

Chapter 2

Prediction of Structure and Antagonist Binding Site:

Human CCR1

A condensed version of this chapter will be submitted as a manuscript to Journal of the American Chemical Society.

Abstract

This chapter details the computational methods used and the results obtained in predicting the structure of and antagonist binding site for the human chemokine receptor 1 (hCCR1). In this work, we used the MembStruk v4.0 computational procedure to construct a structural model of the hCCR1, which incorporates both helical rotational and translational orientation optimization procedures. To validate the hCCR1 model, we utilized the HierDock protocol to determine the binding site for a class of antagonists. The well-studied antagonist BX471 binds in the region between transmembrane (TM) helices 2, 3, 4, 5, 6, and 7. Recent mutation analysis of a similar antagonist showed that residues Tyr41 (TM1), Tyr113 (TM3), and Glu287 (TM7) are involved in binding of this antagonist. These residues are located within our predicted binding site of BX471 and its derivatives.

We have also predicted the binding site of several other antagonists, and the trend in calculated binding energies is compared to experimentally known binding affinities. The results indicate a very high degree of correlation ($R^2 = 0.886$), further validating the predicted structure. The structure has also been used for a virtual ligand screening calculation, where $\sim 62,000$ compounds from a database mixed with top known binders for hCCR1 have been screened. In this virtual ligand screen, the experimental binders to hCCR1 dock and score in the top 5% of hits. Thus this validated structure can further be used to design subtype specific antagonists for CCR1 or for designing common antagonists for CCR1 and other chemokine receptors in animal models such as mouse or rat.

Computational Methods

Force Fields

Protein receptor calculations used the DREIDING Force Field (DFF)¹ with CHARM22 charges². The Cell Multipole Method (CMM)³ in MPSim⁴ was used to calculate all non-bond interactions.

Jaguar v4.0 was used to perform Hartree-Fock (HF) level quantum mechanical calculations to extract electrostatic potential (ESP) fitted charges on the isolated ligands⁵. These ligand charges have been used for docking calculations.

All solvent (water) calculations used the Analytical Volume Generalized Born (AVGB) approximation to the Poisson-Boltzmann (PB) continuum solvation model⁶.

The DFF along with the charge equilibration (QEq)⁷ method was utilized to describe lipids.

Structure Prediction

MembStruk v4.0 was used to predict the structure of hCCR1. A general outline of this procedure is as follows:

- I. Transmembrane (TM) domain length prediction
- II. TM bundle assembly and optimization of helical translation and rotation
- III. Individual helical bend and kink optimization

- IV. Rigid body molecular dynamics (RBMD) of helical bundle in explicit lipid bilayer to equilibrate the system
- V. Full structure optimization

Originally developed in our laboratory⁸⁻¹⁰, MembStruk is now an automated program which has been used to successfully predict the structure of numerous GPCRs¹¹⁻¹⁵, requiring as input only the target structure's primary sequence. However, the output is analyzed at every stage, and accordingly each step is discussed in detail below.

I. TM domain length prediction

The seven TM domains of hCCR1 are predicted using the TM2ndS procedure, in conjunction with output from a multiple sequence alignment (MSA). To obtain the MSA, we performed an NCBI BLAST¹⁶ query with hCCR1 (Figure 2-1) to obtain protein sequences for local alignment. 76 of those sequences with a bit score above 200 (corresponding identity ranging from 22-90%) were then selected for global MSA using ClustalW¹⁷. It is interesting to note that the functionalities of these 76 sequences varied from all chemokine, angiotensin, bradykinin, μ and δ opioid, leukotriene, purinergic, and somatostatin receptors. The 76 sequences selected are provided for reference in Appendix A.

This MSA was then analyzed by the program TM2ndS, which calculated a complete hydrophobicity profile for hCCR1 (Figure 2-2). This is based on the concept that the maximum region of hydrophobicity is near the center (and in fact, the geometric

plane) of the membrane. Extending this principle we are able to calculate the consensus hydrophobicity (average hydrophobicity of each sequence position in the MSA), and subsequently average hydrophobicity at each sequence position over a range of window sizes (12-20). Ignoring the amino and the carboxy termini, the average hydrophobicity values are used to determine the baseline shown in Figure 2-2, which serves as a threshold for identifying the seven TM regions (Figure 2-3).

II. TM bundle assembly and optimization of helical translation and rotation

The predicted TM regions are then used to build seven canonical α -helices with extended side chains. To orient the helical axes in space, we assemble the bundle and position the helices based on the 7.5Å electron density map of frog rhodopsin¹⁸. It is imperative to realize that we do not extract any additional experimental information from this structure. Rather, we developed a novel conformational search approach to determine various rotational and translational orientations of the helices.

As mentioned briefly in section I, we optimize the relative translation of the seven helices utilizing the principle that the location of the residues with maximum hydrophobicity should be in the middle of the lipid bilayer. The hydrophobic maximum for each helix is determined from the hydrophobicity plots mentioned in the previous section. The hydrophobic maximum of each helix is then aligned to a plane. Since this maximum value, or hydrophobic center, will be different for each helix (Figure 2-3), we are now able to translate each helix to fit this plane.

Once the helices are aligned along their axes according to their calculated hydrophobic centers, the next step is to optimize their rotational orientation. This is accomplished in two ways: 1) net hydrophobic moment reorientation (used for TMs 1-2, 4-7) and 2) energy minimization (TM 3). In the first method, the hydrophobic moment of the third of the helix about the hydrophobic maximum is calculated, and the helix is reoriented to orient this hydrophobic moment vector outside facing the lipid. Since every TM (except 3) interacts with the lipid membrane, this method seems to be appropriate for the purpose of rotational orientation. Figures 2-4 and 2-5 show the net hydrophobic moments before and after this rotation, respectively.

In the case of TM 3 (middle of the helical bundle), we use the RotMin energy minimization procedure to optimize the helical rotation in the field of the other six helices. We rotate through a grid of rotation angles (360° in 5° increments) for the main chain, reassign side chains, and minimize the potential energy using conjugate gradients to a tolerance of 0.5 kcal/mol/\AA in RMS force. Shown in Figure 2-6 is the resulting energy scan profile, with a local minimum at -65° . After making this rotation, we observe a number of stabilizing polar interactions involving TM3 (detailed in the Results and Discussion section below). The scanning of rotational angles over 360° enables sampling of different low energy conformations for the TM bundle and helps going over activation barriers for these conformational transitions.

III. Individual helical bend and kink optimization

In this step, we attempt to refine our structure by allowing for local helical distortions, such as bends and kinks. That is, letting prolines and neighboring glycines, along with charged residues, introduce bends and kinks. This is accomplished by subjecting each helix to torsional molecular dynamics (MD) at 300K for 500 ps. The additional benefit of performing MD is that we are able to optimize side chains in the bundle, apart from observing the fact that the helical bends are indeed dynamical in nature in the protein. We proceed with the lowest potential energy structure and minimize it using conjugate gradients.

IV. RBMD of helical bundle in explicit lipid bilayer to equilibrate the system

The motivation behind this part of the MembStruk procedure is to optimize helical translations, angles, and rotations in explicit lipid bilayer environment. Here we add two layers of explicit lipid (52 molecules of dilauroylphosphatidyl choline) around the TM bundle. We then perform RBMD (both helices and lipids are treated as rigid bodies) at 300K for 50 ps, in 1 fs time steps.

V. Full structure optimization

Finally, a full atom conjugate gradient minimization is performed on the protein in vacuum.

Function Prediction Methods for Structure Validation

As there is no experimentally determined structure for comparison, we use hCCR1 function prediction as an independent method to verify the integrity of our calculated structure. Specifically, we aim to predict ligand binding sites and trends in binding energies for a series of antagonists. We do this by employing the HierDock procedure, which has been validated for a number of globular and membrane-bound proteins^{9, 19-20}. Although the ligand binding site is unknown for most GPCRs, HierDock has proven to be remarkably successful in locating antagonist binding sites. The general methodology is outlined below:

I. Binding site determination via coarse-grain docking

We scan the entire protein by mapping the molecular surface with autoMS (DOCK 4.0 utility), then use SPHGEN (DOCK 4.0 utility)²¹ to generate spheres representing the empty volume of the receptor. For hCCR1 this led to a total of 8674 spheres partitioned over 40 (10Å)³ boxes (Figure 2-7). However, these are all putative binding sites and this data requires a detailed analysis to yield the correct binding site. At this stage, we exclude from our data set binding regions extending beyond the conformational space of the protein (i.e., loops), or near the intracellular and membrane boundaries.

For this analysis, we first perform coarse-grain docking by using DOCK to generate and score 500 conformations of the ligand in each box, of which the top 10%

(50) are filtered using a buried surface cutoff of 80%. This step is followed by protein-fixed, ligand-movable minimization for each of the 50 ligand conformations in each box, where again another 10% (5) of lowest energy structures are saved. Finally, we determine which boxes contain the highest number of these conformations, and rank this data by frequency of occurrence across a range of buried surface cutoffs (Table 2-1). For hCCR1, we proceeded with a total of 313 spheres merged from regions 1, 4, and 31 (Figure 2-8).

II. Ligand docking to binding site

We now perform a new sampling of ligand conformations in the binding site determined from merging the three regions. This involves docking the rigid ligand and scoring 1000 conformations. As before, the top 10% (100) of structures are kept and subjected to protein-fixed, ligand movable minimization. In the resulting ensemble, the minimized ligand conformations that satisfy the buried surface area cutoff criterion of 80% are kept for the next step (another 10%, or 10 total). These 10 structures are then minimized completely with all atoms movable using conjugate gradients. The ligand binding energy (BE) is also calculated for each complex according to the following equation:

$$\text{BE} = \text{Potential Energy (free ligand in solvent)} - \text{PE (ligand in protein)}$$

The ligand energy in solvent (water) is calculated using the AVGB continuum solvation method. These structures are ranked one last time according to BE, and the best five are used for subsequent loop addition and side chain rotamer optimization.

III. Loop addition and cavity optimization

In earlier studies, we added extracellular (EC) and intracellular (IC) loops immediately after the RBMD step in MembStruk. However, we believe that the loop involvement in binding comes in the modulation of ligand into the binding site and the loops might close after the ligand binds in the TM region. Thus we add the EC and IC loops (using the Modeler v6.2 software) after the ligand has been docked and optimized in the binding site. After addition, loop side chain rotamers were reassigned using SCWRL²², and the structure minimized first with TM regions fixed and loops movable, followed by an all-atom movable minimization. Next, the side chain conformations for all residues within 6Å of the bound ligand were reassigned using the SCREAM²³ side chain optimization program. The resulting structure was then optimized by all atom conjugate gradient minimization, yielding the final docked hCCR1-ligand complex.

Results and Discussion

Structural Features of hCCR1: Interhelical Interactions

I. Location of Conserved Residues and Comparison to BR

A number of important polar helical interactions (salt bridges, hydrogen bonds, etc.) are observed in the predicted hCCR1 structure. Figure 2-9 shows a bidentate distant interaction between Lys94 (TM2) and Glu287 (TM7), the distances (6.18 and 6.77 Å) suggestive of water mediation of a salt bridge. Another bidentate salt bridge is between Glu120 (TM3) and His293 (TM7), with distances of 3.34 and 5.37 Å (Figure 2-10). Such interactions between Glu122 (TM3) and His211 (TM5) have been observed in BR. There also exists a hydrogen bond network between the very highly conserved residues (across many GPCRs) Asn52 (TM1), Asp80 (TM2), and Asn297 (TM7). The distances reflected in Figure 2-11 also seem to indicate water would be involved in these interactions, as in BR²⁴.

Trp158 (TM4) is another highly conserved residue (across many GPCRs) that is hydrogen bonding with the side chain of Ser79 (TM2), depicted in Figure 2-12. In the crystal structure of BR there is a hydrogen bond between this Trp on TM4 and a TM2 Asn (corresponding to Ser79 of hCCR1). As shown in Figure 2-13, the hCCR1 corollary to the conserved TM6 WXPFF motif in biogenic amines¹¹ is WTPYN, of which Asn256 (TM6) makes two tight hydrogen bonds to Tyr113 (TM3). Illustrated in Figure 2-14 is another type of stabilizing interaction (sigma-pi) observed near the intracellular region

between Tyr72 (TM2) and the conserved TM7 residue Tyr301. Figure 2-15 shows that there appears to be a possible salt bridge between Arg64 (TM1) and Asp130 (TM3), near the junction between TM3 and IC2. This Arg is part of the well-known DRY motif, which is highly conserved across nearly all GPCRs. Recent studies have shown these residues are of critical importance in the regulation of signaling function in chemokine receptors²⁵.

The hCCR1 has 18% overall sequence identity to BR, and 22% in the TM region. The main chain RMS is 5.91 Å, and an overlay of the two structures is shown in Figure 2-16. The hCCR1 structure clearly exhibits bending in every helix with the exception of TM3. As in BR, these bends occur due to the presence of helix breakers, and mapping each bend reveals the direct influence of each TM proline: Pro38 (TM1), Pro39 (TM1), Pro88 (TM2), Pro167 (TM4), Pro211 (TM5), Pro254 (TM6), and Pro298 (TM7). Comparing to BR, there are three conserved prolines that also cause similar bends: Pro170 (TM4), Pro215 (TM5), and Pro267 (TM7). Table 2-2 provides both the helical tilt (angle between maximum hydrophobicity plane and mean helical axis) and bend (angle between helical midpoint and helical top and bottom) angles for each helix.

Predicted Binding Sites of Antagonists

For our validation, we docked a series of known antagonists with established binding affinities (not provided to us until after our calculations were complete) to our predicted hCCR1 structure. This library consisted of five ligands, only one of which is

non-proprietary and published in the literature as R-*N*-[5-chloro-2-[2-[4-[(4-fluorophenyl)methyl]-2-methyl-1-piperazinyl]-2-oxoethoxy]phenyl]urea, or BX471 (Figure 2-17)²⁶. The other four are proprietary compounds provided by Berlex AG, hereafter referred to as C1-C4. With their permission, we have reproduced these structures in Figure 2-18.

These compounds share a common molecular backbone, with different substituents on either aromatic ring. While building these ligands, we determined the pK_a of the piperazinyl tertiary amine nitrogen (indicated in Figure 2-17) using the prediction module in Jaguar v4.0. Interestingly, the pK_a was calculated to be 5.7 across the compound library and as a result we docked only the deprotonated species.

The general hCCR1 antagonist binding domain involves contributions from TMs 3-7, as shown in Figure 2-19. In order to quantify these interactions, we calculate the ligand binding energy (outlined above). However, to assess the contribution of individual TM residues to this energy, we determine the FF interaction energies within the 5 Å binding cavity. Then, we further decompose these values by hydrophobic (vDW) and ionic (Coulomb and hydrogen bond) contributions.

I. Predicted Binding Site for BX471 ($K_i = 1$ nM)

BX471 is a known high affinity antagonist for hCCR1, and the contacts in our binding cavity support this observation²⁷. Contributions to the binding energy are made from residues on TMs 3-7. As seen in Figure 2-20, a core group of residues are responsible for stabilizing the ligand via ionic interactions. Although Gly207 (TM5) is

not globally polar in nature, there is localized side chain hydrogen bonding which serves to anchor the chlorine-substituted aromatic ring. Likewise, the urea functionality is strongly stabilized by localized main and side chain ionic interactions with Gly168 (TM4).

Of the polar residues, Ser110 (TM3) contacts the ester carbonyl and the adjacent oxygen, albeit at larger than ideal hydrogen bond distances of 4.66 and 3.93 Å. The fluorine substituted aromatic is clearly stabilized by hydrogen bonding to Tyr291 (TM7) at a distance of 3.87 Å. Table 2-3 shows the magnitude of these interactions, and the quantitative information also suggests that the bidentate Gly168 contact is the most important polar interaction, followed by Gly207/Ser110 (almost identical contribution), and then Tyr291. It is important to note that in the predicted structures the hydrogen bonds are normally longer than those observed in crystal structures.

In the case of hydrophobic interactions, BX471 appears to be packed quite well on all sides: Leu203 (TM5) to the left, Gly168 and Leu109 (TM3) on top, Val283 (TM7) and Leu260 (TM6) on the right (Figure 2-21), and Tyr113 (TM3) and Tyr114 (TM3) below (Figure 2-22). As seen in Table 2-3, Tyr113 contributes most favorably to the binding energy, as it stacks really well beneath the plane of the ligand piperazine ring. This is followed by Leu109 and Leu260, which directionally pack the hydrophobic regions of the piperazine ring. The most important point to note here is that summation of the total contribution energy of the binding cavity yields -37.17 kcal/mol, or 94.7% of the total binding energy (-39.25 kcal/mol). After addition of the ligand solvation energy of

-1.96 kcal/mol (Table 2-4), we can account for almost 100% of the total calculated binding energy through cavity effects and solvation. This is critical in being able to account for binding differentials across a series of different ligands (see below).

Finally, it is interesting to note that ligand binding does not appear to perturb the interhelical polar and hydrophobic interactions in the apo-protein, thus reaffirming our classification of BX471 as an antagonist.

II. Predicted Binding Site for C1 ($K_i = 20$ nM)

The 5Å binding cavity of C1 involves contributions from TMs 2, 3, 5, 6, and 7 (Figure 2-23). Again, the major contributor to the binding energy is Tyr113, which also seems to weakly stabilize the ester carbonyl and the adjacent oxygen. However, the direct plane to plane stacking interaction is a bit weaker due to the ligand being shifted away from Tyr113. Compared to the BX471 binding site, we see that the urea functionality is substituted by bulky methoxy groups in this ligand, and therefore there is no strong polar interaction with Gly168 (which is not even present in the 5Å binding cavity). Instead, there are relatively weak longer-range interactions with the hydrophobic groups in TM5: Gly207, and Leu208. A redeeming feature of this complex is the pi-pi interaction between Phe264 (TM6) and the methoxy substituted aromatic ring.

The fluorine substituted aromatic ring appears to disrupt the TM2-7 salt bridge in the apo protein (Figure 2-9) by interacting with Lys94 (TM2). In general, this is an undesirable feature for a high affinity antagonist, and accordingly we would predict the binding energy/affinity to be “weaker” than BX471. As shown in Table 2-4, this is indeed

the case, as the binding energy for this complex is -30.57 kcal/mol. As before, including the solvation term accounts for 98% of the total binding energy.

III. Predicted Binding Site for C2 ($K_i = 370$ nM)

The 5Å binding cavity of C2 involves contributions from TMs 3-7, as shown in Figure 2-24. The favorable pi-pi interaction between Phe264 and the methoxy substituted aromatic ring is observed, just as in C1. Also, we see the same degree of hydrophobic interactions around these groups as we did in C1, albeit with Leu203 (TM5) instead of Leu208 (although Gly207 is indeed contributing). Ser110 (TM3) appears in this binding site, and couples to the ester moiety much weaker than it did in BX471. The Tyr113 stacking interaction is still favorable, but considerably weaker in this structure than both BX471 and C1.

We also observe a perturbation of the same TM2-7 salt bridge, although interaction with Glu287 is the responsible party in this docked site. This is a reasonable interaction, as the only difference between C1 and C2 is a change from fluorine to hydroxyl in the aromatic ring. A hydrogen-hydrogen interaction between the hydroxyl and the side chain hydrogens of Lys94 (TM2) would be repulsive, yet the hydroxyl-interaction with the acidic functionality of Glu287 (TM7) is indeed viewed as favorable, and has a greater magnitude of ionic coupling. Accordingly, we rate this binder worse (lower affinity) than both BX471 and C1, and comparing to experiment we see that this is indeed the case (Table 2-4). Finally, comparing the total cavity energy to the calculated

binding energy yields a ratio of 98%, indicating that the differentials so far stem from the binding cavity (as the solvation energies are very similar).

IV. Predicted Binding Site for C3 ($K_i = 2700$ nM)

The 5Å binding cavity of C3 involves contributions from TMs 1-7 (Figure 2-25). As before, we clearly see the pi-pi interaction between Phe264 (TM6) and the methoxy substituted aromatic ring. The methoxy substituents are again stabilized by hydrophobic packing with Leu203 (TM5) and Gly207 (TM5), with the addition of Leu260 (TM6) in this binding site. Interestingly, Leu260 was also observed in the binding site of BX471, although it was packing along the side of the piperazine ring. This indicates a shift in the relative position of the ligand within the binding site, likely a result of accommodating the steric bulk of the methoxy substituents. Another intriguing feature is that while the ester interaction with Ser110 (TM3) is actually stronger than what we observed in C2, there are a number of repulsive interactions in the cavity which serve to ultimately decrease the binding energy. The most dominant of these repulsive interactions are with Trp90 (TM2) and Tyr291 (TM7). The lone pair of electrons on the indole nitrogen of Trp90 is clearly repulsed by the lone pair on the now cyano nitrogen substituent.

Furthermore, the docked conformation of C3 disrupts the TM2-7 salt bridge by interacting very strongly with the positively charged Lys94, suggesting again that it is a weaker antagonist. Although the contribution of Tyr113 (TM3) is about the same as that seen in C2, the overall binding energy is only slightly lower due in large part to these counter-mediating effects. As seen in Table 2-4, the experimental affinities follow this

general trend, but reflect a greater magnitude (10-fold) of change between C2 and C3.

Finally, comparison of cavity and ligand binding energies shows that in our calculations, 100% of this differential is accounted for by binding site interactions.

V. Predicted Binding Site for C4 ($K_i = >10000$ nM)

The 5Å binding cavity of C4 involves contributions from TMs 1, 2, 3, 5, 6, and 7 (Figure 2-26). The binding site looks similar to that of C1, in that Gly207 (TM5) and Leu208 (TM5) pack around the tri-methoxy substituted aromatic ring. Ser110 is actually repulsive in Coulombic energy, and so the nearest partner for the ester carbonyl seems to be either the main chain of Leu117 (TM3) or Tyr113. However, these are long range and very weakly coupled, with the Tyr113 contribution being the lowest we have seen in this series. The only structural difference with respect to the other three antagonists is introduction of a methoxy substituent on the other aromatic ring. This is the only electropositive substituent in this series, and accordingly there are electrostatic clashes with the hydrogens of Tyr291 (TM7) and Lys94 (TM2). The presence of Glu287 (TM7) helps alleviate this repulsion by a comparatively smaller amount (0.3 kcal/mol attractive vs. 1.5 kcal/mol repulsive). However, this again involves disrupting the stable salt bridge in the apo protein, effectively reducing its efficacy as an antagonist. As with the other antagonists, the binding energy is 99% recovered by summation of the cavity and solvation energies (Table 2-4).

VI. Summary of Binding Differentials

Across this series of antagonists, the main differential in binding energy stems from the sequentially diminishing vdW contribution of Tyr113 (with the notable exception of C2 and C3). Although the general binding mode of each ligand is still in the same horizontal plane, this seems to indicate that the ligand has tilted away from the stabilizing stacking conformation seen in BX471. We believe this perturbation is a way of accommodating the bulky methoxy substituents not present in BX471. Coupled with the fact that the urea group has been eliminated, we can account for both the decreased interaction with Tyr113, and loss of the locally polarized Gly168 side chain anchor. In each case the total cavity energies are at least 98% of the calculated binding energies, indicating that these differentials can indeed be attributed to subtle changes in hydrophobic packing and ionic coupling. Finally, we observe that many of these weaker-binding antagonists disrupt stable interactions in the apo protein.

The experimental binding affinities and calculated binding energies are presented for comparison in Table 2-4. Our theoretical data follows the exact same trend when compared to experiment, and this is represented in a plot in Figure 2-27. With a correlation factor of 0.886, we are confident that our hCCR1 structure has been validated via function prediction.

Antagonist Binding Site Comparison to Mutation Experimental Results

Pease and coworkers have recently identified UCB 35625 (Figure 2-28) as a potent inhibitor of hCCR1 at nM concentrations²⁸. Their view of ligand binding is that this antagonist stabilizes the inactive conformation of hCCR1, rendering it incapable of inducing signal transduction. In an effort to establish whether or not the ligand binding domain spanned TM residues, they constructed a homology model of hCCR1 based on the crystal structure of bovine rhodopsin. Using this model, a total of 33 TM residues were chosen for single point mutation, and then expressed in a leukocyte cell line for migration assay (response to inhibition of the endogenous agonist by UCB 35625). Nearly all mutants were expressed in reasonable yield, comparable to wild type hCCR1 (Figure 2-29A). Data from the cell migration assay clearly shows that the Y41A, Y113A, and E278Q mutants failed to competitively bind UCB 35625 (Figure 2-29B)²⁹.

These experimental results implicated the aforementioned TM residues in ligand recognition, prompting the researchers to manually place UCB 35625 in a binding pocket spanning their respective helices (1-3-7). It is interesting to note that Tyr113 makes contacts with ligands in our BX471 binding site, and Tyr41 is within 5.25Å of BX471 (Glu287 within 5Å). Yet, the mode of interaction with both Tyr113 and Glu287 (shown in Figure 2-30) is quite different. The docked ligands in our work stack above the plane of the aromatic ring as opposed to hydrogen bonding with the OH substituent of Tyr113. Also, UCB 35625 is a positively charged quaternary amine, and as such is capable of strong ionic coupling with Glu287 (our ligands are neutral). Nonetheless, it is noteworthy

that our most important ligand binding contributor, Tyr113, has also been experimentally shown to directly affect antagonist binding.

Virtual Ligand Screening

Our collaborator at Berlex AG, Dr. Sabine Schlyer, extended this validation one step further and used our model to conduct a virtual ligand screen for purposes of drug lead discovery. This procedure involved enriching the Maybridge Database with a set of 38 known hCCR1 binders of varying affinities (1 nM-1 μ M). The resulting library of 62,038 compounds was then docked using the GOLD software package, followed by scoring of dock energies and buried surface areas. The top 95% of this subset was then subjected to ligand minimization (25 steps), and the best 50% of binders were all full-atom minimized for an additional 250 steps. After determination of the final binding energy, 63% of all the known binders appeared in the top 2% of all compounds screened, whereas all the binders were recovered in the top 5% of screened ligands.

This is a very exciting development in that it applies an independent method of structural validation. The hCCR1 structure presented here is only minimally optimized after ligand binding (just side chains in the 5Å binding cavity), and yet the results of this virtual screen clearly indicate that all known experimental binders (across a wide range of chemical functionalities) bind favorably to our predicted structure as well. The next step in this validation is an actual experimental high-throughput ligand screening process with a similar library of ligands.

Conclusions

We have presented here a structural model for the hCCR1, obtained via MembStruk v4.0. This procedure involved detailed refinement and optimization at every stage, and ultimately yielded a structure with a number of chemically logical and stable interhelical contacts. This structure was then validated using our HierDock protocol, which involved determination of the ligand binding site and subsequent docking. Five antagonists with known binding affinities were docked to our structure, and our computational results are in excellent agreement with the experimental trend in binding affinities. Further validation was realized by way of independently performed virtual ligand screening, where all known hCCR1 binders ranked amongst the best when docked to our structure. These results strongly support validation of our methodology, and give credence to our structural model of hCCR1.

References

- [1] Mayo, S. L., Olafson, B. D., and Goddard, W. A. III (1990) *J. Phys. Chem.* **94**, 8897-8909.

- [2] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) *J. Phys. Chem. B* **102**, 3586-3616.

- [3] Ding, H. Q., Karasawa, N., and Goddard, W. A. III (1992) *Chem. Phys. Lett.* **97**, 4309-4315.

- [4] Lim, K-T., Brunett, S., Iotov, M., McClurg, R. B., Vaidehi, N., Dasgupta, S., Taylor, S., and Goddard, W. A. III (1997) *J. Comput. Chem.* **18**, 501-521.

- [5] Trabanino, R. J. (2004) Ph.D. thesis, California Institute of Technology.

- [6] Zamanakos, G. (2001) Ph.D. thesis, California Institute of Technology.

- [7] Rappé, A. K., and Goddard, W. A. III (1991) *J. Phys. Chem.* **95**, 3358-3363.

- [8] Floriano, W. B., Vaidehi, N., Singer, M., Shepherd, G., and Goddard, W. A. III (2000) *Proc. Natl. Acad. Sci. USA* **97**, 10712-10716.
- [9] Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and Goddard, W. A. III (2002) *Proc. Natl. Acad. Sci. USA* **99**, 12622-12627.
- [10] Trabanino, R. J., Hall, S. E., Vaidehi, N., Floriano, W. B., Kam, V. W. T., and Goddard, W. A. III (2004) *Biophys. J.* **86**, 1904-1921.
- [11] Freddolino, P. L., Kalani, M. Y., Vaidehi, N., Floriano, W. B., Hall, S. E., Trabanino, R. J., Kam, V., and Goddard, W. A. III (2004) *Proc. Natl. Acad. Sci.* **101**, 2736–2741.
- [12] Kalani, M. Y., Vaidehi, N., Freddolino, Floriano, W. B., Hall, S. E., Trabanino, R. J., Kam, V., and Goddard, W. A. III (2004) *Proc. Natl. Acad. Sci.* **101**, 3815–3820.
- [13] Hall, S. E., Floriano, W. B., Vaidehi, N., and Goddard, W. A. III (2004) *Chem. Senses* **29**, 595-616.
- [14] Floriano, W. B., Vaidehi, N., and Goddard, W. A. III (2004) *Chem. Senses* **29**,

- 269-290.
- [15] Hummel, P., Vaidehi, N., Floriano, W. B., Hall, S. E., and Goddard, W. A. III (2005) *Prot. Sci.* **14**, 703-710.
- [16] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucl. Acids Res.* **25**, 3389-3402.
- [17] Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucl. Acids Res.* **22**, 4673-4680.
- [18] Schertler, G. F. X. (1998) *Eye.* **12**, 504-510.
- [19] Kekenes-Huskey, P. M., Vaidehi, N., Floriano, W. B., and Goddard, W. A. III (2003) *J. Chem. Phys. B* **107**, 11549-11557.
- [20] Datta, D., Vaidehi, N., Zhang, D., and Goddard, W. A. III (2004) *Prot. Sci.* **13**, 2693-2705.
- [21] Ewing, T. A., and Kuntz, I. D. (1997) *J. Comput. Chem.* **18**, 1175-1189.
- [22] Bower, M., Cohen, F. E., and Dunbrack, R. L. (1997) *J. Mol. Biol.* **267**, 1268-1282.

- [23] Kam, V. W. T., Vaidehi, N., and Goddard, W. A. III (2005) unpublished results.
- [24] Palczewski, K., Kumasaka, T., Hori, T., Behnke, C., Motoshima, H., Fox, B., Trong, I., Teller, D., Okada, T., Stenkamp, R., Yamamoto, M., and Miyano, M. (2000) *Science* **289**, 739-745.
- [25] Damaj, B. B., McColl, S. R., Neote, K., Songqing, N., Ogborn, K. T., Hebert, C. A., and Naccache, P. H. (1996) *FASEB J.* **10**, 1426-1434.
- [26] Hesselgesser, J., Ng, H. P., Liang, M., Zheng, W., May, K., Bauman, J. G., Monahan, S., Islam, I., Wei, G. P., Ghannam, A., Taub, D. D., Rosser, M., Snider, R. M., Morrissey, M. M., Perez, H. D., and Horuk, R. (1998) *J. Biol. Chem.* **273**, 15687-15692.
- [27] Onuffer, J. J., and Horuk, R. (2002) *Trends Pharmacol. Sci.* **23**, 459-467.
- [28] Sabroe, I., Peck, M. J., Van Keulen, B. J., Jorritsma, A., Simmons, G., Clapham, P. R., Williams, T. J., and Pease, J. E. (2000) *J. Biol. Chem.* **275**, 25985-25992.
- [29] Lopes de Mendonca, F., da Fonseca, P. C. A., Phillips, R. M., Saldanha, J. W., Williams, T. J., and Pease, J. E. (2005) *J. Biol. Chem.* **280**, 4808-4816.

Table 2-1: Boxes 1, 4, and 31 (of 40) contain the highest ranking ligand conformations across a range of buried surface cutoffs.

<u>Buried Surface Cutoff (%)</u>	80	70	60	50
<u>Best Boxes</u>	1, 4, 31	1, 4, 31	1, 17, 34	17, 34, 12

Table 2-2: Bends and tilts for each helix in hCCR1.

<u>Helix Number</u>	<u>Bend Angle (degrees)</u>	<u>Tilt Angle (degrees)</u>
1	9.7	30.2
2	28.5	39.7
3	10.7	29.7
4	9.0	0.0
5	5.9	23.7
6	6.9	15.5
7	15.2	17.5

Table 2-3: Decomposition of calculated binding energy (-39.25 kcal/mol) for BX471 docked to hCCR1.

<u>Residue (ordered by decreasing total contribution)</u>	<u>vdW Energy (kcal/mol)</u>	<u>Coulomb Energy (kcal/mol)</u>	<u>Total Energy (kcal/mol)</u>
Tyr113	-15.1	-0.24	-15.3
Gly168	-0.74	-4.92	-5.66
Leu109	-3.54	0.10	-3.44
Ser110	-2.59	-1.22	-3.81
Leu260	-2.19	0.18	-2.01
Gly207	-0.46	-1.28	-1.74
Tyr114	-1.11	-0.30	-1.41
Thr286	-1.10	-0.13	-1.23
Leu203	-0.95	-0.21	-1.16
Tyr291	-0.38	-0.35	-0.73
Val283	-0.64	-0.04	-0.68

$$\Sigma = -37.17 \text{ kcal/mol}$$

Table 2-4: Correlation between calculated binding energies (BE) and experimental binding affinities (K_i) for hCCR1 docked ligand library.

<u>Ligand</u>	<u>5Å Cavity Energy (kcal/mol)</u>	<u>Ligand Solvation Energy (kcal/mol)</u>	<u>Cavity Energy + Solvation Energy (kcal/mol)</u>	<u>Total Ligand Binding Energy (kcal/mol)</u>	<u>Expt. K_i (nM)</u>
BX471	-37.17	-1.96	-39.13	-39.25	1
C1	-27.93	-1.91	-29.84	-30.57	20
C2	-24.41	-2.08	-26.49	-27.15	370
C3	-24.45	-1.88	-26.33	-26.68	2700
C4	-22.15	-2.03	-24.18	-25.30	>10000

```
>UniProt/Swiss-Prot|P32246|CCR1_HUMAN C-C chemokine receptor type 1
METPNTTEDYDTTTTEFDYGDATPCQKVNERAFGAQLLPPLYSLVVFVIGLVGNILVVLVLV
QYKRLKNMNTSIYLLNLAI SDLLFLFTLPFWIDYKLDWVFGDAMCKILSGFYTGLYSE
IFFIILLTIDRYLAIVHAVFALRARTVTFGVITSIIIWALAILASMPGLYFSKTQWEFTH
HTCSLHFPHESLREWKL FQALKLNLFGLVLP LLVMIICYTGI IKILLRRPNEKKS KAVRL
IFVIMIIFFLFWTPYNTILISVFQDFLFTHECEQSRHLDLAVQVTEVIAYTHCCVNPVI
YAFVGERFRKYLRQLFHRRVAVHLVKWLPFLSVDRLERVSSTSPSTGEHEL SAGF
```

Figure 2-1: hCCR1 primary sequence in FASTA format.

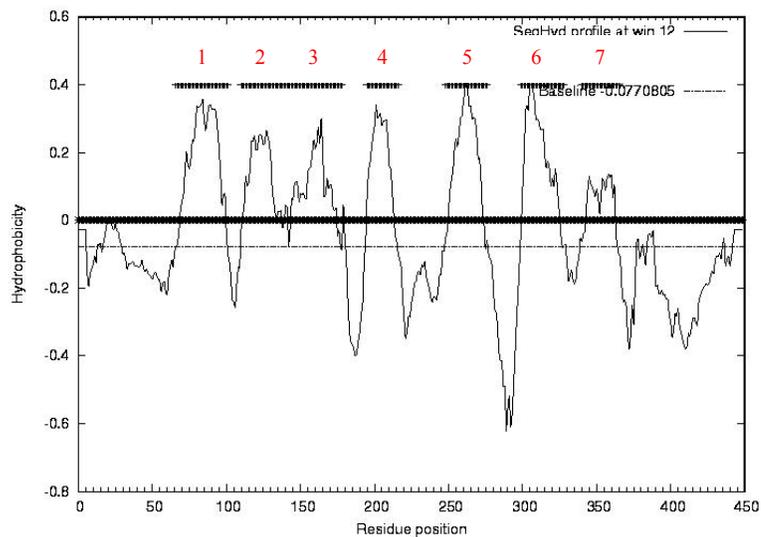


Figure 2-2: hCCR1 hydrophobicity profile.
Indicated in red are TM domains 1-7 at window size 12.

NT:	1	METPNTTDEDYDTTTEFDYGDATPCQKVNER	
TM1:	31	AFGAQLLPPLYSLVVFVIGLVGNILVVLVLVQY	62
IC1:		KRLKN	
TM2:	68	MTSIYLLNLAISDLLFLFTLPFWIDYKDKDD	98
EC1:		WVFGD	
TM3:	104	AMCKILSGFYTYTGLYSEIFFIILLTIDRYLAIVH	137
IC2:		AVFALRARTVTF	
TM4:	150	GVITSIIIWALAILASMPGLYFSK	173
EC2:		TQWEFTHHTCSLHFPHESLREW	
TM5:	196	KLFQALKLNLFGLVLPPLVMIICYTGIIK	224
IC3:		ILLRRPNEKKSKAV	
TM6:	239	RLIFVIMIIFFLFWTPYNLTILISVFQD	266
EC3:		FLFTHECEQSRH	
TM7:	279	LDLAVQVTEVIAYTHCVNPNVIYAFVG	305
CT:		ERFRKYLRQLFHRRVAVHLVKWLPFLSVDRLERVSSTSPSTGEHELSAGF	355

Figure 2-3: hCCR1 TM regions.

NT = N-terminus, CT = C-terminus, IC = intracellular, and EC = extracellular domains. Highlighted in red are the hydrophobic center predictions for each helix.

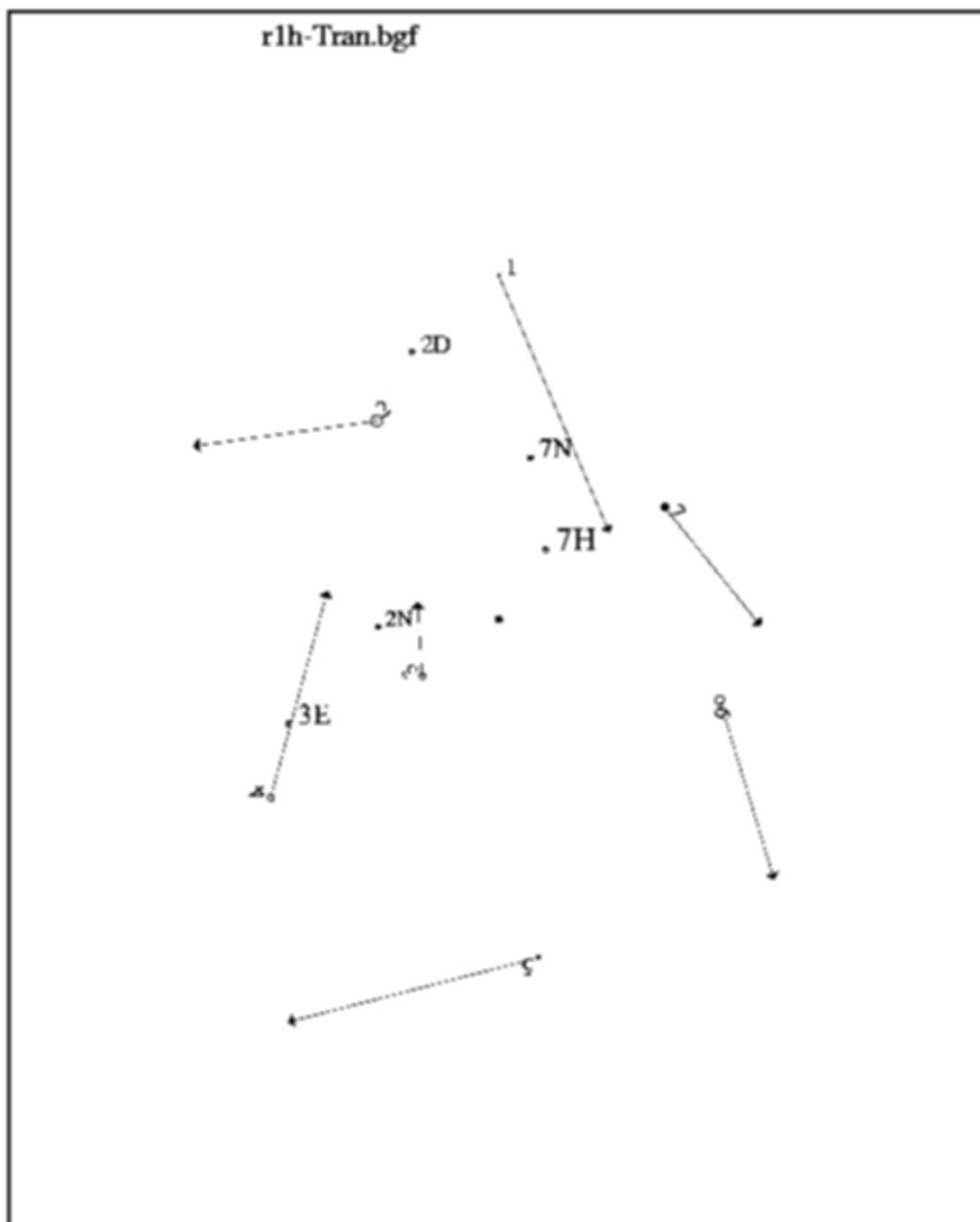


Figure 2-4: hCCR1 bundle before rotational optimization.

The numbers 1-7 refer to the TM regions, and the direction of the arrow indicates the net hydrophobic moment of the helix. Note that the moments are randomly oriented. The lipid membrane, not shown here, is packed uniformly around the bundle.

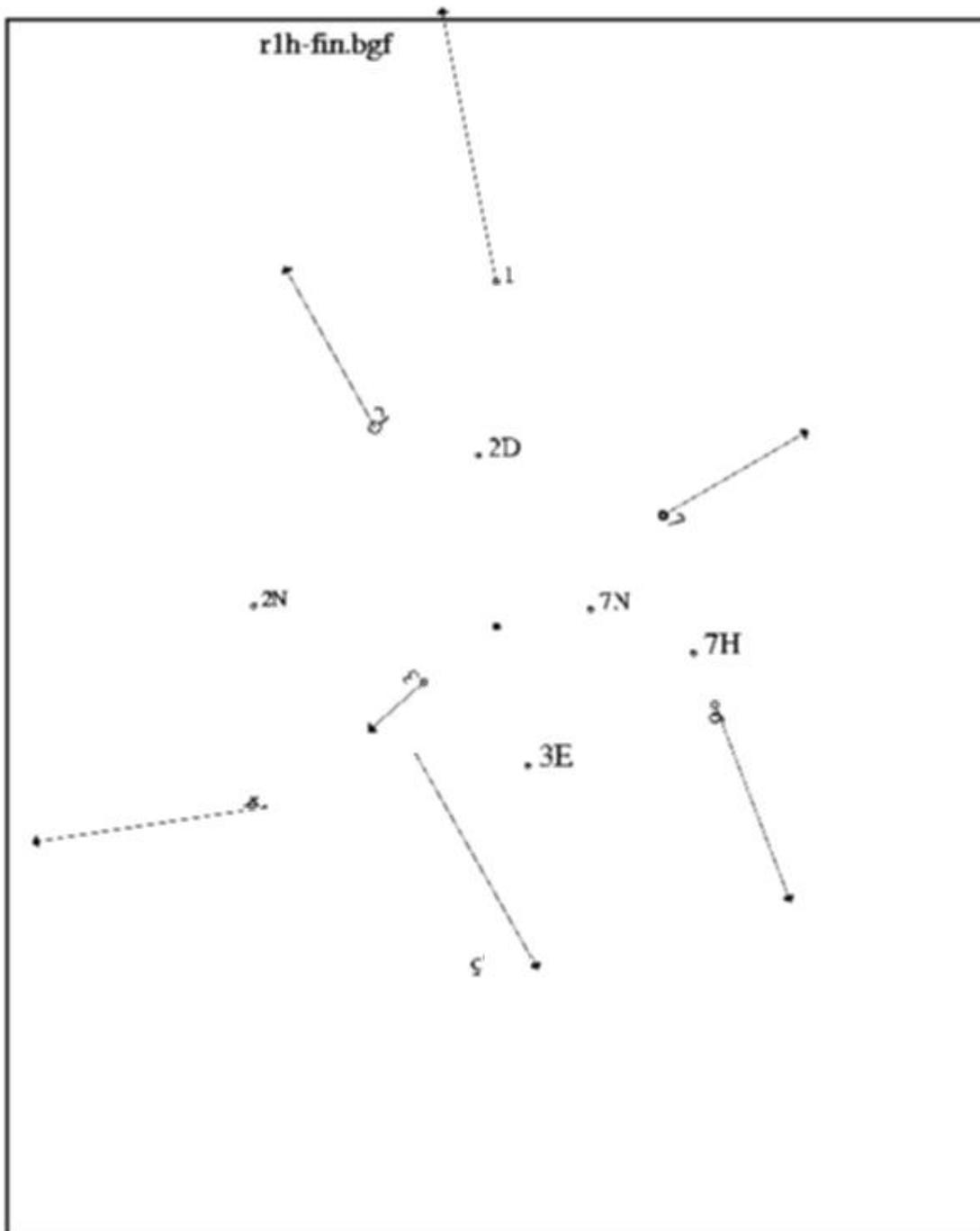


Figure 2-5: hCCR1 bundle after rotational optimization.

After rotational optimization, the hydrophobic moments are now oriented outward, facing the lipid membrane.

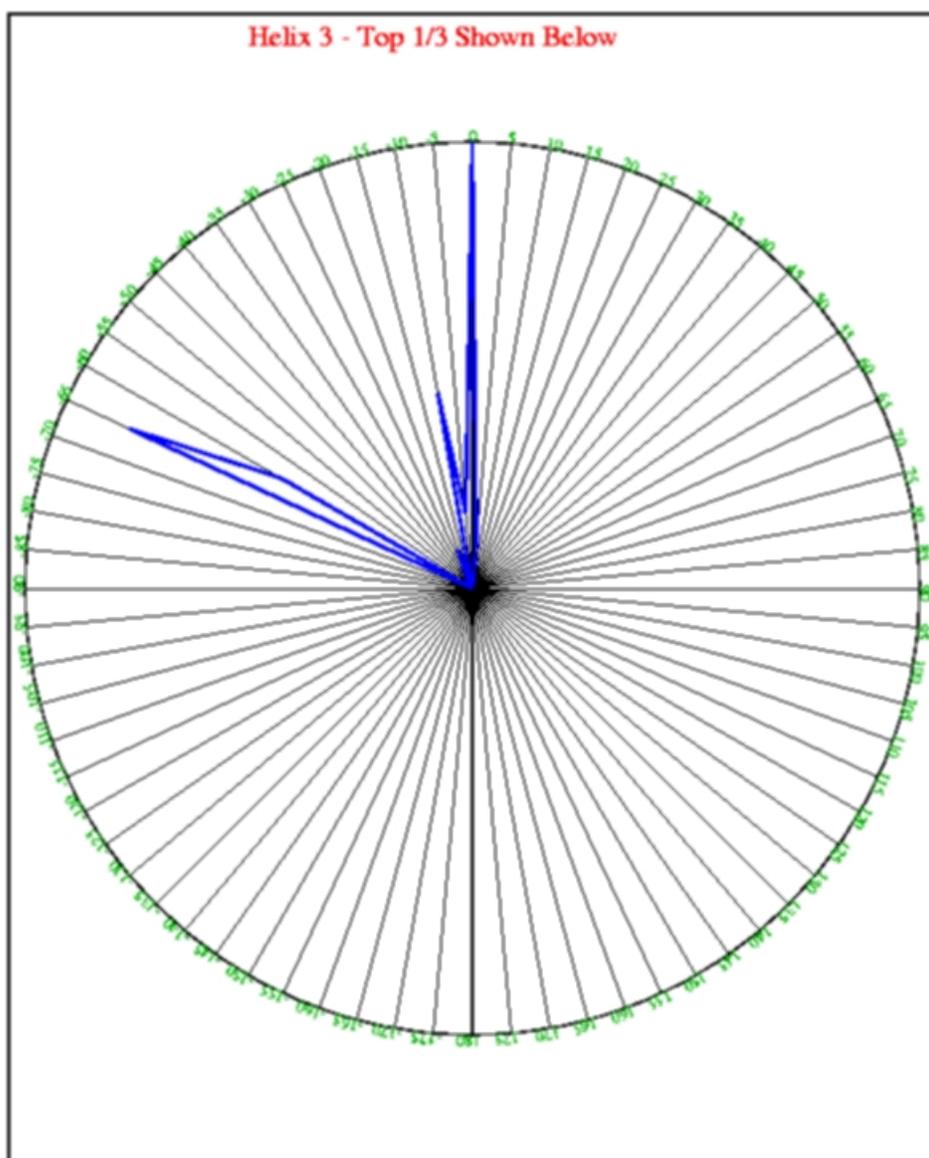


Figure 2-6: hCCR1 TM3 rotational optimization using RotMin.

After conducting a global energy scan for TM3, a minimum at -65° is observed. Subsequent rotation leads to a number of stabilizing polar interactions with other helices in the bundle.

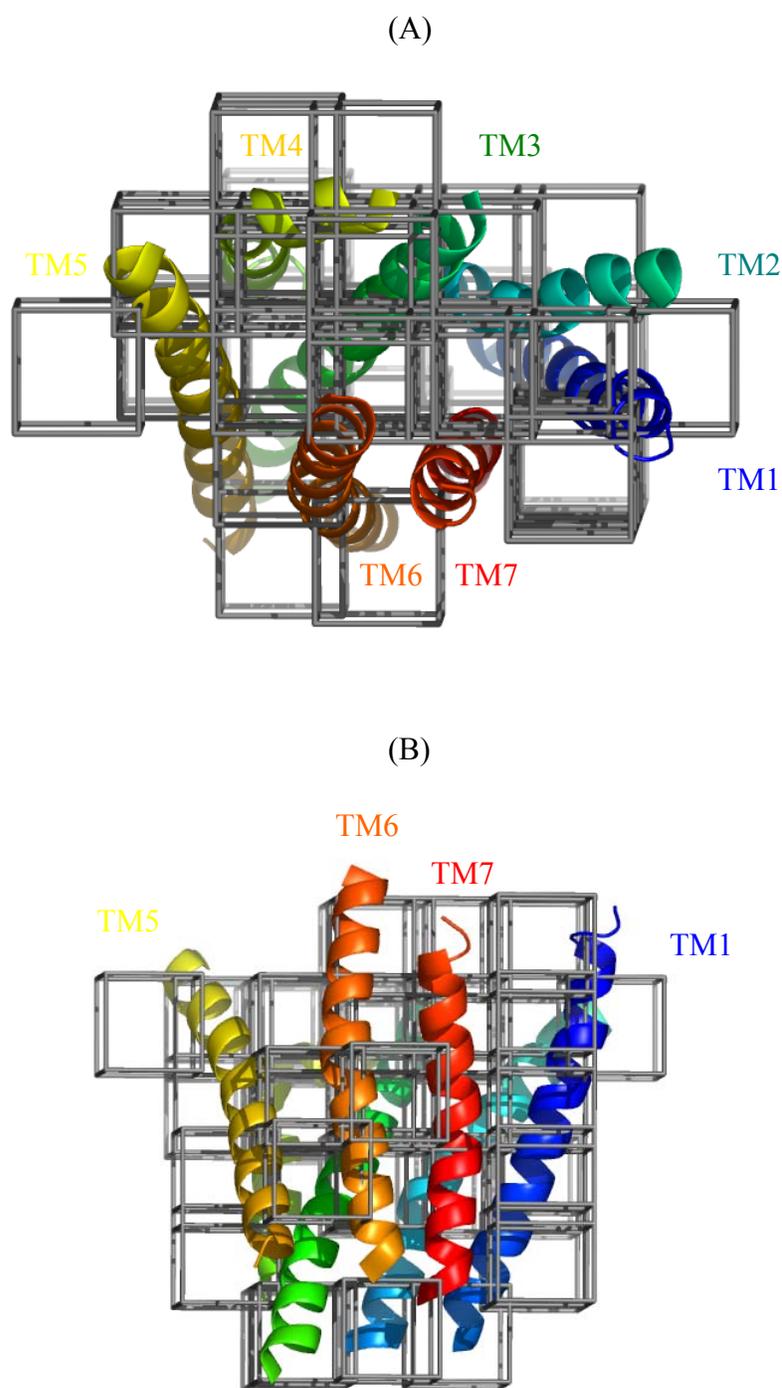


Figure 2-7: hCCR1 putative binding sites.

After scanning the molecular surface of the receptor, 40 boxes (each 1000 \AA^3) containing a total of 8674 spheres are generated. Shown in (A) is the top view, and (B) the side view. Note the convention used here is the extracellular side on “top” and thus the intracellular interface is below. Also provided is the color code used for helices in the bundle.

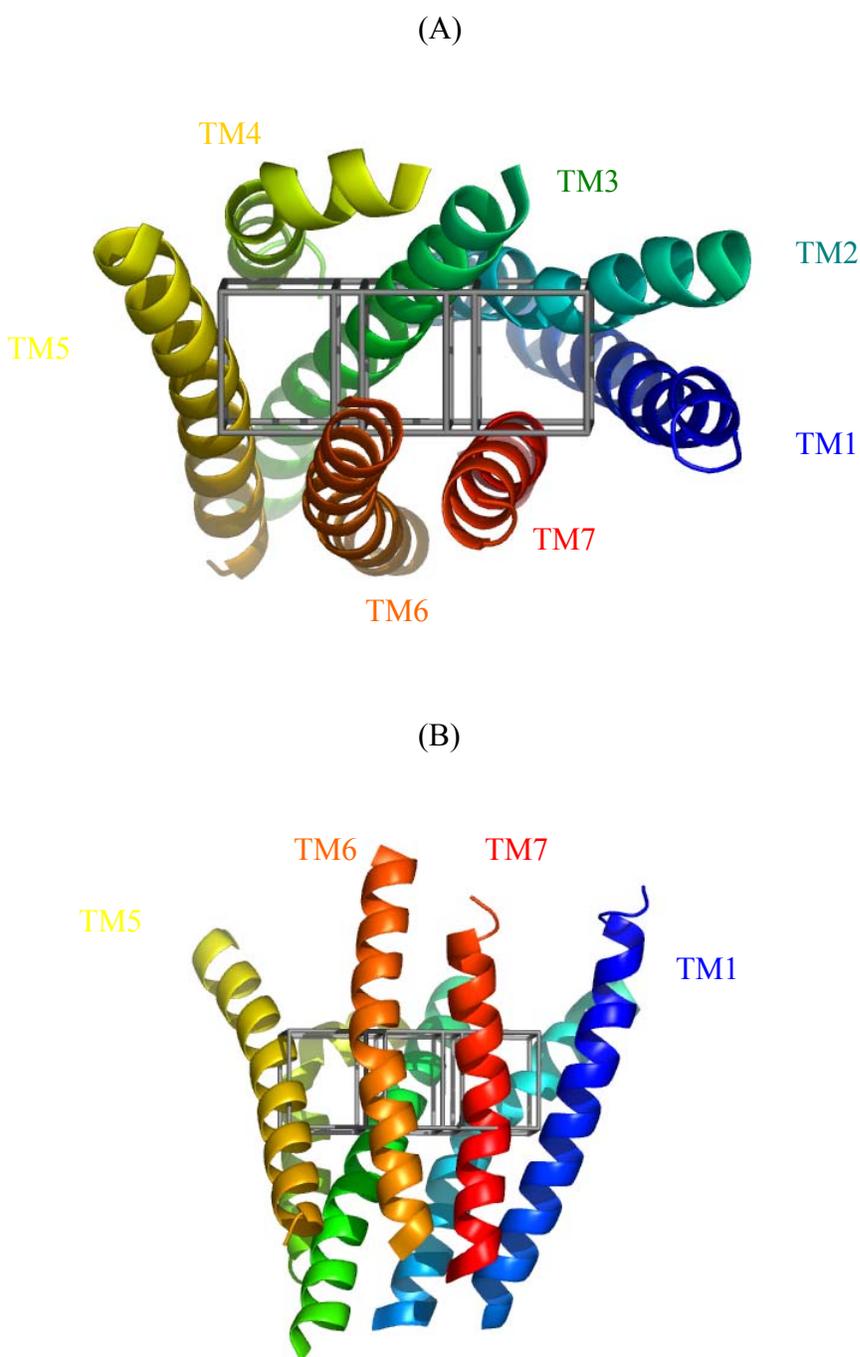


Figure 2-8: hCCR1 binding sites for docking.

From the 40 putative binding sites, we determine our refined binding site to be a horizontal mode across 3 contiguous regions (1, 4, and 31). Shown in (A) is the top view, and (B) the side view.



Figure 2-9: hCCR1 salt bridge (TM2-7).

Water mediated salt bridge between Lys94 (TM2) and Glu287 (TM7).

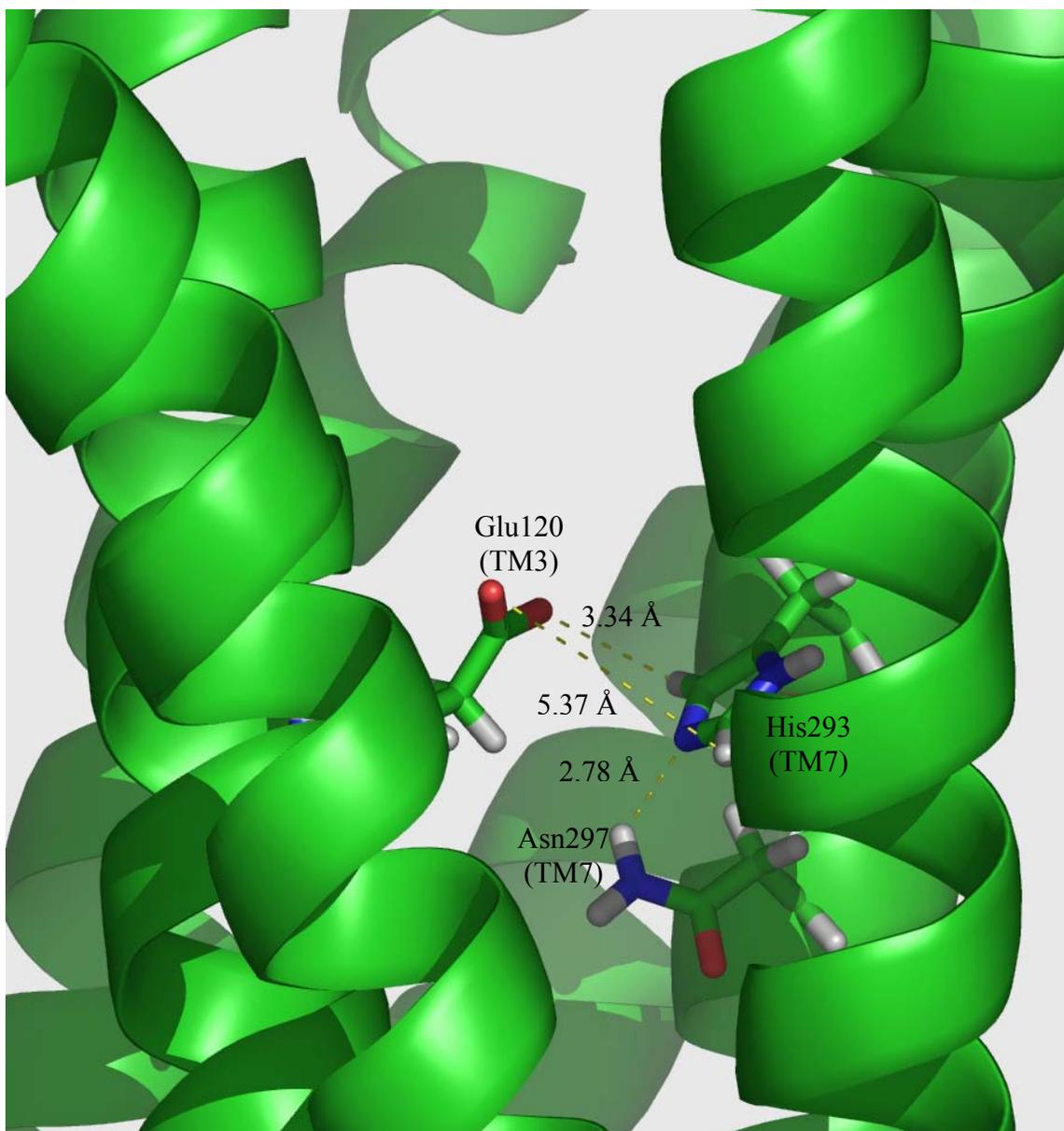


Figure 2-10: hCCR1 salt bridge (TM3-7) and intra-helical hydrogen bond. Salt bridge between Glu120 (TM3) and His293 (TM7). Also shown is a possible intra-helical hydrogen bond between His293 and Asn297.

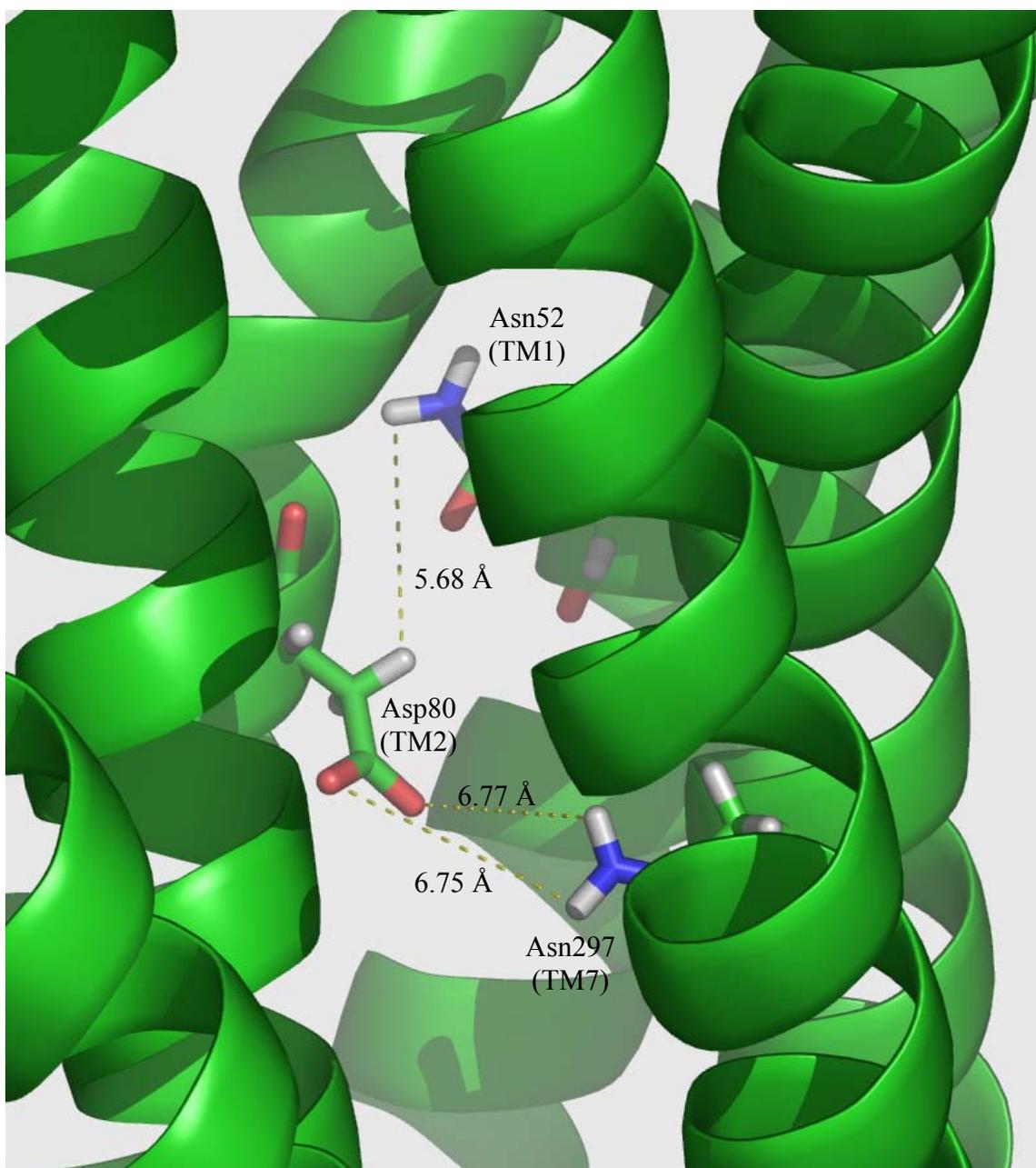


Figure 2-11: hCCR1 inter-helical hydrogen bond network (TM1-2-7).

Hydrogen bond between Asn52 (TM1) and Asp80 (TM2). The Asp80-Asn297 (TM7) hydrogen bonds are likely water mediated. These are conserved residues across many GPCRs.

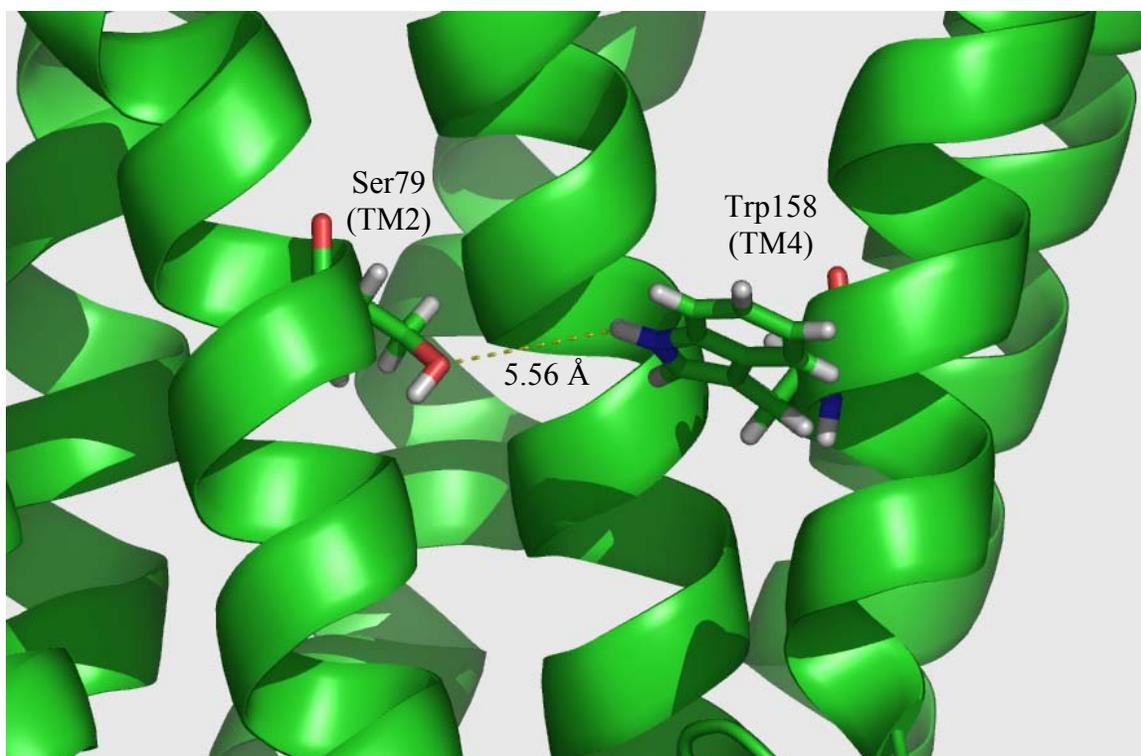


Figure 2-12: hCCR1 inter-helical hydrogen bond (TM2-4).

Hydrogen bond between Ser79 (TM2) and Trp158 (TM4). Trp158 is the most conserved TM4 residue across all GPCRs.

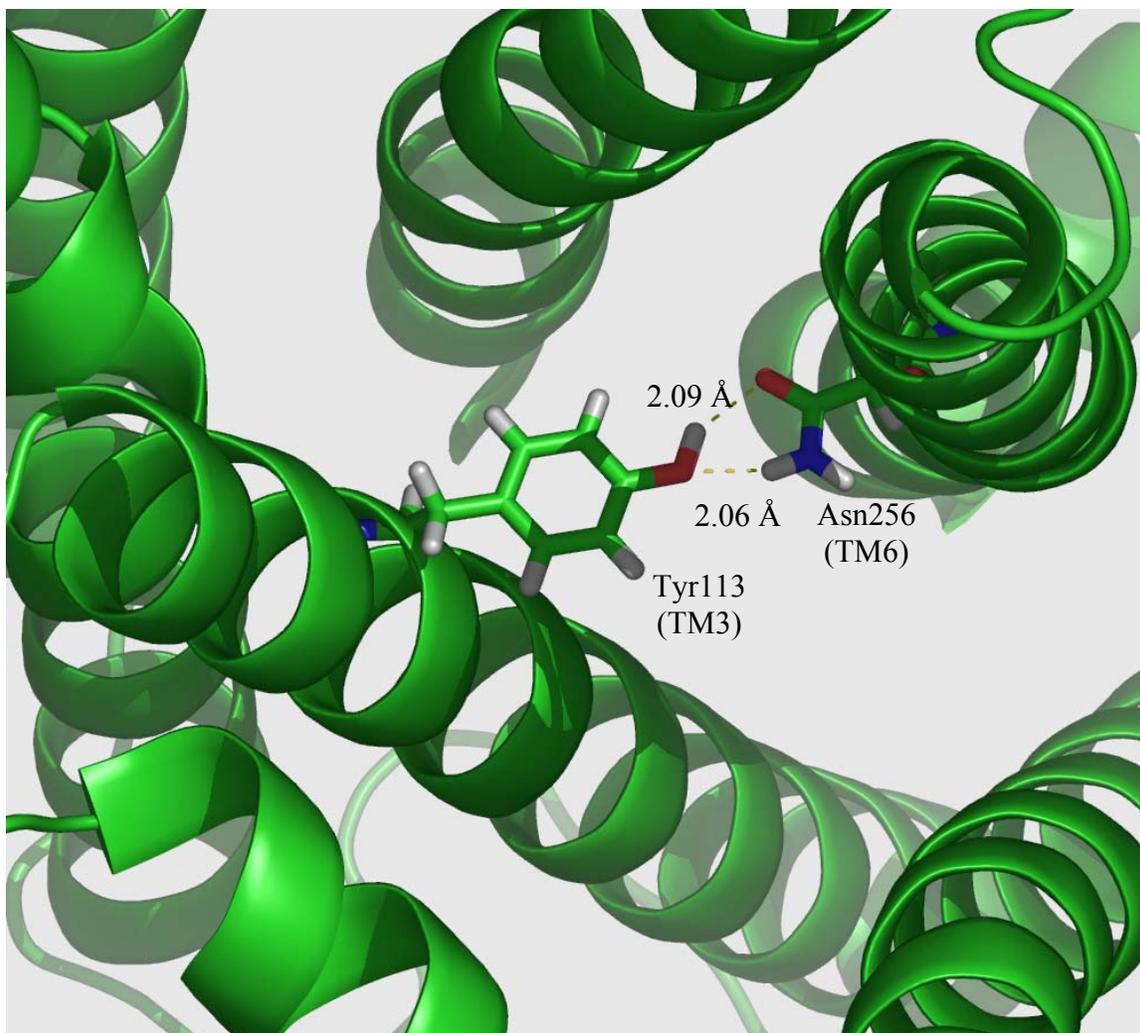


Figure 2-13: hCCR1 inter-helical hydrogen bonds (TM3-6). Hydrogen bonding between Tyr113 (TM3) and Asn256 (TM6). Asn256 is part of the WXPFF motif (WTPYN in hCCR1) in biogenic amine GPCRs.

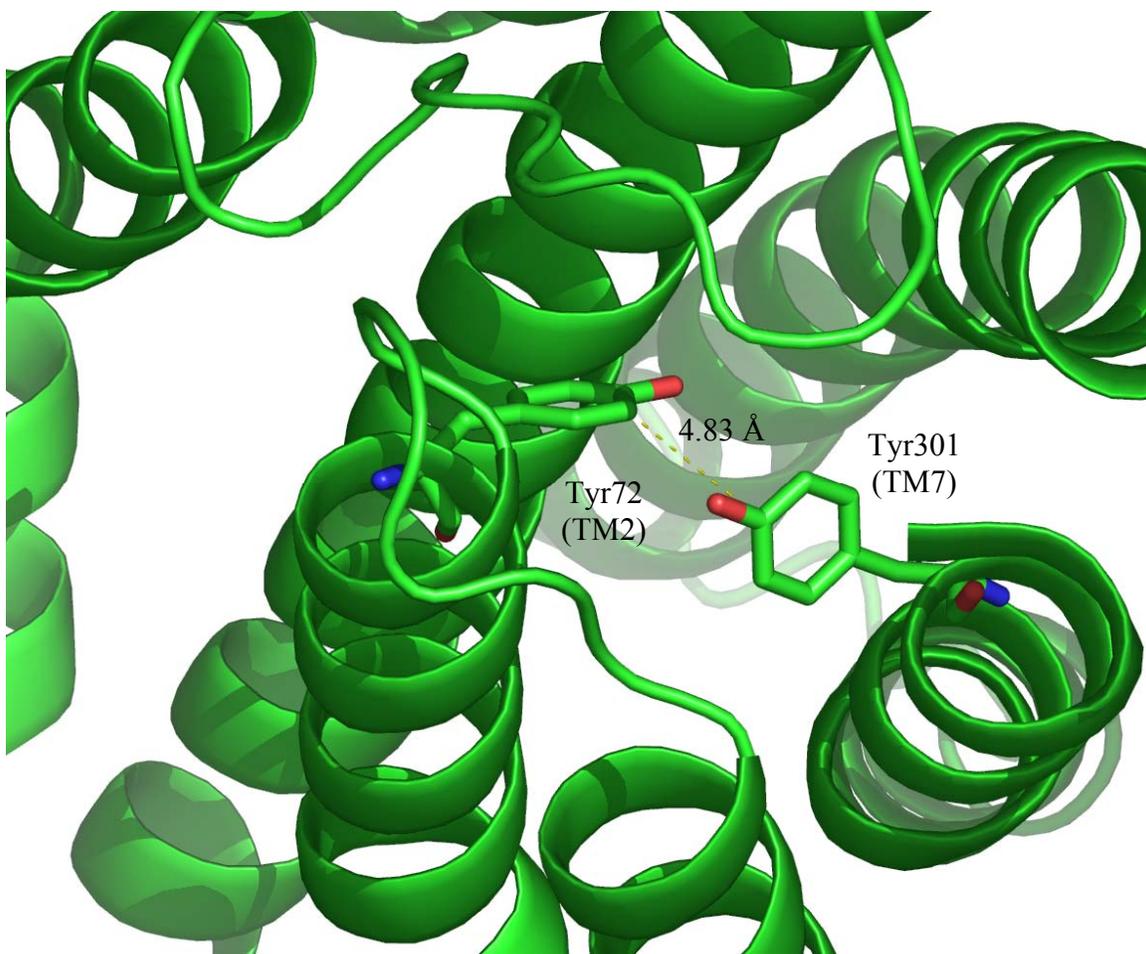


Figure 2-14: hCCR1 inter-helical sigma-pi interaction (TM2-7).

Stabilizing sigma-pi interaction between Tyr72 (TM2) and Tyr301 (TM7). Tyr301 is a conserved TM7 residue. In this figure, the intracellular region is the top face.

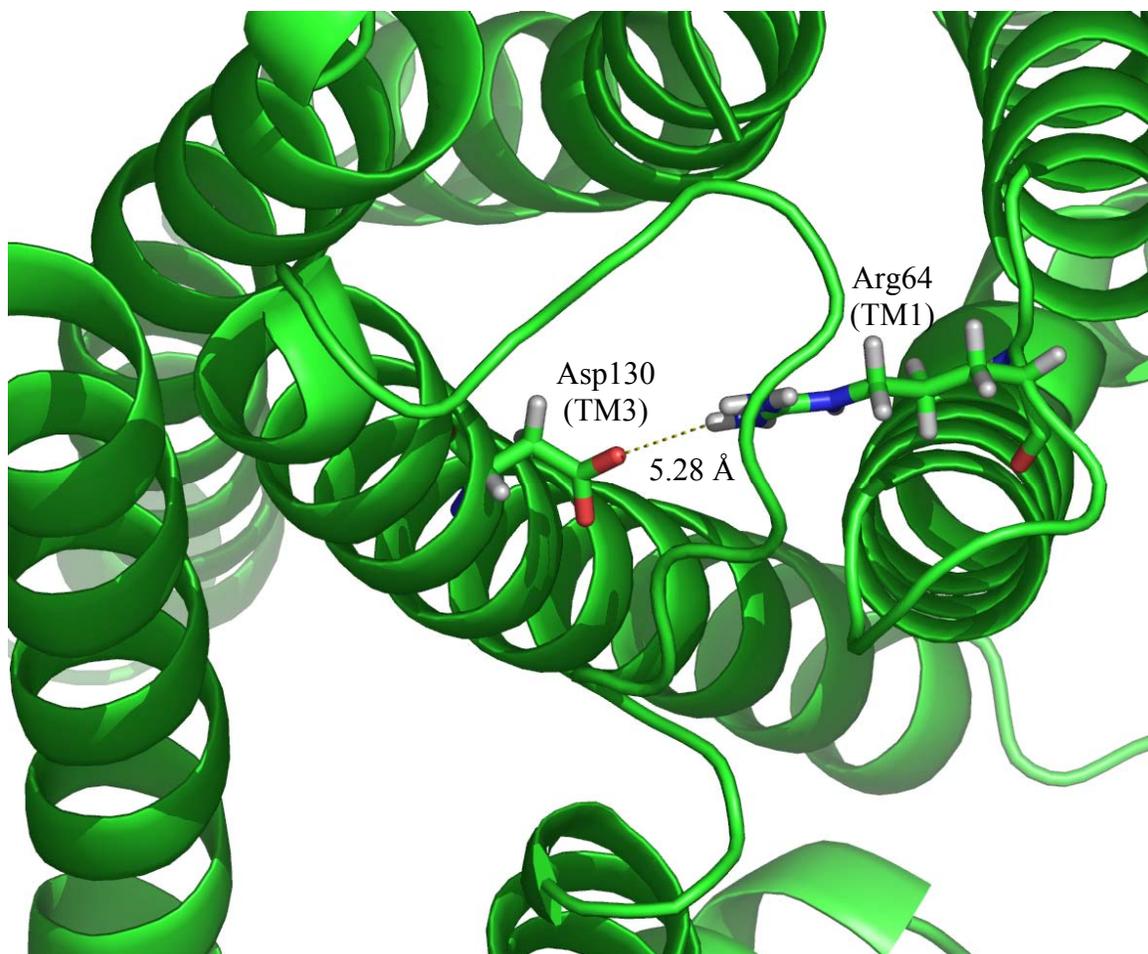


Figure 2-15: hCCR1 inter-helical DRY salt bridge (TM1-3).
Possible salt bridge between Arg64 (TM1) and Asp130 (TM3) of the conserved DRY motif. In this figure, the intracellular region is the top face.

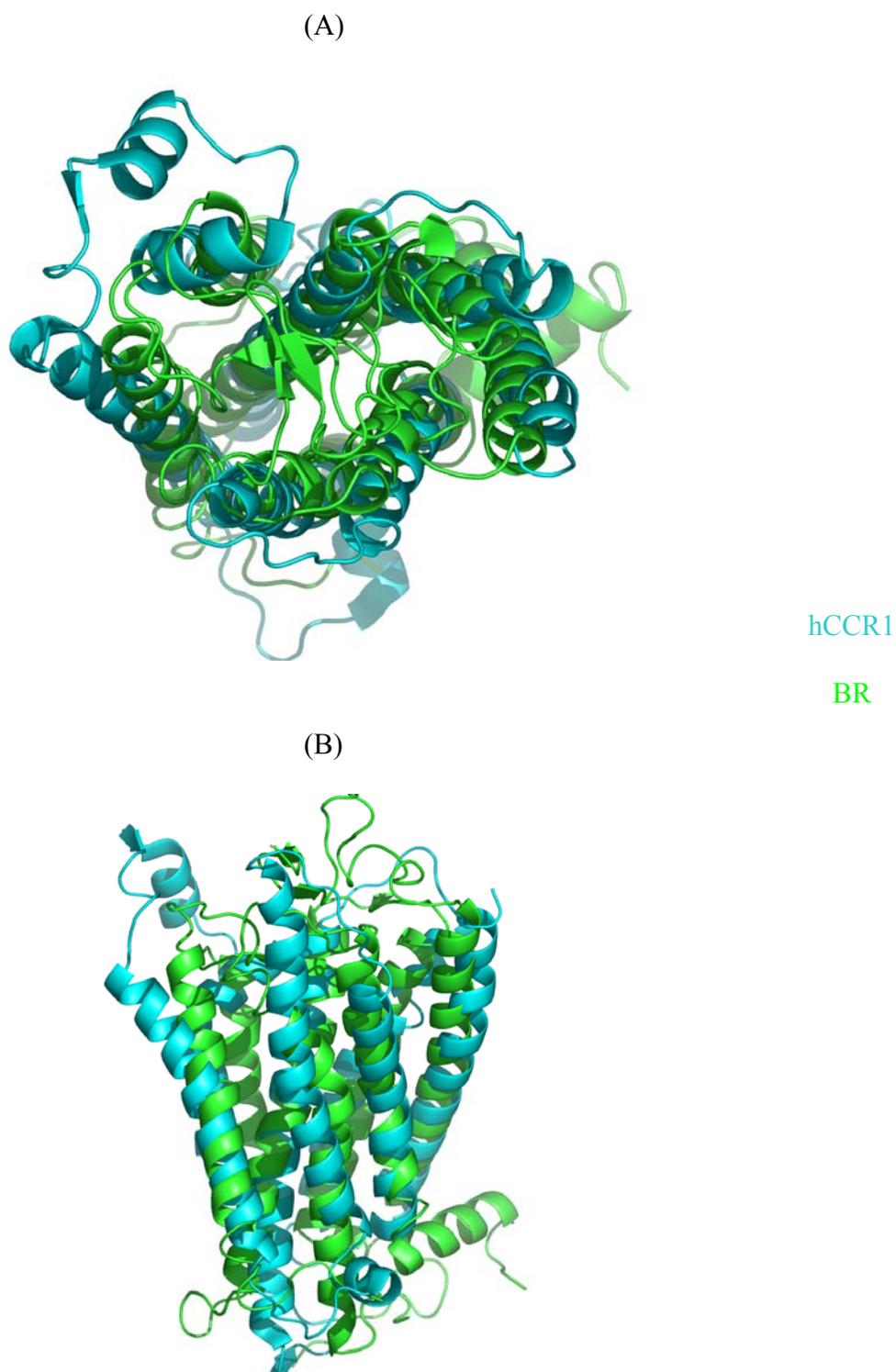


Figure 2-16: Overlay of the hCCR1 and BR structures.

Predicted hCCR1 structure superimposed on experimental BR structure (PDB #1F88).
The main chain RMSD is 5.91 Å.

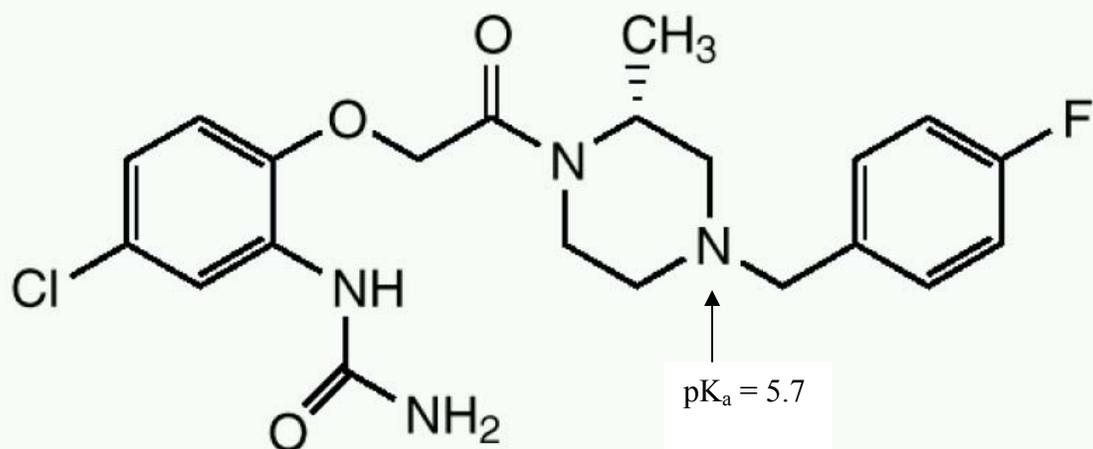


Figure 2-17: Structure of BX471 antagonist.

hCCR1 antagonist with binding affinity of $K_i=1$ nM. Also known as R-N-[5-chloro-2-[2-[4-[(4-fluorophenyl)methyl]-2-methyl-1-piperazinyl]-2-oxoethoxy]phenyl]urea.

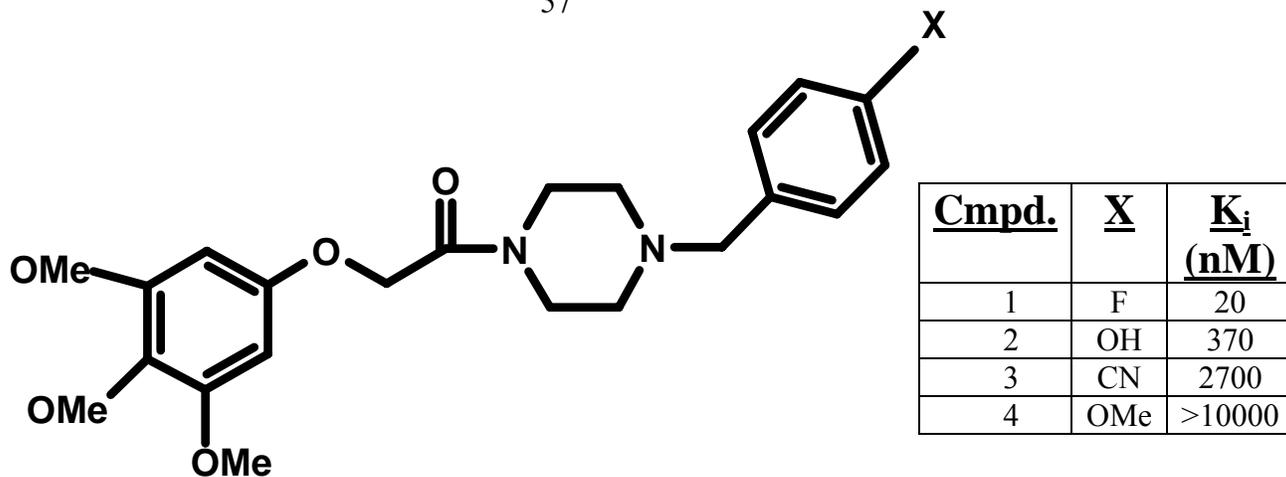


Figure 2-18: Structural template for proprietary antagonists.

Template for a class of hCCR1 antagonists with varying binding affinities (see table above). Data courtesy of Berlex AG.

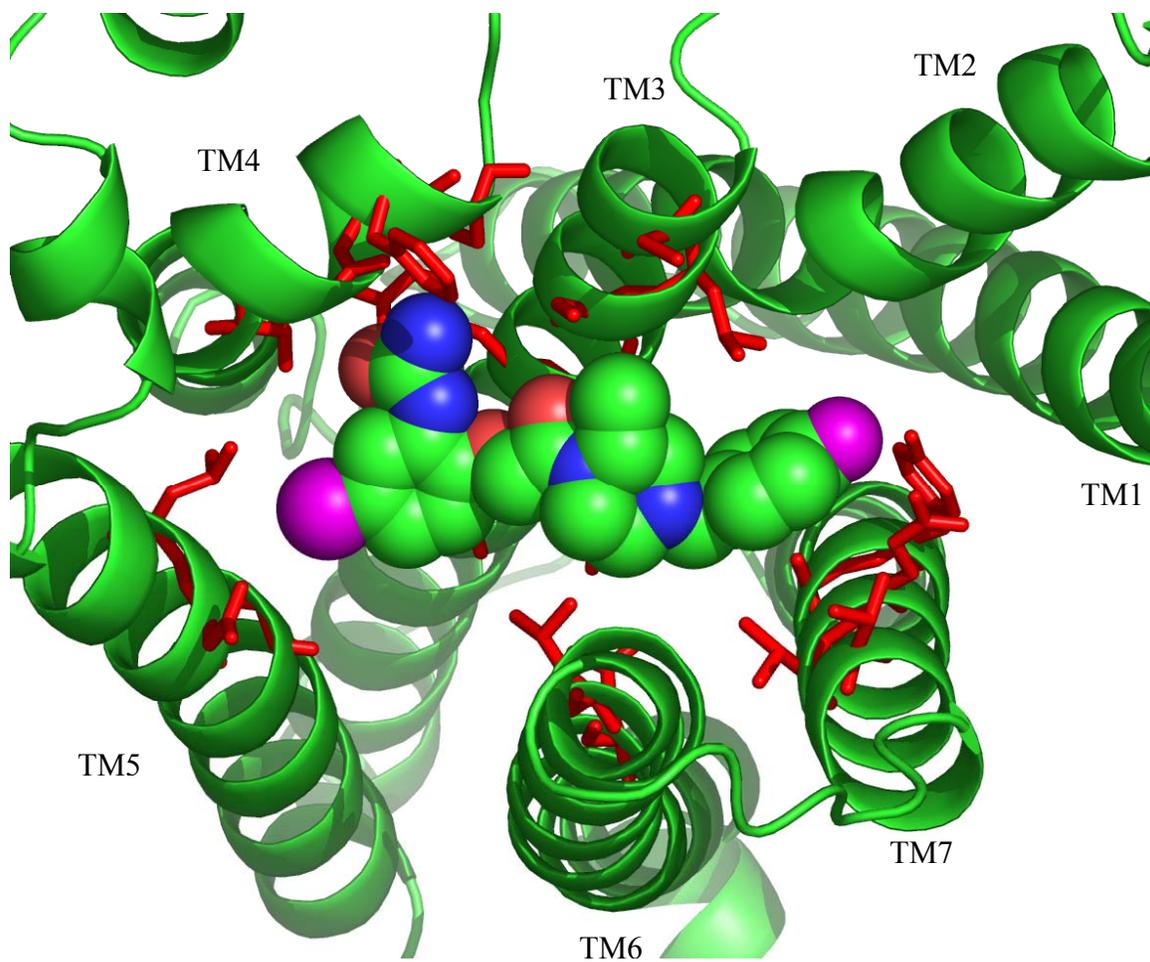


Figure 2-19: TM binding motif for CCR1 antagonists.
BX471 docked in the binding site. Shown in red are residues from TMs 3-7, which are responsible for anchoring the ligand in this cavity.

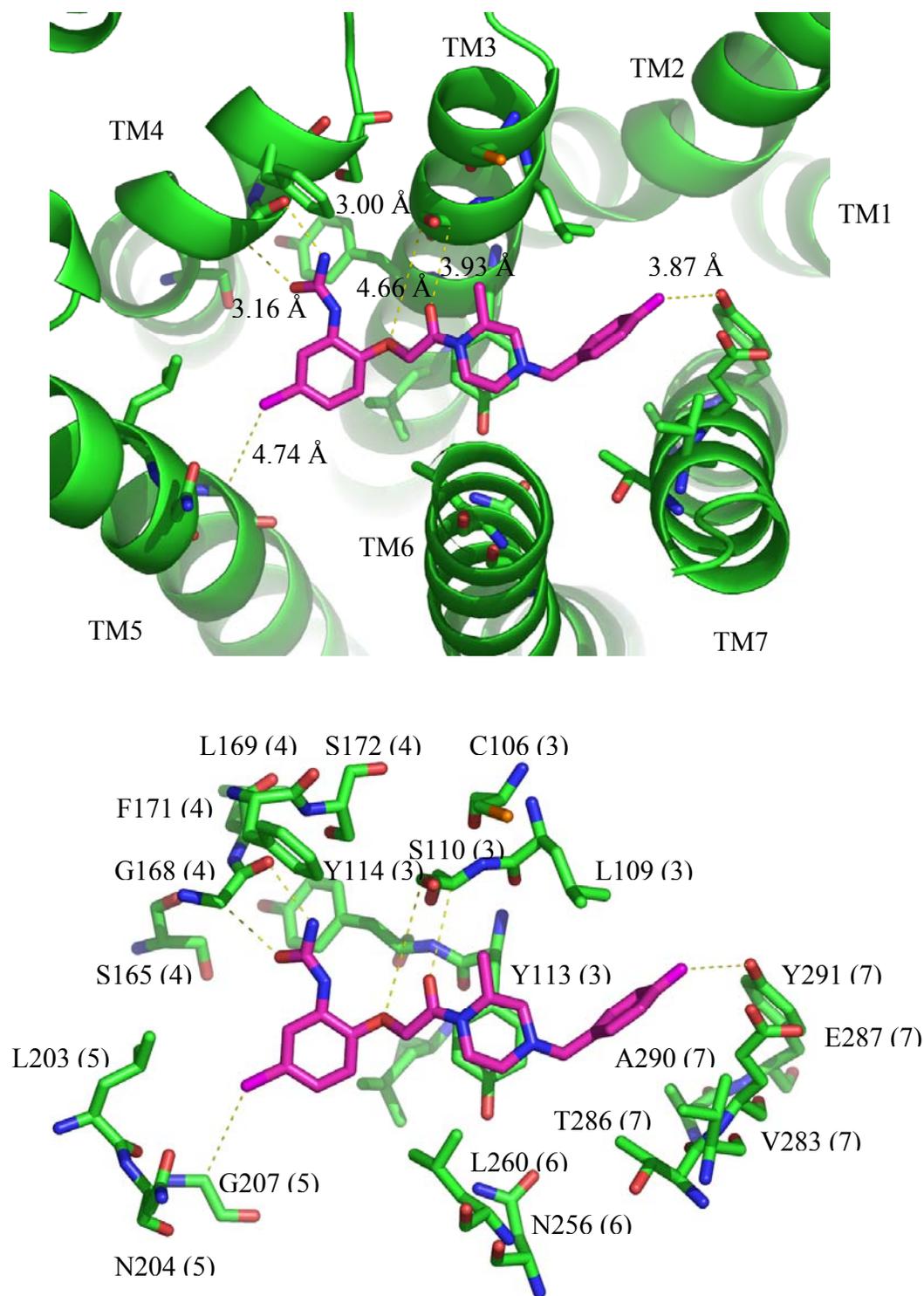


Figure 2-20: Ionic interactions in the 5Å BX471 binding cavity (top).

The ligand is anchored by ionic interactions with Gly168, Gly207, Ser110, and Tyr291. The number in parenthesis indicates the helix number, and the distances shown are between heavy atoms. The magnitude of these interactions is quantified in Table 2-3.

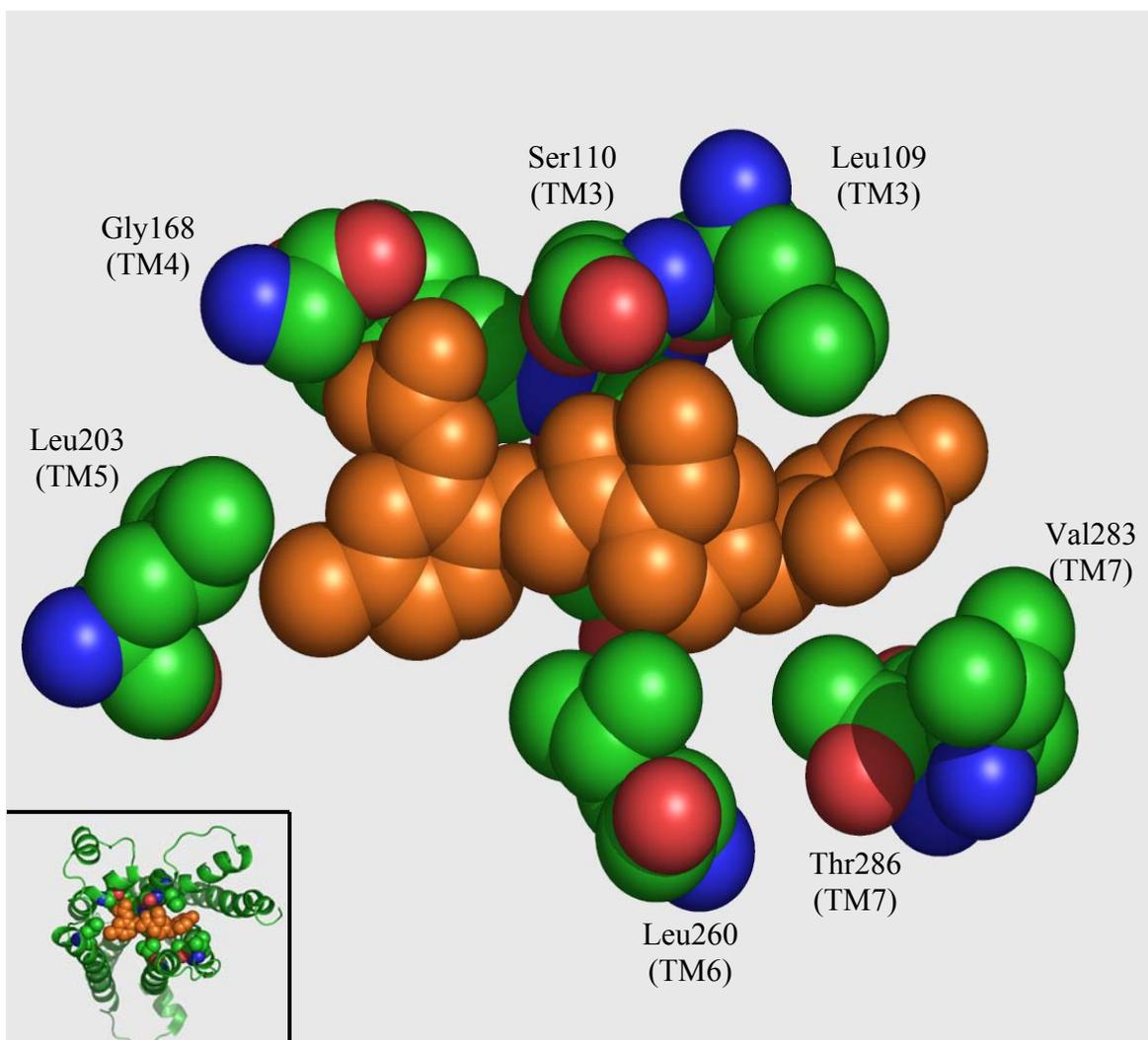


Figure 2-21: Hydrophobic interactions in the BX471 binding cavity (top). BX471 is colored orange, and spheres are used to show the degree of favorable hydrophobic packing around the ligand (within 5 Å). The picture in the lower left-hand corner is provided for perspective (i.e., looking at the bundle through the extracellular domain).

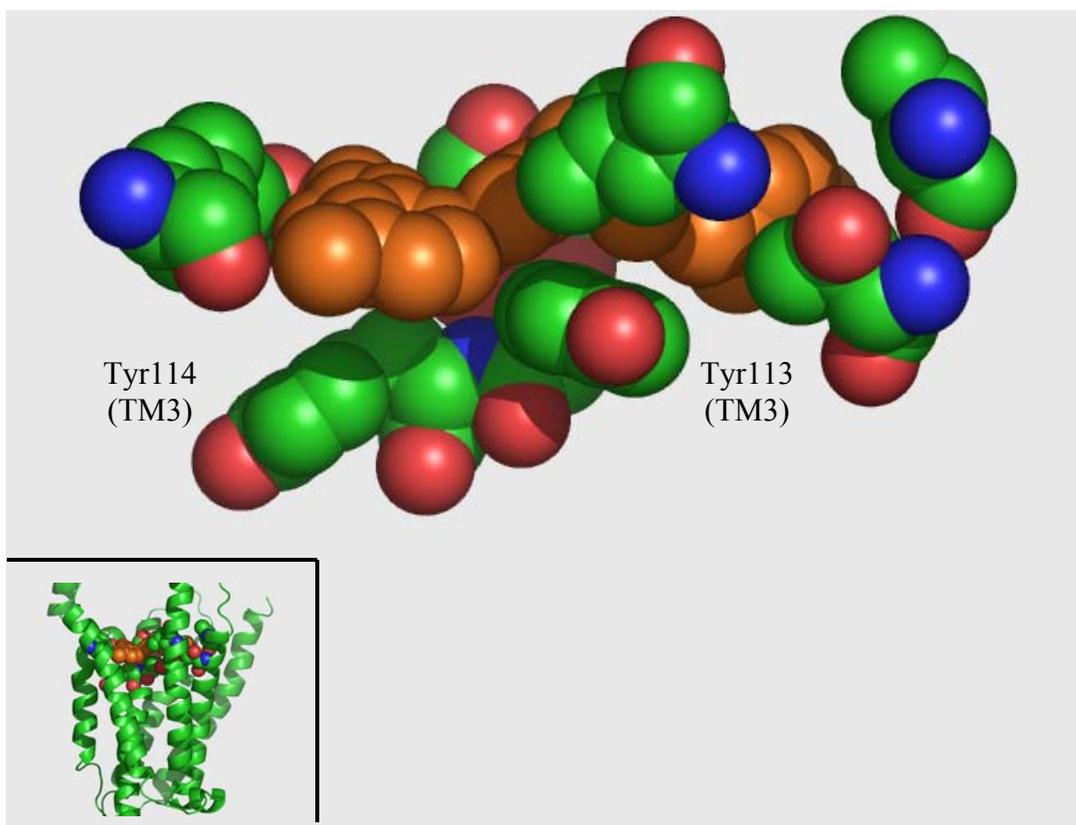


Figure 2-22: Hydrophobic interactions in the BX471 binding cavity (side). BX471 is colored orange, and spheres are used to show the degree of favorable hydrophobic packing around the ligand (within 5 Å). Note the stacking of Tyr113 beneath the plane of the piperazine ring, and Tyr114 below the substituted aromatic ring. The picture in the lower left-hand corner is provided for perspective (i.e., looking at the bundle through the membrane).

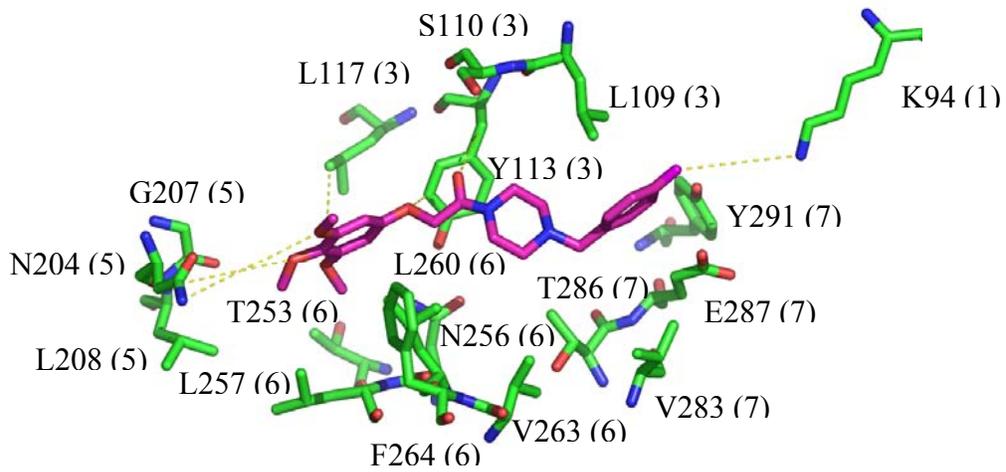
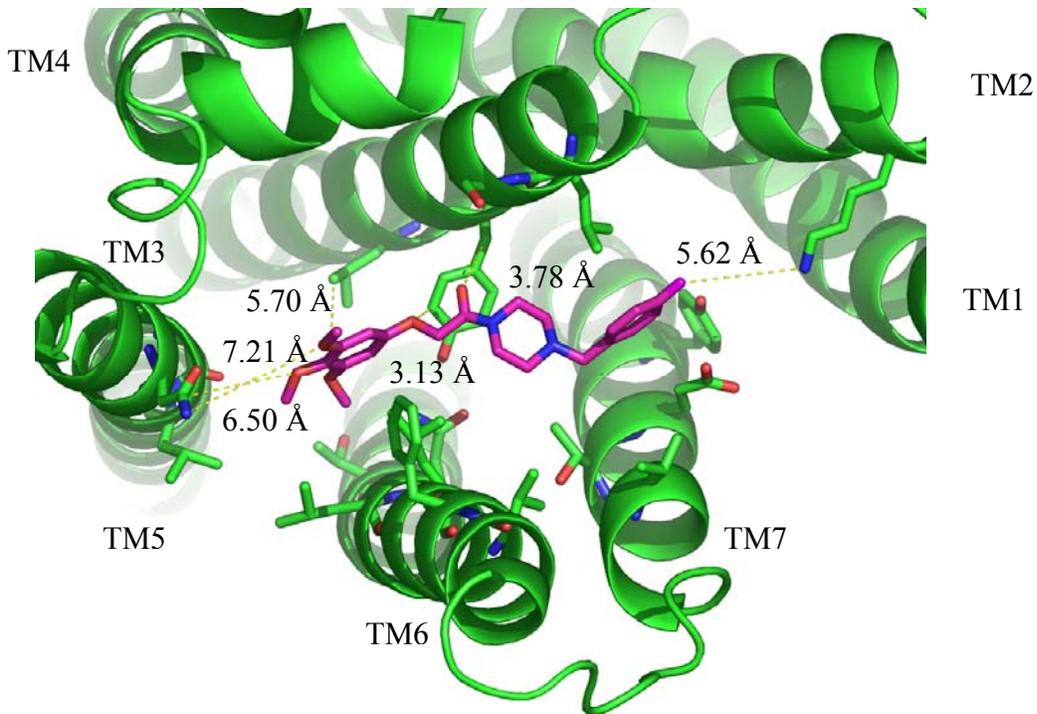


Figure 2-23: Interactions in the C1 binding cavity (top).
Ligand backbone shown in magenta.

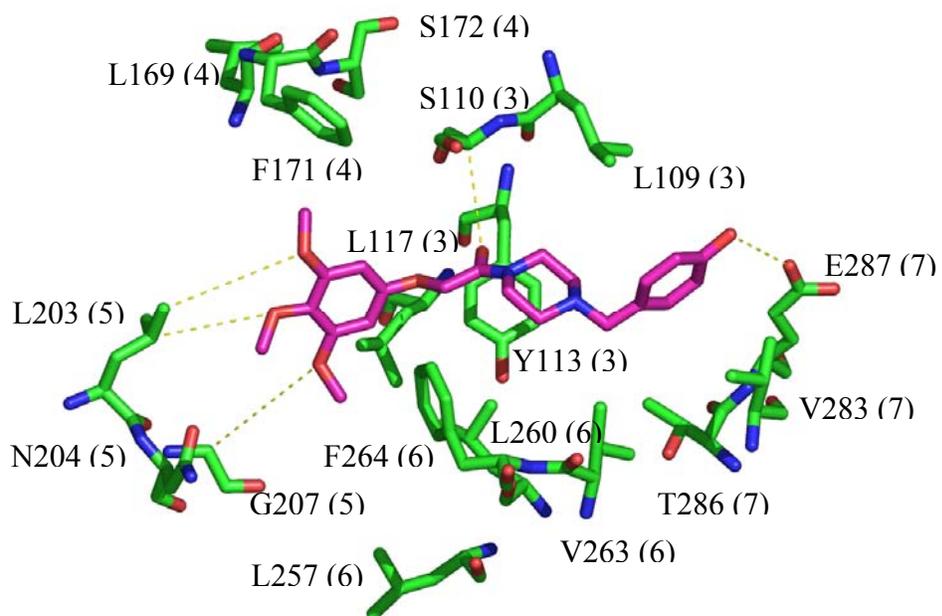
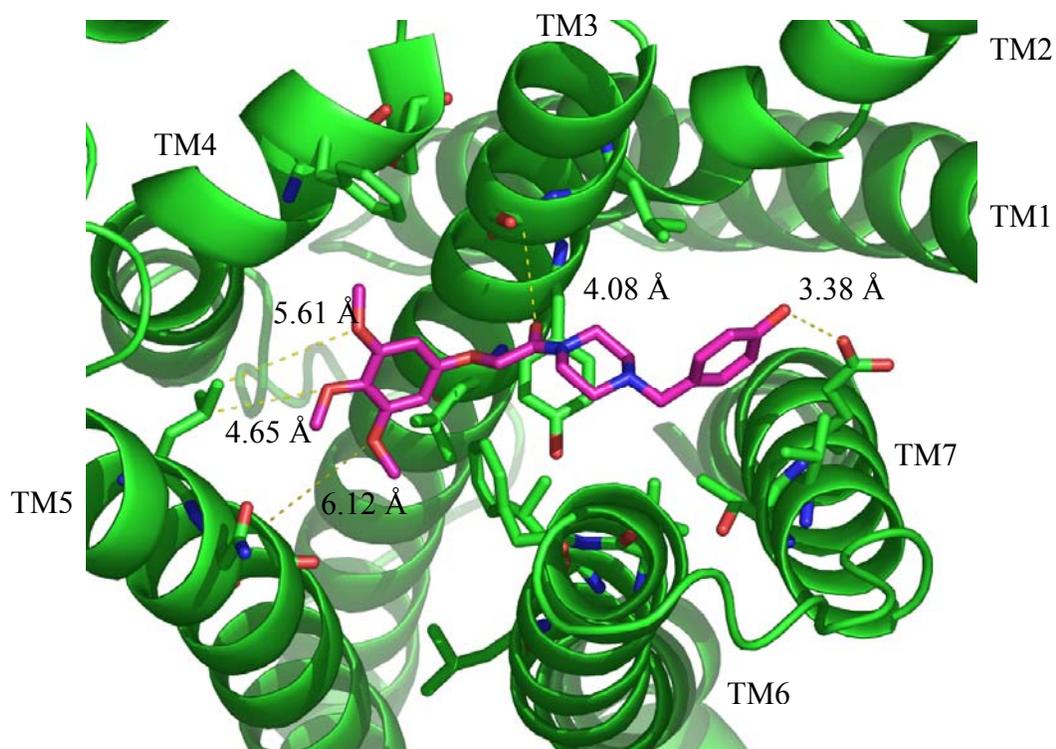


Figure 2-24: Interactions in the C2 binding cavity (top).
Ligand backbone shown in magenta.

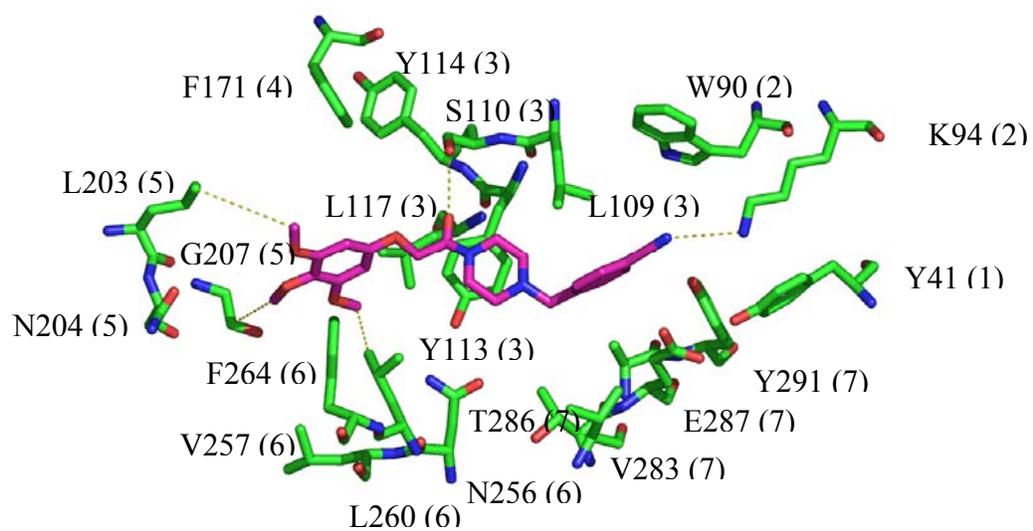
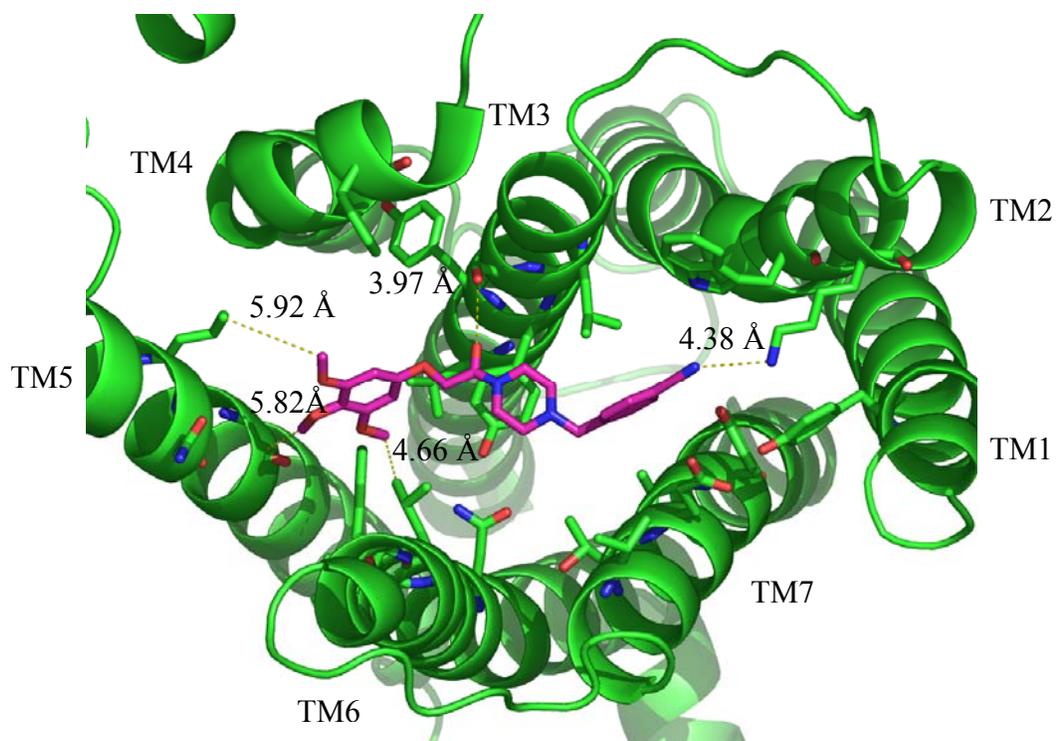


Figure 2-25: Interactions in the C3 binding cavity (top).
Ligand backbone shown in magenta.

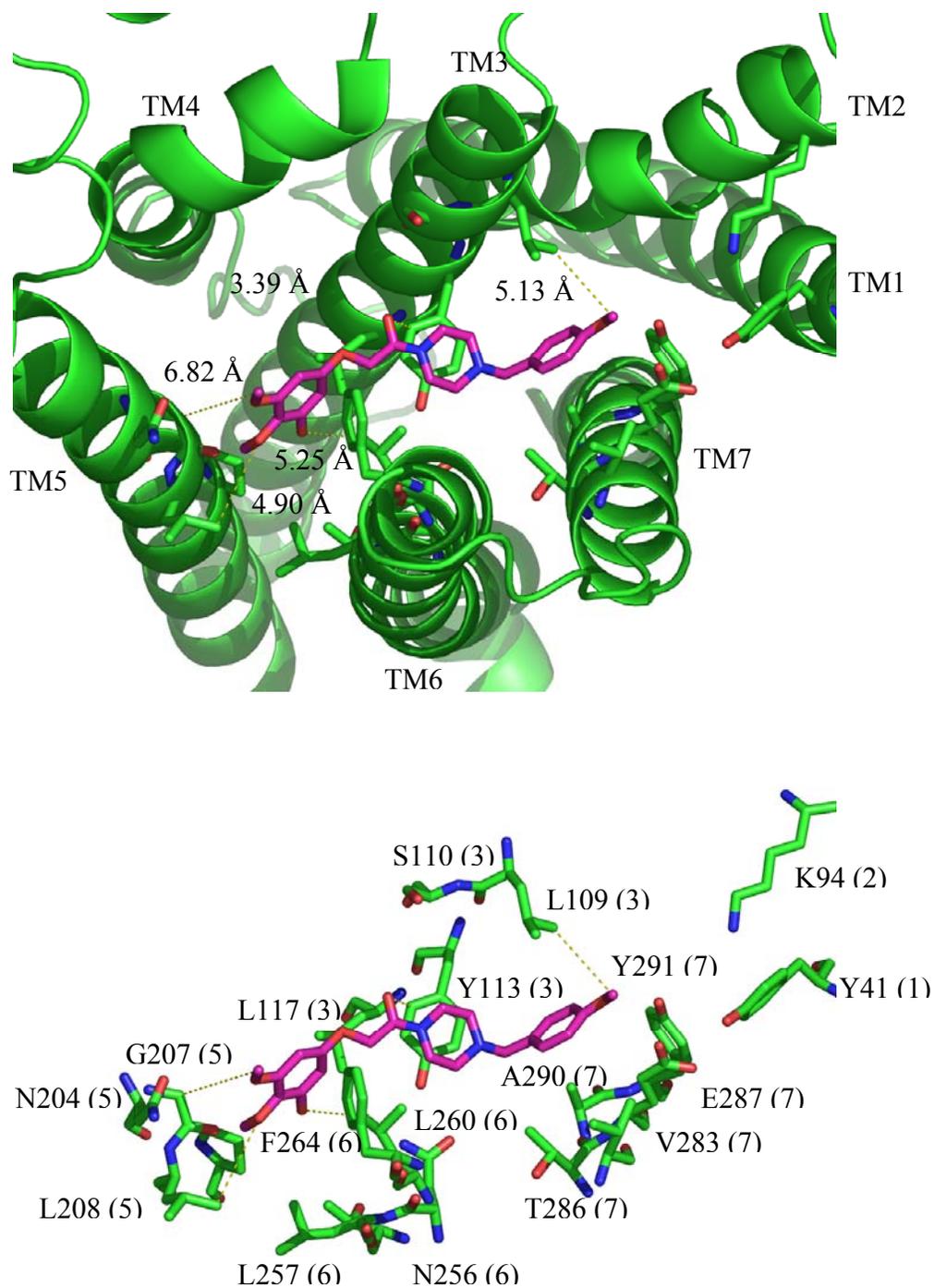


Figure 2-26: Interactions in the C4 binding cavity (top).
Ligand backbone shown in magenta.

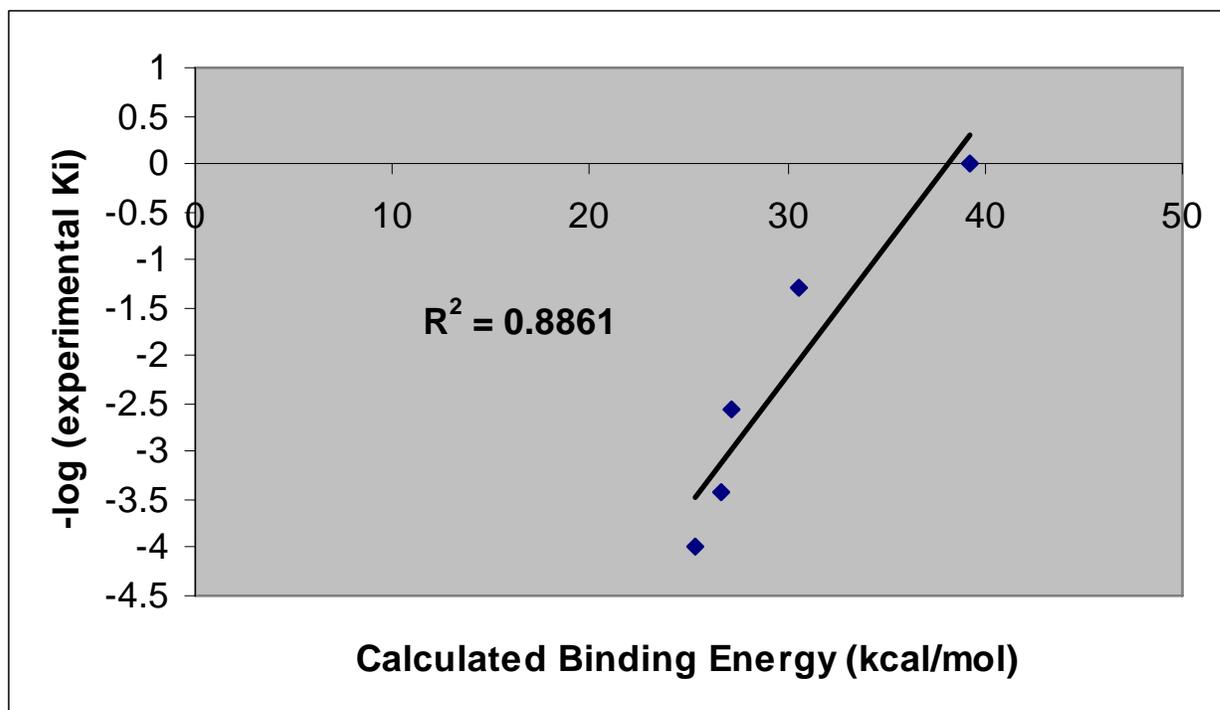


Figure 2-27: Correlation between experimental (K_i) and theoretical (BE) data. Plot of the experimental binding affinity vs. the theoretical binding energy for docked ligand library. Data taken from Table 2-4.

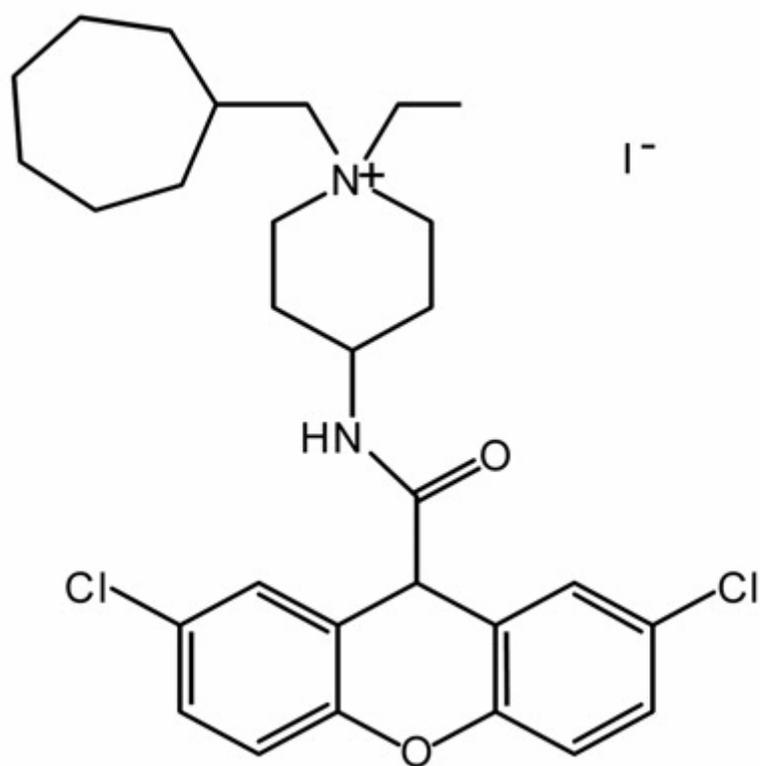


Figure 2-28: Structure of UCB 35625.

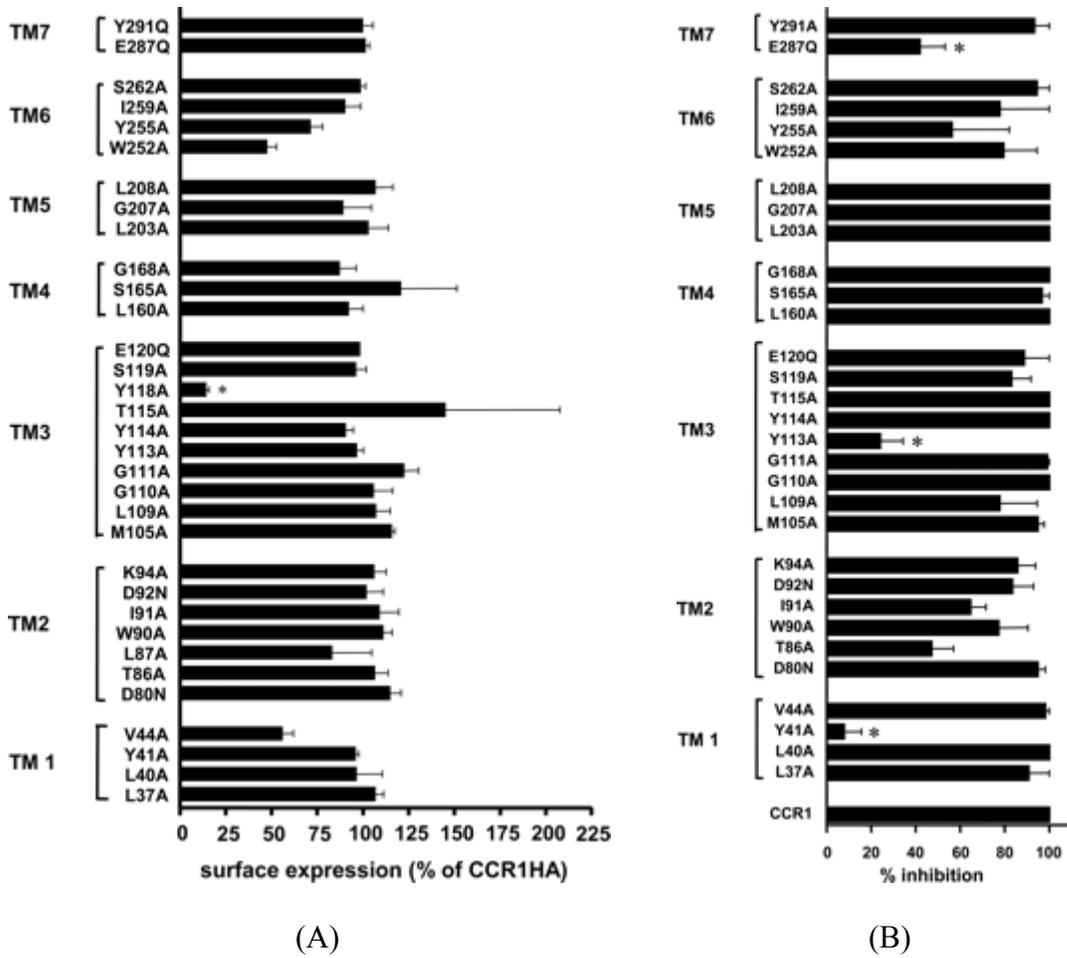


Figure 2-29: Mutant expression (A) and cell migration assay (B) results.
 Reproduced with permission from Lopes de Mendonca, F., et al., (2005) *J. Biol. Chem.* **280**, 4808-4816.

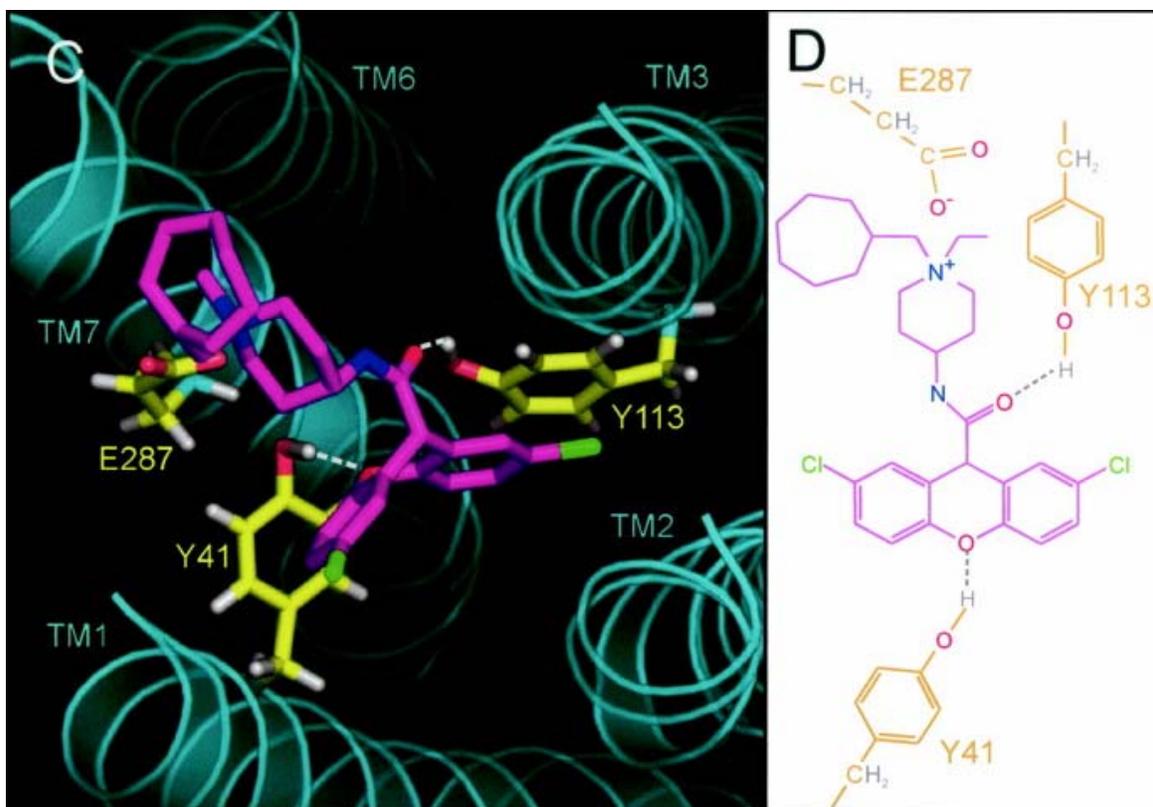


Figure 2-30: Binding mode of UCB 35625 in hCCR1 homology model.

Experimental data implicates Tyr41, Tyr113, and Glu287 in ligand recognition. Both Tyr113 and Glu287 are present in our BX471 binding site, and Tyr41 is within 5.25Å. Reproduced with permission from Lopes de Mendonca, et al., (2005) *J. Biol. Chem.* **280**, 4808-4816.