Supporting Information

Kirkpatrick et al. 10.1073/pnas.1218051109

SI Methods

Unless specified otherwise, all minimizations referenced were conjugate-gradient minimizations performed in MPSim under the Dreiding force field (1, 2). Likewise, all of the molecular dynamics were performed in NAMD 2.6 using CHARMM22 charges for the protein and CHARMM27 charges for the lipids (3–5). The waters were modeled using TIP3P (6).

1. Structure Prediction of Transmembrane Domain. The 3D structure of glucagon-like peptide 1 receptor (GLP1R) transmembrane (TM) bundle was predicted using MembStruk (version 4.3). The details of MembStruk were described elsewhere (7). Here, we outline the procedure, highlighting aspects relevant to GLP1R or aspects that were additionally improved.

1.1. Prediction of TM regions. The TM region was predicted using the TM2nsS method (8). We searched for the sequences of the family B G protein-coupled receptors (GPCRs) from the UnitProtKB/Swiss-Prot database (9). The 166 sequences included 76 members of the LN-TM7 and 18 members of the Methuselah subfamilies. Only seven sequences (belonging to the receptors for glucagon/glucagon-like peptide and glucose-dependent insulinotropic polypeptide) showed a sequence identity higher than 40%. Among 166 sequences, we took 65 sequences with pairwise sequence identity to GLP1R of >20% to calculate the hydropathy analysis curve. A multiple sequence alignment was performed with ClustalW (10), which was used as input to TM2ndS for hydrophathy analysis. In this alignment, the long amino terminus of the sequence was excluded. The hydrophobic center for each TM was determined by the position bisecting the area of each peak in the curve. To define the clear boundary of TM regions, we carried out a second round of seven TM predictions, where the sequence of each TM core (20 aa around the hydrophobic center) was used as a query in BLAST search. Under a high-gap penalty in BLAST search, the sequences with >50% identity were identified from the set of the family B GPCR sequences. The final refined TM region and its hydrophobic center for each of the seven TM domains were determined from this second round of prediction.

1.2. Assembly of TM helical bundle. We built a canonical α-helix for each TM region and assembled these seven helices into a template generated from the predicted structure of human prostaglandin D2 (PGD2) receptor fully equilibrated in explicit lipids and water solvent (11). As expected, all receptors available from our family A GPCR structural database were distant in sequence from the GLP1R. The human prostaglandin PGD2 receptor was chosen, because it was the only hormone receptor in our database. Here, we assumed that the arrangement of TM helices of the GLP1R would be similar to the arrangement of the family A GPCR. The (x, y) coordinate, the tilt angle with respect to the z axis, and the azimuthal angle needed for definition of arrangement of TM α-helices as shown for frog rhodopsin (12) were calculated as follows: the xy midplane and z axis were determined by diagonalizing the matrix of the moment of inertia for the heavy atoms comprising the lipid bilayers where the PGD2 receptor was embedded. The $+z$ direction pointed to the extracellular (EC) region. The center of helix bundle Cα-atoms was set to origin, and the x axis was defined as the axis from the center of the bundle to the center of helix 2. Under this coordinate system, the (x, y) coordinates of each helix center and the tilt/azimuthal angles of each helix (the helix axis was determined from the moment of inertia of Cα-atoms) were computed for the PGD2 receptor template. The seven hydrophobic centers for GLP1R were all in the x–y plane with these (x, y) coordinates, and the seven helices were inclined accordingly. Each helix was rotated about its axis, and therefore, its hydrophobic moment pointed to the membrane.

We then carried out 200 ps of molecular dynamics at 300 K without solvent or lipid to allow the conformation of each individual helix to bend or kink as appropriate. The molecular dynamics were run in MPSim under NVT conditions with the Dreiding force field. We selected the lowest potential energy snapshot after 100 ps. Using this conformation, the net hydrophobic moment vector was calculated from the middle 15 residues around the hydrophobic center. Each helix was rotated, and therefore, this hydrophobic moment vector pointed away from the center of helix bundle.

Next, the orientation of the helix was further examined with energy-based optimization. The rotational orientation of each helix was scanned over 360° in 30° increments, and at each orientation, the side chains were placed using the SCREAM program. Here, we used the coarse energy scoring function that was combined with the penalty score derived from the hydrophobicity scale and the force field-based SCREAM energy function (13). From this scanning step, we chose two orientations for TM1, -2, -3, -4, -5, and -7 and three orientations for TM6, leading to the 192 combinatorial orientations. These 192 combinations were ordered by the number of interhelical hydrogen bonds, showing a consistent orientation for TM2, -3, and -7 on the top. These structures consistently showed hydrogen bonds between R190(TM2) and N240(TM3) and between E247(TM3) and Q394(TM7). Additionally, the orientation of TM1 was scanned, and the best orientation was chosen based on the number of interhelical hydrogen bonds. This process was motivated by previous experimental studies reporting that interhelical hydrogen bonding drives strong interactions in membrane proteins (14, 15). The orientation of TM4 was selected where the aromatic residues were inside the bundle and well-packed with the adjacent helices. TM5 and -6 were then scanned combinatorially, and the best rotations were chosen based on the side chain–side chain hydrogen bond energy.

The final helix bundle was subjected to conjugate gradient minimization to an rms force threshold of 0.5 kcal/mol per angstrom. Two layers of explicit lipid molecules were then added to the bundle, and this 7-helix–lipid complex was optimized to achieve the proper packing using rigid body molecular dynamics in MPSim with the Dreiding force field for 50 ps. The final equilibrated structure was then minimized to an rms force threshold of 0.3 kcal/mol per angstrom.

2. Structure Prediction of the N Terminus of GLP1R/Ligand Complex. The structures of the N terminus of GLP1R (nGLP1R) and ligand were modeled separately and then combined. The details are described elsewhere, and we outline the procedure briefly.

2.1. Determination of the 3D structure for nGLP1R. The structure of nGLP1R was determined by homology modeling with the NMR structure of mouse Corticotropin-releasing factor (CRF) receptor 2 (Protein Data Bank ID code 2JND), which consists of 19 conformations (16). We scored these 19 structures using the potential energy in the Dreiding force field and selected the lowest 3 structures for equilibration in a periodic box with explicit water solvent. The final template structure was determined by considering the energy and possible conserved structural motifs among class B GPCRs (i.e., the salt bridge between Asp65 and Arg101) and the Cα rms with the NMR structures (17).

We used Modeller9v1 (18) to construct homology models for residues 45–130 of nGLP1R. The alignment of sequences between the template (the mouse CRF receptor) and the query (nGLP1R)

was determined from the multiple sequence alignment, including seven other secretin-like family B GPCRs with N-terminal sequences that show >30% identity. The side chains were replaced using the SCREAM program (19) and optimized with 100 steps of conjugate gradient minimization. This final optimized structure was then equilibrated in the explicit water box. The water box was chosen to extend by ~7 Å beyond the solute in each direction. The whole system was then neutralized by adding Na^+ ions. The solvent for each system was minimized with conjugate gradients for 5,000 steps and then equilibrated for 10 ps of molecular dynamics at 310 K using the Langevin thermostat with a damping coefficient of 5 ps^{-1} with all of the coordinates of the solute fixed. This system was then minimized without any constraint and equilibrated for 500 ps at 310 K and a pressure of 1 atm (using the Langevin piston method). All simulations used time steps of 1 fs, with electrostatic interactions computed using the Particle Mesh Ewald method.

2.2. Docking of ligands into nGLP1R. The NMR structure of Exendin-4 (Exe4; Protein Data Bank ID code 1JRJ) was optimized with conjugate gradient minimization and equilibrated for 500 ps in explicit water solvent at 310 K and a pressure of 1 atm using the Langevin piston method (20). We docked the optimized Exe4 ligand into nGLP1R (residues 45–130) using the ZDOCK rigid docking program (21). By assuming that the Exe4 would bind to a region similar to the region of mouse CRF receptor in contact with its peptide ligand astressin, we filtered the initial 2,000 configurations down to 131 configurations (16). After side chain optimization and conjugate gradient minimization, we selected the top five configurations based on the force field energy, each of which was then fully minimized to an rms force of 0.3 kcal/mol per angstrom. The lowest energy configuration was equilibrated in explicit water solvent with a harmonic constraint of 5 kcal/(mol $Å^2$) for the backbone atoms as described previously.

3. Combination of nGLP1R/Exe4 Complex with the TM Bundle. To combine our nGLP1R/ligand complex with the TM bundle, we needed to grow protein residues 131–145. We ran secondary structure predictions on these residues in the programs nnPredict, GORIV, HNN, SOPMA, PSIPRED, Porter, and Jpred (22–28). Based on secondary structure predictions (six consensus predictions among seven predictions), we found that the residues 137–145 were helical. The conformation of this segment was predicted using Modeller9v1 under helix restraint. The overall combined structure was built by matching the common helical parts of residues 137–145 in the nGLP1R/Exe4 complex and the TM1 region. We first generated 1,000 conformations of the residues 131–145 and selected ones with correct helix chirality (right-handed) and the helical conformation preserved (not unraveled). Then, we found the conformations where the ligand Exe4 was located properly inside the TM pocket without any bumping into helices. We examined the interactions between receptor and ligand and chose the conformations where the ligand key residues were involved (29). We carried out minimization and annealing molecular dynamics for the residues of the ligand contacting TM regions and having random coil conformation (residues 1–7). The energy minimization and annealing molecular dynamics were run in MPSim for three cycles [50 K→600 K→50 K (50-K step; 1 ps run for each)].

4. Prediction of Loop Structures. EC1 and EC2 were first modeled by using Modeller9v1. EC1 and EC2 were 29- and 14-aa long, respectively. According to the secondary structure predictions from APSSP2 and PSIPRED servers, EC1 was found to be helical from residues 215–224 (26, 30). We generated 1,000 conformations under helix restraint in Modeller9v1 for EC1 and EC2. The restraints included a distance restraint between both Y205 on EC1 and F6 of the Exe4 ligand as well as a disulphide bond between EC1 (residue 226) and EC2 (residue 296) (31–33). (Because we were building both GLP1 and Exe4 in this process and no corresponding data existed for Exe4, we used information from studies that were performed on GLP1.) We chose the conformations where the predicted parts were helical and right-handed (723 of 1,000 conformations). The final three candidates were selected, for which the conformation was compact and not touching the presumed lipid regions and the key residues of the EC1 and the ligand showed the favorable contacts (or at least close contacts). The contacts emphasized were with M204, Y205, D215, and R227 (34, 35). The loops were optimized by side chain replacement with SCREAM, energy minimization of the EC1 and EC2 only, and then, three-cycle annealing molecular dynamics [50 K→600 K→50 K (50-K steps; 1 ps run for each)] in MPSim. Based on the potential energy and maximizing the total number of contacts, the best structure of EC1 and EC2 was chosen. Next, the 10-aa EC3 loop was modeled. Among the 1,000 conformations generated from Modeller9v1, we chose two candidates where the loop was not touching the presumed lipid regions and likely to contact T7 of the ligand that was known as one of the key residues (32).

The remaining three intracellular (IC) loops were modeled all together. The lengths of the IC1, -2, and -3 loops were 5, 9, and 16 aa, respectively; 200 conformations were generated, and the loops were chosen that sat in a closed conformation on top of the TM region. The EC3 and IC loops were optimized together through SCREAM, energy minimization, and three-cycle annealing molecular dynamics [50 K→600 K→50 K (50-K steps; 1 ps run for each)] in MPSim. Based on the potential energy and maximizing contacts, the final structure was chosen.

5. Relaxation in Explicit Membrane and Water Solvent. After predicting the full structure of the GLP1R/Exe4 complex, we embedded this structure in a periodic infinite membrane, solvated the system with explicit water, and equilibrated with molecular dynamics at 310 K. The system size was 75 × 75 × 117 Å with 6,493 solute atoms, 5,092 lipid atoms, 40,302 water atoms, and 5 Na^+ ions. We used palmitoyl-oleoyl-phosphatidylcholine to form the lipid bilayers. We made the receptor with the acetylated N terminus and the *N*-methylamide C terminus. Before full equilibration, the solvent molecules were equilibrated first at 310 K for 100 ps. Then, the whole system was equilibrated by gradually increasing the temperature to 100, 200, and 310 K, with a 500-ps molecular dynamics run performed at each temperature. Finally, the full molecular dynamics simulation was carried out for 20 ns, and a constant pressure of 1 atm was maintained by using the Langevin piston method.

6. Incorporating the Crystal Structure. After the 2008 crystal structure of the human nGLP1R in complex with the antagonist Exe4(9-39) was revealed (Protein Data Bank ID code 3C5T), we incorporated it into our GLP1R model (36). This crystal structure was aligned to our relaxed GLP1R bound to Exe4 structure using the rmsd Trajectory Tool of the Visual Molecular Dynamics (VMD) program (37). This alignment was done by the backbone atoms of those ligand residues that were well-resolved in the crystal structure—9–33. The resolved crystal residues (9–33 for Exe4 and 28–131 for GLP1R) were then substituted for those residues in our structure. Five residues on either side of the newly connected residues were minimized to 0.5 rms force with conjugate gradient minimization in MPSim with Dreiding force field. Then, the entire complex was minimized to 0.5 rms force. This procedure allowed us to retain our TM bundle conformation and binding site, while adding in the crystal information.

7. Optimizing the New Structure. ***7.1. Molecular dynamics.*** To optimize our models further, we ran full-lipid and full-solvent molecular dynamics. Each complex was inserted into a fully equilibrated hydrated palmitoyl-oleoyl-phosphatidylcholine lipid bilayer having a cell size of 75 × 75 Å. All lipids within 1 Å and all waters within 5 Å were removed. The complex was then solvated using the solvate package of VMD (37). To have a net charge of zero,

three sodium atoms were added to the Exe4 structure using the autoionize feature of VMD (37). Both structures then underwent 250-ps conjugate gradient minimization with a 1-fs time step. This procedure was done with the protein and ligand kept fixed. They remained fixed as the systems were heated to 310 K and equilibrated for 500 ps under NPT (constant number of particles, pressure, and temperature) conditions. Next, both systems were minimized for 250 ps. Finally, the structures underwent NPT dynamics for 18 ns.

7.2. Simulated annealing. On inspecting the dynamics runs, it was decided to optimize the TM region binding sites further by simulated annealing in MPSim with the Dreiding force field. The proteins and ligands were taken from their molecular dynamics runs, and everything within 5 Å of the ligand was allowed to be flexible, along with the loops. These structures underwent minimization to 0.5 rms with a maximum of 1,000 steps, and they were then heated from 50 to 600 K in 50° increments, with 0.1 ps (100 steps) spent at each temperature five times. The final structure was again minimized to 0.5 rms.

7.3. More molecular dynamics. After annealing, the two structures were reinserted into membranes, and the above procedure for starting NPT molecular dynamics was repeated for 10 ns. The trajectories were analyzed at 1-ns intervals, with each protein complex being minimized to 0.5 rms force using MPSim with the Dreiding force field. A representative snapshot was chosen for each run for discussion here.

8. Bihelix and Superbihelix on the MembStruk TM Bundle. With the goal of optimizing our MembStruk TM bundle, which had been through around 20 ns of molecular dynamics, we ran Bihelix on the TM region (38). This process sampled the η-angles from 0° to 360° in 30° increments. The top 10 helix orientations are shown in Table S1. All zero structure (our starting MembStruk structure) shows up as number nine. We then ran Superbihelix on all 10 of these structures, with the top two results shown in Table S2 (39). We sampled both the η- and ϕ-angles of each of the 10 structures from −30° to 30° in 15° increments. The second structure is almost identical to the starting structure, with only two −15° changes to ϕ for helices 2 and 4. Because this structure was so close to our starting structure, we decided to continue with the molecular dynamics-relaxed structure for additional work. The number one case, which deviates more from the all zero, is not discussed in this paper.

1. Lim KT, et al. (1997) Molecular dynamics for very large systems on massively parallel computers: The MPSim program. *J Comput Chem* 18(4):501–521.
2. Mayo SL, Olafson BD, Goddard WA (1990) Dreiding—a generic force-field for molecular simulations. *J Phys Chem* 94(26):8897–8909.
3. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.
4. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616.
5. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30(10):1545–1614.
6. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.
7. Trabanino RJ, et al. (2004) First principles predictions of the structure and function of g-protein-coupled receptors: Validation for bovine rhodopsin. *Biophys J* 86(4): 1904–1921.
8. Trabanino RJ, et al. (2004) First principles predictions of the structure and function of g-protein-coupled receptors: Validation for bovine rhodopsin. *Biophys J* 86(4): 1904–1921.
9. Gasteiger E, et al. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31(13):3784–3788.
10. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.
11. Li YY, et al. (2007) Prediction of the 3D structure and dynamics of human DP G-protein coupled receptor bound to an agonist and an antagonist. *J Am Chem Soc* 129(35):10720–10731.
12. Unger VM, Hargrave PA, Baldwin JM, Schertler GFX (1997) Arrangement of rhodopsin transmembrane alpha-helices. *Nature* 389(6647):203–206.
13. Bray JK, Goddard WA, 3rd (2008) The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists. *J Mol Graph Model* 27(1):66–81.
14. Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization: The two-stage model. *Biochemistry* 29(17):4031–4037.
15. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM (2000) Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat Struct Biol* 7(2): 154–160.
16. Grace CR, et al. (2007) Structure of the N-terminal domain of a type B1 G protein-coupled receptor in complex with a peptide ligand. *Proc Natl Acad Sci USA* 104(12): 4858–4863.
17. Parthier C, Reedtz-Runge S, Rudolph R, Stubbs MT (2009) Passing the baton in class B GPCRs: Peptide hormone activation via helix induction? *Trends Biochem Sci* 34(6): 303–310.
18. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.
19. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. *J Chem Theory Comput* 4(12):2160–2169.
20. Neidigh JW, Fesinmeyer RM, Prickett KS, Andersen NH (2001) Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry* 40(44):13188–13200.
21. Chen R, Li L, Weng ZP (2003) ZDOCK: An initial-stage protein-docking algorithm. *Proteins* 52(1):80–87.
22. Kneller DG, Cohen FE, Langridge R (1990) Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* 214(1):171–182.
23. Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol* 266:540–553.
24. Lin K, Simossis VA, Taylor WR, Heringa J (2005) A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* 21(2):152–159.
25. Levin JM, Robson B, Garnier J (1986) An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett* 205(2):303–308.
26. Buchan DW, et al. (2010) Protein annotation and modelling servers at University College London. *Nucleic Acids Res* 38(Web Server):W563–W568.
27. Pollastri G, McLysaght A (2005) Porter: A new, accurate server for protein secondary structure prediction. *Bioinformatics* 21(8):1719–1720.
28. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36(Web Server):W197–W201.
29. Adelhorst K, Hedegaard BB, Knudsen LB, Kirk O (1994) Structure-activity studies of glucagon-like peptide-1. *J Biol Chem* 269(9):6275–6278.
30. Raghava GPS (2002) APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. *CASP5* A-132.
31. López de Maturana R, Treece-Birch J, Abidi F, Findlay JB, Donnelly D (2004) Met-204 and Tyr-205 are together important for binding GLP-1 receptor agonists but not their N-terminally truncated analogues. *Protein Pept Lett* 11(1):15–22.
32. Adelhorst K, Hedegaard BB, Knudsen LB, Kirk O (1994) Structure-activity studies of glucagon-like peptide-1. *J Biol Chem* 269(9):6275–6278.
33. Mann RJ, Al-Sabah S, de Maturana RL, Sinfield JK, Donnelly D (2010) Functional coupling of Cys-226 and Cys-296 in the glucagon-like peptide-1 (GLP-1) receptor indicates a disulfide bond that is close to the activation pocket. *Peptides* 31(12): 2289–2293.
34. López de Maturana R, Treece-Birch J, Abidi F, Findlay JB, Donnelly D (2004) Met-204 and Tyr-205 are together important for binding GLP-1 receptor agonists but not their N-terminally truncated analogues. *Protein Pept Lett* 11(1):15–22.
35. Xiao Q, Jeng W, Wheeler MB (2000) Characterization of glucagon-like peptide-1 receptor-binding determinants. *J Mol Endocrinol* 25(3):321–335.
36. Runge S, Thøgersen H, Madsen K, Lau J, Rudolph R (2008) Crystal structure of the ligand-bound glucagon-like peptide-1 receptor extracellular domain. *J Biol Chem* 283 (17):11340–11347.
37. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.
38. Abrol R, Bray J, Goddard WA (2012) Bihelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. *Proteins: Structure, Function, and Bioinformatics* 80(2):505–518.
39. Abrol R, Griffith AR, Bray JK, Goddard WA 3rd (2012) Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. *Membrane Protein Structure: Methods and Protocols (Methods in Molecular Biology)*, eds Vaidehi N, Klein-Seetharaman J (Humana, New York), Vol 914, pp 237–254.

Table S1. Top 10 Bihelix structures for GLP1R (η)

H1	H2	H3	H4	H5	H6	H7
90	0	0	0	0	0	270
0	0	0	0	0	0	270
30	0	0	0	0	0	270
0	0	0	0	0	30	0
30	0	0	0	0	0	0
0	0	0	0	270	0	0
90	0	0	0	0	0	0
270	0	0	0	0	0	270
0	0	0	0	0	0	0
240	0	0	0	0	0	270

Red conformation is the starting structure.

Table S2. Top two Superbihelix structures

ϕ							η						
H1	H2	H3	H4	H5	H6	H7	H1	H2	H3	H4	H5	H6	H7
−15	−30	0	0	0	0	15	75	15	0	0	0	15	240
0	−15	0	−15	0	0	0	0	0	0	0	0	0	0

Table S3. Interhelical hydrogen bonds for the GLP1R/Exe4 structure

Donor	Acceptor	Energy
TM3-TM6 ionic lock		
R348(IC3)	E247(TM3)	−45.4
Coupling of TM2-TM3-TM6		
N182(TM2)	E247(TM3)	−10.7
R190(TM2)	N240(TM3)	−15.1
TM1-TM2-TM7 interaction network		
R176(TM2)	E408(TM7)	−52.9
Y152(TM1)	Q394(TM7)	−4.7
T149(TM1)	E387(TM7)	−8.1
H180(TM2)	S163(TM1)	−3.5
EC loop couplings		
R299(EC2)	D222(EC1)	−46.2
W297(EC2)	D222(EC1)	−6.1
C296(EC2)	D222(EC1)	−5.5
H374(EC3)	M303(EC2)	−6.4

We did not include interactions between residues within the same TM or between a TM and its adjoining loop. Energies are in kilocalories per mole and shown simply to give an idea of relative strength. Red indicates that the residue is fully conserved among class B1 GPCRs, green indicates that it is partially conserved, and blue means that it is conserved among some class B1 GPCRs.

Table S4. Full unified cavity analysis of the GLP1R/Exe4 binding site

GLP1R residue	Location	VDW	Coulomb	Hbond	Total
K202	EC1	4.69	−61.5	−7.98	−64.79
K197	2	6.15	−53.75	−6.01	−53.6
K383	7	4.41	−49.6	−5.58	−50.77
R134	N	10.78	−49.19	−12.24	−50.65
R40	N	3.77	−44.13	−6.39	−46.74
R121	N	−1.33	−34.39	−0.03	−35.74
K38	N	−0.25	−31.8	0	−32.05
R190	2	2.7	−33.79	0	−31.09
R43	N	−0.44	−26.51	0	−26.95
R227	EC1	−0.15	−26.75	0	−26.91
K113	N	−0.14	−24.39	0	−24.53
R44	N	−0.04	−24.46	0	−24.5
K130	N	−0.07	−22.17	0	−22.24
R102	N	−0.04	−20.61	0	−20.65
R299	EC2	−0.04	−20.33	0	−20.37
K288	4	−0.03	−20.18	0	−20.21
R376	EC3	−0.15	−19.56	0	−19.71
R131	N	−0.02	−18.75	0	−18.78
R380	EC3	−0.17	−17.68	0	−17.85
R48	N	−0.01	−16.68	0	−16.69
R64	N	−0.01	−14.67	0	−14.68
R310	5	0	−12.79	0	−12.79
R348	IC3	0	−12.56	0	−12.56
Q210	EC1	−3.71	−4.88	−3.61	−12.19
K351	6	0	−11.96	0	−11.96
R170	1	0	−11.68	0	−11.68
R176	2	0	−10.81	0	−10.81
R326	5	0	−10.05	0	−10.05
R267	4	0	−9.59	0	−9.59
K346	IC3	0	−9.26	0	−9.26
K342	IC3	0	−9.18	0	−9.18
K334	IC3	0	−0.11	0	−9.11
K336	IC3	0	−8.52	0	−8.52
L32	N	−7.12	−1.3	0	−8.42
W203	EC1	−7.38	0.01	0	−7.37
W214	EC1	−6.33	−0.8	0	−7.13
S31	N	−2.18	−4.88	0	−7.06
M204	EC1	−6.64	−0.35	0	−6.99
W39	N	−5.33	−1.3	0	−6.64
H212	EC1	−4.2	−2.11	0	−6.31
T35	N	−3.06	−1.87	0	−4.92
P90	N	−3.25	−0.9	0	−4.15
W91	N	−3.85	−0.09	0	−3.95
V36	N	−2.49	−1.2	0	−3.69
C236	3	−1.02	−2.6	0	−3.62
Q213	EC1	−0.72	−2.89	0	−3.61
L89	N	−2.35	−1.2	0	−3.55
Q211	EC1	−3.57	0.56	−0.49	−3.5
T29	N	−3.03	−0.41	0	−3.43
L232	3	−2.87	−0.51	0	−3.38
S193	2	−1.08	−2.15	0	−3.23
T391	7	−0.61	−1.33	−1.36	−3.2
Y69	N	−3.63	0.68	0	−2.95
Y205	EC1	−3.5	0.91	0	−2.59
T149	1	−1.79	−0.65	−0.05	−2.49
V229	3	−2.12	−0.2	0	−2.32
Q37	N	−0.38	−1.79	0	−2.17
L228	3	−1.43	−0.51	0	−1.93
L217	EC1	−1.07	−0.8	0	−1.87
F390	7	−1.03	−0.74	0	−1.77
A153	1	−0.43	−1.32	0	−1.76
L384	7	−1.46	−0.29	0	−1.75

Table S4. Cont.

GLP1R residue	Location	VDW	Coulomb	Hbond	Total
F143	N	−1.23	−0.49	0	−1.72
S206	EC1	−1.55	−0.06	0	−1.62
W33	N	−0.83	−0.77	0	−1.6
A209	EC1	−0.74	−0.85	0	−1.59
L123	N	−1.61	0.04	0	−1.57
Y152	1	−0.65	−0.83	0	−1.47
Y88	N	−3.13	1.74	0	−1.39
M233	3	−2.38	1.07	0	−1.31
A239	3	−0.08	−1.13	0	−1.21
Y42	N	−0.16	−1	0	−1.17
L118	N	−0.15	−0.99	0	−1.15
V194	2	−1.12	−0.02	0	−1.14
Y235	3	−0.34	−0.77	0	−1.12
Q221	EC1	−0.18	−0.91	0	−1.09
E127	N	−1.29	14.62	−0.65	12.69
E139	N	5.11	16.23	−5.71	15.63
E128	N	11.2	16.4	−10.48	17.12
E387	7	2.87	36.66	−5.47	17.12

All interactions that are over −1 kcal/mol or include a hydrogen bond are shown. Energies are in kilocalories per mole.

Table S5. Top 10 hydrophobic interactions between GLP1R and Exe4

GLP1R residue	Location	Strongest interacting Exe4 residue	VDW	Coulomb	Hbond	Total	Experimental study?
L32	N	V18, V19, F22, P36	−7.12	−1.3	0	−8.42	Mutation and crystal
W203	EC1	L10	−7.38	0.01	0	−7.37	
W214	EC1	M14	−6.33	−0.8	0	−7.13	
M204	EC1	M14	−6.64	−0.35	0	−6.99	Mutation
W39	N	F22, L26, P31	−5.33	−1.3	0	−6.64	Crystal
P90	N	I23	−3.25	−0.9	0	−4.15	Mutation and crystal
W91	N	I23	−3.85	−0.09	0	−3.95	Crystal
Y69	N	L26	−3.63	0.68	0	−2.95	Mutation and crystal
Y205	EC1	L10, M14	−3.5	0.91	0	−2.59	Mutation
Y88	N	F22	−3.13	1.74	0	−1.39	Mutation and crystal

All energies are given in kilocalories per mole and summed over the GLP1R residue interactions. The residues are ordered by their total interaction energy.

Table S6. Residues in the binding site with their mutation data

GLP1R residue	Location	Experiment	Ligand interaction	Type of interaction
L32	N	L32A(41), crystal (37)	V18, V19, F22, P36	Hydrophobic
T35	N	T35A(41), crystal (37)	V19, F22	Hydrophobic
V36	N	V36A(41), crystal (37)	F22	Hydrophobic
W39	N	Crystal (37)	F22, L26, P31	Hydrophobic
Y69	N	Y69A(41), crystal (37)	L26	Hydrophobic
Y88	N	Y88A(41), crystal (37)	F22	Hydrophobic
L89	N	L89A(41), crystal (37)	I23	Hydrophobic
P90	N	P90A(41), crystal (37)	I23	Hydrophobic
W91	N	Crystal (37)	I23	Hydrophobic
R121	N	R121A(41), crystal (37)	K27	Hbond
L123	N	L123A(41), crystal (37)	I23	Hydrophobic
E127	N	E127A/E127Q(41), crystal (37)	K27	Hbond
E128	N	E128A/E128Q(41), crystal (37)	R20	Hbond
T149	1	T149M(42)	H1	Hbond
Y152	1	Y152A(43)	E3	Polar
R190	2	R190A(43)	E3	Polar
K197	2	K197A(43)	E3	Hbond
M204	EC1	M204A/Y205A (35)	M14	Hydrophobic
Y205	EC1	M204A/Y205A (35)	L10, M14	Hydrophobic
Y235	3	Y235A(43)	E3	Polar
K288	4	K288A(44)	E3	Polar
R310	5	R310A(43)	E3	Polar
E387	7	E387A(43)	H1	Hbond
T391	7	T391A(43)	H1	Hbond

All energies are given in kilocalories per mole.

Table S7. Suggested mutations for GLP1R or Exe4

Mutation	Location	Motivation	Change in interaction energy
Q213K	Protein	New Hbond with D9	−49.4
S11W	Ligand	New hydrophobic interaction with W214	−1.8
WT	WT	WT	0
W214Y	Protein	Lose M14 hydrophobic interaction	1.8
W203Y	Protein	Lose L10 hydrophobic interaction	3
M14Q	Ligand	Lose hydrophobic interactions	4.6
W203N	Protein	Lose L10 hydrophobic interaction	4.8
W214N	Protein	Lose M14 hydrophobic interaction	7.5
K383Q	Protein	Lose E3 Hbond	45.2
K383A	Protein	Lose E3 Hbond	57.3
K202Q	Protein	Lose E17 salt Bridge	60.2
R134A	Protein	Lose E16 salt bridge	60.7
R134Q	Protein	Lose E16 salt bridge	61.9
K202A	Protein	Lose E17 salt bridge or Hbond	68.2
K12A	Ligand	Lose E139 salt bridge	80.4

All energies are given in kilocalories per mole. Yellow row corresponds to wild-type.