Multi-State Modeling of G-protein Coupled Receptors at Experimental Accuracy

Lim Heo[1] and Michael Feig[1, *]

[1] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

*Corresponding author:

603 Wilson Road, Room 218 BCH

East Lansing, MI 48824, USA

mfeiglab@gmail.com

+1-517-432-7439

ABSTRACT

The family of G-protein coupled receptors (GPCRs) is one of the largest protein families in the human genome. GPCRs transduct chemical signals from extracellular to intracellular regions via a conformational switch between active and inactive states upon ligand binding. While experimental structures of GPCRs remain limited, high-accuracy computational predictions are now possible with AlphaFold2. However, AlphaFold2 only predicts one state and is biased towards the inactive conformation. Here, a multi-state prediction protocol is introduced that extends AlphaFold2 to predict either active or inactive states at very high accuracy using state-annotated templated GPCR databases. The predicted models accurately capture the main structural changes upon activation of the GPCR at the atomic level. The models were also highly successful in predicting ligand binding poses via protein-ligand docking. We expect that high accuracy GPCR models in both activation states will promote understanding in GPCR activation mechanisms and drug discovery for GPCRs. At the time, the new protocol paves the way towards capturing the dynamics of proteins at high-accuracy via machine-learning methods.

INTRODUCTION

High-resolution protein structures provide detailed mechanistic information at the atomistic level about biological function and serve as starting points for structure-based drug design to develop small molecules that control protein behavior.[1, 2] Such high-resolution structures have been acquired by experimental methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryo-electron microscopy. An increasingly powerful alternative is computational protein structure prediction. Predictions based on homology templates have long provided confident models when there is an experimentally determined structure for a close homolog.[3, 4] More recently, protein structure prediction based on machine learning methods such as AlphaFold2 (AF2)[5] and RoseTTAFold[6] has also become able to generate accurate models for essentially any sequence, even when homologs are not available. These methods mainly rely on neural network models that have learned how to deduce inter-residue relationships based on co-evolutionary couplings together with high-resolution structure generation modules trained on known experimental structures. The accuracy of the resulting models may approach experimental accuracy, and at least for some applications, the computational models may be sufficient for further studies. However, limitations remain with respect to capturing structural dynamics that can lead to multiple conformations. Proteins often possess multiple conformational states to perform their biological roles, whereas prediction methods are generally trained to predict a single, native state for a given sequence. As may be expected, models produced by AF2 during CASP14 typically varied little, although different models for one target could match multiple conformational states accurately.[7, 8] Given the overall success of AF2 in producing high-accuracy models, the question of whether and how multiple states can be predicted for a given protein needs further investigation and is the subject of this study.

G-protein coupled receptors (GPCRs) comprise one of the largest protein families in the human genome.[9] They are involved in various biological roles such as behavior regulation and regulation of immune system activity.[1, 2, 10, 11] They are subdivided into several classes according to sequence homology and functional similarity: class A (rhodopsin), B1 (secretin), B2(adhesion), C (glutamate), F (frizzled), T (taste 2).[12] Dysregulation of GPCRs can cause diseases including cardiovascular disease[11] and Parkinson's disease.[10] GPCRs are attractive targets for drug discovery, and indeed, around one third of drugs approved by the U. S. Food and Drug Administration (FDA) target GPCRs.[1, 13] GPCRs are integral membrane proteins with a common topology that consists of seven transmembrane helices. GPCRs generally function by transmitting chemical signals from the extracellular to the intracellular region. The transmission is usually triggered by the binding of agonists on the extracellular side. The binding alters the conformation of a GPCR from the inactive to an active state. The active state can then be sensed on the intracellular side via binding of a transducer molecule, an active G-protein or a β-arrestin. The details

of the activation mechanism vary depending on the class of the GPCR and they are still under investigation via experimental and computational approaches.[14-20]

Experimental structures of GPCRs are available since the first GPCR structure, for bovine rhodopsin, was resolved in 2000.[21] The number of experimental GPCR structures has steadily increased since.[2] As of July 29, 2021, 574 GPCR structures have been deposited to the protein data bank (PDB).[22] However, a majority of GPCR structures are still unknown. Known structures comprise only 99 out of 403 human non-olfactory receptors. Moreover, even although GPCRs can have multiple conformational states, only 58 and 62 GPCRs were determined in active and inactive states, respectively, and only 30 GPCRs have experimental structures in both activation states. The lack of high-resolution GPCR structures has made it difficult to understand GPCR activation mechanisms and enable structure-based drug design for GPCRs. Computational protein structure prediction methods targeted at GPCRs have been used to generated models where experimental structures are not available. Among them, GPCR-I-TASSER[23], RosettaGPCR[24], and GPCRdb[25, 26] have resulted in structure databases of the whole GPCR proteome via computational modeling. The Zhang group designed a specialized version of I-TASSER[27] for GPCRs.[23] It was based on using GPCR structure-specific features with I-TASSER's template-based and -free modeling pipeline. RosettaGPCR performed template-based modeling using Rosetta to predict GPCR structures in the inactive state.[24] GPCRdb maintains multiple databases related to GPCRs including experimental structures of GPCR with activation state annotations.[25, 26] GPCRdb also contains predictions for GPCRs in active and inactive forms via homology modeling by using templates in the corresponding states. However, the accuracy of these GPCR models has not been rigorously benchmarked, especially with respect to reproducing differences between active and inactive states and in light of the recent advances in structure prediction accuracy.[25]

Here, we focus on using AF2 for modeling GPCRs, especially with respect to accurately predicting both active and inactive states for a given GPCR. As described below in more detail, the regular AF2 protocol can predict accurate GPCR models, more accurate than template-based modeling, but it was not possible to predict models in multiple states for most of the benchmarked GPCRs. The predicted models were in only one GPCR state with a preference for inactive states. To overcome this limitation, we describe here a modified protocol that enables the modeling of multiple states via AF2. Briefly, this was accomplished by using state-annotated structure databases together with state specific structural templates and a modification of the MSA input features. While the approach is general in principle, we applied it here to generate accurate models of both, active and inactive states of a given GPCRs. The protocol was tested for GPCRs where experimental structures are available and the resulting

GPCR models were further examined in the context of predicting ligand binding poses. The new protocol for multi-state modeling via AF2 and results are described in more detail in the following.

RESULTS AND DISCUSSION

Modeling of GPCRs using AF2

The original AF2 showed outstanding performance in protein structure prediction during CASP14[5, 7, 28, 29] where target proteins were mostly globular soluble proteins but also included three transmembrane proteins.[8] As benchmarked here, we find that GPCRs, one of the major transmembrane protein families, can also be predicted accurately using AF2. For recently experimentally determined GPCR structures that were not included during AF2 training, 28 out of 55 (51%) GPCRs were predicted at an accuracy of better than 1.5 Å accuracy in terms of Cα-RMSD for the transmembrane helices (TM-RMSD); the median accuracy was 1.47 Å. (**Figure S1**) For comparison, those GPCR structures were also modeled via template-based modeling (TBM) with the same structural templates from the PDB70 database[30] used in the AF2 predictions. As expected, the resulting models were less satisfactory than the AF2 predictions. Only 3 out of 55 (5%) structures had RMSD values below 1.5 Å for the TM domain, and the median TM-RMSD was 2.71 Å.

AF2 only modeled either the active or inactive state for a given GPCR sequence, though, not both. Interestingly, when its performance was benchmarked on human GPCRs that were experimentally determined in both states, the top-ranked models were more likely to be in the inactive state rather than the active state. (**Figure S2**) Out of 28 multi-state human GPCRs, 22 models were closer to their inactive state experimental structures than to their active structures. As a consequence, when comparing AF2 models strictly against the experimental structures, AF2 predictions were highly accurate for the inactive structures with a median TM-RMSD of 1.22 Å while 22 out of 28 (79%) GPCRs were predicted within 1.5 Å in TM-RMSD. In contrast, the AF2 models were dissimilar to active state GPCR structures with a median TM-RMSD of 2.05 Å. For comparison, the median TM-RMSD between active and inactive state structures was 2.15 Å. This bias may originate from the larger number of experimental GPCR structures in inactive states (178) than active states (60) that were available for training.

When multiple output models with different input feature parameters for AF2 were generated, predictions for most GPCRs remained very similar (**Figure S3**). However, for some cases, the resulting models varied more, including structures that were close to either the active or inactive state (**Figure S3**). For example, the human muscarinic acetylcholine receptor M2 (human ACM2; UniProt ID P08172) was

predicted in both states. The top-ranked and fifth-ranked models were in the inactive and active states with a TM-RMSD of 0.64 and 0.71 Å with respect to experimental structures in the corresponding states. Based on this analysis, we inferred that although AF2 is not designed *per se* to generate ensembles consisting of multiple, functionally relevant states, the AF2 neural network model can in principle predict them.

Using either multiple sequence alignments (MSA) or structural templates as input features was enough for high accuracy structure prediction using AF2, based on an ablation study. (**Figure S1** and **S2**). There was little loss in accuracy when structural templates were not used. The median TM-RMSD was 1.55 Å for recently determined human GPCRs. *vs.* 1.47 Å using the original AlphaFold protocol. This finding is consistent with a previous analysis of AF2 for targets where sufficiently deep MSAs are available.[5] The proteins in our benchmark sets have at least thousands of homologous sequences in their MSA. On the other hand, when using input features from templates, but without MSAs - essentially this is "template-based modeling" via AlphaFold - it was still possible to predict overall high accuracy models. The median TM-RMSD was 1.87 Å, better than template-based models with the same templates using MODELLER[31] with a median TM-RMSD of 2.71 Å. (**Figures S1** and **S2**). However, a bias towards more accurate modeling of inactive states vs. active states remained.

Multi-state modeling of GPCRs

The analysis above suggests that simply using templates according to a desired functional state as input to AF2 is not sufficient to model both active and inactive states at the highest accuracy. Using activation state-annotated GPCR databases instead of the standard template database, PDB70, also improved results only very slightly, if any. However, when MSA input features were removed and state-annotated GPCR databases were used, it became possible to generate highly accurate models for both active and inactive state structures of GPCRs (**Figure 1** and **Figures S1–6**). In this manner, the activation state-annotated structural template databases could be used to guide the AF2 modeling network towards a specific activation state based on homology, but without interference from MSA-based contacts that are still biased by conformational state preferences according to training. This strategy led to greatly improved active state predictions while maintaining the previously achieved high accuracy for inactive state structures (**Figure 1a**). Most of the GPCRs (24 out of 28, 86%) were predicted within 1.5 Å TM-RMSD and a median of 1.11 Å for the active state. For inactive states, structures were predicted with a median TM-RMSD of 1.28 Å and 21 out of 28 models predicted at high-accuracy (*i.e.*, < 1.5 Å TM-RMSD). For comparison, 21 out of 28 models were also predicted at high-accuracy using the original AF2 protocol.

Moreover, among the 28 predicted models for each active and inactive state, 27 models were closer to the experimental structures for the corresponding activation state. Similarly, among the recently determined GPCR structures, many of the GPCRs (35 out of 55, 63%) could be predicted with an accuracy of better than 1.5 Å in TM-RMSD with a median of 1.16 Å. (*c.f.* 1.47 Å using the original AF2 protocol), again because of more accurate modeling of active state GPCR conformations.

The model accuracy with the tested modeling protocols was compared with a previous method, GPCRdb.[25] (**Figure S7**) To be comparable with our evaluation, models were retrieved from an archive (2018 April) of the method's model database. Among the recently experimentally determined GPCRs, there were 44 common GPCRs, so the comparisons were made with those GPCRs in terms of TM-RMSD. We find that our template-based models using a single template were slightly worse with a median TM-RMSD of 2.34 Å than the corresponding models in the GPCRdb generated based on multiple templates with a median TM-RMSD of 2.05 Å. However, AF2-based predictions were clearly more accurate, especially when our multi-state modeling protocol was applied. The resulting models were more accurate than the GPCRdb models for all GPCRs except one and the median TM-RMSD of 1.12 Å indicated significant overall improvement.

We further assessed whether detailed aspects of structural changes upon activation of the GPCR could be captured with the multi-state modeling protocol. The movement of TM5 on the extracellular side when transitioning between active and inactive states was successfully modeled. (red arrow in **Figure 1b**) Also, structural changes of TM6 on the intracellular side, which enables G-protein binding, was described very accurately (red arrow in **Figure 1c**). Beyond TM helix rearrangements, other detailed structural features correlated with the activation mechanism could be captured as well (**Figure 2**). There are several known characteristic structural changes upon activation for class A GPCRs.[19] For example, the changes during activation for the human muscarinic acetylcholine receptor M2 (UniProt ID P08172, ACM2_HUMAN) were captured very accurately up to atomistic detail in the predicted active and inactive states when compared to the experimental structures (**Figure 2**). One important feature is the transmission switch involved in signal initiation (**Figure 2a**). The switch includes the reorientation and repacking of some side chains, which was captured well by the predicted models. Similarly, other characteristic features were also accurately predicted including the collapse of the sodium ion pocket, the opening of hydrophobic lock, changes in the NPxxY motif, the rewiring of microswitch residue I^{6x37}, and changes in the DRY motif (**Figure 2b–f**), all of which are essential components of the signal transduction pathway from the extracellular to the intracellular region.

We also predicted all human non-olfactory GPCRs in active and inactive states using our multi-state modeling protocol. We could evaluate the model accuracy for GPCR models that have experimentally

determined structures as a function of the predicted LDDT (pLDDT) by the AF2 machine learning model and the maximum sequence identity of the used templates. (**Figure 3**) There was a clear relationship between the accuracy and the pLDDT as presented in Jumper *et al.*[5] For a model that had a pLDDT higher than 90, it was likely to be accurate; 88% (37 out of 42) and 80% (28 out of 35) of the active and inactive state models had less than 1.5 Å in TM-RMSD, respectively. Among the GPCRs without experimental structures for each state, 73% (240 out of 327) and 64% (201 out of 316) of the active and inactive models are expected to be accurate as they had predicted pLDDTs higher than 90. In contrast, the pLDDT for models by original AF2 was less informative for the simultaneous assessment of multiple states. (**Figure S8**) Presumably this is because these models were similar to one of the states but not the other one. Model accuracy is also related to the maximum template sequence identity. When any template which has a sequence identity higher than 20% was used as an input for the active state modeling, its resulting model usually had a high quality; 88% (37 out of 42) of the active state models had less than 1.5 Å in TM-RMSD. The metric has been widely used to assess what is likely successful template-based modeling. For example, GPCR models predicted using templates with sequence identities higher than 20–40% were assumed to be confident.[24, 32, 33] Likewise, it may be used to infer the success of modeling prior to the active state modeling via our multi-state modeling protocol. The structural template dependence was introduced here because information from structural templates was one of the most important input features for our protocol as MSA information was discarded. For the inactive state models, we could not conclude that there is a clear template similarity dependence since there were not enough GPCRs that had low template homology. We note, that there was no such template-accuracy relationship for the original AF2 protocol. The reason is likely that structural templates have little effect when there are enough sequences in the input MSA as shown in the ablation study.

Usefulness of high-accuracy modeling of multiple states

An important question is how higher accuracy translates into practical advantages when models are used further. One application of high-accuracy GPCR models is for the prediction of protein-ligand complexes. GPCRs are one of the major target proteins for approved drugs and remain extremely attractive targets for new drugs.[1, 13] During the initial steps of drug development, computer-aided drug design relies on the prediction of accurate ligand poses and the estimation of ligand binding affinities. This requires the structure of the target protein for being able to dock potential ligands. Because protein-ligand docking is very sensitive to the structure of the binding pocket, high-accuracy models are essential for docking success. In GPCRs the ligand pockets have a different shape depending on its activation state. Therefore, to design a ligand that targets a certain activation state of a GPCR, a high-accuracy protein model for that

state is required. To validate the effectiveness of the high-accuracy state-specific GPCR models resulting from the multi-state protocol for protein-ligand docking, ligands from experimental structures were docked to the corresponding GPCR models. (**Figure 4**) For active state conformations, the protein-ligand docking success with models predicted by our protocol was the highest among all computational models and only slightly worse than with experimental structures (self-docking). Around 50% of the ligands were successfully docked to active-state conformations within 2 Å in terms of ligand heavy-atom RMSD. On the other hand, protein-ligand docking results were less satisfactory with the original AF2 models given a 20% success ratio. As discussed above, many original AF2 models were more similar to inactive-state structures than active-state structures, and this diminished the accuracy of binding site predictions. (**Figure S9**) As an example, in **Figure 4b and c**, a ligand (N-ethyl-5'-carboxamido adenosine, NEC, from PDB ID 5G53) is shown after docking to protein structures for the human Adenosine receptor A2a (UniProt ID P29274, AA2AR_HUMAN). The experimental structure of the target protein is in the active state. With a model in the active state from our multi-state modeling protocol, it was possible to dock the ligand with very high accuracy based on a ligand heavy-atom RMSD of 0.47 Å. The protein model has a very accurate binding site structure with 0.49 Å Cα-RMSD in the binding site along with an overall highly accurate structure according to 1.39 Å in TM-RMSD to the experimental structure. On the other hand, the original AF2 model was modeled in the inactive state, and the docking result to this model was poor with a ligand heavy-atom RMSD of 7.84 Å. As the model was in the inactive state, some of the binding pocket residues were misplaced. Especially $W246^{6x48}$ (indicated by a red arrow in **Figure 4b**), which acts as the transmission switch activated by agonist binding, was in its position of inactive state. This misplacement narrowed the ligand binding pocket, so that the ligand could not be docked into the correct binding pocket. Using a template-based model using a template in active-state, it did not perform as well either. The template-based model was also modeled at relatively high-accuracy with a binding site Cα-RMSD of 0.71 Å. However, many side-chains in the binding pockets were oriented incorrectly and, therefore, protein-ligand docking resulted in a poor ligand pose with a ligand heavy-atom RMSD of 6.06 Å.

For antagonist ligands bound to inactive GPCR states, the protein-ligand docking results with predicted models were also more successful with the multi-state protocol (37% success) than with the original AF2 (27% success), but they were overall less successful than docking to experimental structures, where the docking success rate was 80%. The worse docking performance for models in the inactive state is not explained by reduced overall model qualities, since the overall TM-RMSD for active and inactive states was high and comparable with the multi-state protocol. Instead, the structural similarity of the ligand binding pocket (**Figure S9**) was the determining factor since active state models had higher

binding site accuracy than inactive state models. As a result, protein-ligand docking was more successful for active state models.

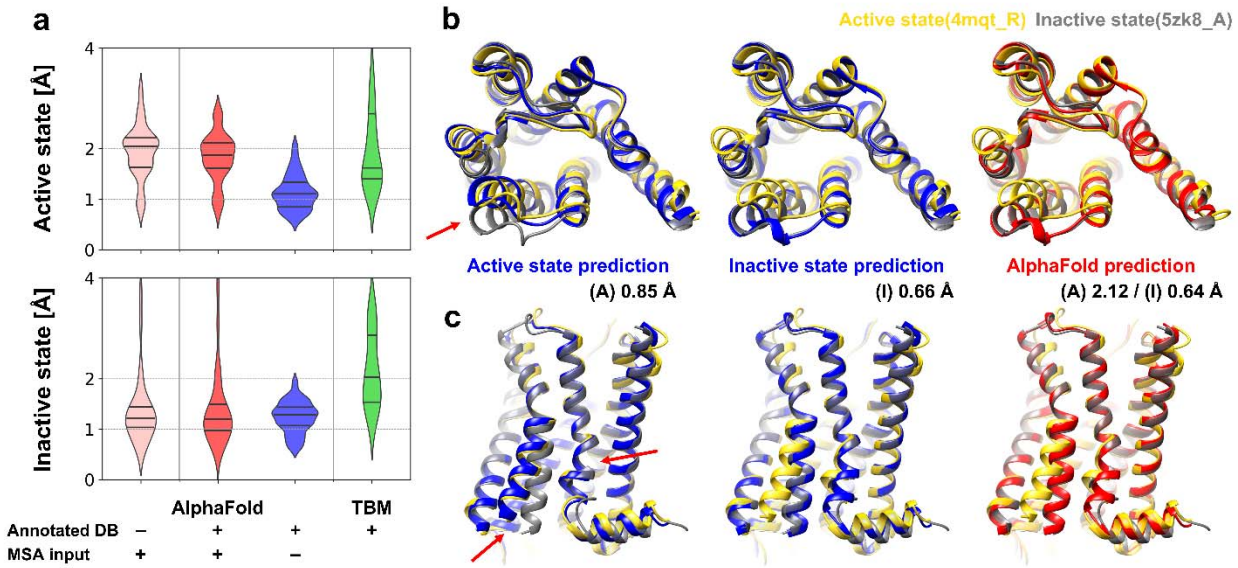

Figure 1. Modeling of active and inactive states of human G-protein coupled receptor structure (GPCR). (a) Modeling accuracies for human GPCRs that have both active and inactive state experimental structures using various modeling protocols. Cα-RMSDs are measured with respect to active and inactive forms of transmembrane helices. Distributions of modeling accuracies are shown as violin plots with black lines indicating three quartiles. (b and c) Multi-state modeling of human muscarinic acetylcholine receptor M2 (UniProt ID P08172, ACM2_HUMAN). (b) View from the extracellular region; (c) Side-view. Models predicted as active and inactive forms using customized GPCR databases via AF2 without MSA input features are shown in blue. A predicted model using AF2 with MSA input features and the standard PDB70 database is shown in red. Experimental structures of active state (yellow, PDB ID 4mqt_R) and inactive state (gray, PDB ID 5zk8_A) are compared to the predictions. Selected conformational differences between states are highlighted by red arrows.

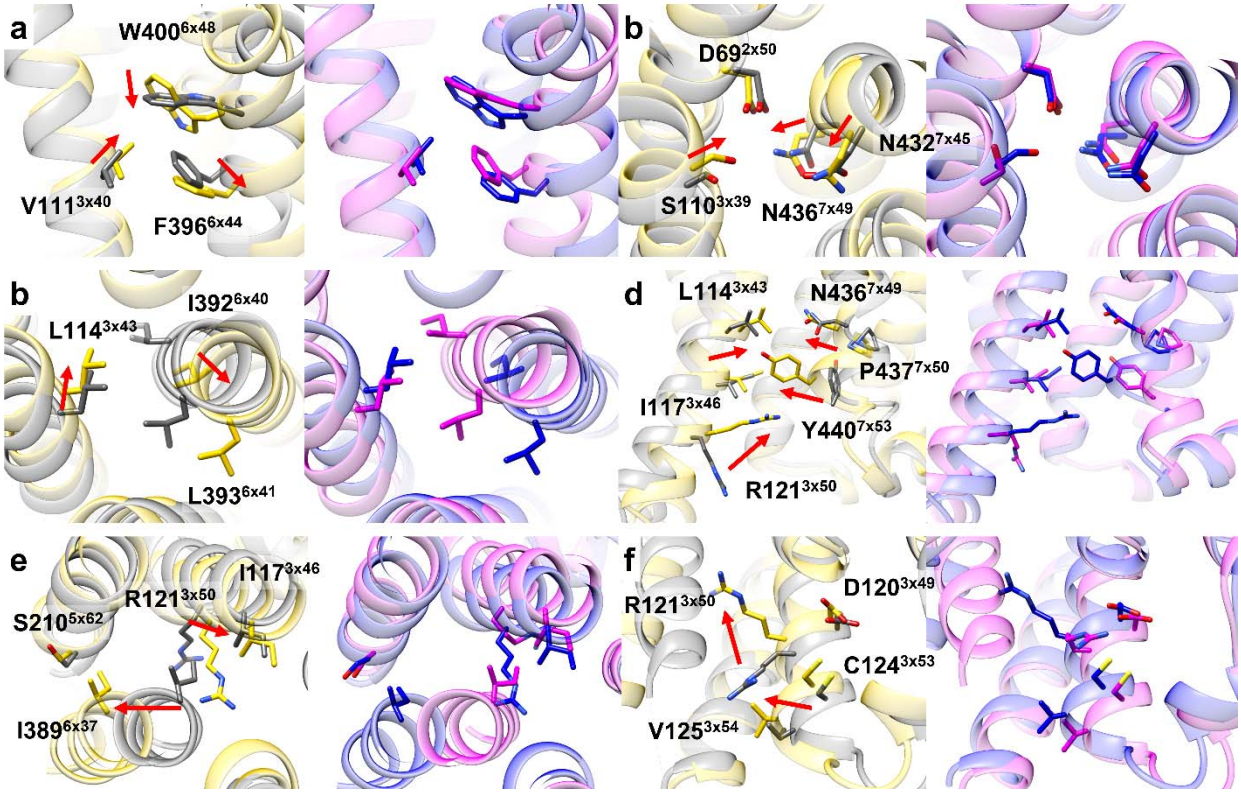

Figure 2. Detailed structural changes upon activation of human muscarinic acetylcholine receptor M2 (UniProt ID P08172, ACM2_HUMAN). Experimentally determined active (PDB ID 4mqt_R) and inactive (PDB ID 5zk8_A) state structures are shown in yellow and grey, respectively. Predicted active and inactive state model structures are shown in blue and magenta, respectively. Key sidechains for the activation mechanisms are shown as sticks with their residue numbers and the GPCRdb numberings, and the direction of changes is indicated by red arrows. (a) signal initiation step via transmission switch, (b) collapse of a sodium ion pocket, (c) opening of hydrophobic lock, (d) structure changes around the NPxxY motif, (e) rewiring of microswitch residue I389 6x37, and (f) structure changes around the DRY motif.

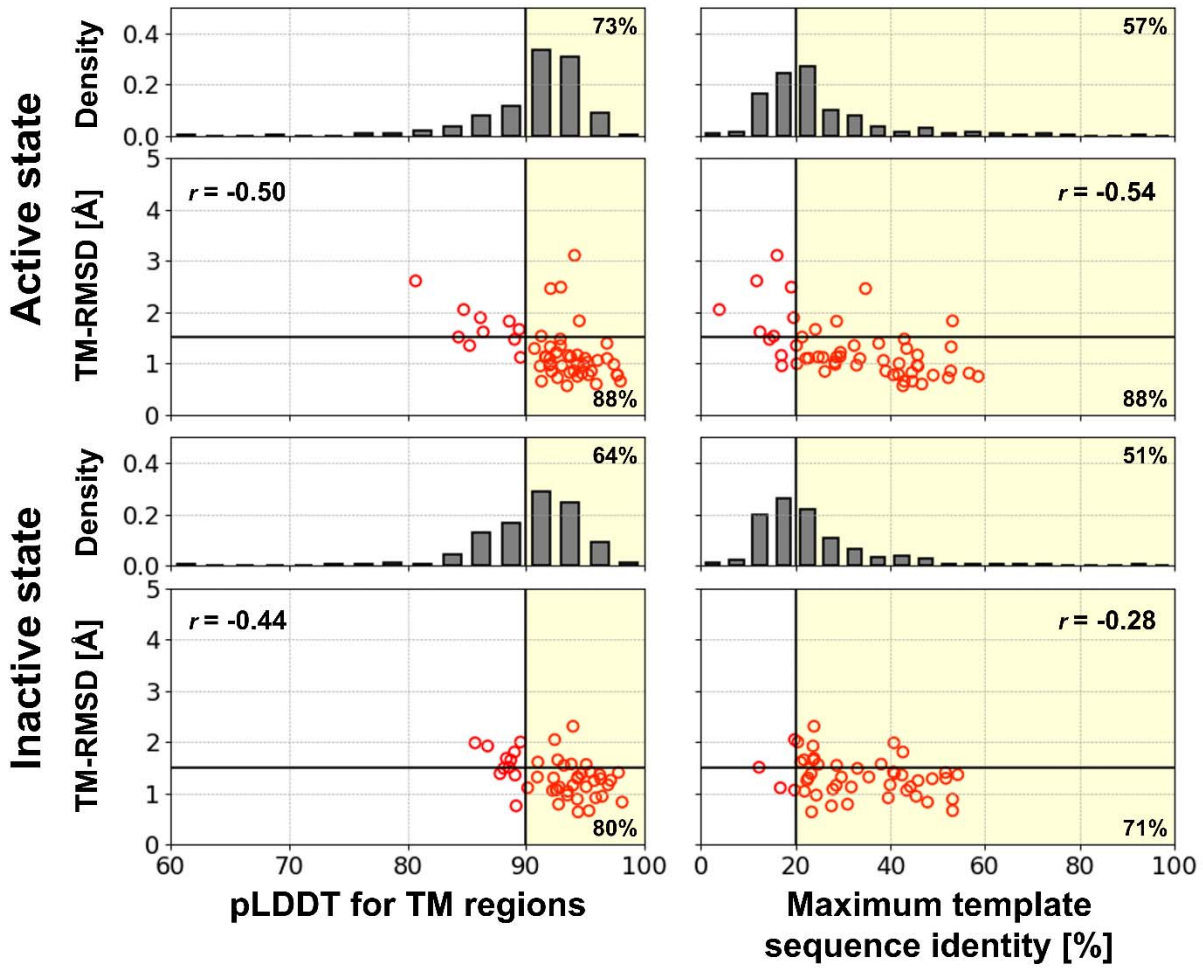

Figure 3. Estimated model accuracies for the multi-state modeling protocol. Relationships between TM-RMSD and the pLDDT or the maximum sequence identity of the structural templates are shown as red circles. Pearson correlation coefficient for each relationship is shown in the box and denoted as "*r*=". A model that has a pLDDT higher than 90 or is predicted using a template with a sequence identity of more than 20% was likely to be accurate (TM-RMSD < 1.5 Å). Those selection criteria are shown as yellow background color, and the ratios of accurate models among the selected models are noted at lower right corner. For human GPCRs without experimentally determined structures for each state, distributions of the pLDDT and the maximum sequence identity of the structural templates are shown as grey histograms. The percentages of models that are likely to be accurate assessed by each selection criteria are noted at the top right corner.

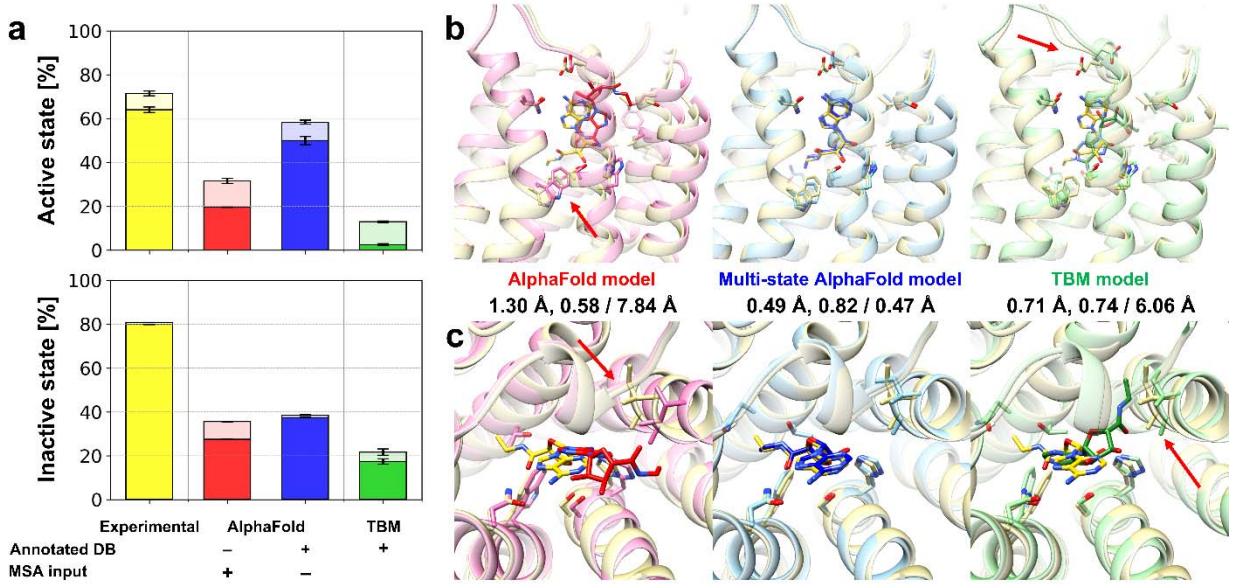

Figure 4. Protein-ligand docking on predicted GPCR models. (a) Protein-ligand docking success ratios for GPCR model structures using various modeling protocols. Docking simulations were performed five times independently for each structure using AutoDock Vina. A docking simulation is considered as success when the ligand RMSD is lower than 2 Å (opaque bars) and 3 Å (transparent bars). Success ratios and their standard errors are shown as bar charts and overlaid error bars, respectively. An example of protein-ligand docking using different models is shown in the side-view (b) and the top-view (c). An agonist ligand (N-ethyl-5'-carboxamido adenosine, NEC, from PDB ID 5G53) was docked to protein structure models for human Adenosine receptor A2a (UniProt ID P29274, AA2AR_HUMAN). Protein structures are shown in transparent cartoon representations, while docked ligand conformations are depicted as sticks. The experimental protein-ligand complex structure is shown in yellow. Protein-ligand docking results using GPCR models with the original AF2, our multi-state modeling protocol, and template-based modeling are shown in red, blue, and green, respectively. Some issues that led to incorrect protein-ligand docking results are indicated by red arrows. The binding site Cα-RMSD, the lDDT for the transmembrane region, and the resulting docking accuracy in terms of ligand heavy-atom RMSD are shown under the model names.

CONCLUSIONS

Here, we examined AF2 for the modeling of GPCRs in active and inactive conformations. While AF2 clearly generates more accurate models of GPCRs than template-based modeling, the models typically only model one functional state, usually the inactive state, presumably because of training based on the PDB where inactive state GPCR structures are overrepresented. A modified protocol is described here that can guide AF2-based high-accuracy modeling towards a specific activation state. In this multi-state modeling protocol, AF2 was used with activation state-annotated GPCR structure databases instead of a general PDB structure database and without MSA input features to avoid learned biases towards inactive GPCR states. With this protocol, it became possible to model both active and inactive states at the same high accuracy as the original AF2 models for just one GPCR state, usually the inactive conformation. The resulting models capture key structural changes upon activation/deactivation of the GPCR at the atomic level. In addition, the multi-state models improved correctly predicting ligand binding poses using protein-ligand docking, reaching almost the same success rate as with experimental structures, especially for active states. The multi-state modeling approach introduced here can be extended in principle to other protein families such as kinases as long as experimental structures in multiple states are available to form state-specific template databases and we expect that the approach described here is a first step towards capturing not just native structures but conformational dynamics at high accuracy via machine learning-based approaches.

METHODS

Multi-state protein modeling protocols

Several protein structure prediction protocols were tested to model GPCRs as active and inactive states. First, the original pipeline of AF2[5] was used to predict GPCR structures. The input consists of target protein sequences, MSAs, and structure templates; as output five models are predicted. We checked whether AF2 can model both active and inactive states. For a GPCR, structural similarities to active and inactive experimental structures were evaluated. The ability of being able to model multiple states was judged based on some of the five models being closer to active state than inactive state while others were closer to the other state. In addition, we intended to guide AF2 to model a specific activation state structure using activation state-annotated GPCR structure databases. A database was built for each active or inactive state. The details about how databases were constructed are described in the following section. Protein structure prediction for a state was guided by the input database for the state, as a replacement of the PDB70 database. [30] Databases were used in two ways. One approach was to simply replace the PDB70 database for the structure template search. In another approach, residues in the input MSAs were modified to gaps for sequence positions which were aligned to selected structure templates from the activation state-annotated databases. (**Algorithm S1**) Furthermore, an ablation study was performed to better understand the role of each input feature. The original AF2 method was compared with three variants: one without MSAs, one without structure templates, and one without both MSAs and structure templates. Finally, as a reference, structure prediction was performed via template-based modeling using MODELLER[31] with the identical structure templates and the sequence alignments used for the AF2 predictions.

Building activation state-annotated GPCR structure databases

Template structure databases for each active and inactive state GPCRs were built based on the GPCRdb[25, 26] activation state annotation. The database building procedure was based on the official procedure of building customized HHsearch[30] database. A list of PDB IDs was collected from the GPCRdb for either active or inactive states. GPCR sequences were extracted from mmCIF files of each PDB entry. If there were multiple chains of GPCRs for a PDB entry, a preferred chain selected by the GPCRdb was used. To remove redundant sequences, the extracted sequences were clustered by MMseqs2[34] with a sequence identity cutoff of 100% and a sequence coverage of 100%. For each representative sequence for a cluster, a multiple sequence alignment was generated using HHblits[30] by searching homologous sequences against the UniClust30 database with two iterations and HHblits default options. The generated multiple sequence

alignments were post-processed to build a HHsearch database for a given state. As of July 29, 2021, there were 224 and 309 experimentally determined structures for active and inactive state structures, respectively. And they resulted in 161 and 206 unique entries for the activation state-annotated GPCR structure databases after removing proteins with identical sequences.

Benchmark tests

Two benchmark sets of human GPCRs were used to evaluate the different protocols for multi-state modeling of proteins. One set was composed of human GPCRs that were experimentally determined as both active and inactive states structures. As of July 29, 2021, 28 human GPCRs belonged to the set, and the list of proteins is summarized in **Table S1**. Since AF2[5] was trained on experimental structures that were determined by April 30, 2018, those structures could have been implicitly memorized during machine learning. However, nine structures out of 28 GPCRs in the multi-state set were newly discovered after that date. To further validate the applicability to structures that were not used for training, we introduced another set of human GPCRs that were recently experimentally determined in either active or inactive states. For the proteins in the set, none of the structures had been experimentally determined by April 30, 2018, so they were not used for AlphaFold training. There were 55 GPCRs for this set, and they are summarized in **Table S2**. For the benchmark tests, close homologous structures that have a sequence identity higher than a cutoff of 70% were excluded from the template lists. For the model accuracy evaluation in terms of TM-RMSD, the Cα-RMSD was evaluated using transmembrane helices, whose definition was taken from the GPCRdb.[25, 26] An experimental structure was considered as a reference structure if it had a resolution of 4.0 Å or higher. If there were multiple experimental structures for an activation state of a GPCR, the closest result was reported.

All human non-olfactory GPCRs were modeled as active and inactive state structures. The list of the GPCRs was retrieved from GPCRdb.[25, 26] (https://gpcrdb.org/alignment/targetselection) Since some of the GPCRs have non-transmembrane domains, we modeled the transmembrane (TM) domain only to focus on its activation state. The UniProt topology annotations were used to define the TM domains. Up to 100 residues towards N- and C-termini from the first and seventh TM helices were additionally considered for modeling not to lose any of their interactions with the TM domain and relevant TM helix residues because the annotations are based on a TMHMM prediction.[35]

Protein-ligand docking using AutoDock Vina

AutoDock Vina[36] was used to perform protein-ligand docking on the predicted GPCR structures. Protein structures were prepared using the "prepare_receptor' tool of the ADFR suite.[37] Ligand structures were extracted from the experimental structures and prepared for docking. They were first converted to the Tripos Mol2 File format by adding hydrogens at pH=7 and converted to the PDBQT file for AutoDock Vina. All torsion angles were set to be rotatable except for torsion angles around amide and guanidinium bonds, which are the default option for the "prepare_ligand" tool of ADFR suite. Atomic charges of ligand atoms were assigned using the Gasteiger charge model.[38] Protein-ligand docking was carried out at the binding pocket for a ligand. For an experimental protein structure, the center of the cubic search space of docking was located at the geometrical center of the ligand from its bound structure, and the width of the search space was 20 Å for each axis. For predicted GPCR structures, the same search space was used after superimposition to the experimental structure. Protein-ligand docking was performed using AutoDock Vina with an exhaustiveness of 32 and repeated five times for a protein structure and a ligand pair. Ligand heavy-atom RMSDs with consideration of symmetry were calculated to evaluate the accuracy of protein-ligand docking using GalaxyDock.[39]

The protein-ligand docking success ratio was related to protein model qualities for each protein structure prediction protocol and each activation state of the GPCR. We used TM-RMSD and lDDT[40] as the overall protein model accuracy metrics for backbone atoms and additional consideration of sidechains, respectively. In addition to them, the binding site Cα-RMSD was used to measure the binding site accuracy. The binding site residues were defined as residues whose atoms were within 8 Å from the ligand in the experimental structure.

Availability

The scripts used for this study, activation annotated GPCR databases, and predicted models in either activation states are available at https://github.com/huhlim/alphafold-multistate.

ACKNOWLEDGEMENTS

This research was supported by National Institutes of Health Grant R35 GM126948. We appreciate Martin Steinegger for help with building and utilizing customized HHsearch databases.

AUTHOR CONTRIBUTIONS

LH and MF designed the research, LH performed and analyzed the work, and LH and MF jointly wrote the manuscript.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

FIGURE LEGENDS

Figure 1. Modeling of active and inactive states of human G-protein coupled receptor structure (GPCR). (a) Modeling accuracies for human GPCRs that have both active and inactive state experimental structures using various modeling protocols. Cα-RMSDs are measured with respect to active and inactive forms of transmembrane helices. Distributions of modeling accuracies are shown as violin plots with black lines indicating three quartiles. (b and c) Multi-state modeling of human muscarinic acetylcholine receptor M2 (UniProt ID P08172, ACM2_HUMAN). (b) View from the extracellular region; (c) Side-view. Models predicted as active and inactive forms using customized GPCR databases via AF2 without MSA input features are shown in blue. A predicted model using AF2 with MSA input features and the standard PDB70 database is shown in red. Experimental structures of active state (yellow, PDB ID 4mqt_R) and inactive state (gray, PDB ID 5zk8_A) are compared to the predictions. Selected conformational differences between states are highlighted by red arrows.

Figure 2. Detailed structural changes upon activation of human muscarinic acetylcholine receptor M2 (UniProt ID P08172, ACM2_HUMAN). Experimentally determined active (PDB ID 4mqt_R) and inactive (PDB ID 5zk8_A) state structures are shown in yellow and grey, respectively. Predicted active and inactive state model structures are shown in blue and magenta, respectively. Key sidechains for the activation mechanisms are shown as sticks with their residue numbers and the GPCRdb numberings, and the direction of changes is indicated by red arrows. (a) signal initiation step via transmission switch, (b) collapse of a sodium ion pocket, (c) opening of hydrophobic lock, (d) structure changes around the NPxxY motif, (e) rewiring of microswitch residue I389 6x37, and (f) structure changes around the DRY motif.

Figure 3. Estimated model accuracies for the multi-state modeling protocol. Relationships between TM-RMSD and the pLDDT or the maximum sequence identity of the structural templates are shown as red circles. Pearson correlation coefficient for each relationship is shown in the box and denoted as "*r*=". A model that has a pLDDT higher than 90 or is predicted using a template with a sequence identity of more than 20% was likely to be accurate (TM-RMSD < 1.5 Å). Those selection criteria are shown as yellow background color, and the ratios of accurate models among the selected models are noted at lower right corner. For human GPCRs without experimentally determined structures for each state, distributions of the pLDDT and the maximum sequence identity of the structural templates are shown as grey histograms. The percentages of models that are likely to be accurate assessed by each selection criteria are noted at the top right corner.

Figure 4. Protein-ligand docking on predicted GPCR models. (a) Protein-ligand docking success ratios for GPCR model structures using various modeling protocols. Docking simulations were performed five times independently for each structure using AutoDock Vina. A docking simulation is considered as success when the ligand RMSD is lower than 2 Å (opaque bars) and 3 Å (transparent bars). Success ratios and their standard errors are shown as bar charts and overlaid error bars, respectively. An example of protein-ligand docking using different models is shown in the side-view (b) and the top-view (c). An agonist ligand (N-ethyl-5'-carboxamido adenosine, NEC, from PDB ID 5G53) was docked to protein structure models for human Adenosine receptor A2a (UniProt ID P29274, AA2AR_HUMAN). Protein structures are shown in transparent cartoon representations, while docked ligand conformations are depicted as sticks. The experimental protein-ligand complex structure is shown in yellow. Protein-ligand docking results using GPCR models with the original AF2, our multi-state modeling protocol, and template-based modeling are shown in red, blue, and green, respectively. Some issues that led to incorrect protein-ligand docking results are indicated by red arrows. The binding site Cα-RMSD, the lDDT for the transmembrane region, and the resulting docking accuracy in terms of ligand heavy-atom RMSD are shown under the model names.

REFERENCES

1. Yang, D. et al. G protein-coupled receptors: structure- and function-based drug discovery. *Signal Transduct Target Ther* **6**, 7 (2021).
2. Congreve, M., de Graaf, C., Swain, N.A. & Tate, C.G. Impact of GPCR Structures on Drug Discovery. *Cell* **181**, 81-91 (2020).
3. Baker, D. & Sali, A. Protein structure prediction and structural genomics. *Science* **294**, 93-96 (2001).
4. Marti-Renom, M.A. et al. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **29**, 291-325 (2000).
5. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
6. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).
7. Jumper, J. et al. Applying and improving AlphaFold at CASP14. *Proteins* (2021).
8. Kinch, L.N., Schaeffer, R.D., Kryshtafovych, A. & Grishin, N.V. Target classification in the 14th round of the critical assessment of protein structure prediction (CASP14). *Proteins* (2021).
9. Gloriam, D.E., Fredriksson, R. & Schioth, H.B. The G protein-coupled receptor subset of the rat genome. *BMC Genomics* **8**, 338 (2007).
10. Schwarzschild, M.A., Agnati, L., Fuxe, K., Chen, J.F. & Morelli, M. Targeting adenosine A2A receptors in Parkinson's disease. *Trends Neurosci* **29**, 647-654 (2006).
11. Bristow, M.R. Treatment of chronic heart failure with beta-adrenergic receptor antagonists: a convergence of receptor pharmacology and clinical cardiology. *Circ Res* **109**, 1176-1194 (2011).
12. Bjarnadottir, T.K. et al. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* **88**, 263-273 (2006).
13. Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schioth, H.B. & Gloriam, D.E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* **16**, 829-842 (2017).
14. Mattedi, G., Acosta-Gutierrez, S., Clark, T. & Gervasio, F.L. A combined activation mechanism for the glucagon receptor. *Proc Natl Acad Sci U S A* **117**, 15414-15422 (2020).
15. Wright, S.C. et al. A conserved molecular switch in Class F receptors regulates receptor activation and pathway selection. *Nat Commun* **10**, 667 (2019).
16. Weis, W.I. & Kobilka, B.K. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu Rev Biochem* **87**, 897-919 (2018).
17. Alhadeff, R., Vorobyov, I., Yoon, H.W. & Warshel, A. Exploring the free-energy landscape of GPCR activation. *Proc Natl Acad Sci U S A* **115**, 10327-10332 (2018).
18. Dror, R.O. et al. Activation mechanism of the beta2-adrenergic receptor. *Proc Natl Acad Sci U S A* **108**, 18684-18689 (2011).
19. Zhou, Q. et al. Common activation mechanism of class A GPCRs. *Elife* **8** (2019).
20. Latorraca, N.R., Venkatakrishnan, A.J. & Dror, R.O. GPCR Dynamics: Structures in Motion. *Chem Rev* **117**, 139-155 (2017).
21. Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745 (2000).
22. Berman, H.M. et al. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).
23. Zhang, J., Yang, J., Jang, R. & Zhang, Y. GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. *Structure* **23**, 1538-1549 (2015).
24. Bender, B.J., Marlow, B. & Meiler, J. Improving homology modeling from low-sequence identity templates in Rosetta: A case study in GPCRs. *PLoS Comput Biol* **16**, e1007597 (2020).

25. Pandy-Szekeres, G. et al. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res* **46**, D440-D446 (2018).
26. Kooistra, A.J. et al. GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res* **49**, D335-D343 (2021).
27. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8 (2015).
28. Kinch, L.N., Pei, J., Kryshtafovych, A., Schaeffer, R.D. & Grishin, N.V. Topology evaluation of models for difficult targets in the 14th round of the critical assessment of protein structure prediction. *Proteins* (2021).
29. Pereira, J. et al. High-accuracy protein structure prediction in CASP14. *Proteins* (2021).
30. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).
31. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815 (1993).
32. Miszta, P. et al. GPCRM: a homology modeling web service with triple membrane-fitted quality assessment of GPCR models. *Nucleic Acids Res* **46**, W387-W395 (2018).
33. Mobarec, J.C., Sanchez, R. & Filizola, M. Modern homology modeling of G-protein coupled receptors: which structural template to use? *J Med Chem* **52**, 5207-5216 (2009).
34. Steinegger, M. & Soding, J. Clustering huge protein sequence sets in linear time. *Nat Commun* **9**, 2542 (2018).
35. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580 (2001).
36. Trott, O. & Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455-461 (2010).
37. Ravindranath, P.A., Forli, S., Goodsell, D.S., Olson, A.J. & Sanner, M.F. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLoS Comput Biol* **11**, e1004586 (2015).
38. Gasteiger, J. & Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **36**, 3219-3228 (1980).
39. Shin, W.H. & Seok, C. GalaxyDock: protein-ligand docking with flexible protein side-chains. *J Chem Inf Model* **52**, 3225-3232 (2012).
40. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722-2728 (2013).