The Predicted Ensemble of Low-Energy Conformations of Human Somatostatin Receptor Subtype 5 and the Binding of Antagonists

Sijia S. Dong,[a] Ravinder Abrol,[a, b] and William A. Goddard, III*[a]

Human somatostatin receptor subtype 5 (hSSTR5) regulates cell proliferation and hormone secretion. However, the identification of effective therapeutic small-molecule ligands is impeded because experimental structures are not available for any SSTR subtypes. Here, we predict the ensemble of low-energy 3D structures of hSSTR5 using a modified GPCR Ensemble of Structures in Membrane BiLayer Environment (GEnSeMBLE) complete sampling computational method. We find that this conformational ensemble displays most interhelical interactions conserved in class A G protein-coupled receptors (GPCRs) plus seven additional interactions (e.g., Y2.43–D3.49, T3.38–S4.53, K5.64–Y3.51) likely conserved among SSTRs. We then predicted the binding sites for a series of five known antagonists, leading to predicted binding energies consistent with experimental results reported in the literature. Molecular dynamics (MD) simulation of 50 ns in explicit water and lipid retained the predicted ligand-bound structure and formed new interaction patterns (e.g., R3.50–T6.34) consistent with the inactive μ-opioid receptor X-ray structure. We suggest more than six mutations for experimental validation of our prediction. The final predicted receptor conformations and antagonist binding sites provide valuable insights for designing new small-molecule drugs targeting SSTRs.

1. Introduction

Somatostatins (SSTs) are regulatory peptides involved in inhibition of a number of endocrine and exocrine secretion functions through somatostatin receptors, which are G protein-coupled receptors (GPCRs).[1] SSTs regulate the secretion of factors such as insulin and growth hormone. All five SST receptor subtypes (SSTRs) are able to down-regulate cell proliferation, but they vary in a number of other functions such as the regulation of ion channels.[1a] Thus, their effects on cell proliferation and apoptosis are of interest for developing nonpeptidic agonists to enhance tumor growth suppression.[2] The subtype 5, SSTR5, forms a heterodimer with SSTR2 and presents enhanced cell growth inhibition ability.[2a]

Of the five human (h) SSTRs, hSSTR5 is the only receptor subtype that has different affinities for the two endogenous ligands SST-14 and SST-28, which are cyclic peptides with 14 and 28 residues, respectively.[3] SSTR5 has a higher affinity for SST-28, which is shown to suppress glucagon-like peptide-1 (GLP-1) secretion much more effectively than SST-14.[4] Therefore, hSSTR5 antagonists are potentially useful for the treatment of diabetes. Indeed, it has been shown that certain nonpeptidic antagonists are able to improve glucose tolerance in rodent models of type 2 diabetes.[5]

Two peptide-based somatostatin mimics, octreotide and vapreotide, have been commercialized to treat various diseases or conditions such as metastatic carcinoid tumors and esophageal variceal bleeding. In recent years, a number of small-molecule agonists and antagonists have also been published.[6] However, to the best of our knowledge, none have passed clinical trials. We expect that designing small-molecule ligands with improved potency and greater selectivity would be useful to minimize off-target side effects.

For the rational design of better hSSTR5 ligands, it is essential to know the molecular details of the receptor binding pockets for both the active and inactive states. Since no experimental structures are available for any of the SSTRs, we used computational methods that have proved effective for other GPCRs.[7]

Apart from the biological functions described above, a recent study has identified hSSTR5 as one of the most valuable templates for homology modeling of nonorphan and nonolfactory class A GPCRs, which represent the majority of the class A GPCRs (highest sequence identity sum and a percentage of sequences for accurate models value of 31%).[8] This has increased the significance of obtaining an hSSTR5 structure.

To predict the 3D structures of hSSTR5, we developed a modified version of GPCR Ensemble of Structures in Membrane BiLayer Environment (GEnSeMBLE)[9] complete sampling computational method, which aims to predict the 10 to 25 lower-energy protein structures likely to play a role in activation upon

[a] S. S. Dong, Dr. R. Abrol, Prof. W. A. Goddard, III
Material & Process Simulation Center (MC 139-74)
California Institute of Technology
Pasadena, CA 91125 (USA)
E-mail: wag@wag.caltech.edu

[b] Dr. R. Abrol
Departments of Medicine & Biomedical Sciences
Cedars-Sinai Medical Center
Los Angeles, CA 90048 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.201500023.
binding various ligands. GenSeMBLE has been successfully applied in the prediction of inactive-state structures of GPCRs such as cannabinoid receptor type 1 (CB1),[8a] adenosine A3 receptor (hAA3R),[8b] olfactory receptor OR1G1,[8c] bitter taste receptor Tas2R38,[8d] the GLP-1 receptor (GLP1R),[8e] and C-C chemokine receptor type 5 (CCCR5).[8f] For CB1 and hAA3R, the active states were also identified successfully. In particular for CB1, it has been demonstrated that a single mutation dramatically changes the binding site so that the receptor becomes completely inactive, whereas another single mutation makes it essentially completely active.[8a,b] Indeed these same methods successfully predicted what single mutations would interconvert these activities.[8a,b]

Herein, we focus on predicting the ensemble of structures for hSSTR5, using the modified version of the GenSeMBLE method, and their binding with several known antagonists.

2. Results and Discussions

2.1. GPCR structure predictions

Our predictions of the ensemble of low-energy 3D structures for hSSTR5 followed the general GenSeMBLE method, which is described in detail in the Experimental Section. However, we introduced a hierarchical SuperBiHelix/SuperComBiHelix sampling procedure (coarse and fine sampling) described below. This new sampling procedure provides a more systematic means for predicting inactive-state and active-state conformations of the target protein. An overview of the new GenSeMBLE procedure applied to hSSTR5 structure prediction is provided in Figure 1.

To prepare the starting structures for the GenSeMBLE procedure, we first carried out PredicTM and secondary structure predictions to determine which residues are in the seven helical transmembrane domains (TMDs). The PredicTM result and the final assignment of each TMD are shown in Figures S1 and S2 in the Supporting Information. Then we carried out multiple sequence alignments between hSSTR5 and the GPCRs with X-ray structures available, which identified human nociceptin receptor (hOPRX), mouse μ-opioid receptor (mOPRM), and human κ-opioid receptor (hOPKR) as the best candidate templates to model the hSSTR5 structure. To determine the shapes of the helices, we used OptHelix and homology modeling. Then, the TMD bundle of hSSTR5 was assembled based on the helix positions of each template. A total of 15 starting structures with different helical shapes and positions were generated.

Among the six parameters \(\{x, y, h, D, \varphi, \eta\}\) that uniquely define the orientation of a rigid TMD, the hydrophobic center (HPC) residue \(h\) and the Cartesian coordinates \((x, y)\) of the HPC, were taken from the template. Among the helical tilts and rotations \((\theta, \phi, \eta)\), the helical rotations \(\eta\) were first sampled using the BiHelix method with a sampling range of \(\Delta \eta\) from \(0^\circ\) to \(360^\circ\) and a step size of \(30^\circ\).

The top 10 structures from the BiHelix step are shown in Table S2 in the Supporting Information, where we see that all 10 structures use homology helix shapes. Since all three templates were represented in the top 10, we used the rotations for the best candidate from each template in the next step, SuperBiHelix. SuperBiHelix optimizes tilts together with rotations \((\theta, \phi, \eta)\) based on the best structures from BiHelix. For each of these, we first carried out a coarse sampling step \(\Delta \theta = 0, \pm 15^\circ; \Delta \phi = 0, \pm 45^\circ, \pm 90^\circ; \Delta \eta = 0, \pm 30^\circ\), other selected angles) from angles optimized in BiHelix. This sampled at least \((3 \times 5 \times 3)^7 \approx 374\) billion configurations from which we built and optimized the lowest 2000 7-helix bundles. This was done for all three starting templates (mOPRM, hOPRK, hOPRX), with the sampling space for each template summarized in Table S3 in the Supporting Information, and the resulting lowest energy structure for each template shown in Table 1.

To predict the structures for inactive states, we selected the lowest energy structure with a TM3–TM6 ionic lock for each template. Then we carried out a finer SuperBiHelix sampling \(\Delta \theta = 0, \pm 15^\circ; \Delta \phi = 0, \pm 15^\circ, \pm 30^\circ; \Delta \eta = 0, \pm 30^\circ\). This examined \((3 \times 5 \times 3)^7 \approx 374\) billion configurations from which we built and optimized the lowest 2000 7-helix bundles.

As shown in Table 2, the top 10 structures from this fine SuperBiHelix sampling all come from the mOPRM template except for the one ranked 5th, which is from hOPKR. Thus, we focused on structures using the mOPRM template in the sub-
sequent procedures. The structure ranked 4th by the average energy of Ctotal, ClnterH, NTotal, and NinterH (E\textsubscript{tot}) has the largest number of interhelical hydrogen bonds among the top 10 among which the following are conserved from the inactive states across class A GPCR X-ray structures (Figure 2a):

- TM3–TM6 (3–6) ionic lock (R137\textsuperscript{*}–K245\textsuperscript{*})
- TM1–TM2–TM7 (1–2–7) network (N58\textsuperscript{150}–S297\textsuperscript{46}–D86\textsuperscript{250}, N300\textsuperscript{46}–D86\textsuperscript{250})

In addition, there are seven strong interactions involving residues that are conserved in all or most hSSTRs (Figure 2b):

- Y78\textsuperscript{241}–D136\textsuperscript{249}
- N122\textsuperscript{35}–A85\textsuperscript{149} (A85\textsuperscript{249} backbone)
- T126\textsuperscript{138}–S167\textsuperscript{241}
- K227\textsuperscript{147}–Y138\textsuperscript{147}
- Y294\textsuperscript{241}–D119\textsuperscript{241}
- R113\textsuperscript{22}–L174\textsuperscript{46} (L174\textsuperscript{60} backbone; valine in hSSTR1, 4)
- K245\textsuperscript{46}–F306\textsuperscript{255} (F306\textsuperscript{255} backbone; R254\textsuperscript{32}–F315\textsuperscript{15} in hSSTR3)

Therefore, we considered this structure as the most promising inactive-state candidate, denoted InactiveConf1. We also found an important polar interaction that seems likely only for hSSTR3 and hSSTR5 because the other hSSTRs do not have a tyrosine on 7.35:

- Y286\textsuperscript{28}–N271\textsuperscript{58}

In addition, we found one interaction that we expect to be unique to hSSTR5:

- T117\textsuperscript{30}–S171\textsuperscript{47}–

These interactions are shown in Figure 2b.

In order to obtain a diverse set of low-energy protein structures, we selected two other protein conformations from the 25 lowest energy predicted hSSTR5 structures (listed in Table S4 in the Supporting Information). Here, we selected the two that have the largest root-mean-square deviation (RMSD) with InactiveConf1 and with each other. These two are labeled InactiveConf2 (ranked 7th in Table 2, and 6th in Table S4 in the Supporting Information) and InactiveConf3 (ranked 16\textsuperscript{th} in Table S4 in the Supporting Information).

To obtain structures that might be candidates for active states (with TM6 well separated from TM3), we carried out a finer sampling starting from the best structures from the coarse sampling that satisfy specific structural criterion, which is described below. The resulting optimal active-state structures are named ActiveConf1 and ActiveConf2.

To distinguish potential active-state from inactive-state structures, we defined R\textsubscript{36}, the distance between the intracellular ends of TM3 and TM6, to be the shortest distance between the backbone atoms of the four residues at the intracellular ends of TM3 and TM6. We selected the lowest energy structure with an R\textsubscript{36} value of 4 Å larger than the R\textsubscript{36} value of 7.18 Å from the inactive-state structure InactiveConf1. The active-inactive R\textsubscript{36} value difference of 4 Å was chosen because the X-ray struc-
tures for both human \(\beta_2\)-adrenergic receptor (h\(\beta_2\)AR) and bovine rhodopsin display this feature. Then we carried out a finer sampling of \((\theta, \phi, \eta)\) on this selected structure to obtain the first putatively active conformation, ActiveConf1. Substituting the TM6 shape in this starting structure with the TM6 from the homology model with active h\(\beta_2\)AR X-ray structure followed by finer sampling gives the second putatively active conformation ActiveConf2 (for details, see the Supporting Information). These will be described and used in a subsequent paper that validates this new hierarchical sampling method and predicts active structures for hSSTR5.

A summary of all structures used in the following antagonist binding study is in Table 3.

2.2. Antagonist binding

To validate the structures predicted for the hSSTR5 receptor, we predicted the binding site and energy for five known small-molecule antagonists to the five predicted protein structures (InactiveConf1,2,3; ActiveConf1,2). The antagonists chosen for the docking were selected from a series of benzoxazole piperidines screened by Martin and co-workers,\(^{[7c]}\) which exhibit a wide range of binding affinities while retaining the same structural scaffold of the ligand (Figure 3). The molecules are labeled “Mx”, with “x” preserving the numbering scheme from Ref. \([7c]\). We selected M38, M40, M42, M59, M60 based

---

Table 3. Summary of structures from fine SuperBihelix/SuperCombiHelix sampling and docked by the antagonists.

<table>
<thead>
<tr>
<th>Structure</th>
<th>(\Delta \phi ) [°]</th>
<th>(\Delta \psi ) [°]</th>
<th>(\Delta \chi ) [°]</th>
<th>(\epsilon_{\text{DMM}}) [kcal mol(^{-1})]</th>
<th>(R_m) [Å]</th>
<th>RMSD [Å]</th>
<th>Inactive-Conf1</th>
<th>Inactive-Conf2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inactive-Conf1</td>
<td>0 0 0 0 0 105 0 -15 0 0 0 0 0 -30 0 0 0 -350.7 7.18 0.0</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive-Conf2</td>
<td>0 0 0 -15 60 30 15 0 0 0 0 0 60 0 0 -345.7 7.59 1.7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive-Conf3</td>
<td>0 0 0 -15 60 -30 15 0 0 0 0 0 30 0 0 -328.9 7.56 1.4</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active-Conf1</td>
<td>0 0 0 0 0 0 -60 -15 0 0 0 0 0 30 -90 0 -313.2 10.71 - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active-Conf2</td>
<td>-15 0 0 -15 0 -30 30 0 0 0 0 0 30 -30 0 -306.6 10.20 - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[a] The \(\Delta \theta\) values for H1, H2, H3, H4, H5 and H7 are all zero for the cases listed. The angles \(\Delta \phi, \Delta \psi\) and \(\Delta \chi\) are deviations from the respective initial homology templates.
on their experimentally determined binding affinities ($K_i$) for hSSTR5—M59 is the most potent derivative reported, exhibiting a $K_i$ value of 3 nm, while the other compounds selected exhibit a diverse range of binding affinities (from 23 nm to over 1000 nm).

The experimental binding affinity, predicted binding site and energy for the five antagonists chosen are summarized in Table 4. To determine the best pose for each antagonist, we allowed each antagonist to select its preferred conformations out of the docking results to all five predicted protein structures. For each ligand, we selected the optimal pose for the ligand’s preferred binding mode using the best unified-cavity (UCav) energy in comparing the ligand binding. The UCav energy ranked the five ligands as M59 (best) < M60 < M40 < M42 < M38 (worst), while the binding energy calculated from $\Delta G_1 - \Delta G_2 = RT \ln \left( K_{i2}/K_{i1} \right)$ using experimental binding constants gives the rank order as M59 < M60 < M38 < M40 < M42. Thus, only M38 is an outlier. The UCav energies range over a factor of nine of the binding energies calculated from the experimental binding constants, which range over 3.57 kcal mol$^{-1}$ (Table S5 in the Supporting Information).

To further investigate how well our predicted binding energies correlate with the experimental results, we plotted the UCav energy against the negative logarithm of the experimental binding constants (Figure 4). This shows that the UCav energies for the optimal poses of the antagonist series correlate with the experimental p$K_i$ values with a linear regression coefficient of 0.78. This suggests that UCav captures the essential aspects of the relative binding affinities of these antagonists.

We also find that M42 and M60 favor the InactiveConf2 conformation while the other three favor InactiveConf1, which means InactiveConf1 and InactiveConf2 could possibly be two inactive conformations selected by the antagonist series. This suggests that the structures of hSSTR5 predicted to be more stable are more likely to be in the inactive state than the less stable ones.

Table 4 shows the ligand interaction diagram (LID) of the best pose for each antagonist. We find that the antagonists predominantly bind with a pocket defined by TMDs 1–2–3–6–7. In these best poses, all antagonists form a salt bridge between their positively charged piperidine amine group and D119$^{3,22}$. This aspartic acid on TM3 is conserved in all somatosensory receptors.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>$K_i$ [nm]</th>
<th>UCav E [kcal mol$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>M59</td>
<td>3</td>
<td>-118.48</td>
</tr>
<tr>
<td>M60</td>
<td>23</td>
<td>-115.00</td>
</tr>
<tr>
<td>M38</td>
<td>113</td>
<td>-86.35</td>
</tr>
<tr>
<td>M40</td>
<td>6 (95, -86.35)</td>
<td>-86.88</td>
</tr>
<tr>
<td>M42</td>
<td>7 (6.4, -115.00)</td>
<td>-118.48</td>
</tr>
</tbody>
</table>

Figure 4. Relationship between the UCav energy ($E$) from docking and the experimental p$K_i$ values of the antagonist series. $R^2$ is the coefficient of determination. The arrow represents the possible direction of change of the M42 data point.
tatin receptors, and mutagenesis studies have shown it is essential in SST binding by forming an electrostatic interaction with a positively charged group in SST. Therefore, our result has further confirmed that D119 mediates after the MD simulation (Section 2.3, below). Unlike D119, R39 is not conserved in any hSSTRs. This might explain why the antagonists with polar groups at R1 or R2 are extremely selective towards the receptor subtype S. Mutating R39\textsuperscript{1,31} to a negatively charged residue, a nonpolar residue, or serine (as in hSSTR1, 2, 3) should be able to test this hypothesis.

Other residues playing an important role in these high-affinity antagonists are polar residues N100\textsuperscript{1,44}, Q123\textsuperscript{1,39}, N268\textsuperscript{1,49} and S293\textsuperscript{7,42}, and nonpolar residues V43\textsuperscript{3,35}, Y47\textsuperscript{1,39}, Y89\textsuperscript{5,53}, W261\textsuperscript{1,48}, F264\textsuperscript{5,51}, F265\textsuperscript{5,52}, V290\textsuperscript{7,39} and Y294\textsuperscript{7,41}. Residues shared by M59, M60, M38 and M40 are polar residues Q123\textsuperscript{3,36} and S293\textsuperscript{7,42}, and nonpolar residues Y89\textsuperscript{5,53}, W261\textsuperscript{6,48}, F264\textsuperscript{5,51}, F265\textsuperscript{5,52}, V290\textsuperscript{7,39} and Y294\textsuperscript{7,41}. This is shown in the pharmacophore mapping in Table 4. The residues interacting with the strongly binding antagonists that are missing in the predicted pose for the nonbinding molecule M42 are F265\textsuperscript{5,52} and V290\textsuperscript{7,39}. Since M42 binding is not experimentally detected, we can deduce that F265A and/or V290A mutations might cause the other antagonists to have a decreased affinity towards hSSTR5.

In the predicted binding poses, M38 and M40 both form a \( \pi \cdots \pi \) stacking interaction between the benzoazole and W261\textsuperscript{4,48}, but both lack the salt bridge with R39\textsuperscript{1,31} that is found in the predicted interactions for M59 and M60. Therefore, without the strong electrostatic interaction constraining the ligand position, the weaker \( \pi \cdots \pi \) interaction becomes a dominating force of the ligand with the protein. Although M38 does not have a stronger binding in docking than M40 and M42 as predicted, its exposed chlorine group might lower the binding energy once solvation is taken into account.

Apart from the polar groups, the ligand size also plays a role in determining the binding affinities. Figure 5a,b shows that the ethoxy group at the R4 position in M59 prevents the phenyl group of M59 from being parallel to the hydrophobic plane of the GPCR as in the M60 pose. The ethoxy group has directed the M59 phenyl head to go deeper into the binding pocket and reach more polar and nonpolar residues than M60. This explains why M59 has a higher affinity than M60 for hSSTR5.

2.3. Molecular dynamics (MD) simulation

In order to anneal and validate our predicted structure, we carried out a 50 ns MD simulation of the system with the protein embedded in explicit lipid and water box starting with the predicted structure of M59-bound InactiveConf1. The RMSD analysis of the trajectory (Figures S3 and S4 in the Supporting Information) and fluctuation of \( R_{50} \) (Figure 6) all suggest that the protein starts to rearrange to a different state at \( \sim 41 \) ns. Such fluctuations between slightly different states of the GPCR along the trajectory are typical in GPCR MD simulations during which water is diffusing into and throughout the protein, modulating various hydrogen bonds and other interactions.

The hydrogen-bond distances for various interactions along the trajectory are shown in Figures S5–S13 in the Supporting Information. The constancy of these interactions suggest that the overall protein structural features from the region of 33 ns to 41 ns and that of later times are similar with most structural features maintained at the end of 50 ns trajectory. Thus we consider these structural features to provide a reliable representation of the structure.

### Table 4. (Continued)

<table>
<thead>
<tr>
<th>Ligand</th>
<th>( K_i ) [nm]</th>
<th>UCav ( E ) [kcal mol(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>M40</td>
<td>524</td>
<td>–92.88</td>
</tr>
<tr>
<td>M42</td>
<td>&gt; 1000</td>
<td>–87.04</td>
</tr>
</tbody>
</table>

[a] The structure given immediately below the data for each ligand represents the best docked pose ligand interaction diagram (LID). The LIDs were generated using Maestro 9.3. The 3D visualization of these poses is shown in Figure 5. For M59 and M60, the ligand position is clearly dominated by the hydrogen-bond interactions: green; polar interactions: blue; hydrogen bonds (cutoff distance: 2.5 \( \AA \)); purple arrows; \( \pi \cdots \pi \) stacking: straight green lines.
Figure 5. Predicted 3D structures of the best docking pose of a) M59, b) M60, c) M38, d) M40 and e) M42. Each pose is presented in both the side view and the top view. Ligand carbon: purple, protein carbon: cyan.
We find that the ionic lock between R137$^{3.50}$ and E243$^{6.30}$ on the intracellular end of TM6 breaks after 20 ns (Figure S14 in the Supporting Information). Then, R137$^{3.50}$ establishes electrostatic interactions with D136$^{3.49}$, while E243$^{6.30}$ forms a salt bridge with R241$^{6.28}$ on intracellular loop 3 (IC3). The strong electrostatic interaction with the loop explains the changes in the interatomic distance between R137$^{3.50}$ and E243$^{6.30}$ (Figure S14 in the Supporting Information). D136$^{3.49}$ also makes a polar interaction with R151 on intracellular loop 2 (IC2), as shown in Figure 7. Similar polar interaction patterns were also observed in the X-ray crystallographic structure of mOPRM, coupling D164$^{3.49}$ and R165$^{3.50}$ in the DRY motif, and coupling D164$^{3.49}$ and R179 on IC2 (Figure 7).

[12] The X-ray structure for h$\beta_2$AR was also found to have analogous patterns (R131$^{3.50}$ interacts with D130$^{3.49}$ rather than E268$^{6.30}$, and at the same time D130$^{3.49}$ has polar interaction with S143 in IC2).

[13] The observation that the salt bridge between the intracellular sides of TM3 and TM6 is not formed in the antagonist-bound mOPRM structures supports our observation for the predicted hSSTR5. In addition, all other interhelical interactions between side chains found in the apo hSSTR5 structure remain intact except for the one involving D119$^{3.32}$, because D119$^{3.32}$ now engages in interaction with the ligand. Among the interhelical interactions, K227$^{5.64}$–Y138$^{3.51}$ as well as T117$^{3.30}$–S171$^{4.57}$ become water-mediated.

Figure 6 shows how $R_{\infty}$ changes during the 50 ns MD. It remains in the inactive state range 87% of the time if we set the inactive/active cut-off to be 8.0 Å, and 71% of the time if we set the cut-off at 7.5 Å. We do not find much rotation of TM3 or TM6 relative to each other because the closest backbone atoms between the intracellular side of TM3 and TM6 remain to be between the Cα atoms of R137$^{3.50}$ and T247$^{6.34}$ 81% of the time. In addition, the polar interaction between R137$^{3.50}$ and T247$^{6.34}$ side chains converges to a water-mediated hydrogen bond with a length of ~4.5 Å although the bond length starts from 5.2 Å and quickly drops to 1.9 Å (Figure S11 in the Supporting Information). This is an intriguing result because the polar interaction R165$^{3.50}$–T279$^{6.34}$ is also found in the antagonist-bound mOPRM X-ray structure, and mutating T279$^{6.34}$ to lysine (which most likely breaks this polar interaction) can result in a constitutively active mOPRM receptor. [12] We can therefore infer that the distance between R137$^{3.50}$ and T247$^{6.34}$ is also critical in determining hSSTR5 activity, and mutating T247$^{6.34}$ to a lysine residue is likely to give a constitutively active hSSTR5. More importantly, since both R137$^{3.50}$ and T247$^{6.34}$ are conserved in all hSSTRs, this hypothesis might be extendable to all hSSTRs.

Analysis of the changes of several protein–ligand interactions during the MD simulation finds that all the protein–ligand salt bridges become water-mediated during the dynamics. The salt bridge with D119$^{3.32}$ starts to be water-mediated after 4.47 ns, and the one with R39$^{1.31}$ starts to be water-mediated after 2.41 ns. Although there are fluctuations during the 50 ns process, from 33 ns to 41 ns, the distance between M59 piperidine amine nitrogen atom and D119$^{3.32}$ carboxylic acid...
oxygen atom fluctuates around 4.5 Å (Figure S12 in the Supporting Information). The distance between the carboxylic acid oxygen atom of M59 and the amine nitrogen atom of R39 fluctuates more vigorously as shown in Figure S13 (see Supporting Information), but this could be because R39 becomes part of the loop during the dynamics. Although the protein has a structural shift right after 41 ns, and the protein can be in different states a few nanoseconds before and after 41 ns, the two states have many similar antagonist-bound inactive-state characteristics. Therefore, we conclude that the MD simulation retains the character of an antagonist-bound inactive-state structure.

3. Conclusions

In this study, we predicted the ensemble of low-energy structures of hSSTR5 and found plausible binding sites for a series of antagonists with a common scaffold but a diverse set of binding constants. We obtained binding energies consistent with the experimental binding constants. Also, these structures exhibit a TM3–TM6 coupling associated with an inactive GPCR. This indicates that predicted structures InactiveConf1 and InactiveConf2 are reasonable hSSTR5 inactive-state structures. In addition, we have identified residues that might be critical in antagonist binding to hSSTR5, and the results are able to rationalize the order of experimentally determined binding affinities for the five antagonists in the series. Furthermore, the MD simulations show that our antagonist-bound InactiveConf1 structure gains features consistent with those experimentally found in closely related GPCRs. We also introduced an approach aimed at systematically sampling structures in which TM6 is well separated from TM3 as candidates for active structures in addition to sampling small TM3–TM6 separation inactive structures.

In conclusion, this study provides structural information for the understanding of antagonist binding to hSSTR5 that will likely be useful in designing new small-molecule antagonists for hSSTR5. We have also provided structural features that are possible to be extended to other hSSTRs.

4. Experimental Section

4.1. hSSTR5 structure prediction

4.1.1. PredicTM and secondary structure prediction: determining the seven transmembrane domains (TMDs) and any helical extensions beyond the membrane

We carried out multiple sequence alignments using MAFFT method over all GPCRs having a sequence identity greater than 8.8% (from BLAST) with hSSTR5. Then we used the hydrophobicity values from the White and von Heijne scales to predict the hydrophobicity along the target sequence. Then, we removed noise in the hydrophobicity profile by using the median hydrophobicity values obtained from averaging windows ranging from seven residues to 21 residues. Regions with a hydrophobicity value above zero in the final smooth hydrophobicity profile for hSSTR5 (Figure S1 in the Supporting Information) are defined as "raw" TMDs, leading to exactly seven continuously positive regions expected to correspond to the seven TMDs buried inside the membrane.

The X-ray structures for GPCRs often find helical extensions of the TMDs well beyond what would correspond to the boundary of the membrane (for example, in squid opsin, TM5 and TM6 are helical 25 Å beyond the membrane). To identify these helical extensions protruding from the membrane for each helix, we predicted helix propensity using a cross comparison of consensus results from protein secondary structure prediction servers Porter, SSpro, APSSP2, Jpred, and PSIPRED, all of which predict helical regions using trained neural networks. The raw results are in Figure S2 in the Supporting Information. The final TM helical domains extended from the raw TMDs are denoted as "cap" regions, as indicated in Figure S2 (see Supporting Information).

In PredicTM, we specified the hydrophobic center (HPC) of each helix by one of two criteria:

- "Rawmid" takes HPC to be the geometric midpoint of the raw TMD.
- "Area" integrates the hydrophobicity over the raw TMD, and takes HPC to be the centroid (half the total area on each side).

The HPCs for all chains were taken to be in the same x–y plane (the midplane of the lipid membrane bilayer).

4.1.2. Template selection

The sequence alignment from PredicTM identified three GPCRs for which X-ray structures were available that had the highest sequence identity over the TM region with hSSTR5:

- nociceptin receptor (OPRX HUMAN, hOPRX, 46.79%)
- µ-opioid receptor (OPRM MOUSE, mOPRM, 44.62%)
- κ-opioid receptor (OPRK HUMAN, hOPRK, 40.33%)

The next closest GPCR was human CXC chemokine receptor type 4 (CXCR4 HUMAN, hCXCR4, 32.13%). Thus, for exploring a diverse set of relatively high sequence identity templates, we used hOPRX (PDB ID: 4EA3) and hOPRM (PDB ID: 4DKL) and hOPRK (PDB ID: 4DJH) as templates for our structure predictions. There was no experimental structure available at the time of this study for δ-opioid receptor (OPRD MOUSE). A summary of the selected sequence identity comparison is in Table S1 in the Supporting Information.

4.1.3. Predicting the helical shape: OptHelix versus homology modeling

The predicted shape of the TMD is to be used in the future step that determines the tilts and rotations of the TMD. Two methods were used to predict the shape of TMDs:

- OptHelix generates the helical shape features using energy minimization and molecular dynamics (MD) starting with an α-helix based on the peptide sequence in which residues other than proline, glycine, alanine, serine, and threonine are replaced with alanine. Then after dynamics, the average structure is mutated to the correct sequence.
- Homology to the template shape. Here we start with the backbone from the template (usually from an X-ray structure), mutate it to the new sequence, optimize side chains using Side Chain Rota- mer Energy Analysis Method (SCREAM), and minimize the TMD.

The detailed description of these two methods can be found in the Supporting Information.
4.1.4. Assembling the bundle

With the shape of each TMD determined, the next step is to assemble the helices into a bundle.

Assuming each TMD to be rigid, six parameters uniquely define the orientation of the TM helices: 1) the HPC residue $h$ (which is taken to be at $z=0$ so that all TMDs have their HPC on the same plane); 2) & 3) the Cartesian coordinates ($x,y$) of the HPC; 4) the inclination angle $\theta$ of the helical axis relative to the z-axis; 5) the azimuthal angle $\phi$; 6) the rotation angle $\eta$ of the helix around its own helical axis. Except for TM3, we define the reference point for the rotation angle $\eta$ using the most conserved residue in each TM, which is denoted as n.50 in the Ballesteros numbering scheme. For TM3, we chose 3.32 rather than 3.50 because 3.32 is closer to the center of the helix and is also well conserved.

The hydrophobic centers used in assembling OptHelix helices were obtained using PredicTM (based either on the “area” or “rawmid” criterion as described in Section 4.1.1). The other parameters were obtained using PredicTM (based either on the “area” or “rawmid” criterion as described in Section 4.1.1). The other parameters ($x,y,\theta,\phi,\eta$) were all based on the template protein structure from the Orientations of Proteins in Membranes (OPM) database.\(^{26}\) We chose helical shapes from both the “minrmsd” (minimum root mean square deviation to the average structure) and “mineng” (minimum energy) criteria (for detailed definitions, see the Supporting Information). Thus, for each of the three template proteins, we generated a total of four structures based on OptHelix, for a total of 12. For homology helices, we selected ($x,y,h,\theta,\phi,\eta$) from the template.

Now GENSeMBLE starts with the parameters ($x,y,h,\theta,\phi,\eta$) of a starting structure. It then optimizes first $\eta$ using the BiHelix method,\(^{27}\) and then optimizes ($\theta, \phi, \eta$) using the SuperBiHelix\(^{28}\) method, as described in Section 4.1.5.

4.1.5. Determining the optimum helical rotations and tilts

4.1.5.1. Selecting the optimum helical rotations ($\eta$) using the BiHelix method

Mutating from the template sequence to the target sequence, hSSTRS in this case, is likely to make dramatic changes in some of interhelical interactions. Thus, the first step of GENSeMBLE is to sample all changes in the rotation angles, $\Delta\eta$, from 0° to 360° with a step size of 30°. This leads to $12^2 \approx 35$ million combinations. Rather than construct 7-helix bundles for all 35 million, the BiHelix method simplifies the problem by considering the 12 pairs of interacting helices independently. Thus for each pair, we considered $12^2 = 144$ cases, for each of which we optimized the residue side chain conformations with SCREAM. This set of $12 \times 144 = 1728$ numbers was used to estimate the energies for all 35 million combinations. Then we took the lowest 1000 combinations to analyze in the CombiHelix step.

In the CombiHelix step for each of the 1000 combinations from BiHelix, we built the full 7-helix bundle, reoptimized the side chains using SCREAM and minimized for 10 steps. Then the total energies from these 1000 were ordered, and the lowest energy cases were kept for consideration of the optimum tilt angles. We considered four ways to compare the energies:

- The total energy of the charged protein (CTotal)
- The interhelical energy (CInterH), which neglects the intrahelical energy of each chain
- The total energy (NTotal)
- The interhelical energy (NInterH) of the neutralized protein

We find that the isolated net charges on aspartic acid, glutamic acid, lysine, and arginine particularly on the external surfaces of the protein can cause what we consider to be artifacts in the energetics. Thus we neutralize these charged residues by adding or subtracting a proton for surface residues and transferring a proton within each salt bridge. This leads to two sets of energy: NTotal and NInterH.

From many previous studies, we found that the most reliable scoring criterion for identifying the best structures is to combine these four criteria. Thus each conformation was ranked by $E_{\text{comb}}$, which is the average energy of CTotal, CInterH, NTotal, and NInterH. This averaging method puts more weight on interhelical energies than on intrahelical energies. We have validated that for the known X-ray structures, and this procedure correctly identifies the known X-ray rotation angles as the lowest energy structures.\(^{28}\)

4.1.5.2. Selecting the optimum helical tilts ($\theta, \phi$) and rotations ($\eta$) using the SuperBiHelix method

Our previous studies showed that, starting with the X-ray structure of one GPCR, we could not match the structure of a different GPCR without allowing both the helix rotations and helix tilts to change. To make this search practical, we developed the SuperBiHelix method.\(^{28}\) Starting with the optimum rotation angles from BiHelix, we first carried out a coarse sampling step ($\Delta\theta = 0, \pm 15^\circ$; $\Delta\phi = 0, \pm 45^\circ, \pm 90^\circ$; $\Delta\eta = 0, \pm 30^\circ$; selected angles from BiHelix), which involved more than $(3 \times 5 \times 3) \approx 374$ billion configurations. The selected angles of $\Delta\eta$ apart from 0 and $\pm 30^\circ$ were those appearing more than twice in top 20 structures of the BiHelix result or appearing in top 3, and are shown in Table S3 in the Supporting Information. The energies for all these configurations were estimated using the BiHelix energies but combined into three groups of quad helices as explained by Bray and coworkers.\(^{28}\) For the top 2000 combinations of tilts and rotations, we built the 7-helix bundles, optimized the side chains and selected the best case based on the energy ranking.

The subsequent finer SuperBiHelix sampling was based on the selected coarse sampling resulting structures, and the sampling range was $\Delta\theta = 0, \pm 15^\circ$; $\Delta\phi = 0, \pm 15^\circ, \pm 30^\circ$; $\Delta\eta = 0, \pm 30^\circ$. Again, the top 2000 combinations were built into 7-helix bundles.

4.2. Docking of antagonists

4.2.1. Ligand preparation

Our strategy for docking (GenDock) is to select a diverse set of low-energy ligand conformations for each of which we sample a complete set of poses. The initial structure of M59 was built based on the X-ray crystallographic structure of a molecule, M48, which has the same benzoxazole piperidine scaffold.\(^{21}\) To decrease the torsional sampling space, we replaced the ethoxy groups in M59 with methoxy groups for the sampling of conformations. This conformational search involved six rotatable $\eta$-values from $\eta_0$ to 360° with a step size of 30°. This leads to $12^7 \approx 35$ million combinations. The subsequent finer SuperBiHelix sampling was based on the selected coarse sampling resulting structures, and the sampling range was $\Delta\theta = 0, \pm 15^\circ$; $\Delta\phi = 0, \pm 15^\circ, \pm 30^\circ$; $\Delta\eta = 0, \pm 30^\circ$. Again, the top 2000 combinations were built into 7-helix bundles.

4.2.1. Ligand preparation

Our strategy for docking (GenDock) is to select a diverse set of low-energy ligand conformations for each of which we sample a complete set of poses. The initial structure of M59 was built based on the X-ray crystallographic structure of a molecule, M48, which has the same benzoxazole piperidine scaffold.\(^{21}\) To decrease the torsional sampling space, we replaced the ethoxy groups in M59 with methoxy groups for the sampling of conformations. This conformational search involved six rotatable $\eta$-values from $\eta_0$ to 360° with a step size of 30°. This leads to $12^7 \approx 35$ million combinations. The subsequent finer SuperBiHelix sampling was based on the selected coarse sampling resulting structures, and the sampling range was $\Delta\theta = 0, \pm 15^\circ$; $\Delta\phi = 0, \pm 15^\circ, \pm 30^\circ$; $\Delta\eta = 0, \pm 30^\circ$. Again, the top 2000 combinations were built into 7-helix bundles.

705 conformations generated from the previous step. The clustering criterion is RMSD of 0.5 Å. Then we added the terminal methyl groups back to M59m, rotated the O–C bond in the ethoxy groups and generated five possible M59 conformations from each M59m conformation. Similarly, we modified the M59 structures to obtain structures for the other antagonists.

The charge distribution was obtained by the Mulliken population analysis using B3LYP/6–311G** in Jaguar 7.6.[31]

4.2.2. Scanning the complete set of poses for each ligand conformation

The DarwinDock procedure modifies the receptor structure to replace the six types of hydrophobic residues by alanine, and then samples the complete set of poses (–500 000) for regions that could potentially bind a ligand. To do this sampling, the potential binding region is filled by SphGen with “spheres” having 2 Å overlaps with each other and the spheres classified into “boxes” of 10 Å sides. Boxes containing 75 or more spheres were kept. For docking purpose, we have discarded all spheres except for those that are in the extracellular half of the GPCR TMDs and are not potentially in contact with the membrane lipids (i.e. are in the interior of the GPCR helix bundle).

4.2.3. Docking procedure

For each of the five protein structures and for each ligand conformation, we generated 200 000 poses without energy evaluation aiming at providing a complete set of poses. The poses were clustered into –7 300 Voronoi families based on RMSD and the binding energy of the family head evaluated. Then for the top 10% of families, we evaluated the energy for all children. Then we selected the top 50 based on each three energy scores: polar energy, hydrophobic energy, and total energy. Then for these 150, we dealanized (mutating alanine back to the original hydrophobic residues) and optimized the side chains using SCREAM. Then the protein–ligand complexes (poses) were subject to minimization. Then a simulated annealing was performed on the lowest energy 15 complexes before another minimization was done. All final poses were scored together by unified-cavity (UCav) energy. The UCav energy of a particular pose is defined as the binding energy of the ligand of this pose and the union of the binding pocket (cavity) of all poses. A “cavity” is defined as residues within 5 Å of the ligand of a particular pose.

4.3. Molecular dynamics (MD)

4.3.1. Loop building

The protein structure prediction and ligand docking described in Section 4.1 and 4.2 were all done on TM bundles without the connecting intracellular (IC) and extracellular (EC) loops. In order to do MD simulations on the best protein–ligand complexes from previous steps (best M59–InactiveConf1 pose), we constructed the loops connecting the TMDs as follows. The IC2, IC3 and EC3 loops have high sequence identity with the template mOPRM, so these loops were constructed using homology methods, mutating the aligned residues in mOPRM according to the hSSTR5 sequence. Loop EC2 has a disulfide bond with TM3, which is similar to that in mOPRM. The disulfide bond induces loop EC2 in mOPRM to have a hairpin structure, but hSSTR5 has fewer residues in the hairpin. Thus, we constructed EC2 of hSSTR5 by cutting off the extra residues and mutating the remaining residues for EC2 of mOPRM. The remaining loops, IC1 and EC1 were built with a Monte Carlo technique that grows geometrically allowed loop structures from the two fixed TM ends. Then we added the C terminus of hSSTR5 up to the C terminus of Helix 8 (C320) by attaching Helix 8 of the template after aligning their NPxxY motifs followed by mutating to hSSTR5. In addition, we added the N terminus from residue 36–38. Minimization of 500 steps was then carried out on the final structure while keeping the TMDs fixed, and then the whole structure was subjected to minimization of 500 steps.

4.3.2. Building the lipid/water environment

Using Visual Molecular Dynamics (VMD),[34] this final protein structure from the step above was inserted into a 75 Å × 75 Å lipid bilayer structure. The system was then placed into a water box with a total of ~8300 water molecules in the 15 Å and 25 Å thick space on the extracellular and intracellular sides of the lipid bilayer. In addition, Na+ and Cl− ions were placed into the system for a physiological NaCl concentration and a neutral system. The final system had a total of 43 104 atoms.

4.3.3. MD simulation

We used the CHARMM22 force field for the protein, and the CHARMM27 force field for lipids and water as implemented in NAMD 2.6.[35] The conjugate gradient method was used in minimization. The Nose–Hoover Langevin piston pressure control was used in the dynamics. The MD protocol had four steps:

1) With the protein and ligand fixed, the lipids and water molecules in the system were minimized for 5000 steps using the conjugate gradient method.

2) With the protein and ligand fixed, the lipids and water molecules were equilibrated at 310 K and 1 atm for 1.5 ns using the NPT ensemble. This allowed the water molecules to diffuse into the ligand–protein system to modify the interactions.

3) The whole system (43 104 atoms) was then minimized for 5000 steps using the conjugate gradient method.

4) The whole system was heated from 0 K to 310 K in –30 ps and then equilibrated for 50 ns using the NPT ensemble.

Acknowledgements

The authors thank Adam R. Griffith for providing the docking program GenDock and Vaclav Cvicek for providing the Monte Carlo loop-building program. Support from donors to the Materials and Process Simulation Center (MISC), California Institute of Technology (USA) is acknowledged.

Keywords: computational chemistry · docking · G protein-coupled receptors (GPCRs) · molecular dynamics · protein structures · somatostatin receptor subtype 5 (SSTR5)
