DOI: 10.1002/cmdc.201402087

Predicted Ligands for the Human Urotensin-II G Protein-Coupled Receptor with Some Experimental Validation

Soo-Kyung Kim,[a] William A. Goddard, III,*[a] Kyu Yang Yi,[b] Byung Ho Lee,[b] Chae Jo Lim,[b] and Bartosz Trzaskowski[c]

Human Urotensin-II (U-II) is the most potent mammalian vasoconstrictor known.[1] Thus, a U-II antagonist would be of therapeutic value in a number of cardiovascular disorders.[2] Here, we describe our work on the prediction of the structure of the human U-II receptor (hUT_2R) using GEnSeMBLE (GPCR Ensemble of Structures in Membrane BiLayer Environment) complete sampling Monte Carlo method. With the validation of our predicted structures, we designed a series of new potential antagonists predicted to bind more strongly than known ligands. Next, we carried out R-group screening to suggest a new ligand predicted to bind with 7 kcal mol^{-1} better energy than 1-{2-[4-(2-bromobenzyl)-4-hydroxypiperidin-1-yl]ethyl}-3-(thieno[3,2-*b*]pyridin-7-yl)urea, the designed antagonist predicted to have the highest affinity for the receptor. Some of these predictions were tested experimentally, validating the computational results. Using the pharmacophore generated from the predicted structure for hUT_2R bound to ACT-058362, we carried out virtual screening based on this binding site. The most potent hit compounds identified contained 2-(phenoxymethyl)-1,3,4-thiadiazole core, with the best derivative exhibiting an IC_{50} value of 0.581 μM against hUT_2R when tested in vitro. Our efforts identified a new scaffold as a potential new lead structure for the development of novel hUT_2R antagonists, and the computational methods used could find more general applicability to other GPCRs.

Introduction

Urotensin-II (U-II) isolated from goby fish urophysis is a C-terminal conserved cyclic peptide (ETPDCFWKYCA) with a disulfide constraint.[3] The human isoform of U-II, an undecapeptide (AGTADCFWKYCV), is the most potent mammalian vasoconstrictor known—10 to 100 times more potent than endothelin-1.[1,3] The identification of this peptide as the endogenous ligand for the orphan G protein-coupled receptor (GPCR), originally designated GPR14,[1] which is mainly expressed in the cardiovascular system, has stimulated interest in developing novel nonpeptidic human U-II receptor (hUT_2R) agonists and antagonists that might be of therapeutic value in cardiovascular disorders characterized by increased vasoconstriction, myocardial dysfunction, and even atherosclerosis.[2]

The sequence WKY for hUT_2R appears to be very important for biological activities, whereas the disulfide bridge of hUT_2R is of minor importance. Alanine scans of the native undecapeptide revealed that the $W^7K^8Y^9$ triad is required for receptor recognition and activation.[4] Correlating this pharmacophore study, all reported compounds of hUT_2R antagonists contain a basic amino group and at least two aromatic moieties.[5] The recently described P5U superagonist, with higher affinity than U-II at hUT_2R as well as higher activity in the rat thoracic aorta assay, where a cysteine residue in position 5 is replaced by penicillamine (β,β-dimethylcystein), adopts a well-defined type II′ beta-hairpin structure as determined by nuclear magnetic resonance (NMR) studies.[2] Structure–activity relationship (SAR) studies indicate that the replacement of the tryptophan residue with the corresponding D-isomer switches the activity from agonist to partial agonist. Similarly, partial agonists are obtained by replacement of the lysine residue with ornithine, which lacks one methylene unit. The simultaneous presence of a D-tryptophan residue in position 7 and an ornithine residue in position 8 leads to the potent antagonist urantide.[6] The pharmacophoric distances of key residues, W^7, K^8, Y^9, for the hUT_2R interaction, provided the first successful application of virtual screening (VS) based on a pharmacophore generated from SAR and NMR studies on the physiological peptidic ligand of its receptor.[4]

In order to understand the species selectivity of human and rat UT_2R, we previously studied a hUT_2R-selective antagonist, ACT-058362 (hUT_2R K_d: ~4 nM vs rUT_2R K_d: ~1500 nM),[7] and a non-selective arylsulphonamide-based antagonist,[8] SB-706375 (hUT_2R K_d: ~9 nM vs rUT_2R K_d: ~21 nM).[2] We investigated the binding site of ACT-058362 and SB-706375 on both human and rat UT_2R to explain the dramatic (~400-fold) lower

[a] Dr. S.-K. Kim, Prof. W. A. Goddard, III
Materials & Process Simulation Center, California Institute of Technology
Beckman Institute (BI);Code (139-74)
1200 E. California Blvd., Pasadena, CA 91125 (USA)
E-mail: wag@wag.caltech.edu

[b] Dr. K. Y. Yi, Dr. B. H. Lee, C. J. Lim
Division of Drug Discovery Research
Korea Research Institute of Chemical Technology
Yuseong-gu, Daejeon 305-343 (Republic of Korea)

[c] Prof. B. Trzaskowski
Chemistry Department, University of Warsaw
Krakowskie Przedmieście 26/28 Warszawa, 00-927 (Poland)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.201402087.

affinity of ACT-058362 for rUT_2R and the similar (~10 nM) affinity of SB-706375 for both UT_2Rs. We predicted that ACT-058362 binds in the transmembrane (TM) 3456 region while SB-706375 binds in the TM 1237 region. These predicted sites explained the known differences in binding ACT-058362 to the rat and human receptor, while explaining the similar binding of SB-706375 compound.

In order to provide a structural model for understanding the experimental results and for predicting new ligands for hUT_2R, we report now the results of applying the "GPCR ensemble of structures in membrane bilayer environment" (GEnSeMBLE) complete sampling Monte Carlo method, which samples ~1.7 trillion combinations of TM helix rotations (η) and tilts (θ, ϕ) for each of 12 cases of 9 templates from experiment (X-ray) or previous calculations to select an ensemble of 10 low-energy packings to be used to then predict the binding site and energy for both known and new ligands.

With the aim of providing a basis for rational design of new hUT_2R antagonists, we first used the DarwinDock Monte Carlo complete sampling method to predict the binding site to hUT_2R of four known hUT_2R selective antagonists: ACT-058362, SB-706375, GSK-1440115, and GSK-1562590, which will hereafter be referred to as ACT, SB, GSK1, and GSK2, respectively.

1) ACT-058362: a 4-ureido-quinoline palosuran; 1-[2-(4-benzyl-4-hydroxy-piperidin-1-yl)ethyl]-3-(2-methyl-quinolin-4-yl)urea sulfate salt.[8]
2) SB-706375: a arylsulphonamide; 2-bromo-4,5-dimethoxy-*N*-[3-(*R*)-1-methyl-pyrrolidin-3-yloxy]-4-trifluoromethylphenyl)-benzene sulfonamide hydrochloride.[9]
3) GSK-1440115: (4′-[(1*R*)-1-[[(6,7-dichloro-3-oxo-2,3-dihydro-4*H*-1,4-benzoxazin-4-yl)acetyl](methyl)amino]-2-(4-morpholinyl)ethyl]-4-biphenylcarboxylic acid trifluoroacetate.[10]
4) GSK-1562590: 4′-[(1*R*)-1-[[(6,7-dichloro-3-oxo-2,3-dihydro-4*H*-1,4-benzoxazin-4-yl)acetyl](methyl)amino]-2-(1-pyrrolidinyl)ethyl]-3-biphenylcarboxamide hydrochloride.[10]

In each case, we considered the lowest 10 structures from the ensemble of hUT_2R structures. Using this binding site, we proceeded to predict modifications of GSK1 expected to improve binding, and we used virtual screening with the resulting pharmacophore to suggest a new scaffold expected to improve binding.

Results

GEnSeMBLE predictions of apo-protein structures for hUT_2R

Details of the GEnSeMBLE predictions are given in Appendix A in the Supporting Information. For hUT_2R, we considered the following nine X-ray structure templates: human nociceptin/orphanin FQ receptor (hNOPR),[11] mouse delta opioid receptor (mOPRD),[12] mouse mu opioid receptor (mOPRM),[13] kappa opioid receptor (hOPRK),[14] human D_3 dopamine receptor (hD_3DR),[15] turkey β_1 adrenergic receptor ($t\beta_1AR$),[16] human adenosine A_{2A} receptor ($hAA_{2A}R$),[17] human β_2 adrenergic receptor ($h\beta_2AR$),[18] human sphingosine 1-phosphate (hS1P1R),[19] human chemokine $CXCR_4$ ($hCXCR_4$),[20] human H_1 histamine receptor (hH_1HR),[21] rat M_3 muscarinic receptor (rM_3MR),[22] human M_2 muscarinic receptor (hM_2MR),[23] and bovine rhodopsin (bRho).[24] We selected these nine templates based on high sequence identity to hUT_2R from our GEnSeMBLE analysis (TM sequence identities in parentheses): hNOPR (36.0%), mOPRD (34.5%), mOPRM (34.4%), hOPRK (32.8%), hD_3DR (31.3%), $hAA_{2A}R$ (27.9%), $h\beta_2AR$ (27.3%), $hCXCR_4$ (27.1%) and bRho (21.8%). We also selected the templates with lower than 30% sequence identity for the diversity of templates and the validation of our methods. For families in which more than one GPCR is known, we chose the structure with the highest sequence identity.

To obtain the optimum packings of the hUT_2R helices in these templates, we first considered rotating each of the 12 transmembrane domains (TMD) by increments of 15° over the range of ±60°, a total of $(9)^7 = 4.7$ million configurations using the BiHelix method. This was carried out for the following 12 cases:

- Nine cases using homology templates: 1. hNOPR, 2. mOPRD, 3. mOPRM, 4. hOPRK, 5. hD_3DR, 6. $hAA_{2A}R$, 7. $h\beta_2AR$, 8. $hCXCR_4$, 9. bRho.
- One case using the area hydrophobic center (HPC) method to position the helices with respect to a common membrane midplane: 10. $h\beta_2AR$ GEnSeMBLE.
- One case using the rawmid HPC method to position the helices with respect to a common membrane midplane: 11. $h\beta_2AR$ GEnSeMBLE.
- One case that we had previously predicted using an early version of GEnSeMBLE, which had been subjected to 1 ns of MD after inserting in the membrane and water box: 12. hUT_2R MembStruk.[25]

For each of these 12 cases, we selected the lowest energy 2000 combinations out of the 4.7 million possibilities to build into 7-helix bundles (CombiHelix) for which we optimized the side chains followed by some minimization. Then we compared the interhelical interaction energies for all 12*2000 = 24000 cases using an energy scoring in which each amino acid side chain is neutralized (NiH). We finally selected the best 20 to consider optimizing the tilts of the helices.

From these 20 configuration, we selected six based on considerations of energy and diversity, and considered 1.7 trillion combinations of the rotation angles (η) and tilt angles (θ, ϕ) to select the best 2000 to pack into 7-helix bundles. Using an energy scoring in which each amino acid side chain is neutralized (NiH), we selected from these 12000 cases. All ensembles of the top 20 are from the mOPRD template (Table 1), suggesting a close correspondence of UT_2R with mOPRD, which may be responsible for the observation that a side effect of pain relieving drugs that target opioid receptors is to activate hUT_2R.

Figure 1 shows the final best structure for hUT_2R, which derives from mOPRD only by η variations of −15° for TM 5, θ deviations of −10° for TM 4, and ϕ changes of −15° for TM 3–5. However, many of the other low-energy structures have larger deviations that may play a role in activation.

Table 1. Top 20 predicted structures of the human urotensin-II receptor (hUT_2R) selected as the lowest energy configures from the 12000 final structures from the SuperBiHelix analysis of the 6 BiHelix cases (shaded in gray in Table S2 and S3 in the Supporting Information).[a]

Apo-#	θ							ϕ							η							BiHelix	RMSD
	H1	H2	H3	H4	H5	H6	H7	H1	H2	H3	H4	H5	H6	H7	H1	H2	H3	H4	H5	H6	H7	Rank	[Å]
1	0	0	0	10	0	0	0	0	0	−15	−15	−15	0	0	0	0	0	0	−15	0	0	Top1	1.07
2	0	0	0	10	0	0	0	0	0	−15	−15	15	0	0	0	0	0	0	−15	0	0	Top1	1.16
3	0	0	0	0	0	0	0	0	0	0	0	−15	0	0	0	0	0	0	−30	0	0	Top1	0.70
4	0	0	0	0	10	0	0	0	0	0	0	−15	0	0	0	0	0	−15	15	0	0	Top9	1.13
5	0	0	0	0	10	0	0	0	0	−15	−30	−15	15	0	0	0	0	0	−15	0	0	Top9	1.41
6	0	0	0	0	10	0	0	0	0	0	−15	15	15	0	0	0	0	−15	−15	0	0	Top9	1.21
7	0	−10	0	−10	10	0	0	−15	−15	15	−15	−15	−30	−15	−30	−30	−15	15	0	15	0	Top1	2.09
8	0	0	0	0	10	0	0	0	0	−15	−15	−15	15	0	0	0	0	0	−15	0	0	Top9	1.40
9	0	0	0	0	10	0	0	0	0	0	−15	−30	15	0	0	0	0	0	−30	0	0	Top1	1.48
10	0	0	0	0	10	0	0	0	0	−15	0	−15	15	0	0	0	0	−15	−15	0	0	Top9	1.42
11	0	0	0	0	10	0	0	0	0	0	−15	−30	15	0	0	0	0	−15	−30	0	0	Top9	1.49
12	0	0	0	0	0	0	0	0	0	0	15	15	0	0	0	0	15	30	−30	0	0	Top1	0.87
13	0	−10	0	−10	10	0	0	−15	−15	0	15	0	0	0	15	−15	−15	15	−15	0	0	Top9	1.73
14	0	0	0	0	0	0	0	0	0	−15	−15	−15	15	0	0	0	0	−15	−30	0	0	Top9	1.05
15	0	0	0	0	0	0	0	0	0	0	−30	−15	0	0	0	0	0	15	−30	0	0	Top1	0.76
16	0	0	0	0	0	0	0	0	0	0	−15	15	0	0	0	0	0	15	−30	0	0	Top1	0.73
17	0	0	0	0	10	0	0	0	0	0	−30	−15	15	0	0	0	0	−15	−15	0	0	Top9	1.19
18	0	0	0	10	10	0	0	0	0	−15	−15	−30	30	0	0	0	0	0	−30	0	0	Top1	1.72
19	0	0	0	0	10	0	0	0	0	0	−15	−15	15	0	0	0	0	−15	15	0	0	Top9	1.18
20	0	0	0	0	0	0	0	0	0	−15	−15	−30	15	0	0	0	0	0	0	0	0	Top9	1.09

[a] This SuperBiHelix analysis considered: $\pm 10^\circ$ sampling of θ tilt angles, $\pm 30^\circ$ sampling of both ϕ and η angles by 15° increments, leading to a total of $(3\times5\times5)^7 \sim 13$ trillion combinations. The structures were ordered by the average rank from four energy scorings: charge interhelical (CiH), charge total (CTot), neutral interhelical (NiH), and neutral total (NTot). The top 20 structures identified all came from the mOPRD template.

The low-lying packing structures have important interhelical couplings. Classical interhelical H-bonding networks were found in TM 1,2–7 (N1.50, D2.50 and N7.49) between D2.50 and N7.49 except for apo-1 and between N1.50 and D2.50 in close proximity of 5 Å among the top 20. However, none of the top 20 structures contain a salt-bridge between R3.50 and D(E)6.30 in the D/ERY region that is observed in many other class A GPCRs. This is expected because residue 6.30 is K273 instead of D or E.

Instead, we find that R155$^{4.42}$ makes a salt bridge with E147$^{3.49}$ coupling TM 3 and TM 4 in the D/ERY region, as do apo-5, 8, 10, 11, and 14 out of the top 20 structures. There are H-bonding networks among S85$^{2.38}$, Y89$^{2.42}$ and E147$^{3.49}$, coupling TM 2 with TM 3 in the D/ERY region between S2.38 and E3.49 (except for apo-7, 11, 13, 17, 19) and between Y2.42 and E3.49 (except for apo-11, 12, 15, 17, and 19 out of the top 20 structures). TM 2–4 H-bonding networks—either between N92$^{2.45}$ and R166$^{4.42}$ for apo-1, 2, 7, 13, 18 or between S85$^{2.38}$ and R166$^{4.42}$ for apo-3-6, 8-11, 1, 17, 19—in the top 20 structures are expected to make these configurations inactive.

Some structures like apo-5, 8, and 10 have an extra TM 3–5 H-bonding network between R148$^{3.50}$ and Y231$^{5.58}$. Some structures have additional TM 3–5 interactions between H135$^{3.37}$ and S219$^{5.46}$ at the middle of TMD for apo-2, 14, 20, as well as between S145$^{3.47}$ and Y231$^{5.58}$ only for apo-5 in the proximity of the D(E)RY region. Apo-5-11, 14, 17, 19 and 20 structures have TM 2–6 among Y87$^{2.40}$ and R255$^{6.32}$. Finally, all structures except for apo-13 have TM 1–7 interaction between D47$^{1.38}$ and R294$^{7.32}$.

Predicted binding sites for the predicted hUT_2R structures

Predicting the binding site of ACT to hUT_2R

We used the DarwinDock complete sampling Monte Carlo method to predict ~50000 poses for each of the seven conformations of ACT to finally select the best binding site (for details, see Experimental Section). We did this for each of the top 10 low-energy 7-helix bundles predicted from SuperBiHelix (all based on mOPRD). We found that the best structure binds to apo-10 of the ensemble based on the best binding energy from all ensemble docking conformations.

As shown in Figure 2A, critical components of the binding are: 1) a salt bridge from the protonated nitrogen to D130$^{3.32}$; 2) an H-bond of the carbonyl oxygen atom with T305$^{7.42}$; 3) the phenyl group in a hydrophobic pocket formed by I107$^{2.60}$ and L126$^{3.28}$; 4) the indole ring in a hydrophobic pocket formed by L212$^{5.39}$ and, L215$^{5.42}$, F216$^{5.43}$, I220$^{5.47}$ and W275$^{6.48}$.

To validate this predicted binding pose for ACT, we matched five ACT analogues to the low-energy predicted structures of hUT_2R from DarwinDock and minimized. Two of the binding modes of ACT63 and ACT97 (the number from the conformational search) have a salt bridge at D130$^{3.32}$, which is the major anchoring point in hUT_2R. The binding mode of ACT63, which has a *cis* amide bond in the ureido group, shows additional interactions at L215$^{5.42}$ with the terminal protonated amine in the indoline ring, while the binding mode of ACT179, which has a *trans* conformation, and ACT211 with opposite orientation, has no additional H-bonding.

We examined five ACT analogues (**ACT-1** to **ACT-5**) to assess how well our predicted structure explains the experimental SAR (Table 2). Matching the five ACT analogues into the three

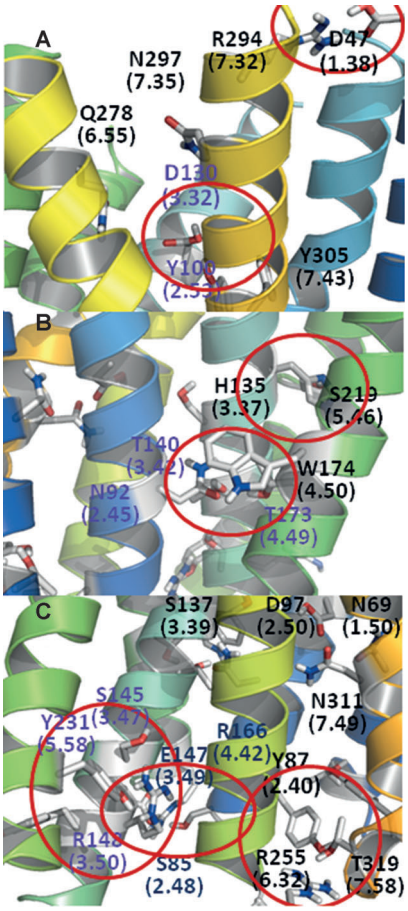


Figure 1. Modelled H-bonding networks of human urotensin-II receptor (hUT_2R) based on the mouse delta opioid receptor (hUT_2R-mOPRD)[12] in upper (A), middle (B), and lower (C) transmembrane (TM) regions. The salt bridge or H-bonding is shown in a red circle. We used Ballesteros–Weinstein numbering consisting of the TM helix number followed by residue number relative to the highly conserved residue in the helix, numbered as 50.

best binding poses, ACT63, ACT97, and ACT211, we found that the binding mode of ACT97 has the best correlation ($r^2 = 0.83$) compared with the binding modes of ACT63 ($r^2 = 0.18$) and ACT211 ($r^2 = 0.33$). Thus, only this binding mode of ACT97 agrees with the experimental observations.

From our docking study, we find that the protonated amine in the corresponding part of K^8 interacts with $D130^{3.32}$ while the indoline ring of W^7 interacts with $L215^{5.42}$ and $W275^{6.48}$. For the Y^9 position, the terminal phenyl group forms a pi–pi interaction with $Y111^{2.64}$.

Docking of SB to hUT_2R

We predicted the binding site and energy of the SB antagonist to the top 10 low-energy predicted structures of hUT_2R–mOPRD from SuperBiHelix. The observed IC_{50} value for SB against the human reception is 0.073 μM, three times higher than the experimental IC_{50} value of ACT ($IC_{50} = 0.024$ μM). The conformational search of SB generated 17 conformations out of 360 conformations with the diversity of 2.5 and 1.5 Å RMSD. We selected six extended conformations for the docking. The best pose of SB bound to the hUT_2R–mOPRD leads to the structure in Figure 2B with the following features: 1) a salt bridge between the protonated amine and $D110^{3.32}$, the major anchoring point for SB to hUT_2R; 2) an H-bonding at $Y306^{7.43}$ with the O atoms of the sulfonamide group; 3) a hydrophobic interaction at the terminal methoxy phenyl group with hydrophobic residues ($F111^{2.64}$ and $Y298^{7.36}$); 4) a pi–pi interaction at the center phenyl group with aromatic residues $F274^{6.51}$ and $F127^{3.29}$.

Docking of GSK1 and GSK2 to hUT_2R

We predicted the binding site and energy of GSK1 ($IC_{50} = 0.048$ μM) to the top 10 low-lying packing protein conformations from SuperBiHelix, predicting the structure shown in Figure 3A. The best pose is from the conformation number 40 out of seven different ligand conformations and WT (wild-type) 10 out of top 10. The predicted structures have the following features: 1) a salt bridge at $D130^{3.32}$ to the protonated amine; 2) an H-bonding between the carbonyl O atom and $Y111^{2.64}$; 3) an aromatic–aromatic interaction at $F127^{3.29}$ (-5.63 kcal mol^{-1}), $Y298^{7.35}$ (-4.04 kcal mol^{-1}), $F274^{6.51}$ (-2.86 kcal mol^{-1}), and $F131^{3.33}$ (-2.72 kcal mol^{-1}) with the biphenyl group; 4) a hydrophobic interaction at the hydrophobic residues, $F131^{3.33}$, $M134^{3.36}$, $L212^{5.39}$, $L215^{5.42}$, and $W275^{6.48}$ with the benzoxazolinone ring.

The binding of GSK1 (UCavE = -42.09 kcal mol^{-1}) is 2.72 kcal mol^{-1} less favorable than the binding of ACT (UCavE = -44.81 kcal mol^{-1}), which is consistent with the observed two-times better binding of ACT. We also predicted the binding site and energy of GSK2 ($IC_{50} = 0.004$ μM) to 10 low-lying packing protein conformations from SuperBiHelix, predicting the structure shown in Figure 3B and displaying the same interactions as GSK1.

Predicting new ligands using multiple R-group optimization

Starting from the predicted best binding site of **ACT-1** bound to the apo-10 conformation of hUT_2R from mOPRD, we searched multiple R groups, considering the optimization for three positions as shown in Table 3. Since a CH_3 group in the R^1 position is better than H, we considered the functional groups of CH_3, CH_2CH_3, $CH=CH_2$, and CH_2OH in the R^1 position. We found that single substitution at R^1 with any of the four substituents increased the binding affinity. Since a Br group in the R^2 group is slightly better than H in the R^1 position, we considered OCF_3, CF_3, N=C=S and I in the R^2 position as plausible analogues of Br. We found that for single substitution at the R^2 position, the CF_3 and OCF_3 groups are best. To optimize the R^3 position, we considered CH_3, Cl, I, and phenyl groups to providing maximum diversity. We found that for single substitution at the R^3 position, I, Cl, and CH_3 are better than H (**ACT-1**).

To suggest new ligands with improved binding, we then considered nine di-substituted derivatives, and finally 22 candidates in Table 3 with up to triple excitations. Based on this R-

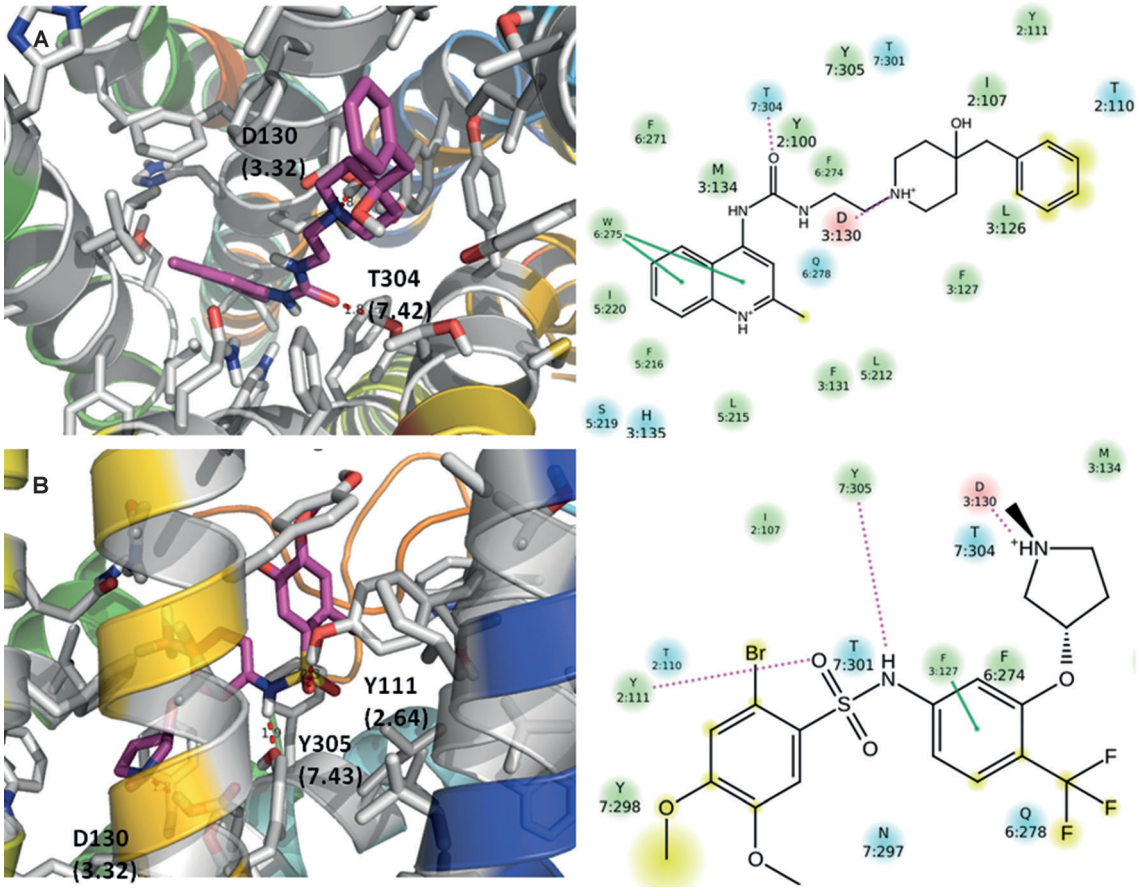
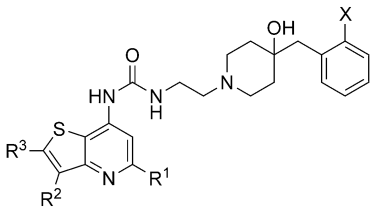


Figure 2. The docking result of human urotensin-II receptor (hUT_2R)-selective ligands at the hUT_2R-mOPRD model derived using the mouse delta opioid receptor (mOPRD) as a template.[12] The binding site and 2D ligand interaction diagram, determined using the Maestro program[46] for A) ACT-058362 and B) SB-706375. We used Ballesteros–Weinstein numbering consisting of the TM helix number followed by residue number relative to the highly conserved residue in the helix, numbered as 50. H-bonding is indicated by the dotted lines.

Table 2. The experimental IC_{50} and predicted binding energy of ACT-058362 analogues.[a]



Compd	R^1	R^2	R^3	X	IC_{50} [μM]	UCavE [kcal mol^{-1}]	BE [kcal mol^{-1}]
ACT-058362	H	H	H	H	0.015 ± 0.008	−37.08	−51.94
ACT-1	H	H	H	Br	0.021 ± 0.007	−35.42	−63.28
ACT-2	Me	H	H	H	0.026 ± 0.003	−36.37	−61.41
ACT-3	H	H	H	H	0.030 ± 0.010	−39.15	−65.80
ACT-4	H	H	Me	H	0.069 ± 0.048	−37.45	−62.48
ACT-5	H	Br	H	H	3.607 ± 0.966	−23.62	−64.58

[a] Unified cavity energy (UCavE) and binding energy (BE) were calculated; BE = ${}^{complex}E - {}^{protein}E - {}^{ligand}E$). IC_{50} data shown are the mean ± standard deviation of three separate experiments.

group screening for **ACT-1**, we suggest that CH=CH_2, CH_2OH, and CH_3 in the R^1 position and I, Cl, and CH_3 in the R^3 position are the best candidates for improved hUT_2R interactions (Figure 4). Comparing the binding of **ACT-1** and the other analogues, we predict that these compounds might bind up to ~7 kcal mol^{-1} better energetically.

Predicting new ligands using virtual screening

To identify novel hUT_2R-selective lead compounds, we performed virtual screening using the predicted pharmacophore of ACT to three sets of hUT_2R structures, hUT_2R_fRho, hUT_2R_$CXCR_4$ and hUT_2R_mOPRD. Experimental tests were carried out for the hUT_2R_fRho results. Thus, to compare with experimental results, we will discuss just the results for hUT_2R_fRho.

We used LigandScout 3.0[26] to for search new leads for hUT_2R ligands. Based on the pharmacophore score, we selected 5036 hit molecules from the ZINC database that includes 19.6 million compounds, each with 100 conformations. For each of these 5036 compounds, we used our predicted protein structure with our scoring methods to select the best 200 poses from which we selected the best 10 molecules using two alternative scoring methods. These are listed in Table S1 and Figure S1 of the Supporting Information. Of these top 10 from each scoring method, we selected the six predicted to be best for experimental study, leading to the compounds shown

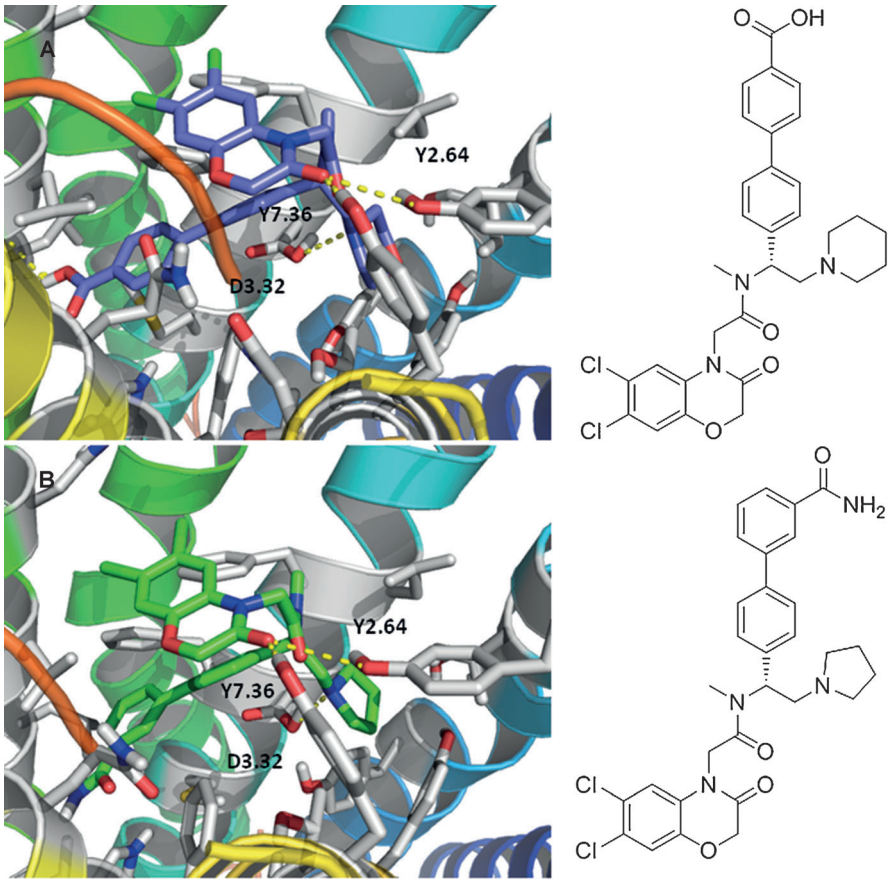


Figure 3. The binding site and 2D ligand interaction diagram, determined using the Maestro program, for A) GSK-1440115 (GSK1) and B) GSK-1562590 (GSK2) at the hUT_2R-mOPRD model derived using the mouse delta opioid receptor (mOPRD) as a template.[12]

in Table 4. The best measured affinity for hUT_2R leads to 89.7% binding inhibition at 10 μM concentration. The top two ranked compounds by calculated UCavE value have similar structures and exhibit the most potent inhibition of hUT_2R; compounds #1 and #2 (ZINC02322315) in Table 4 were determined to have IC_{50} values of 1.40 μM and 0.581 μM, respectively.

Discussions

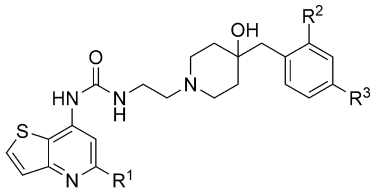
To predict the structures of antagonists bound to hUT_2R that might be useful in developing improved ligands, we applied the GEnSeMBLE method that samples ~13 trillion potential packings (considering rotations and tilting of all seven TMD) to each of nine GPCR templates to finally select the best 10 that might be important in binding antagonists or activation by agonists.

The best 20 7-helix bundles all derive from the mOPRD homology template, which has the second highest sequence identity in TM (34.5%) among the nine templates evaluated. The H-bonding networks are highly dependent on the template chosen. The hUT_2R-mOPRD structure displays an extra H-bonding that is not found in the other two templates: TM 2–6–7 among $Y87^{2.40}$, $R255^{6.32}$ and $T319^{7.58}$, and TM 1–7 between $D47^{1.38}$ and $R294^{7.32}$. Thus, the final best structure from the mOPRD template has the best intermolecular H-bonding networks compared with the other two templates, as indicated by the better energies in the SuperBiHelix analysis.

When we examined the H-bonding networks for all 20 ensembles of the hUT_2R-mOPRD template, we found the classical H-bonding networks in TM 2–4 (N2.45 and W4.50) known for many members of the rhodopsin family A GPCRs (the top 20 also shared this interaction except for apo-7, 13, 17, and 18). We also found hUT_2R-specific interactions: TM 2–4 networks in the D(E)RY region among $S85^{2.48}$, $R166^{4.42}$, and $E147^{3.49}$. The H-bonding network between S2.38 and E3.49 was shown for all except for apo-7, 11, 13, 17, and 19, and the interaction between Y2.42 and E3.49 was shown for all except for apo-11, 12, 15, 17, and 19. In addition, there are TM 2–4 H-bonding networks either between $N92^{2.45}$ and $R166^{4.42}$ for apo-1, 2, 7, 13, 18 or between $S85^{2.38}$ and $R166^{4.42}$ for apo-3-6, 8-11, 1, 17, 19 in the top 20, and TM 3–5 H-bonding networks between $R148^{3.50}$ and $Y231^{5.58}$ for apo-2, 14, 20, as well as between $S145^{3.47}$ and $Y231^{5.58}$ only for apo-5 in the proximity of the D(E)RY region. However, none of the top 20 contain a salt bridge between R3.50 and D(E)6.30 in the D/ERY region, which is observed in many other class A GPCRs. This is expected because residue 6.30 is K273 instead of D or E. Instead, we find that $R155^{4.42}$ makes a salt bridge with $E147^{3.49}$, coupling TM 3 and TM 4 for apo-5, 8, 10, 11, and 14 structures. Thus, diverse ensemble structures were stabilized by various H-bonding networks, and conserved interactions between TM 3 and TM 6 in the D/ERY motif in class A GPCRs were replaced by TM 2–4 networks.

To determine the binding mechanism of hUT_2R antagonists, we docked various hUT_2R-selective antagonists, ACT, SB, GSK1, and GSK2. From the docking of ACT, SB, GSK1, and GSK2 with the apo-10 low-energy predicted structures from SuperBiHelix on hUT_2R-mOPRD, we find major interactions at $D130^{3.32}$ with the protonated nitrogen through a salt-bridge interaction, indicating that such interaction is necessary for hUT_2R recognition. The presence of both aromatic and positively charged group leads to compounds with similar interactions. However, additional H-bonding depends on the ligand type. The main binding difference observed in three structures is an additional H-bond. For ACT, H-bonding forms between the carbonyl oxygen

Table 3. The predicted binding energies from R-group screening of **ACT-1**. Compounds are ordered by unified cavity energy, thus we suggest #32 and 25 as the best candidates.[a]



#	R^1	R^2	R^3	UCavE [kcal mol^{-1}]	SnapBE [kcal mol^{-1}]
32	CH_2OH	Br	I	−42.47	−61.30
17	Me	Br	I	−41.77	−60.17
27	$CH=CH_2$	Br	I	−39.81	−61.09
25	$CH=CH_2$	Br	Me	−39.51	−64.97
16	Me	Br	Cl	−39.50	−56.87
22	Et	Br	I	−39.47	−61.06
31	CH_2OH	Br	Cl	−39.16	−57.15
3	$CH=CH_2$	Br	H	−38.24	−62.87
29	CH_2OH	CF_3	H	−38.22	−55.12
26	$CH=CH_2$	Br	Cl	−38.10	−63.12
30	CH_2OH	Br	Me	−37.96	−57.28
15	Me	Br	Me	−37.80	−56.75
11	H	Br	I	−37.77	−54.67
4	CH_2OH	Br	H	−37.33	−55.64
1	Me	Br	H	−37.19	−54.98
19	Et	CF_3	H	−36.82	−55.72
14	Me	CF_3	H	−36.49	−53.56
21	Et	Br	Cl	−36.49	−57.08
20	Et	Br	Me	−35.57	−57.45
28	CH_2OH	OCF_3	H	−35.36	−53.75
2	Et	Br	H	−35.23	−55.89
13	Me	OCF_3	H	−35.05	−53.05

[a] Unified cavity energy (UCavE) and binding energy (BE) were calculated; BE = ${}^{complex}E - {}^{protein}E - {}^{ligand}E$).

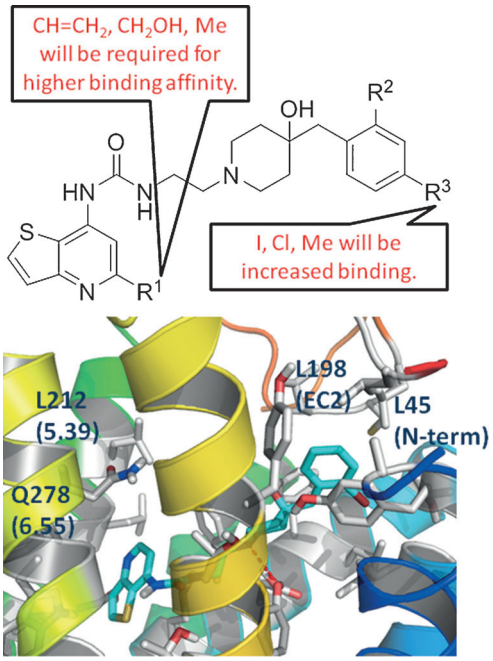


Figure 4. Prediction of improved substituents from R-group screening (top) based on the predicted binding site of ACT-**1** at human urotensin-II receptor (hUT_2R) (bottom). N-term: N-terminal and EC2: the second extracellular loop.

atom and T305$^{7.42}$. However, an additional H-bond is shown at the sulfonamide group of SB with Y306$^{7.43}$/Y111$^{2.64}$ and the carbonyl atom of GSK with Y298$^{7.36}$/Y111$^{2.64}$. The binding site of three classes of hUT_2R antagonists is overlapped at D130$^{3.32}$. Except for SB, which mainly interacts at TM 1–2–7, the main interaction of ACT and the GSK compounds forms at TM 3–5. The benzopyran ring of both GSK1 & 2 binds at the extra sites close to upper TM 7 and the second extracellular loop (EC2). Thus, our docking study predicted a different binding mode for three classes of hUT_2R antagonists, displaying a different spatial disposition of the pharmacophoric elements.

For ligand optimization, we used R-group screening (Table 3), with the predicted structure for **ACT-1** in our predicted hUT_2R structure to select the compounds shown in Figure 4, which we suggest might exhibit improved bind over **ACT-1** to hUT_2R by ~7 kcal mol^{-1}. According to the information from the docking experiments, we could extend the hydrophobic interactions at three leucine residues (45, 198, and 212) from TM 5 and loops.

For lead generation, based on the binding site for **ACT-1**, we used our best ligand–protein structure as the basis for virtual screening. This led to the identification of a new scaffold: 2-(phenoxymethyl)-1,3,4-thiadiazole. The best virtual screening hit has an IC_{50} value of 0.581 μM (Table 4), representing a new lead for the development of hUT_2R antagonists. Since most known hUT_2R antagonists have poor pharmacology profiles, new leads will provide the opportunity to develop improved compounds as selective renal vasodilator in renal ischemia.

Conclusions

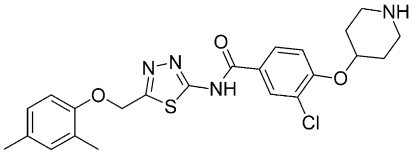
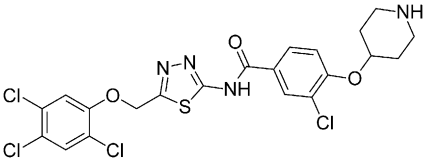
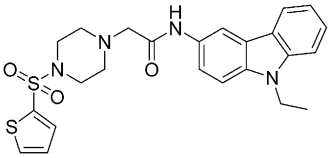
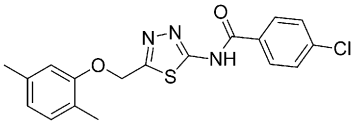
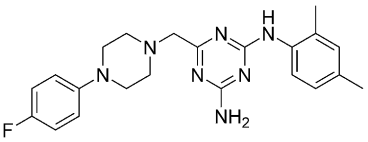
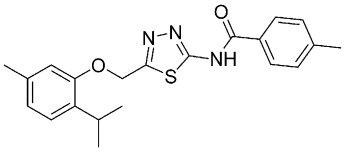
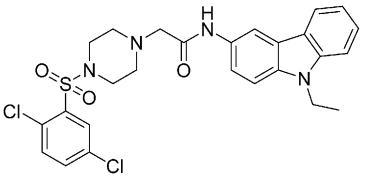
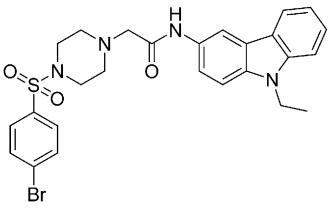
This study shows the importance of considering the ensemble of the best 10 to 20 models of the 7-helix bundle for a GPCR, as exemplified using hUT_2R, a validated target for the treatment of renal ischemia. Thus, the best antagonist binding site was found in the 10th structure of the apo-protein hierarchy. We consider that these GEnSeMBLE and DarwinDock techniques, which estimate the energies for massive numbers of conformations and structures but rapidly decrease them to an ensemble of 10–100 for detailed scoring, will be useful for many other GPCRs, most of which have little experimental data available that is useful for ligand design. Our predicted 3D structures for hUT_2R are expected to provide increased understanding of the binding mechanism of hUT_2R, which should be valuable in developing improved selective antagonists.

Experimental Section

Synthesis of ACT derivatives

The general synthetic route employed for the preparation of ACT derivatives (**ACT-1**–**5**) is outlined in Scheme 1. Thieno[3,2-*b*]pyridinyl carboxylic acids (**1**) were reacted with diphenyl phosphoryl azide

Table 4. The experimentally determined inhibitory activities (inhibition at 10 μM and IC_{50}) of virtual screening hits against human urotensin-II receptor (hUT_2R). Hits were identified by virtual screening against the crystal structure of frog rhodopsin (hUT_2R-fRho) in complex with ACT-058362.[a]

#	Compd	Inhib. @10 μM [%]	IC_{50} [μM]	UCavE [kcal mol^{-1}]	BE [kcal mol^{-1}]
1		89.7	1.40 ± 0.19	−44.20	−47.53
2		87.7	0.59 ± 0.07	−50.88	−58.69
3		17.2	–	−43.79	−59.62
4		0.0	–	−40.49	−47.73
5		0.0	–	−38.33	−52.07
6		0.0	–	−38.22	−46.84
7		20.9	–	−23.05	−54.64
8		13.4	–	−36.68	−52.75

[a] Unified cavity energy (UCavE) and binding energy (BE) were calculated; $BE = {}^{complex}E - {}^{protein}E - {}^{ligand}E$). IC_{50} data shown are the mean ± standard deviation of three separate experiments.

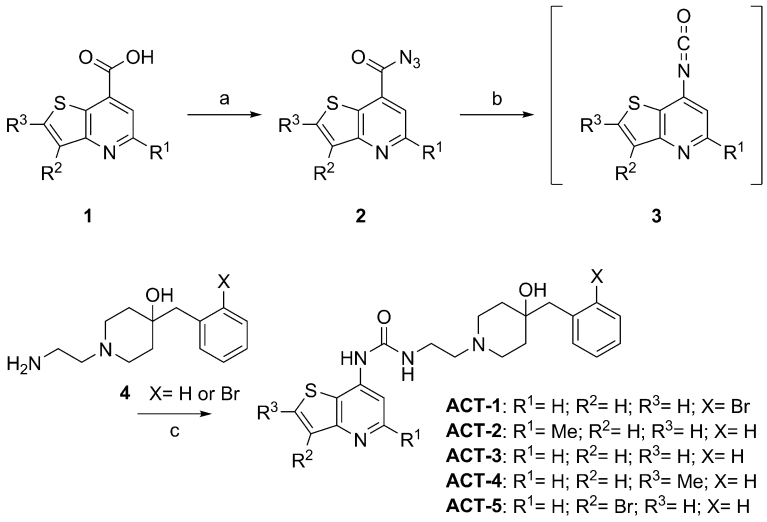
(DPPA) using Et_3N as the base to give corresponding acyl azides **2**. The target ureas were obtained by reaction of thieno[3,2-*b*]pyridinyl isocyanates **3**, which were generated in situ by Curtius rearrangement of azides **2**, with amine **4**.

^{1}H NMR spectra were recorded on a Varian Gemini 200 or Bruker DRX-300 spectrometer. High-resolution mass spectra (HRMS) were recorded on a JEOL JMS-700 mass spectrometer. Column chromatography was carried out using silica gel (230–400 mesh). All solvents and reagents were commercially available and used without purification.

Thieno[3,2-*b*]pyridine-7-carbonyl azide (2): A solution of thieno[3,2-*b*]pyridine-7-carboxylic acid (568 mg, 3.18 mmol) in DMF (10 mL) was treated with DPPA (1.02 mL, 4.77 mmol) and Et_3N (0.89 mL, 6.36 mmol) and stirred at RT for 5 h. The mixture was diluted with water (50 mL), and extracted with EtOAc (2×50 mL). The organic layer was washed with brine (2× 25 mL), dried over anhyd Na_2SO_4, filtered and concentrated in vacuo. Purification of the crude residue by column chromatography on silica gel (EtOAc/*n*-hexane, 1:3) gave **2** as a white solid (195 mg, 30%): ^{1}H NMR (300 MHz, $CDCl_3$): δ = 7.94 (d, J = 5.4 Hz, 1H), 7.83 (d, J = 4.7 Hz, 1H), 7.66 (d, J = 5.4 Hz, 1H), 7.39 ppm (d, J = 4.7 Hz, 1H).

1-[2-(4-Benzyl-4-hydroxypiperidin-1-yl)ethyl]-3-(thieno[3,2-*b*]pyridin-7-yl)urea (ACT-3): A solution of **2** (50 mg, 0.24 mmol) in toluene (3 mL) was heated at 110 °C for 1 h. After cooling to RT, a solution of 1-(2-aminoethyl)-4-benzylpiperidin-4-ol (84 mg, 0.36 mmol) and Et_3N (0.1 mL, 0.72 mmol) in CH_2Cl_2 (3 mL) was added to the reaction mixture, and stirring was continued for 1 h. The mixture was diluted with water (20 mL), and extracted with CH_2Cl_2 (2×20 mL). The organic layer was washed with brine (2×20 mL), dried over anhyd Na_2SO_4, filtered and concentrated in vacuo. Purification of the crude residue by column chromatography on silica gel ($MeOH/CH_2Cl_2$, 20%) gave ACT-**3** as a yellow solid (20 mg, 20%): ^{1}H NMR (300 MHz, CD_3OD): δ = 8.43 (d, J = 3.3 Hz, 1H), 8.01 (d, J = 3.4 Hz, 1H), 7.93 (d, J = 3.3 Hz, 1H), 7.46 (d, J = 3.4 Hz, 1H), 7.20–7.29 (m, 5H), 3.61–3.66 (m, 2H), 3.47–3.51 (m, 2H), 3.17–3.28 (m, 4H), 2.82 (s, 2H), 1.90–1.95 (m, 2H), 1.75–1.81 ppm (m, 2H); ^{13}C NMR (125 MHz, DMSO): δ = 36.1, 36.8, 49.4 (2C), 55.3, 57.6, 60.2, 68.3, 106.6, 121.1, 125.8, 126.2, 128.0 (2C), 130.5, 131.0 (2C), 138.3, 142.6, 148.6, 154.7, 157.6; HRMS (EI): calcud for $C_{22}H_{26}N_4O_2S$ $[M]^+$: 410.1776, found: 410.1796.

1-{2-[4-(2-Bromobenzyl)-4-hydroxypiperidin-1-yl]ethyl}-3-(thieno-[3,2-*b*]pyridin-7-yl)urea (ACT-1): Reaction of 1-(2-aminoethyl)-4-(2-bromobenzyl)piperidin-4-ol (114 mg, 0.37 mmol, 1.5 equiv) as de-



Scheme 1. *Reagents and conditions*: a) diphenyl phosphoryl azide (DPPA), Et_3N, DMF, RT, 5 h; b) toluene, 110 °C, 1 h; c) Et_3N, CH_2Cl_2, RT, 1 h.

scribed for ACT-**3** gave **ACT-1** as a yellow solid (31 mg, 26%): ^{1}H NMR (500 MHz, CD_3OD): δ = 8.41 (d, J = 4.5 Hz, 1H), 8.02 (d, J = 4.5 Hz, 1H), 7.90 (d, J = 5.6 Hz, 1H), 7.54 (d, J = 7.5 Hz, 1H), 7.45 (d, J = 5.6 Hz, 1H), 7.38–7.39 (m, 1H), 7.23 (t, J = 7.0 Hz, 1H), 7.09 (t, J = 7.0 Hz, 1H), 3.87–3.50 (m, 2H), 3.00 (s, 2H), 2.71–2.74 (m, 2H), 2.51–2.55 (m, 2H), 2.36–2.43 (m, 2H), 1.78–1.87 (m, 2H), 1.57–1.61 ppm (m, 2H).

1-[2-(4-Benzyl-4-hydroxypiperidin-1-yl)ethyl]-3-(5-methylthieno-[3,2-*b*]pyridin-7-yl)urea (ACT-2): Reaction of 5-methylthieno[3,2-*b*]pyridine-7-carbonyl azide (77 mg, 0.35 mmol) as described for ACT-**3** gave ACT-**2** as a yellow solid (12 mg, 8%): ^{1}H NMR (500 MHz, CD_3OD): δ = 7.91 (s, 1H), 7.90 (d, J = 5.6 Hz, 1H), 7.41 (d, J = 5.6 Hz, 1H), 7.30 (m, 2H), 7.25 (m, 3H), 3.60 (t, J = 6.0 Hz, 2H), 3.36 (m, 2H), 3.12 (t, J = 6.0 Hz, 2H), 3.09 (t, J = 11.8 Hz, 2H), 2.85 (s, 2H), 2.62 (s, 3H), 2.04 (m, 2H), 1.91 ppm (m, 2H).

1-(2-(4-Benzyl-4-hydroxypiperidin-1-yl)ethyl)-3-(2-methylthieno-[3,2-*b*]pyridin-7-yl)urea (ACT-4): Reaction of 2-methylthieno[3,2-*b*]pyridine-7-carbonyl azide (122 mg, 0.56 mmol) as described for ACT-**3** gave ACT-**4** as a yellow solid (129 mg, 54%): ^{1}H NMR (500 MHz, CD_3OD): δ = 8.31 (d, J = 5.7 Hz, 1H), 7.92 (d, J = 5.7 Hz, 1H), 7.27 (m, 2H), 7.21 (m, 3H), 7.12 (s, 1H), 3.47 (t, J = 6.3 Hz, 2H), 2.94 (m, 2H), 2.78 (s, 2H), 2.76 (t, J = 6.3 Hz, 2H), 2.69 (m, 2H), 2.62 (s, 3H), 1.77 (m, 2H), 1.61 ppm (m, 2H); ^{13}C NMR (125 MHz, DMSO): δ = 16.6, 35.2, 36.3, 49.2 (2C), 55.3, 57.2, 68.0, 106.6, 120.8, 123.7, 126.3, 128.1 (2C), 131.0 (2C), 138.1, 142.0, 144.0, 148.2, 155.0, 158.1; HRMS (EI): calcd for $C_{23}H_{28}N_4O_2S$ $[M]^+$: 424.1933, found: 424.1928.

1-(2-(4-Benzyl-4-hydroxypiperidin-1-yl)ethyl)-3-(3-bromothieno-[3,2-*b*]pyridin-7-yl)urea (ACT-5): Reaction of 3-bromothieno[3,2-*b*]pyridine-7-carbonyl azide (35 mg, 0.12 mmol) as described for ACT-**3** gave ACT-**5** as a yellow solid (12 mg, 20%): ^{1}H NMR (500 MHz, CD_3OD): δ = 8.48 (d, J = 5.5 Hz, 1H), 8.05 (d, J = 5.5 Hz, 1H), 8.00 (s, 1H), 7.29–7.22 (m, 2H), 7.21–7.17 (m, 3H), 3.48 (t, J = 6.0 Hz, 2H), 3.02 (d, J = 9.7 Hz, 2H), 2.83 (t, J = 6.0 Hz, 2H), 2.77 (s, 2H), 2.80–2.71 (m, 2H), 1.81–1.74 (m, 2H), 1.61 ppm (d, J = 13.5 Hz, 2H).

hUT$_2$R binding assay

Filtration-based time-resolved fluorescence (TRF) receptor binding assays with europium-labeled U-II (Eu-UII) were performed in 96-well AcroWell plates with GH polypro (GHP) membrane (PALL Life Sciences, Ann Arbor, USA) by incubating 10 μg per well HEK293-hUT membranes with 2 nM of Eu-UII in a total assay volume of 100 μL. U-II was labeled with europium (Eu) at the N1 position by the PerkinElmer labeling service (Waltham, USA), and hUT$_2$R membranes were prepared from human embryonic kidney (HEK293) cells, stably expressing hUT$_2$R. The assay buffer contained 25 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5% bovine serum albumin (BSA) pH 7.4. Nonspecific Eu-labeled hUT$_2$R binding was determined experimentally by the presence of 1 μM unlabeled U-II. After incubation at room temperature for 90 min, the incubation mixtures were filtered and washed in the automatic vacuum filtration system for filter plates. The europium was dissociated from the bound ligand by the addition of 150 μL of DELFIA enhancement solution (PerkinElmer Oy) and incubated for 15 min with shaking. Dissociated europium created highly fluorescent complexes, which were measured in a multilabel counter with a TRF option (Envision, PerkinElmer). The counter setting was as follows: excitation at 340 nm, 200 μs delay, and emission collection for 400 μs at 615 nm. The extent of antagonism was expressed as % displacement. The IC_{50} value was characterized in an eight-dose response study to generate the compound concentration required to yield 50% displacement.[27]

Computational methods

GEnSeMBLE predictions of apo-protein structures for hUT$_2$R: We have been developing methods for predicting the 3D structures of GPCRs from sequence information without use of homology since 1998,[28] with significant improvements in 2002 and 2004.[29,30] The MembStruk method[30] was validated by a series of applications to human D_2 dopamine,[31] human β_2 adrenergic,[32] human M_1 acetylcholine,[33] human CCR_1 cytokine,[34] mouse MrgC11,[35] and human DP prostaglandin receptor.[36] However, the more recent availability of X-ray structures has allowed us to improve on these techniques significantly, leading to the GEnSeMBLE method.[37]

The GEnSeMBLE method samples a complete set (trillions) of possible packings (rotations and tilts of the helices), all of which are scored using a fast approximate (BiHelix) analysis from which we extract smaller sets (1000s), which are scored more accurately to select the ensemble of ~20 low lying structures expected to play a role in binding ligands and activating the GPCR. The GEnSeMBLE prediction methodology has been described elsewhere,[37] but it is summarize here in brief:

1) **PredicTM**: We use PredicTM to perform a multiple sequence alignment of 1726 GPCR sequences using the MAFFT[38] program and to predict the TM domains for the target GPCR. These TM do-

mains were extended by capping rules, and the secondary structure prediction using PORTER[39] and SSPro,[40] as shown in Figure S2 of the Supporting Information.

2) **Helix shape**: We used two methods to predict the shape of the TM domains: a) OptHelix uses molecular dynamics to predict the structure for each TM region, which is usually helical except for kinks that may be caused by proline residues; b) homologize helices is used rather than OptHelix, particularly when closely related X-ray structures are available. For hUT_2R, we considered the following nine templates, the X-ray structures for human nociceptin/orphanin FQ receptor (hNOPR),[11] mouse delta opioid receptor (mOPRD),[12] mouse mu opioid receptor (mOPRM),[13] human kappa opioid receptor (hOPRK),[14] human D_3 dopamine receptor (hD_3DR),[15] human adenosine A_{2A} receptor ($hAA_{2A}R$),[17] human β_2 adrenergic receptor ($h\beta_2AR$),[18] human chemokine $CXCR_4$ ($hCXCR_4$),[20] and bovine rhodopsin (bRho).[24] The sequence identities of TM regions for these systems are 36.0% for hNOPR, 34.5% for mOPRD, 34.4% for mOPRM, 32.8% for hOPRK, 31.3% for hD_3DR, 27.9% for $hA_{2A}AR$, 27.3% for $h\beta_2AR$, 27.1% for $hCXCR_4$, and 21.8% for bRho. All nine templates were selected based on high sequence identity and diversity of GPCR family.

3) **BiHelix optimization of helix rotations**: Since the sequence identity in TM region (22–36%) was high, we sampled $\pm 60^\circ$ rotation about each TM axis (η angle) by 15° increments, leading to a total of $(9)^7 \sim 4.7$ million packings. To make it practical to evaluate the energies for all 4.7 million packings, we use the BiHelix method in which we first approximate the interactions within the 7-helix bundle by partitioning the interactions into 12 sets of pairwise helix interactions, which are added together (mean field approximation). For each of the 972 ($9\times9\times12$) pairwise interactions, we use SCREAM[41] to optimize the side chains. The BiHelix mean field energies for all packings are used to select the best 2000. Then, we build the full helix bundle for each of these 2000, optimize the side chains for each using SCREAM, and neutralize the charged residues for more accurate energy scoring. The Dreiding D3 force field (D3FF)[42] was used throughout wherever energies were evaluated.

4) **SuperBiHelix simulations optimization of helix tilts and rotations**: Starting with the X-ray tilt angles the BiHelix step always identifies the experimental rotation angles correctly in the BiHelix ensemble,[37] but even for closely related GPCRs, optimization of only the rotation angles does not lead to the X-ray structure for other GPCRs. To do this, we must optimize the tilt angles: θ (tilting away from the *z*-axis) and ϕ (the azimuthal angle of the tilting from the *xy*-plane) angle[43] simultaneous with the rotation angles (η). Since the best rotation angles from BiHelix match the template mOPRD, we considered it sufficient to sample $\pm 10^\circ$ for θ tilt angle with simultaneous sampling by $\pm 15^\circ$ and $\pm 30^\circ$ for both ϕ and η angles, leading to a total of $(3\times5\times5)^7 \sim 10$ trillion combinations. For cases in which the optimum angle is at the boundary of these variations, we do a second round starting from the new best conformation until it is consistent.

When we pack the helices into a 7-helix bundle, we choose the hydrophobic center (HPC) for each TM from the PredicTM analysis to place at $z=0$, for the common mid-membrane plane. SuperBiHelix can also be used to optimize translation along the TM axis simultaneously with optimizing rotations.

5) **Loop generation**: The loops and the N/C terminus including helix 8 for a new structure are modeled through homology with the each template structure. The disulfide bridge between $C123^{3.25}$and C199 for hUT_2R, which is conserved among class A GPCRs, was constructed by homology. Then keeping the 7-TM region fixed, we minimized the loops (up to 1000 steps or down to 0.5 RMS force) followed by quench annealing between 50 K and 600 K each for 0.1 ps for 10 cycles. The lowest energy structures, which were at eight cycles for $hUT_2R\_hCXCR_4$ and at seven cycles for hUT_2R_mOPRD, were selected for further docking. Both the annealed structures had 68 H-bonds (increased from 30 and 45 H-bonds at the beginning of hUT_2R $hCXCR_4$ and hUT_2R mOPRD, respectively).

DarwinDock: Our earlier studies used the HierDock[32a] and MSCDock,[44] which have now been replaced by DarwinDock. For each ligand conformation, DarwinDock iteratively generates ~50 000 poses into the putative binding regions of the bulky-residue-alanized protein. This is followed by the energy scoring of family heads to select the top 10% ordered by total energy. The top 100 conformations are chosen for further optimization. For each of these, we dealanize the protein side chains (using SCREAM) to find the optimum side chains for each of the best 100 poses. Then, we neutralize the protein and ligand by transferring protons appropriately within salt bridges and protonating or deprotonating exterior ligands, followed by further full geometry minimization. We consider that use of these neutral residue charges improves the accuracy for comparing different docked structures. The result is that small changes in geometries of charged ligands far from the binding site can lead to large differential binding energies of 10–30 kcal mol^{-1}. We find that neutralizing these exposed residues removes the sensitivity to details of the distances of charged residues (and counter ions) remote from the active site. This neutralization leads to differential binding energies that are dominated by the local cavity interactions and leads to much smaller solvation energies.

This same procedure was followed for each of seven ligand conformations generated as follows. Starting from the minimized structure of the ACT-058362 (ACT), SB-706375 (SB), GSK-1440115 (GSK1), and GSK-1562590 (GSK2), we performed the conformational search of mixed torsional/low-mode sampling (1000 steps, 100 steps per rotatable bond, 5 kcal mol^{-1} of energy window, 0.5 Å of maximum atom deviation cutoff) using the Maestro software.[45] The low-energy conformations were reminimized by D3FF and clustered by 2.5 and 1.5 Å of RMSD in two steps. For docking, the lowest 7–10 ligand conformations (within 10 kcal mol^{-1} of the best energy) were selected out of 227 for ACT, 360 for SB, and 158 for GSK2. The final docked structure with the best binding energy from all ligand conformations was selected.

DarwinDock has been validated for a number of X-ray co-crystals including three crystal structures of ligand/GPCR complexes: $h\beta_2AR$ (0.4 Å RMSD),[18] $hA_{2A}AR$ (0.8 Å RMSD),[17] and turkey β_1-adrenergic receptor (0.1 Å RMSD).[16] This shows that we can accurately identify ligand binding sites in proteins, which can then be used to optimize the ligands with desirable properties.

Virtual screening: We used LigandScout 3.0[26] to identify new leads for hUT_2R ligands, based on the 3D pharmacophore form derived from our predictions on ACT1. For the pharmacophore, we selected: 1) one pi–pi stacking at $F131^{3.33}$; 2) three hydrophobic interactions of the methyl group at V184 and $L212^{5.39}$, the phenyl ring at I54 and L126, and the phenyl ring from fused ring with $F216^{5.43}$, $I220^{5.37}$, and P271; 3) one H-bond donor at $D130^{3.32}$; and 4) one positive N at $D130^{3.32}$ with 15 exclusion volume.

Supporting Information

The following are available in the Supporting Information: details of GEnSeMBLE prediction of the 3D structure of hUT_2R (Appendix A in the Supporting Information); details of predicting the binding site of ACT-058362 and SB-706375 to hUT_2R (Appendix B in the Supporting Information); hits from virtual screening result of ACT-058362 with hUT_2R from the rhodopsin template (Figure S1); TM region and the secondary structure prediction of hUT_2R (Figure S2); the hydropathy prediction from PredicTM for hUT_2R (Figure S3); the final TM regions for hUT_2R from the PredicTM multiple alignment method with the X-ray structures for nine templates (Figure S4); the predicted structure of hUT_2R with three templates (Figure S5); the H-bonding networks of hUT_2R with three templates (Figure S6); the binding site of ACT-058362 at hUT_2R with three templates (Figure S7); the binding site of SB-706375 at hUT_2R with the three templates (Figure S8), The post virtual screening result of hUT_2R from the template of frog rhodopsin (hUT_2R-fRho) bound with ACT-058362 (Table S1), Sequence identities of the hUT_2R with all predicted structures as well as 14 currently available X-ray structures of GPCRs (Table S2), Top 20 predicted structures for the hUT_2R selected from the total 24000 of the BiHelix analysis for the top 9 templates (Table S3), top 10 predicted structures of the hUT_2R based on three templates: $h\beta_2AR$, $hCXCR_4$, mOPRD (Table S4); top 20 predicted structures of the hUT_2R from the SuperBiHelix analysis of $h\beta_2AR$ and $hCXCR_4$ (Table S5); the x, y, z, η, θ, ϕ, and RMSD (Å) comparison of hUT_2R predicted with the hUT_2R-fRho and $hCXCR_4$ templates compared with that from mOPRD (Table S6).

Acknowledgements

Funding for this project was provided by the Korea Research Institute of Chemical Technology (South Korea).

Keywords: computational chemistry · docking · GPCRs · R-group screening · urotensin II · virtual screening

[1] R. S. Ames, H. M. Sarau, J. K. Chambers, R. N. Willette, R. V. Aiyar, A. M. Romanic, C. S. Louden, J. J. Foley, C. F. Sauermelch, R. W. Coatney, Z. Ao, J. Disa, S. D. Holmes, J. M. Stadel, J. D. Martin, W.-S. Liu, G. I. Glover, S. Wilson, D. E. McNutty, C. E. Ellis, N. A. Eishourbagy, U. Shabon, J. J. Trill, D. V. P. Hay, E. H. Ohlstein, D. J. Bergsma, S. A. Douglas, *Nature* **1999**, *401*, 282 – 286.

[2] A. Carotenuto, P. Grieco, P. Campiglia, E. Novellino, P. Rovero, *J. Med. Chem.* **2004**, *47*, 1652 – 1661.

[3] S. A. Douglas, *Curr. Opin. Pharmacol.* **2003**, *3*, 159 – 167.

[4] S. Flohr, M. Kurtz, E. Kostenis, A. Brkovich, A. Fournier, T. Klabunde, *J. Med. Chem.* **2002**, *45*, 1799 – 1805.

[5] A. Carotenuto, P. Grieco, P. Rovero, E. Novellino, *Curr. Med. Chem.* **2006**, *13*, 267 – 275.

[6] P. Grieco, A. Carotenuto, P. Campiglia, L. Marinelli, T. Lama, R. Patacchini, P. Santicioli, C. A. Maggi, P. Rovero, E. Novellino, *J. Med. Chem.* **2005**, *48*, 7290 – 7297.

[7] S. K. Kim, Y. Li, C. Park, R. Abrol, W. A. Goddard III, *ChemMedChem* **2010**, *5*, 1594 – 1608.

[8] M. Clozel, C. Binkert, M. Birker-Robaczewska, C. Boukhadra, S.-S. Ding, W. Fischli, P. Hess, B. Mathys, K. Morrison, C. Müller, C. Müller, O. Nayler, C. Qiu, M. Rey, M. W. Scherz, J. Velker, T. Weller, J.-F. Xi, P. Ziltener, *J. Pharmacol. Exp. Ther.* **2004**, *311*, 204 – 212.

[9] S. A. Douglas, D. J. Behm, N. V. Aiyar, D. Naselsky, J. Disa, D. P. Brooks, E. H. Ohlstein, J. G. Gleason, H. M. Sarau, J. J. Foley, P. T. Buckley, D. B. Schmidt, W. E. Wixted, K. Widdowson, G. Riley, J. Jin, T. F. Gallagher, S. J. Schmidt, L. Ridgers, L. T. Chirstmann, R. M. Keenan, S. D. Knight, D. Dhanak, *Br. J. Pharmacol.* **2005**, *145*, 620 – 635.

[10] D. J. Behm, N. V. Aiyar, A. R. Olzinski, J. J. McAtee, M. A. Hilfiker, J. W. Dodson, S. E. Dowdell, G. Z. Wang, K. B. Goodman, C. A. Sehon, M. R. Harpel, R. N. Willette, M. J. Neeb, C. A. Leach, S. A. Douglas, *Br. J. Pharmacol.* **2010**, *161*, 207 – 228.

[11] A. A. Thompson, W. Liu, E. Chun, V. Katritch, H. Wu, E. Vardy, X. P. Huang, C. Trapella, R. Guerrini, G. Calo, B. L. Roth, V. Cherezov, R. C. Stevens, *Nature* **2012**, *485*, 395 – 399.

[12] S. Granier, A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, W. I. Weis, B. K. Kobilka, *Nature* **2012**, *485*, 400 – 404.

[13] A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, J. M. Mathiesen, R. K. Sunahara, L. Pardo, W. I. Weis, B. K. Kobilka, S. Granier, *Nature* **2012**, *485*, 321 – 326.

[14] H. Wu, D. Wacker, M. Mileni, V. Katritch, G. Han, E. Vardy, W. Liu, A. A. Thompson, X.-P. Huang, F. I. Carroll, S. W. Mascarella, R. B. Westkaemper, P. D. Mosier, B. L. Roth, V. Cherezov, R. C. Stevens, *Nature* **2012**, *485*, 327 – 332.

[15] E. Y. Chien, W. Liu, Q. Zhao, V. Katritch, G. W. Han, M. A. Hanson, L. Shi, A. H. Newman, J. A. Javitch, V. Cherezov, R. C. Stevens, *Science* **2010**, *330*, 1091 – 1095.

[16] T. Warne, M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. W. Leslie, C. G. Tate, G. F. X. Schertler, *Nature* **2008**, *454*, 486 – 491.

[17] V. P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman, R. C. Stevens, *Science* **2008**, *322*, 1211 – 1217.

[18] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. F. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, R. C. Stevens, *Science* **2007**, *318*, 1258 – 1265.

[19] M. A. Hanson, C. B. Roth, E. Jo, M. T. Griffith, F. L. Scott, G. Reinhart, H. Desale, B. Clemons, S. M. Cahalan, S. C. Schuerer, M. G. Sanna, G. W. Han, P. Kuhn, H. Rosen, R. C. Stevens, *Science* **2012**, *335*, 851 – 855.

[20] B. Wu, E. Y. Chien, C. D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F. C. Bi, D. J. Hamel, P. Kuhn, T. M. Handel, V. Cherezov, R. C. Stevens, *Science* **2010**, *330*, 1066 – 1071.

[21] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. W. Han, T. Kobayashi, R. C. Stevens, S. Iwata, *Nature* **2011**, *475*, 65 – 70.

[22] A. C. Kruse, J. Hu, A. C. Pan, D. H. Arlow, D. M. Rosenbaum, E. Rosemond, H. F. Green, T. Liu, P. S. Chae, R. O. Dror, D. E. Shaw, W. I. Weis, J. Wess, B. K. Kobilka, *Nature* **2012**, *482*, 552 – 556.

[23] K. Haga, A. C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. I. Weis, T. Okada, B. K. Kobilka, T. Haga, T. Kobayashi, *Nature* **2012**, *482*, 547 – 551.

[24] a) K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, R. E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, *289*, 739 – 745; b) D. C. Teller, T. Okada, C. A. Behnke, K. Palczewski, R. E. Stenkamp, *Biochemistry* **2001**, *40*, 7761 – 7772; c) T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, E. M. Landau, Y. Shichida, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5982 – 5987; d) J. Li, P. C. Edwards, M. Burghammer, C. Villa, G. F. X. Schertler, *J. Mol. Biol.* **2004**, *343*, 1409 – 1438; e) T. Okada, M. Sugihara, A. N. Bondar, M. Elstner, P. Entel, V. Buss, *J. Mol. Biol.* **2004**, *342*, 571 – 583.

[25] S.-K. Kim, Y. Li, R. Abrol, J. Heo, W. A. Goddard III, *J. Chem. Inf. Model.* **2011**, *51*, 420 – 433.

[26] a) G. Wolber, A. A. Dornhofer, T. Langer, *J. Comput.-Aided Mol. Des.* **2007**, *20*, 773 – 788; b) C. R. Underwood, P. Garibay, L. B. Knudsen, S. Hastrup, G. H. Peters, R. Rudolph, S. Reedtz-Runge, *J. Biol. Chem.* **2010**, *285*, 723 – 730.

[27] K. S. Oh, S. Lee, B. H. Lee, *Assay Drug Dev. Technol.* **2011**, *9*, 514 – 521.

[28] W. B. Floriano, N. Vaidehi, W. A. Goddard III, M. S. Singer, G. M. Shepherd, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10712 – 10716.

[29] R. J. Trabanino, S. E. Hall, N. Vaidehi, W. B. Floriano, V. Kam, W. A. Goddard III, *Biophys. J.* **2004**, *86*, 1904 – 1921.

[30] N. Vaidehi, W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12622 – 12627.

[31] Y. Kalani, N. Vaidehi, S. E. Hall, W. B. Floriano, R. J. Trabanino, P. L. Freddolino, V. Kam, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3815 – 3820.

[32] a) W. B. Floriano, N. Vaidehi, G. Zamanakos, W. A. Goddard III, *J. Med. Chem.* **2004**, *47*, 56 – 71; b) P. Spijker, N. Vaidehi, P. L. Freddolino, P. A.

Hilbers, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4882 – 4887.
[33] J. Y. Peng, N. Vaidehi, S. E. Hall, W. A. Goddard III, *ChemMedChem* **2006**, *1*, 878 – 890.
[34] N. Vaidehi, S. Schlyer, R. J. Trabanino, W. B. Floriano, R. Abrol, S. Sharma, M. Kochanny, S. Koovakat, L. Dunning, M. Liang, J. M. Fox, F. L. de Mendonca, J. E. Pease, W. A. Goddard III, *J. Biol. Chem.* **2006**, *281*, 27613 – 27620.
[35] a) J. Heo, S.-K. Han, N. Vaidehi, J. Wendel, P. Kekenes-Huskey, W. A. Goddard III, *ChemBioChem* **2007**, *8*, 1527 – 1539; b) J. Heo, N. Vaidehi, J. Wendel, W. A. Goddard III, *J. Mol. Graphics Modell.* **2007**, *26*, 800 – 812.
[36] Y. Li, F. Zhu, N. Vaidehi, W. A. Goddard III, F. Sheinerman, S. Reiling, I. Morize, L. Mu, K. Harris, A. Ardati, A. Laoui, *J. Am. Chem. Soc.* **2007**, *129*, 10720 – 10731.
[37] R. Abrol, J. K. Bray, W. A. Goddard III, *Proteins* **2012**, *80*, 505 – 518.
[38] K. Katoh, K. Kuma, H. Toh, T. Miyata, *Nucleic Acid Res.* **2005**, *33*, 511 – 518.
[39] G. Pollastri, A. McLysaght, *Bioinformatics* **2005**, *21*, 1719 – 1720.
[40] G. Pollastri, D. Przybylski, P. Rost, B. Baldi, *Proteins* **2002**, *47*, 228 – 235.
[41] V. W. T. Kam, W. A. Goddard III, *J. Chem. Theory Comput.* **2008**, *4*, 2160 – 2169.
[42] S. L. Mayo, B. D. Olafson, W. A. Goddard III, *J. Phys. Chem.* **1990**, *94*, 8897 – 8909.
[43] R. Abrol, A. R. Griffith, J. K. Bray, W. A. Goddard III, *Methods Mol. Biol.* **2012**, *914*, 237 – 254.
[44] A. E. Cho, J. A. Wendel, N. Vaidehi, P. M. Kekenes-Huskey, W. B. Floriano, P. K. Maiti, W. A. Goddard III, *J. Comput. Chem.* **2005**, *26*, 48 – 71.
[45] K. S. Watts, P. Dalal, R. B. Murphy, W. Sherman, R. A. Friesner, J. C. Shelley, *J. Chem. Inf. Model.* **2010**, *50*, 534 – 546.
[46] V. M. Unger, P. A. Hargrave, J. M. Baldwin, G. F. Schertler, *Nature* **1997**, *389*, 203 – 206.

Received: March 26, 2014
Published online on July 2, 2014