J Mol Model (2009) 15:53–65
DOI 10.1007/s00894-008-0372-2

ORIGINAL PAPER

Molecular modeling of the three-dimensional structure of GLP-1R and its interactions with several agonists

Fu Lin · Renxiao Wang

Received: 3 July 2008 / Accepted: 22 September 2008 / Published online: 22 October 2008

Abstract Glucagon-like peptide-1 receptor (GLP-1R) is a promising molecular target for developing drugs treating type 2 diabetes. We have predicted the complete three-dimensional structure of GLP-1R and the binding modes of several GLP-1R agonists, including GLP-1, Boc5, and Cpd1, through a combination of homology modeling, molecular docking, and long-time molecular dynamics simulation on a lipid bilayer. Our model can reasonably interpret the results of a number of mutation experiments regarding GLP-1R as well as the successful modification to GLP-1 by Liraglutide. Our model is also validated by a recently revealed crystal structure of the extracellular domain of GLP-1R. An activation mechanism of GLP-1R agonists is proposed based on the principal component analysis and normal mode analysis on our predicted GLP-1R structure. Before the complete structure of GLP-1R is determined through experimental means, our model may serve as a valuable reference for characterizing the interactions between GLP-1R and its agonists.

Keywords GLP-1R · GLP-1R agonist · Homology modeling · Molecular docking · Molecular dynamics

Introduction

Glucagon-like peptide-1 receptor (GLP-1R) is one of the class-B G-protein-coupled receptors (GPCR). It has received intensive studies in recent years as a promising molecular target for developing drugs treating type 2 diabetes [1]. The intrinsic agonist of GLP-1R is glucagon-like peptide-1 (GLP-1), a peptide consisting of 30 amino acid residues (Fig. 1). GLP-1 is secreted mainly by intestinal L cell as a gut hormone, which can induce glucose-dependent insulin secretion and regulate glucagon secretion simultaneously [2]. It however cannot be used as an effective therapy since it has a short half-time of less than two minutes in body due to the rapid degradation by dipeptidyl peptidase-4 [3]. A derivative of GLP-1 developed by Novo Nordisk Ltd., *i.e.*, Liraglutide (Fig. 1), exhibits all of the biological effects of GLP-1 but has a much longer pharmacokinetic half-life around 12 hours in human body [4]. The effects of Liraglutide on blood glucose level last for 24 hours following one injection per day. It is expected to reach market in the middle of 2008.

Some small-molecule agonists of GLP-1R have also been reported recently. Boc5 (Fig. 1) was discovered by Chen et al. [5] from the screening of a compound library against a cell line stably cotransfected with GLP-1R and cAMP responsive reporter. Boc5 was found to increase the intracellular level of cAMP. Such biological response could be blocked by exendin-4, a peptidic antagonist of GLP-1R. Based on this observation, Boc5 was proposed to act on the orthosteric site of GLP-1R, a mechanism similar to that of GLP-1. In contrast, another small-molecule agonist reported by Knudsen et al. [6], namely Cpd1 (Fig. 1) was proposed to act on an allosteric site on GLP-1R since exendin-4 did not antagonize the biological effect of Cpd1. It is amazing that Cpd1 acts as both allosteric activators of the receptor and independent agonists. Potency of GLP-1 was not changed by the allosteric agonists, but affinity of GLP-1 for the receptor was increased.

F. Lin · R. Wang (✉)
State Key Laboratory of Bioorganic Chemistry,
Shanghai Institute of Organic Chemistry,
Chinese Academy of Sciences,
354 Fenglin Road,
Shanghai 200032, People's Republic of China
e-mail: wangrx@mail.sioc.ac.cn

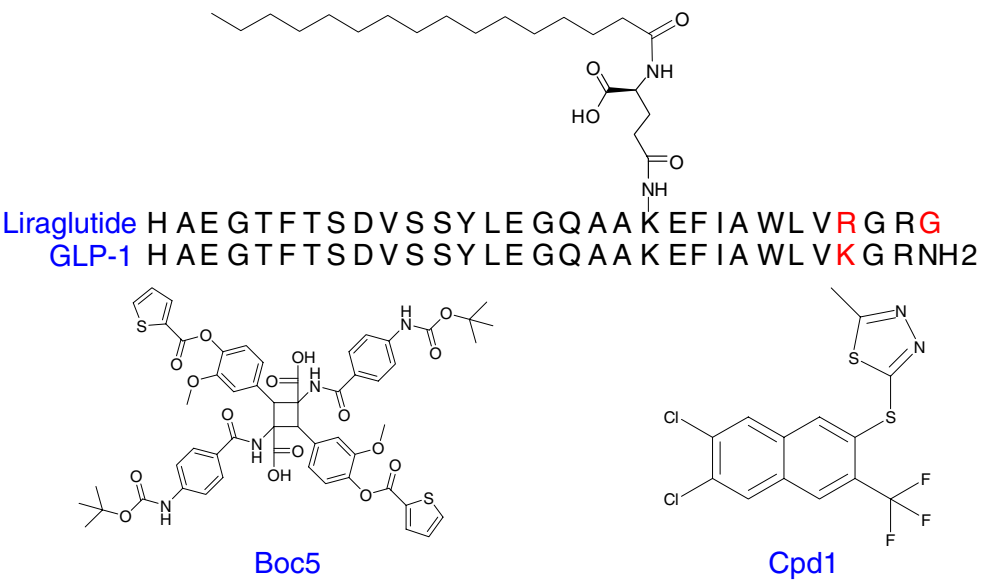

Fig. 1 Chemical structures of several known GLP-1R agonists

Since the complete three-dimensional structure of GLP-1R has not been revealed yet, how these agonists bind to and activate GLP-1R is unclear. This remains as a major obstacle for the further development of potent GLP-1R agonists of pharmaceutical interests. To provide a solution to this problem, we have predicted the three-dimensional structure of *r*GLP-1R and used it to study the binding of GLP-1, Boc5, and Cpd1 through molecular modeling. The initial models of GLP-1R-agonist complexes are constructed by homology modeling and molecular docking. They are then subjected to long-time molecular dynamics (MD) simulations for refinement. The MD simulations are conducted on a hydrated lipid bilayer to mimic the real environment of GLP-1R. Our predicted model can reasonably interpret the results of a number of mutation experiments on GLP-1R and GLP-1. It can also explain the successful modification on GLP-1 made by Liraglutide. Interestingly, our model agrees well with a recently released crystal structure of the extracellular domain of GLP-1R. Details of the computational procedures employed in our study are given in the following sections.

Methods

Homology modeling of the GLP-1R/GLP-1 complex structure

The motivation of our study largely came from the discovery of Boc5, a novel small-molecule GLP-1R agonist [5]. Biological activities of Boc5 were determined based on *rat* GLP-1R. Accordingly, we chose to construct the structural model of *r*GLP-1R to interpret the interactions between GLP-1R and Boc5 more precisely. The amino acid sequence of *r*GLP-1R used in our study was retrieved from PubMed (access ID = P32301). It has a high identity of 90.9% (similarity = 95.9%) with the one of *human* GLP-1R. We thus assume that the binding modes of GLP-1R agonists obtained based on the three-dimensional structures of *r*GLP-1R can be applied to *h*GLP-1R as well. The difference between *r*GLP-1R and *h*GLP-1R is ignored in all of the following steps.

Locations of the extracellular N-terminal domain and the transmembrane domain on the sequence of GLP-1R were predicted by using the PSI-PRED Server [7]. The extracellular domain was predicted to be residues 1–146; while the transmembrane domain was predicted to be residues 147–463. This prediction is basically consistent with the results reported by Xiao et al. in a previous study [8].

The extracellular domain of GLP-1R was modeled through homology modeling mainly by using the known structure of CRF2R (PDB entry 2JND) as template. CRF2R is also one of the class-B GPCRs, which has a sequence identity of 34.5% (sequence similarity = 53.4%) to the extracellular domain of GLP-1R. However, CRF2R is an appropriate template only for modeling the segment between residues 61–117. Several other templates were chosen for modeling other segments on the extracellular domain of GLP-1R, which are summarized in Table 1. Sequence alignment of the extracellular domain of GLP-1R and all of these templates are shown in Fig. 2. All of the templates used in our homology modeling were selected from the Protein Data Bank [9] through sequence similarity searches with the FASTA algorithm [10]. The Modeler module [11] in the Discovery Studio software [12] was employed to generate a total of 30 structural models based on the known three-dimensional structures of these templates. Note that the extracellular N-terminal domain of class-B GPCRs is characterized by three disulfide bonds formed between six conservative cysteine residues [13],

Table 1 Summary of the templates used in the homology modeling of the extracellular N-terminal domain of GLP-1R

Residues on GLP-1R	Templates used in homology modeling			
	PDB code of the template	Residues on the template	Sequence identity	Sequence similarity
1–20	1HZP	220–239	35.0%	60.0%
26–54	2AA2	94–122	41.4%	55.2%
61–117	2JND	45–102	34.5%	53.4%
77–142	2GSX	411–467	28.4%	47.8%

i.e., Cys46-Cys71, Cys62-Cys104, and Cys85-Cys126. Accordingly, these three disulfide bonds were set as constraints in the generation of homology models. All of the resulting models were refined through high-level optimizations by Modeler. They were then visually examined to exclude those models which did not possess the unique structural features of class-B GPCRs. Models on which key residues for agonist binding (as identified in mutation experiments [14–21]) were remote from the agonist binding site were also excluded. The probability density function (PDF) energies computed by Modeler were then adopted as an objective criterion for assessing the quality of the remaining models. The model with the lowest PDF energy was selected for further analysis.

The transmembrane domain of GLP-1R is composed of seven transmembrane helices (TMs). We employed a number of computational methods to predict the locations of these TMs on the GLP-1R sequence. Most of them produced consistent predictions (Table 2) [22–34]. These seven TMs were then modeled by using the known structure of rhodopsin (PDB entry 1U19) as template. The extracellular loops on the transmembrane domain of GLP-1R are relatively long and therefore were also modeled based on appropriate templates. For extracellular loop 1 (residues 211–225), PDB entry 1WMU (residues 11–25) (sequence identity = 40.0%, sequence similarity = 73.3%) was chosen as the template; while for extracellular loop 2 (residues 287–304), PDB entry 1AHP (residues 755–772) (sequence identity = 44.4%, sequence similarity = 61.1%) was chosen as the template. Again, these templates were selected throughout the entire PDB according to the sequence similarity computed by the FASTA algorithm [10]. No qualified template was found for extracellular loop 3 (residues 372–384) though. Thus, this loop, as well as all intracellular loops on the transmembrane domain, was constructed from scratch by using Modeler. Based on these templates, a total of 30 models of the complete transmembrane domain, including all extracellular and intracellular loops, were generated and refined by high-level optimizations by using Modeler. During this process, the disulfide bond between Cys226 and Cys296, which is arguably a conserved feature of class-B GPCRs [35], was set as constraint. All of the resulting models were then visually inspected to exclude those having crossing loop regions or serious internal collisions. Models on which the key residues for GLP-1 binding were remote from the agonist binding site were also excluded. Among the remaining models, the one with the lowest probability density function (PDF) energy computed by Modeler was chosen for further refinement.

Finally, the three basic building blocks, including the extracellular N-terminal domain, the transmembrane domain, and GLP-1, were assembled to produce a complete model of the GLP-1/GLP-1R complex. The structure of GLP-1 in trifluoroethanol/water solution has been determined by NMR techniques (PDB entry 1D0R). A total of

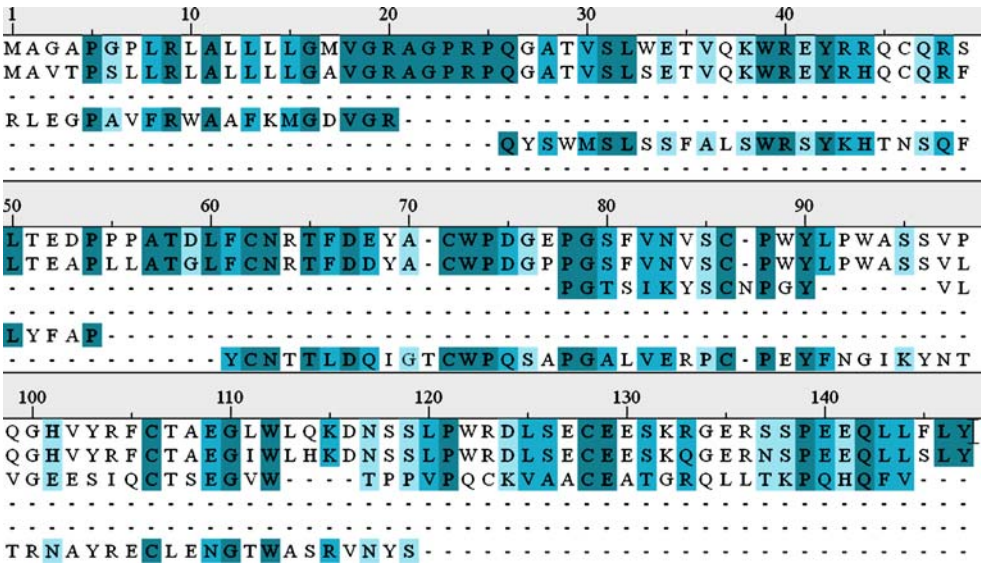

Fig. 2 Sequence alignment of the extracellular domain of *r*GLP-1R with four templates: First line = *Human* GLP-1R (residues 1–145); Second line = *Rat* GLP-1R (residues 1–145); Third line = PDB entry 2GSX (residues 411–467); Fourth line = PDB entry 1HZP (residues 220–239); Fifth line = PDB entry 2AA2 (residues 94–122); Sixth line = PDB entry 2JND (residues 45–102)

Table 2 Predicted locations of the trans-membrane helices (TMs) on GLP-1R

Method	N [a]	TM1	TM2	TM3	TM4	TM5	TM6	TM7
MEMSAT3 [17]	7	143–170 (28) [b]	179–198 (20)	226–253 (28)	269–286 (18)	305–332 (28)	350–369 (20)	381–408 (28)
TMHMM [18]	7	145–167 (23)	179–196 (18)	239–261 (23)	268–287 (20)	307–329 (23)	349–371 (23)	384–406 (23)
DAS [19]	7	144–167 (24)	182–195 (14)	226–257 (32)	270–287 (18)	307–334 (28)	354–367 (14)	395–404 (10)
SMART [20]	7	145–167 (23)	179–196 (18)	239–261 (23)	268–287 (20)	307–329 (23)	349–371 (23)	384–406 (23)
ConPred II [21]	7	146–166 (21)	180–200 (21)	242–262 (21)	268–288 (21)	312–332 (21)	352–372 (21)	388–408 (21)
HMMTOP [22]	7	144–162 (19)	183–201 (19)	228–247 (20)	268–287 (20)	314–333 (20)	354–371 (18)	388–405 (18)
TopPred [23]	7	149–169 (21)	177–197 (21)	241–261 (21)	267–287 (21)	313–333 (21)	352–372 (21)	386–406 (21)
HMM-TM [24]	7	145–167 (23)	178–196 (19)	241–257 (17)	267–287 (21)	307–333 (27)	354–371 (17)	379–405 (27)
SOSUI [25]	6	146–168 (23)	179–201 (22)	242–264 (23)	267–289 (23)	310–332 (23)	—	390–412 (23)
TMAP [26]	6	145–170 (26)	176–204 (29)	232–252 (31)	266–286 (21)	303–329 (27)	—	384–407 (24)
waveTM [27]	5	144–167 (24)	—	228–257 (30)	268–287 (20)	305–333 (29)	—	388–405 (18)
BPROMPT [28]	6	147–167 (21)	185–193 (9)	246–257 (12)	273–285 (13)	311–328 (18)	—	397–403 (7)
SPLIT 4.0 [29]	7	141–170 (30)	176–199 (24)	225–260 (36)	265–289 (25)	303–335 (33)	352–373 (22)	390–406 (17)

(a) Predicted number of TMs.
(b) Numbers in brackets are the lengths of TMs.

20 models are presented in this PDB entry, which can be classified into two categories (Fig. 3a): one is basically a straight helix; while the other is an L-shaped helix with a turn in the middle. Both structures consist of a short random coil (residues 1–7) near the N-terminal. We assume that the binding pose of GLP-1 is one of these two forms. Donnelly et al. [36] demonstrated that Asp198 on the top of TM2 on GLP-1R is essential for the binding of N-terminal of GLP-1. Mutation experiments on Glucagon receptor and Glucagon [37] suggested the same (GLP-1R is homologous to Glucagon receptor with a sequence identity of 45.7% and a sequence similarity of 65.7%). Based on such information, two possible models of the GLP-1/GLP-1R complex were manually assembled (Fig. 3b). Technically, this task was done by anchoring GLP-1 on the top of the transmembrane domain first, and then docking the N-terminal domain to achieve a tightest fit to GLP-1.

On *Model A*, GLP-1 is almost completely buried inside the cavity between the N-terminal domain and the transmembrane domain; while on *Model B*, the N-terminal of GLP-1 is buried inside the same cavity, but the C-terminal is not, due to the turn in the middle of its structure. Some other experimental proofs [38] indicate that the C-terminal of GLP-1 interacts only with the extracellular N-terminal domain of GLP-1R, prompting that *Model B* is more reasonable. Moreover, as indicated by the NMR structure of GLP-1, the C-terminal of GLP-1 exhibits the typical characteristics of an alpha-helix: a hydrophobic side and a hydrophilic side. Mutation experiments on GLP-1 [39, 40] indicate that mutation of the residues on the hydrophobic side causes a loss in binding affinity, while mutation of the residues on the hydrophilic face does not. In *Model B*, the hydrophobic face of the C-terminal of GLP-1 interacts with the N-terminal domain of GLP-1R;

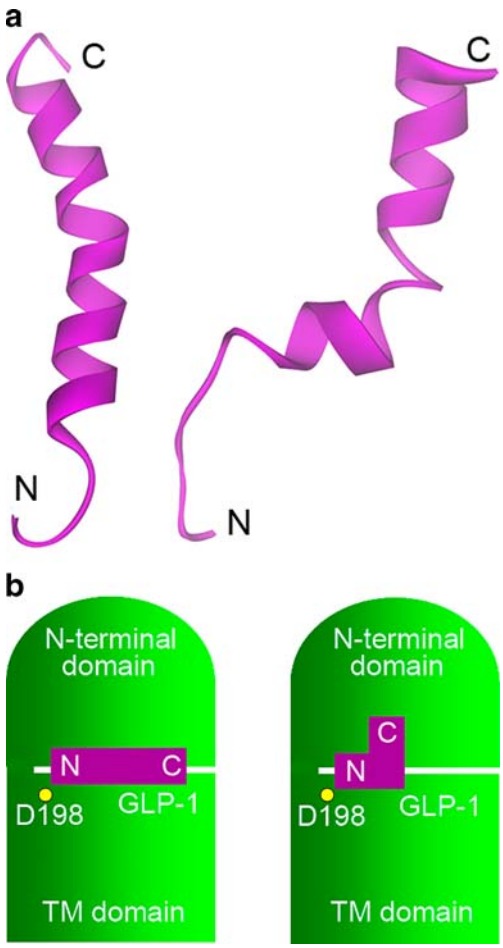

Fig. 3 (*A*) GLP-1 may take two major forms in trifluoroethanol/water solution (PDB entry 1D0R): a straight helix (left) or an L-shaped helix with a knot in the middle (right). (*B*) Illustration of the two corresponding models of the GLP-1/GLP-1R complex

while its hydrophilic side is exposed to solvent. We thus concluded that *Model B* was a more reasonable choice than *Model A*. Only *Model B* was considered in all following analyses.

The PROCHECK program (version 3.5.4) [41] was then applied to check the stereochemical quality of this model. The Ramachandran plot produced by PROCHECK indicates that 80.2% of residues on our model are in the most favored regions (A, B, L), 13.3% of residues are in the additionally allowed regions (a, b, l, p), 4.0% of residues are in generally allowed regions (~a, ~b, ~l, ~p), and only 2.4% of residues are in the disallowed regions. These results indicate that our structural model of the GLP-1/GLP-1R complex is acceptable from the statistical aspect.

Molecular docking of Boc5 and Cpd1 to GLP-1R

The aim of this step is to produce the rough structural models of Boc5/GLP-1R and Cpd1/GLP-1R complexes for further refinements. Boc5 has four chiral centers in its chemical structure. Due to the internal symmetry in its chemical structure, Boc5 has a total number of five possible enantiomers (Fig. 4). It is reasonable to speculate that only one of them is the active form of Boc5. The precise stereochemistry of Boc5, however, was not determined by Chen et al. in their study [5]. To figure out the stereochemistry of Boc5, we constructed the molecular models of all five possible enantiomers, and optimized their structures at the HF/6-31G* level with the Gaussian 03 program [42]. Each enantiomer was then manually docked into the binding site of GLP-1R by using the Sybyl software [43] as a reference to define the desired binding site and also to provide a reasonable starting point for subsequent automatic docking. The manually docked binding pose of each enantiomer was then optimized *in situ* with the parameters of the Tripos force field. Gasteiger-Hückel partial charges were applied on the small-molecule side and the AMBER FF02 charges were applied on the protein side during this process.

The GOLD program (version 3.2) [44] was then employed to perform automated molecular docking in order to explore the possible binding poses of each enantiomer more systematically. The binding pose obtained through manual docking at the previous step was supplied as the starting configuration in each case. The active site was defined as the residues within 10 Å from the input binding pose. Other key parameters used in docking included: population size = 100; number of GA operations = 300,000; mutation rate = 95%; crossover rate = 95%; scoring function = GoldScore and Chemscore. During this process, the GLP-1R structure was kept fixed. A total of 30 top-ranked binding poses of each given enantiomer were output after the docking job was accomplished. These binding poses were then inspected one by one visually. After all of the binding poses of all five possible enantiomers of Boc5 were inspected, it turned out that only one particular enantiomer of Boc5, *i.e.*, Enantiomer 1 in Fig. 4, could fit into the given binding site on GLP-1R reasonably. The best-scored binding pose of this enantiomer, together with the GLP-1R structure, was supplied to subsequent molecular dynamics simulations for refinement.

A structural model of the Cpd1/GLP-1R complex was obtained through basically the same molecular docking procedure with the same set of parameters as described above. Cpd1 does not have a chiral center in its chemical structure, and thus it is not necessary to predict the active enantiomer of it.

Molecular dynamics simulations of GLP-1R complexes

The GLP-1R complexes derived through the previous steps, including GLP-1/GLP-1R, Boc5/GLP-1R, and Cpd1/GLP-1R were subjected to long-time molecular dynamics (MD) simulations on a lipid bilayer in explicit solvent for refinement. In addition, a structural model of GLP-1R in *apo*-form was obtained simply by removing GLP-1 from the GLP-1/GLP-1R complex. This structure was also subjected to the same MD simulation.

A fully hydrated lipid bilayer is necessary in such simulations to soak the transmembrane domain of GLP-1R. In order to construct such a model, we utilized a pre-optimized bilayer model of palmitoyloleylphosphatidylcholine (POPC) molecules, which is publicly available from Prof. Peter Tieleman's website (http://moose.bio.ucalgary.ca/). This bilayer model (a total of 128 POPC molecules), however, was too small to accommodate GLP-1R. We made four copies of this model and then assembled them together by 2×2 on the *xy* plane to produce a larger piece of POPC bilayer. We also extended the *z* dimension of the cubic MD box in order to create enough space to accommodate the extracellular N-terminal domain of GLP-1R in our subsequent simulations. The size of our final MD box was 119×117×159 Å. It contained a bilayer of 512 POPC molecules as well as 52380 water molecules. After full minimization, the whole system was subjected to a 2 *ns* MD simulation under NPT condition with the GROMACS (version 3.3) package [45]. During this process, the SPC water model [46] was applied; while the parameters for POPC molecules were taken from Berger's study [47]. The last configuration on the resulting MD trajectory was retrieved, minimized, and then used in subsequent MD simulations.

After the POPC bilayer was prepared, all of the four GLP-1R structures, *i.e.*, three complexes and one *apo*-form, were processed. In each case, the transmembrane domain of GLP-1R was inserted into the POPC bilayer from the perpendicular direction. All POPC and water molecules

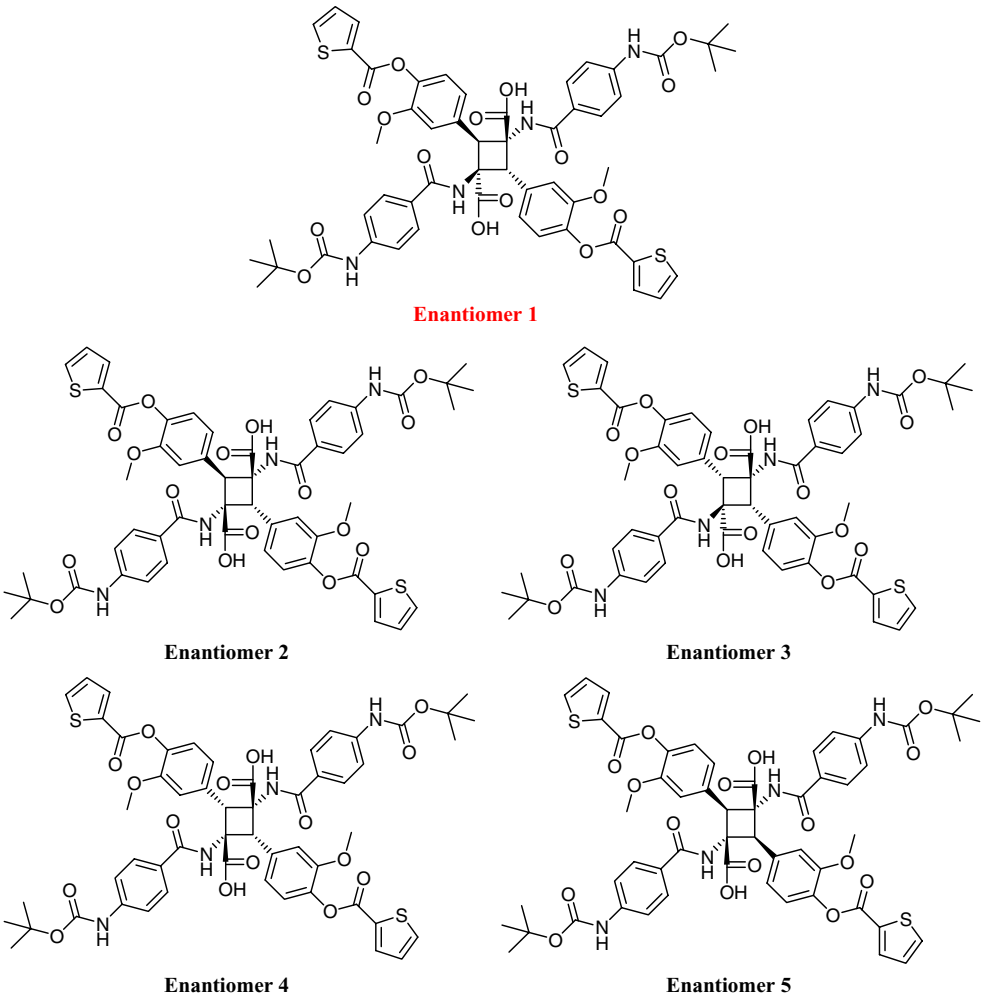

Fig. 4 Five possible enantiomers of Boc5. Enantiomer 1 is predicted to be the active enantiomer

which were in collision with GLP-1R or the agonist molecule were removed. The whole system was then subjected to a stepwise MD simulation as follows:

(a) Minimization of the POPC bilayer with restraints on the rest of the system to remove any large cavity in the POPC bilayer after the insertion of GLP-1R.
(b) Minimization of water molecules with restraints on the rest of the system to fully solvate the POPC bilayer and GLP-1R.
(c) Minimization of the whole system without any restraint to release internal repulsions.
(d) A 1 *ns* MD simulation under NVT condition with restraints on GLP-1R (restraint force constant = 1000 kJ $mol^{-1}nm^{-2}$) to speed the equilibration of the POPC bilayer and water molecules.
(e) A three-step MD simulation under NPT condition with restraints on the backbone of GLP-1R where the restraint force constant was set as 1000 kJ $mol^{-1}nm^{-2}$, 500 kJ $mol^{-1}nm^{-2}$, and 100 kJ $mol^{-1}nm^{-2}$, respectively. Each step was 0.5 *ns* long.
(f) A 4 *ns* MD simulation under NPT condition without any restraint. The Berendsen thermostat and barostat were enabled at this step to make the whole system reach equilibrium more rapidly.
(g) A 16 *ns* long MD simulation under NPT condition without any restraint. The Nose-Hoover thermostat and the Parrinello-Rahman barostat were enabled to obtain correct ensembles of the whole system.

The GROMACS (version 3.3) software [45] was used to perform all of the minimizations and MD simulations mentioned above. The following parameters were applied if not otherwise specified. The temperature was set to 310 *K*. The pressure was set to 1.0 *atm*, which was coupled on the *xy* plane and the *z* direction separately using the semi-isotropic pressure coupling method. The time interval was set to 2 *fs*. The particle mesh Ewald (PME) [48] interpolation order was set to 4. The non-bonded cutoff was set to 14.0 Å, and the lists of non-bonded pairs were updated every 10 steps. The SHAKE algorithm [49] was enabled on all chemical bonds with hydrogen atoms.

Table 3 Normal mode analysis returned by the elNémo server

Mode[a]	Frequency	Collectivity[b]	Cumulative overlap[c]	Amplitude (dq)
mode 7	1.00	0.6424	0.207	232.1179
mode 8	1.21	0.6299	0.226	77.2664
mode 9	1.65	0.6300	0.226	31.7299
mode 10	2.27	0.5440	0.226	2.5908
mode 11	2.37	0.5792	0.283	126.5490
mode 12	2.63	0.2996	0.377	−169.3851
mode 13	2.76	0.4459	0.678	289.9325
mode 14	2.89	0.5321	0.735	121.7800
mode 15	3.12	0.2169	0.754	−57.8483
mode 16	3.37	0.3765	0.792	−93.8932

(a) Only the 10 normal modes with lowest frequencies are displayed here.
(b) The level of collectivity indicates the percentage of residues that are involved in a certain normal mode.
(c) The level of overlap measures the similarity between a desired conformational change and that of a certain normal mode.

Snapshots of the structure under study were saved every 10 *ps* along the entire trajectory.

For each of the four GLP-1R structures, the last snapshot on the MD trajectory was retrieved and further minimized to full convergence. The final model in each case was then used in the interpretation of the binding mode of GLP-1R agonists as well as other analyses.

Normal mode analysis and principal component analysis on GLP-1R structure

Normal mode analysis (NMA) was conducted to analyze the intrinsic motions of the GLP-1R structure. The elNémo online server (http://igs-server.cnrs-mrs.fr/elnemo/index.html) [50] was employed in our study for this purpose. This server is part of the Elastic Network Model, which provides a fast simple tool to compute, visualize and analyze low-frequency normal modes of biological macromolecules. The structural model of GLP-1R in *apo*-form was submitted for NMA analysis; while the GLP-1/GLP-1R complex was submitted as a reference of the desired conformational change. Key parameters used in computation included: DQMIN=-100, DQMAX=100, DQSTEP=20, and NRBL = "*auto*". The default cutoff of eight residues was used to identify elastic interaction ranges. A total of 100 normal modes with the lowest frequencies were requested. Essential features of the top ten low-frequency normal modes, including its frequency, amplitude, collectivity of atom movements, and the overlap with observed conformational changes are summarized in Table 3. The two lowest-frequency normal modes (*i.e.*, modes 7 and 8 in Table 3) are illustrated in Fig. 5.

In order to confirm the outcomes of NMA analysis, we also conducted the principal component analysis (PCA) on GLP-1R in *apo*-form to detect its major conformational motions during MD simulation. The PCA tool provided in the GROMACS package [45] as applied to the last 16 *ns* of the MD trajectory of GLP-1R in *apo*-form. Each snapshot of the GLP-1R structure was fitted to a reference configuration to remove translational and rotational motions. All adjusted snapshots were then used to form a covariance matrix. This matrix was solved to obtain eigenvalues and the corresponding eigenvectors. Projection of the MD trajectory onto two eigenvectors with the largest eigenvalues is plotted in Fig. 6.

Results and discussion

Validation of the predicted GLP-1R structure

As described in the Methods section, our initial model of the GLP-1/GLP-1R complex was constructed through homology

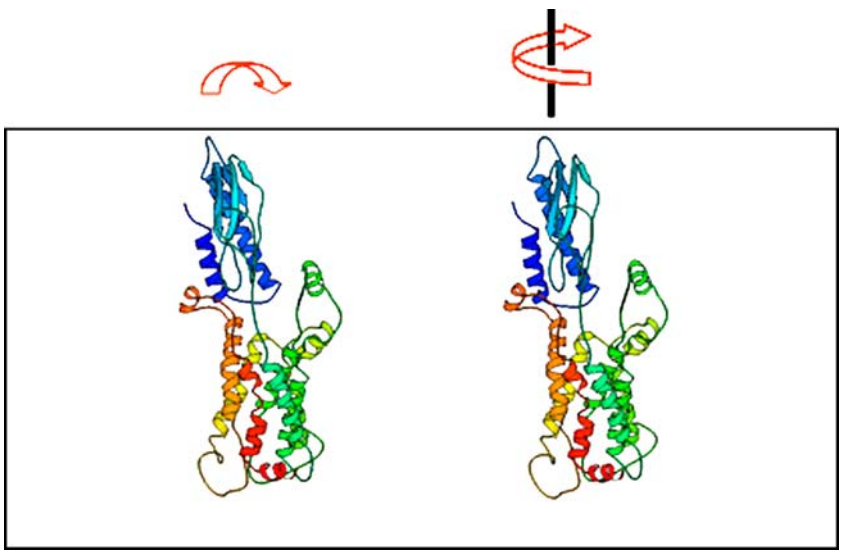

Fig. 5 The two normal modes of GLP-1R with the lowest frequencies (Left: bending of the N-terminal domain toward the transmembrane domain. Right: rotation of the N-terminal domain on the top of the transmembrane domain)

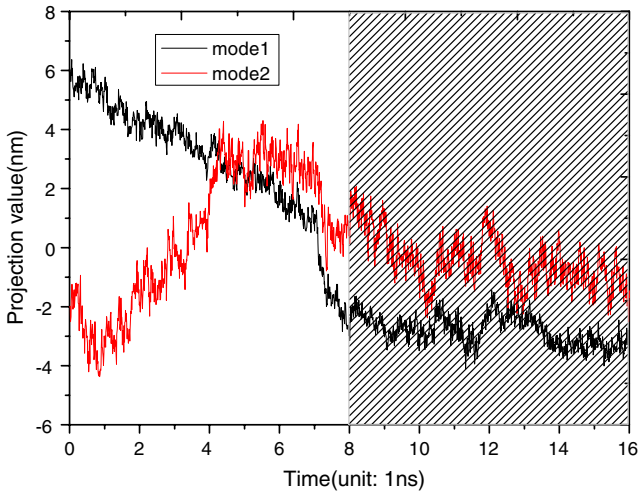

Fig. 6 The two most significant eigenvectors given by PCA along the MD trajectory of the *apo*-form GLP-1R. The *x* coordinate is simulation time; *y* is the value projected onto the two eigenvectors with the largest eigenvalues

modeling and molecular docking. It was then subjected to a long-time MD simulation on a hydrated POPC bilayer to mimic the real environment of GLP-1R as much as possible. The final model of the GLP-1/GLP-1R complex is shown together with the POPC bilayer in Fig. 7. GLP-1 is located in the orthosteric binding site on GLP-1R, which is enclosed by N-terminal domain and first extracellular loop (residues 198–231). Our model of the GLP-1/GLP-1R complex can explain

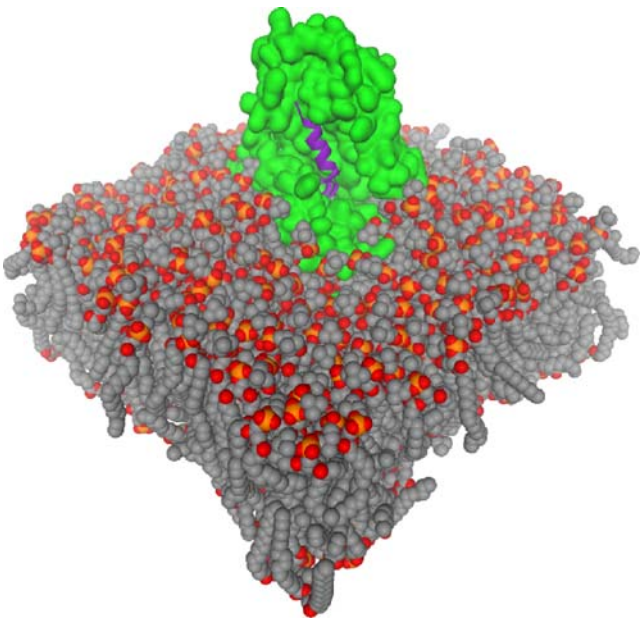

Fig. 7 Overview of the GLP-1/GLP-1R complex after 20 *ns* MD simulation on a hydrated POPC bilayer. GLP-1R (in green) and POPC molecule (in CPK color scheme) are rendered as space-filling models; while GLP-1 is rendered as purple ribbon. Water molecules are not shown in this figure

the roles of a number of key residues for the binding of GLP-1, which are identified by mutation experiments on GLP-1R and GLP-1 [8, 15, 16, 18–21]. In summary (Fig. 8):

(1) His1 on GLP-1 forms a hydrogen bond with Asp198 on GLP-1R [16].
(2) Phe6 on GLP-1 resides in a cavity on the top of several TM helices, forming favorable hydrophobic contacts with the surrounding residues on GLP-1R [18–20].
(3) Tyr13 on GLP-1 is involved in electrostatic interactions with Arg227 and Lys288 on GLP-1R [8, 15].
(4) Phe22, ILe23 and Leu26 on GLP-1 are in favorable hydrophobic contacts with Trp39 on GLP-1R [21].
(5) The primary role of Gly4, Asp9, Glu15, Gln17, Ala18 on GLP-1 is maintaining the correct binding pose of GLP-1 rather than forming any specific interactions with GLP-1R [18–20].

It is critical to validate a theoretical structural model with experimental proofs. Our model of the GLP-1/GLP-1R complex agrees well with the results obtained from these mutation experiments.

Our model of the GLP-1/GLP-1R complex can also explain the success of Liraglutide. Liraglutide is basically a derivative of GLP-1 by adding a long C_{16} hydrocarbon

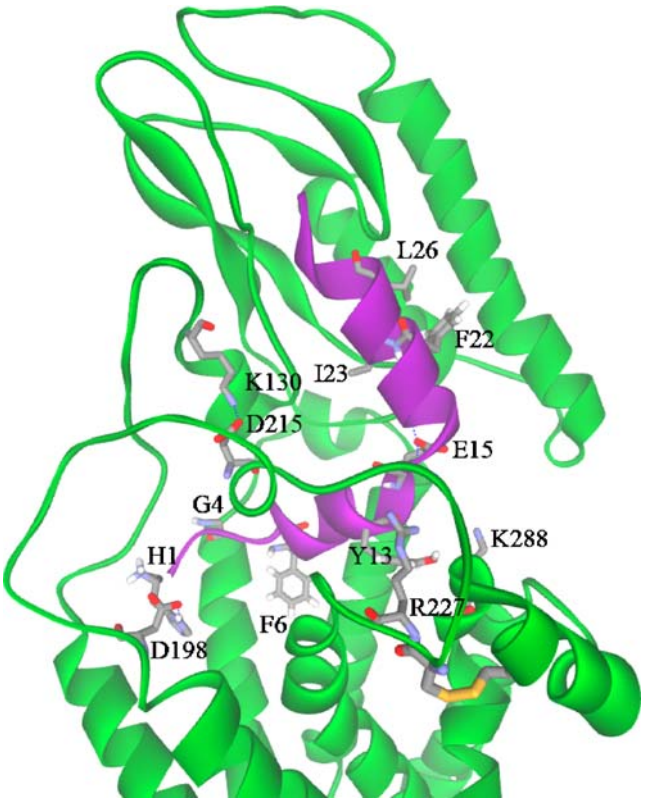

Fig. 8 Illustration of the key residues involved in the binding of GLP-1. The backbones of GLP-1R and GLP-1 are rendered as green and purple ribbons, respectively

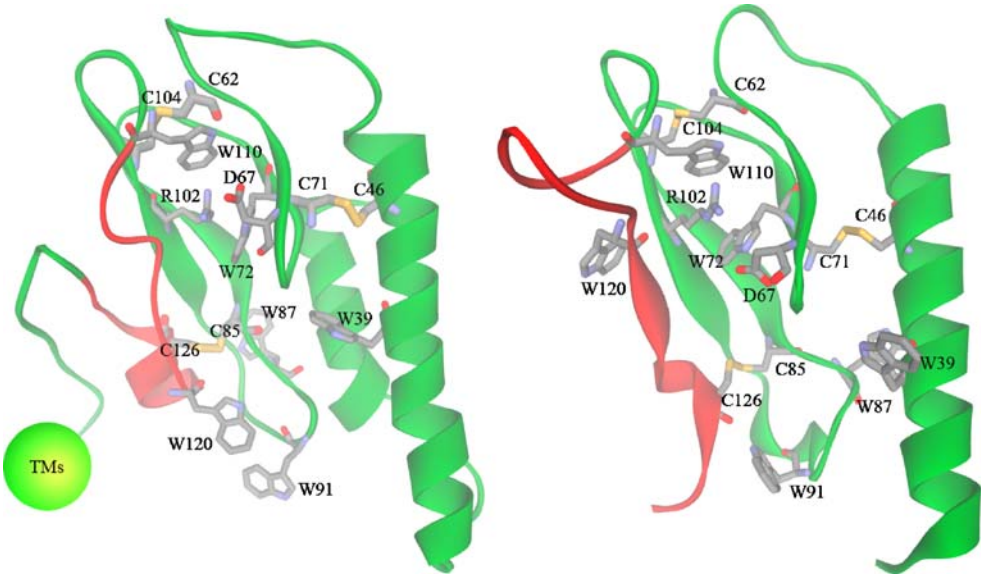

Fig. 9 Our predicted model of *r*GLP-1R (left) and the crystal structure of *h*GLP-1R recently revealed by Runge et al. (right). Only the extracellular domain (residues 1–146) on our model is shown in this figure. The only segment apparently different between our model and Runge's structure is colored in red

chain to the side chain of Lys20 on GLP-1 through a glutamic acid moiety as linker (Fig. 1). It is reasonable to assume that the binding mode of Liraglutide is basically the same as that of GLP-1. Our model of the GLP-1/GLP-1R complex indicates that the side chain of Lys20 on GLP-1 stretches out the binding site of GLP-1R rather than points inside. Thus, developing this side chain is not expected to disturb the binding of this peptidic agonist. This modification, however, apparently disturbs the degradation of Liraglutide by dipeptidyl peptidase-4. In addition, our model suggests that the remaining available carboxylic acid group on this glutamic acid linker forms a salt bridge with the nearby Arg28 on GLP-1R.

When our manuscript was in preparation, a crystal structure (PDB entry 3C5T, resolution = 2.10 Å) of the extracellular domain of *h*GLP-1R in complex with exendin-4 (residues 9–39) was revealed by Runge et al. [51]. This crystal structure of GLP-1R only contains the segment between Ala28 to Arg131 on the extracelluar domain, missing residues 1–27 and 132–146 as well as the entire transmembrane domain. Nevertheless, our predicted model can still be validated by a comparison with this structure. As one can see in Fig. 9, the overall tertiary structure of the extracellular domain in our model agrees well with Runge's crystal structure. The only major difference is observed at the loop region (residues 111–131) close to the joint between the extracelluar domain and the transmembrane domain. We suspect that presence of the rest parts of the extracelluar domain and the transmembrane domain in our model may be the origin of this observed difference. In addition, Runge's crystal was obtained in a solution of *n*-Decyl-β-D-thiomaltoside; whereas our model was refined by MD simulation on a hydrated lipid bilayer. A different media may have an impact on the three-dimensional structure of this loop region. Our model may resemble the structure of GLP-1R in reality more closely.

Besides the correct tertiary structure, it is encouraging to observe that our model also provides sufficient details of the GLP-1R structure. For example, Arg102 not only forms cation-π interaction with Trp72 and Trp110, but also forms a salt bridge with Asp67, which is a conserved feature of class-B GPCRs; Trp39, Cys85, Trp91, and Cys126 form a hydrophobic core; Trp87 locates at the other side of the extracellular domain and thus does not affect the binding of GLP-1 significantly. All of the above features agree with

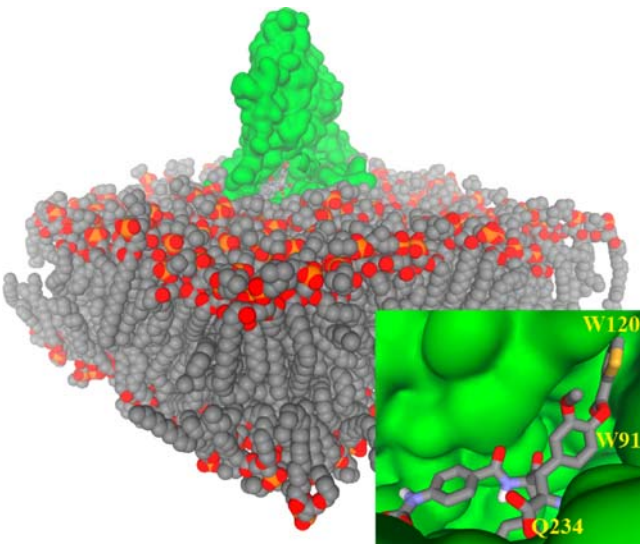

Fig. 10 Overview of the Boc5/GLP-1R complex after 20 *ns* MD simulation on a POPC bilayer in explicit water. In order to illustrate Boc5 inside the binding site of GLP-1R, the first extracellular loop (residues 198–231) on the transmembrane domain, which is in front of Boc5 from this viewing angle, is still rendered in ribbon in this figure

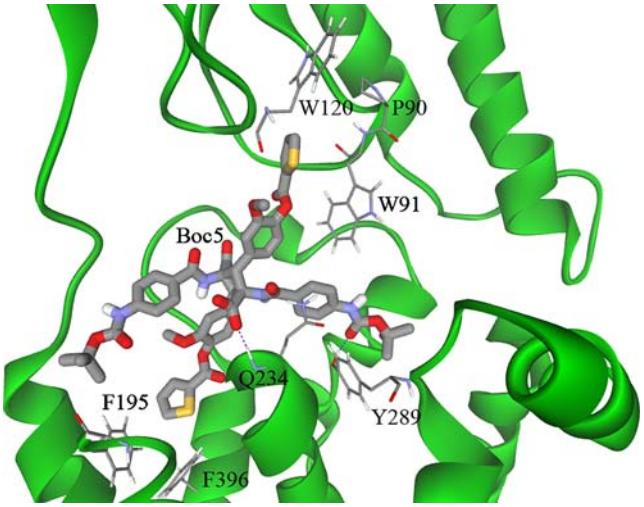

Fig. 11 Illustration of the interactions between Boc5 and GLP-1R. Boc5 forms hydrogen bonds with Q234 and Y289, which are indicated by dashed lines. One thiophene moiety on Boc5 resides in the hydrophobic cavity on the top of seven TM helices, mimicking the role of F6 on GLP-1. For the sake of clarity, the first extracellular loop on the transmembrane domain, which is in front of Boc5 from this viewing angle, is not shown in this figure

Runge' crystal structure. Nevertheless, there is one notable exception: Runge's structure shows that Trp120 is away from the agonist-binding site, which does not explain the complete loss of GLP-1 binding after a W120A mutation [21]. In contrast, our model indicates that Trp120 is part of the binding site for GLP-1 (Fig. 9), which is more reasonable for interpreting the critical role of this residue. In addition, our model indicates that the first extracellular loop on the transmembrane domain of GLP-1R is actually an indispensable part of the binding site for GLP-1. Thus, even if an experimental structure of the complete extracellular domain of GLP-1R becomes available in the future, it still may not be sufficient for unraveling the details in the binding of GLP-1R agonists.

Binding modes of small-molecule GLP-1R agonists

Our molecular docking of all five possible enantiomers of Boc5 revealed that only one of them was able to fit the given binding site (Fig. 10), suggesting that it is the active enantiomer of Boc5 accounting for its biological effects. This initial model of Boc5/GLP-1R complex was then refined by a 20 *ns* MD simulation on a hydrated POPC bilayer. Our final model of the Boc5/GLP-1R complex reveals that the key residues involved in the binding of Boc5 include Pro90, Trp91, Trp120, Trp214, Gln234, and Tyr289 (Fig. 11). In particular, one thiophene moiety on Boc5 resides in a hydrophobic cavity on the top of seven TM helices, mimicking the role of Phe6 on GLP-1. Boc5 also forms hydrogen bonds with Gln234 and Tyr289, both of which locate on the transmembrane domain rather than the extracelluar domain.

Cpd1 was reported as an allosteric agonist of GLP-1R [6]. Apart from the orthosteric agonist-binding site of GLP-1R, another cavity near helices TM5 and TM6 can be identified on our models of GLP-1R. The first and the second extracellular loops (residues 198–229 and 287–309, respectively) on the transmembrane domain are involved in the formation of this cavity. We found that Cpd1 can be reasonably docked into this cavity, suggesting that it is where Cpd1 carries out its function on GLP-1R (Fig. 12). Hydrophobic effect is obviously the driving force for the binding of Cpd1. The key residues involved in this matter include Tyr88, Trp91, Tyr289, and Val370. Our models of the Boc5/GLP-1R and Cpd1/GLP-1R complexes provide a useful guidance if any small-molecule GLP-1R agonists are to be developed based on Boc5 or Cpd1.

Activation mechanism of GLP-1R agonists based on NMA and PCA analysis

Structures of the GLP-1/GLP-1R complex and GLP-1R in *apo*-form after 20 *ns* MD refinements are compared in Fig. 13. The obvious difference is that one helix (residues Thr27-Leu50) partially blocks the binding site of GLP-1R in *apo*-form; while it undergoes an apparent movement upon the binding of GLP-1. This direct comparison prompts that the activation of GLP-1R by agonist binding is related to some intrinsic conformation motions of GLP-1 which occur during this process.

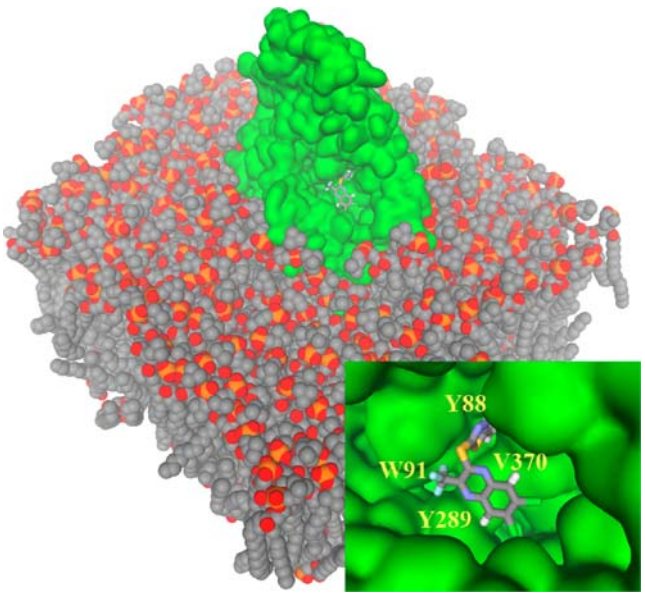

Fig. 12 Overview of the Cpd1/GLP-1R complex after 20 *ns* MD on a POPC bilayer in explicit water. This model suggests that the binding of Cpd1 is mainly through hydrophobic contacts with Y88, W91, Y289, and V370 on GLP-1R

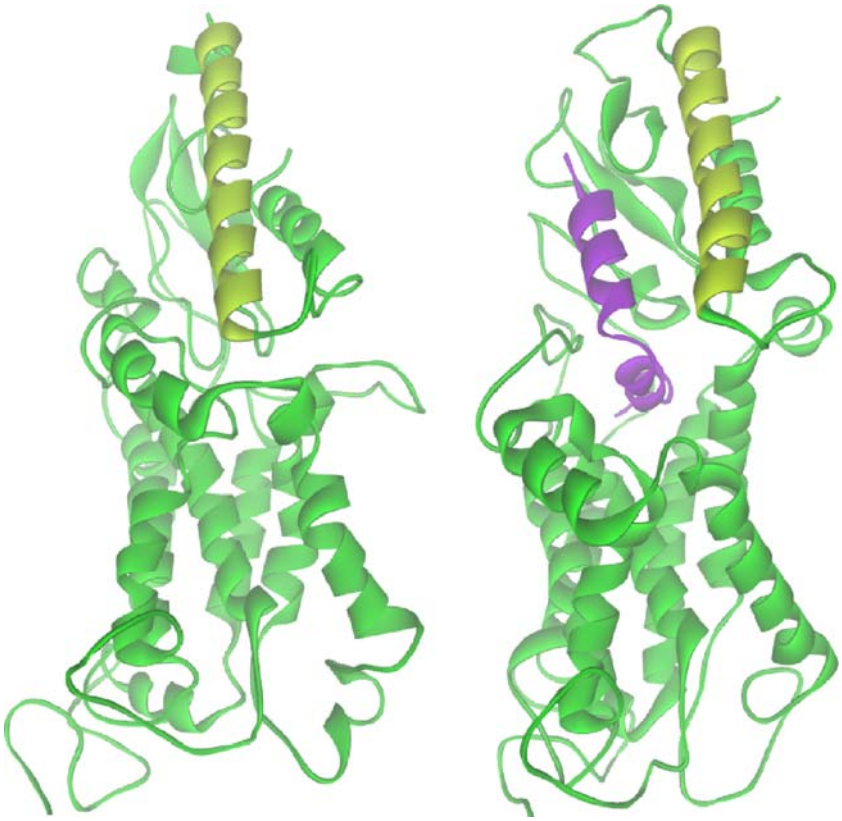

Fig. 13 The *apo*-form of GLP-1R (left) and the GLP-1/GLP-1R complex (right) after 20 *ns* MD simulation. One can see by a comparison of these two structures that the helix colored in yellow-green (residues Thr27-Leu50) partially blocks the binding site of GLP-1R in the *apo*-form, and it undergoes an apparent movement upon the binding of GLP-1 (in purple). The POPC bilayer is not presented in this figure for the sake of clarity

In order to detect such conformational motions, we performed normal mode analysis (NMA) on the GLP-1R structure in *apo*-form by using the structure in *holo*-form as reference. The two lowest-frequency normal modes (*i.e.*, modes 7 and 8 in Table 3) are illustrated in Fig. 5: one features bending of the N-terminal domain toward the transmembrane domain; the other features rotation of the N-terminal domain on the top of the transmembrane domain. Our principal component analysis (PCA) on the MD trajectory of the GLP-1R structure confirms that two distinct states can be identified regarding the two major motions during the transition from the *holo*-form to the *apo*-form (Fig. 13).

Based on the outcomes of NMA and PCA analyses, we can propose an activation mechanism of GLP-1R. There are two major states of the GLP-1R structure: one is the inactive state, in which the orthosteric agonist-binding site is partially blocked as the consequence of the relative motions between the N-terminal domain and the transmembrane domain; while the other is the active state, in which the orthosteric agonist-binding site is fully accessible (Fig. 14). These two states are connected by the low-

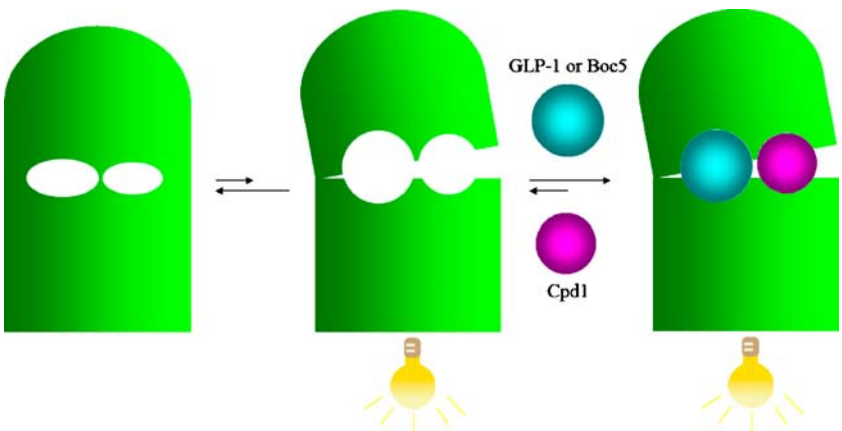

Fig. 14 The proposed activation mechanism of GLP-1R agonists: Binding of GLP-1, Boc5 or Cpd1 makes GLP-1R stay in the active state

frequency conformational motions. The equilibrium between these two states favors the former when no agonist is present. Binding of an agonist, as indicated by our long-time MD simulations of several GLP-1R complexes, makes GLP-1R stay basically in the latter state. This mechanism is received by GLP-1R as the activation signal for triggering down-stream biological processes. In particular, binding of Cpd1 at an allosteric site rigidizes the GLP-1R structure with an open binding site, which explains why binding of Cpd1 promotes the binding of GLP-1 [6].

Conclusions

We have obtained some reasonable structural models of GLP-1R in complex with several peptidic and small-molecule agonists, including GLP-1, Boc5, and Cpd1, through extensive molecular modeling. It is encouraging to find that our models agree well with a recently released crystal structure of the extracellular domain of GLP-1R. Our models, however, provide the complete structure of GLP-1R, including the extracellular domain and the transmembrane domain. Our predicted binding mode of GLP-1 can reasonably interpret the role of a number of key residues on GLP-1R and GLP-1 identified by mutation experiments. It can also explain the successful modification of GLP-1 by Liraglutide. An activation mechanism of GLP-1R agonists is proposed based on NMA and PCA analyses of our GLP-1R structure. Before the complete structure of GLP-1R is determined through experimental means, our model serves as a helpful reference for characterizing the interactions between GLP-1R and its agonists.

Acknowledgment The authors are grateful to the financial supports from the Chinese National Natural Science Foundation (grants 20502031 & 20772149), the Chinese Ministry of Science and Technology (grant 2006AA02Z337), and the Science and Technology Commission of Shanghai Municipality (grants 06PJ14115 & 074319113).

Supporting information available Three-dimensional structures of the GLP-1R models described in this manuscript are available from the authors upon request.

References

1. Murphy KG, Bloom SR (2007) Proc Natl Acad Sci USA 104:689–690. doi:10.1073/pnas.0610679104
2. Drucker DJ (2007) Endocrinology 142:521–527. doi:10.1210/en.142.2.521
3. Demuth HU, McIntosh CH, Pederson RA (2005) Biochim Biophys Acta 1751:33–44
4. Madsbad S, Schmitz O, Ranstam J, Jakobsen G, Matthews DR (2004) Diabetes Care 27:1335–1342. doi:10.2337/diacare.27.6.1335
5. Chen DS, Liao JY, Li N, Zhou CH, Liu Q, Wang MW et al. (2007) Proc Natl Acad Sci USA 104:943–948. doi:10.1073/pnas.0610173104
6. Knudsen LB, Kiel D, Teng M, Behrens C, Bhumralkar D, Kodra JT et al. (2007) Proc Natl Acad Sci USA 104:937–942. doi:10.1073/pnas.0605701104
7. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Nucleic Acids Res 33:36–38. doi:10.1093/nar/gki410
8. Xiao Q, Jeng W, Wheeler MB (2000) J Mol Endoc 25:321–335. doi:10.1677/jme.0.0250321
9. Philip Bourne TNB, Feng ZK, Gilliland G, Jain S, Ravichandran V, Schneider B et al. (2001) Nucleic Acids Res 29:214–218. doi:10.1093/nar/29.1.214
10. Pearson WR (1990) Methods Enzymol 183:63–98. doi:10.1016/0076-6879(90)83007-V
11. Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A (2000) Annu Rev Biophys Biomol Struct 29:291–325. doi:10.1146/annurev.biophys.29.1.291
12. Discovery Studio software (version 2.0) Accelyrs Inc. San Diego CA U.S.A. (2007)
13. Bazarsuren A, Grauschopf U, Wozny M, Reusch D, Hoffmann E, Schaefer W et al. (2002) Biophys Chem 96:305–318. doi:10.1016/S0301-4622(02)00023-6
14. Xiao Q, Jeng W, Wheeler MB (2000) J Mol Endoc. 25:321–335. doi:10.1677/jme.0.0250321
15. Suleiman AS, Donnelly D (2003) FEBS Lett 553:342–346. doi:10.1016/S0014-5793(03)01043-3
16. Maturana RL, Donnelly D (2002) FEBS Lett 530:244–248. doi:10.1016/S0014-5793(02)03492-0
17. Runge S, Gram C, Hans BO, Madsen K, Knudsen LB, Wulff BS (2003) J Biol Chem 278:28005–28010. doi:10.1074/jbc.M301085200
18. Pan CQ, Buxton JM, Yung SL, Tom I, Yang L, Chen HX et al. (2006) J Biol Chem 281:12506–12515. doi:10.1074/jbc.M600127200
19. Adelhorst K, Hedegaard BB, Knudsen LB, Kirks O (1994) J Biol Chem 269:6276–6278
20. Gallwitz B, Witt M, Paetzold G, Wortmann CM, Zimmermann B, Eckart K et al (1994) Eur J Biochem 225:1151–1156. doi:10.1111/j.1432-1033.1994.1151b.x
21. Wilmen A, Eyll BV, Goke B, Goke R (1997) Peptides 18:301–305. doi:10.1016/S0196-9781(96)00321-X
22. Jones DT, Taylor WR, Thornton JM (1994) Biochemistry 33:3038–3049. doi:10.1021/bi00176a037
23. Moller S, Croning MDR, Apweiler R (2001) Bioinformatics 17:646–653. doi:10.1093/bioinformatics/17.7.646
24. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Protein Eng 10:673–676. doi:10.1093/protein/10.6.673
25. Schultz J, Milpetz F, Bork P, Ponting CP (1998) Proc Natl Acad Sci USA 95:5857–5864. doi:10.1073/pnas.95.11.5857
26. Arai M, Mitsuke H, Ikeda M, Xia JX, Kikuchi T, Satake M et al. (2004) Nucleic Acids Res 32:390–393. doi:10.1093/nar/gkh380
27. Tusnady GE, Simon I (1998) J Mol Biol 283:489–506. doi:10.1006/jmbi.1998.2107
28. von Heijne G (1992) J Mol Biol 225:487–494. doi:10.1016/0022-2836(92)90934-C
29. Bagos PG, Liakopoulos TD, Hamodrakas SJ (2006) BMC Bioinform 7:189–206. doi:10.1186/1471-2105-7-189
30. Hirokawa T, Boon-Chieng S, Mitaku S (1998) Bioinformatics 14:378–379. doi:10.1093/bioinformatics/14.4.378
31. Persson B, Argos P (1997) J Protein Chem 16:453–457. doi:10.1023/A:1026353225758
32. Pashou EE, Litou ZI, Liakopoulos TD, Hamodrakas SJ (2004) In Silico Biol 4:0012–0012
33. Taylor PD, Attwood TK, Flower DR (2003) Nucleic Acids Res 31:3698–3700. doi:10.1093/nar/gkg554

34. Juretic D, Zoranic L, Zucic D (2002) J Chem Inf Comput Sci 42:620–632. doi:10.1021/ci010263s
35. Unson CG (2002) Biopolymers 66:218–235. doi:10.1002/bip.10259
36. Maturana RL, Donnelly D (2002) FEBS Lett 530:244–248. doi:10.1016/S0014-5793(02)03492-0
37. Runge S, Gram C, Brauner-Osborne H, Madsen K, Knudsen LB, Wulff BS (2003) J Biol Chem 278:28005–28010. doi:10.1074/jbc.M301085200
38. Maturana RL, Willshaw A, Kuntzsch A, Rudolph R, Donnelly D (2003) J Biol Chem 278:10195–10200. doi:10.1074/jbc.M212147200
39. Suleiman AS, Donnelly D (2003) Br J Pharmacol 140:339–346. doi:10.1038/sj.bjp.0705453
40. Gallwitz B, Witt M, Paetzold G, Wortmann CM, Zimmermann B, Eckart K et al. (1994) Eur J Biochem 225:1151–1156. doi:10.1111/j.1432-1033.1994.1151b.x
41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Cryst 26:283–291. doi:10.1107/S0021889892009944
42. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR et al. farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, V Rotiz J, Stefanov BB, Liu G, Liashenko A, Piskora P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, AlLaham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill MW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Gordon MH, Replogle ES, Pople JA Gaussian, Inc., Pittsburgh PA (2003)
43. The SYBYL software (version 7.2) Tripos Inc., St. Louis, Missouri (2006)
44. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267:727–748. doi:10.1006/jmbi.1996.0897
45. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) J Comput Chem 26:1701–1718. doi:10.1002/jcc.20291
46. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684–3690. doi:10.1063/1.448118
47. Berger O, Edholm O, Jähnig F (1997) Biophys J 72:2002–2013
48. Darden T, York D, Pedersen L (1993) J Chem Phys 98:10089–10092. doi:10.1063/1.464397
49. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341. doi:10.1016/0021-9991(77)90098-5
50. Suhre K, Sanejouand YH (2004) Nucleic Acids Res 32:610–614. doi:10.1093/nar/gkh368
51. Runge S, Thogersen H, Madsen K, Lau J, Rudolph R (2008) J Biol Chem 283:11340–11347. doi:10.1074/jbc.M708740200