2020).
21. D. Wootten, J. Simms, L. J. Miller, A. Christopoulos, P. M. Sexton, Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5211–5216 (2013).
22. X. Zhang *et al.*, Differential GLP-1R binding and activation by peptide and non-peptide agonists. *Mol. Cell* **80**, 485–500.e7 (2020).
23. C. Montrose-Rafizadeh *et al.*, High potency antagonists of the pancreatic glucagon-like peptide-1 receptor. *J. Biol. Chem.* **272**, 21201–21206 (1997).
24. Y. L. Liang *et al.*, Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor-Gs complex. *Nature* **555**, 121–125 (2018).
25. A. B. Bueno *et al.*, Structural insights into probe-dependent positive allosterism of the GLP-1 receptor. *Nat. Chem. Biol.* **16**, 1105–1110 (2020).
26. H. Fan *et al.*, The non-peptide GLP-1 receptor agonist WB4-24 blocks inflammatory nociception by stimulating  $\beta$ -endorphin release from spinal microglia. *Br. J. Pharmacol.* **172**, 64–79 (2015).
27. T. Kawai *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).
28. H. Ma *et al.*, Structural insights into the activation of GLP-1R by a small molecule agonist. *Cell Res.* **30**, 1140–1142 (2020).
29. F. Zhou *et al.*, Molecular basis of ligand recognition and activation of human V2 vasopressin receptor. *Cell Res.* **31**, 929–931 (2021).
30. Z. Cong *et al.*, Molecular insights into ago-allosteric modulation of the human glucagon-like peptide-1 receptor. *Nat. Commun.* **12**, 3763 (2021).
31. S. G. Rasmussen *et al.*, Crystal structure of the  $\beta$ 2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
32. S. Q. Zheng *et al.*, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
33. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
34. J. B. Heymann, Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. *Protein Sci.* **27**, 159–171 (2018).
35. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
36. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
37. E. L. Wu *et al.*, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* **35**, 1997–2004 (2014).
38. O. Guvench *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).
39. K. Vanommeslaeghe *et al.*, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
40. B. Hess, P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
41. K. M. Aoki, F. Yonezawa, Constant-pressure molecular-dynamics simulations of the crystal-smectic transition in systems of soft parallel spherocylinders. *Phys. Rev. A* **46**, 6541–6549 (1992).