

Chapter 1: G-Protein Coupled Receptor Structure Prediction: MembStruk Methods and Validation using Bovine Rhodopsin

1.0 Abstract

The G-protein-coupled receptor superfamily is one of the most important drug targets for pharmaceutical companies. However, there is very little structural information available about these proteins. This paper describes the current methods used in MembStruk for the construction of 3D models for these proteins and the methods used in HierDock for prediction of protein function. Bovine rhodopsin was used as a validation case for these methods, having the only crystal structure available to compare against.

The prediction of the transmembrane (TM) regions is improved compared to the latest crystal structure (1U19), and new methods for the analysis of the helical rotation for the MembStruk structures are shown. The rotational analysis reveals that TM region 3 is involved in a rotation of 90 degrees when the structure changes from an open to closed conformation during binding of cis-retinal (14). The MembStruk structure of rhodopsin has a 1.37 Å RMS compared to the binding site of the crystal structure main chain atoms after scanning for potential binding sites.

2.0 Introduction

Integral membrane proteins are coded on 20-30% of genes (1) in humans and other organisms. These proteins take part in processes such as ion translocation, electron transfer, and transduction of extracellular signals. G-protein-coupled receptors (GPCRs) make up one of the most important classes of transmembrane (TM) proteins being involved in cell communication processes and in mediating such senses as vision, smell, taste, and pain. The signals that activate these proteins are usually chemical in nature,

however for the opsin family, the signal is “visible” light (electromagnetic radiation). The malfunction of GPCRs are implicated in the pathology of many diseases and their progression such as ulcers, allergies, migraine, anxiety, psychosis, nocturnal heartburn, hypertension, asthma, congestive heart failure, Parkinson’s, schizophrenia, and glaucoma (2-3). This makes GPCRs an essential target for drug development. In fact, GPCRs only account for about 3-4% (4) of the human genome, and yet are targets for more than 50% of the drugs in the current market (5).

One of the major challenges in drug development for GPCRs is to design subtype specific drugs. Since GPCRs of one particular function have many subtypes, design of subtype specific drugs calls for structural information on the target GPCRs. Unfortunately, there is very little structure information on GPCRs although these proteins are important drug targets. In fact, there is only one experimental 3D structure for a single GPCR, which is bovine rhodopsin (4-5). The sequence identity to rhodopsin is low for most GPCRs of interest (17 % for dopamine, 14 % for serotonin), so the use of homology modeling for obtaining reliable structures is not a valid option (6).

Due to the difficulty in generating 3-D structures using high resolution X-ray diffraction data or NMR data for GPCRs, it is widely accepted that theory and computation to predict the 3-D structures of GPCRs from first principles can aid the structure-based drug design for many GPCR targets [for example, Strader 1994, Parrill 2000 and many other references for different GPCRs]. Successful protein structure prediction methods for globular proteins generally utilize homology to known structures [7]. This is not practical for GPCRs, since there is just one crystal structure. Moreover, homology-derived models are not reliable when the sequence homology is very low, i.e.,

below 30% or less (in the “twilight zone”) [9-12]. Thus we believe that it is important to explore the viability of MembStruk for predicting structure of GPCRs.

With only one experimental 3D structure available, bovine rhodopsin is the choice for testing and validation of current MembStruk methods. Using the amino acid sequence available, the structure of bovine rhodopsin will be built and compared to the current available structure 1U19 (chain A). During the process of validation, MembStruk was improved with methods for TM prediction and helical rotation that now constitute version 4.30 described in the methods.

3.0 Methods

GPCRs have a well defined, three dimensional topology, with seven helical TM domains, which provides an organizing principle (allowing some of the structural information to be deduced from sequence) that we have used to advantage in developing the MembStruk first principles method [13-14], which uses no information from the high resolution crystal structure of rhodopsin or bacteriorhodopsin. MembStruk is a hierarchical method that optimizes the rotational and translational orientation of the helices at a coarse grain level using Monte Carlo methods and does fine grain optimization in explicit lipid system. MembStruk method has been developed by me, Trabanino, Vaidehi and Goddard and the advancements I have made in the methods are detailed in this chapter.

Membstruk4.3, the current version of this method, consists of several steps as shown in a simple flow chart in Figure 1.1. A short description of the various steps in MembStruk is given below followed by a detailed description of each of these steps in the other sections of this thesis.

1. **TM Prediction:** Predict the seven TM domains using hydrophobicity analysis “**TM2ndS**” [14] combined with information from sequence alignments. The extent of the TM regions are predicted using sequence alignments of sequences varying from 20% to over 90% sequence variability as input. The second step of TM2ndS is to calculate the consensus hydrophobicity for every residue position in the alignment using the average hydrophobicity of all the amino acids in that position over all the sequences in the multiple sequence alignment. Then, we calculate the average hydrophobicity over a *window size (WS)* of residues about every residue position, using WS ranging from 12 to 20. This average value of hydrophobicity at each sequence position is plotted to yield the hydrophobic profile, for WS=14. The baseline for this profile serves as the threshold value for determining the TM regions.
2. **Position of maximum hydrophobicity:** Identify lipid-accessible residues from the sequence alignments (as the less conserved residues) and from analysis of the peaks in hydrophobicity [22, 40] of the hydrophobic residues in the sequence. The position of maximum hydrophobicity called the “hydrophobic center” is calculated as the maximum of each of the seven peaks. In the next section, I have detailed the substantial improvements to this procedure determining the maximum especially for complex cases like for example TM 2 and 3 (see Appendix A), cases where the hydrophobicity peak has double maxima.
3. **Optimization of helical kinks:** Construct canonical helices for the predicted TM segments and optimize the structures of the individual helices, with conjugate gradient energy minimization using the dreiding all atom force field (QEq

charges) followed by molecular dynamics (default) or fast torsional NEIMO dynamics (optional) [41-42]. This optimizes the bends and kinks in each helix.

4. **Assemble the helix bundle:** The helical axes are oriented according to the 7.5 Å electron density map of frog rhodopsin [27], depending on sequence similarity. The relative translational orientation of each helix is based upon forming the best fitting plane of all the hydrophobic centers obtained from step 2. This step is crucial to determining the location of ligand binding sites using the predicted structures.
5. **Monte Carlo Optimization of rotation and translational degrees of freedom:** This step is an important step that optimizes the rotational and translational degrees of freedom of each helix with respect to the other. Here optimize the rotational and translational orientation of the helices using a systematic search algorithm over a grid of rotational angles and translational distances. This step allows the system to surmount energy barriers. Coarse grain optimization of the helical orientations is performed using the net hydrophobic moment of the middle one-third of the helix about their hydrophobic centers.
6. **Optimization of the assembled helical bundle in explicit lipids:** Embed the helix bundle into a lipid bilayer and optimize the composite system. Equilibration of the helix bundle plus lipid bilayer system uses Rigid Body Molecular Dynamics [20-21]. The helix bundle surrounded by lipid bilayers was optimized using rigid body dynamics with DREIDING forcefield [18] and CHARMM22 [19] charges for the protein.

7. **Optimization of the final model:** Construction of the inter-helical loops and disulfide bridges using Whatif [43] or Modeller [32-34]. Optimization of the final model in presence of lipid bilayers.

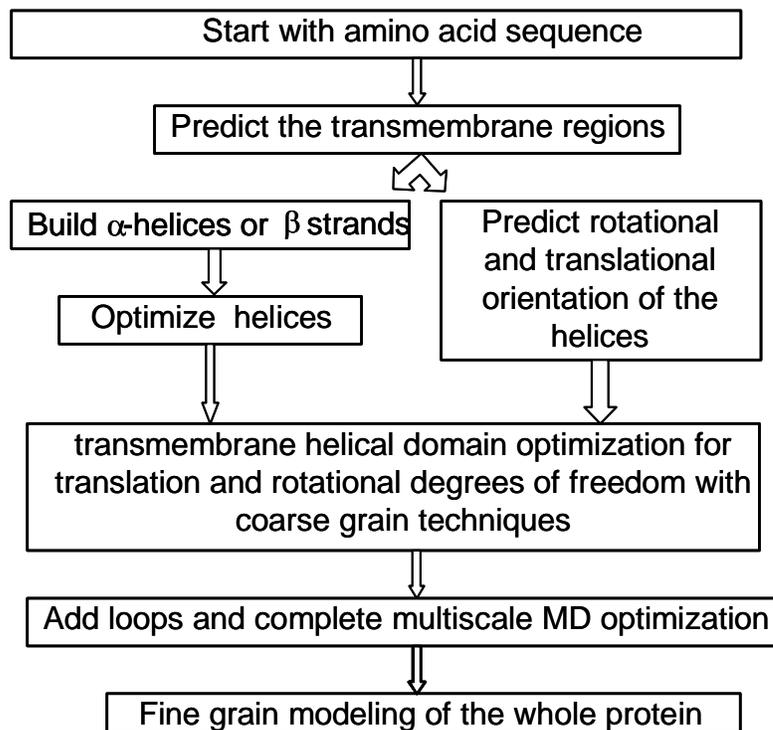


Figure 1 - Flow Chart of MembStruk, the ab initio method for predicting GPCR structures.

3.1 ForceFields and Molecular Simulation Setup

Unless noted differently, all calculations done will use the DREIDING force field (FF) (18) with charges obtained from CHARMM22 (19) for the protein. The van der Waals (vdW) and electrostatic (Coulomb) interactions will be calculated using the cell multipole method (CMM) (20) and a dielectric constant of 2.5 in MPSim (21). The default hydrophobicity scale to be used will be the Eisenberg scale (22). Solvation is done with explicit ions (Na^+ and Cl^-) added next to each charged residue (excepting those forming a saltbridge) and explicit lipid bilayer – dilauroyl phosphatidylcholine.

These ions are given a modified molecular surface of 131.000 (Na⁺) and 143.000 (Cl⁻) to represent these as hydrated ions. Dreiding FF is also used for lipids with charge equilibration charges (31).

Unless noted otherwise in the text, minimization of a structure is always done on a BGF file format of the structure using MPSim. Using the standard force fields with the CMM expansion method set to centroid and at level 2. The non-bond update frequency and cell reallocation frequency are set at 5 steps. The minimization is done using the conjugate-gradient (conjugate-wag) method, and stops after 1000 steps or when a RMS of 0.5kcal/mol/Å in force is reached.

3.2 TM2NDS v4.30 – Prediction of the Bovine Rhodopsin TM Regions

Utilizing the fact that the middle of the protein is helical and immersed in a lipid bilayer (membrane), these regions will have a larger grouping of hydrophobic residues with the hydrophobic center being near the middle of the membrane which is highly hydrophobic (23). TM2ndS looks for these conserved hydrophobic portions of the protein sequence to be assigned the helical TM regions. This is accomplished through the following steps:

3.2.1 Sequence Alignment

Generate a multiple sequence alignment using a NCBI Blast search (24-25) on the SwissProt database and filtering out those hits less than a 200 bit score. ClustalW (26) is then used to align the sequences. This alignment of all sequences is used to filter out similar sequences that tend to cluster in certain percentile regions, so that an even distribution of sequences from 100 to 20 percent homology is kept and used for the final sequence alignment.

3.2.2 *Coarse Hydrophobic Profile TM Prediction*

The SeqHyd hydrophobic profile algorithm is then used to calculate an averaged hydrophobic score across the entire alignment. This score is averaged over a range of nearest neighbors in the sequence alignment, and the size of this range is called the window size (i.e. a range of 5 on each side of the target residue will be window size 10). The profiles generated from window sizes 12 to 30 are analyzed to look for seven distinct hydrophobic peaks using a baseline derived from the global average of the profile (± 0.05 of the baseline choosing the closest to the average). The lowest window size that yields seven peaks is used to determine the ends of the TM regions by finding where the peaks cross the baseline.

3.2.3 *Finer optimization of TM length prediction*

Fine grain prediction of the TM regions is done using the TM2ndS helix capping program (14). This will look for helix breaking residues (Pro) or charged residues, and alter the ends of the TM region up to four residues on each side (minimum 21 residues) to exclude these.

3.2.4 *Determination of Hydrophobic Centers*

The target's alignment hydrophobic profile from step 3.2.2 is now used to find the point of maximum hydrophobicity or "hydrophobic center" in each TM region. The profile is averaged over varying lengths (12 to 30 residues) and the hydrophobic peaks within the regions defined from step 3.2.3 are analyzed. Each TM region has the window sizes ordered into contiguous groups that contain all sizes that are within three residues of the each others peak. Then, each window size within a contiguous group on every TM region is listed and the largest set of window sizes that are continuous is chosen. These

window sizes that give consistent peaks over every TM region are then averaged to determine the TM hydrophobic centers.

3.2.5 *Summed Window Analysis (Optional)*

The final TM predictions are plotted on the lowest window size that contained seven helices from step 3.2.3. This plot is then visually inspected to make sure that there are clear defined regions between TM 2 to 3 and TM 6 to 7, as these areas often have less defined edges (see Appendix A, Graph 3). If there appears to be a discrepancy between the predicted regions and the peaks of the hydrophobicity plots, then the plots for different window sizes of the hydrophobic centers are summed together to produce one hydrophobicity plot for analysis (see Appendix A, Graph 5).

The predicted TM regions are then compared against this summation plot, and new regions are chosen that maintain the core TM residues (altering the ends to a maximum of six residues). These new TM predictions are then run through the fine prediction protocol used in step 3.2.3. These new predictions will then be used to recalculate the hydrophobic centers as done in step 3.2.4.

3.3 *MembStruk v4.30 – Construction of the 3D Protein Structure*

Once a consistent set of TM predictions has been obtained from TM2ndS, the construction of the 3D protein structure begins. The helical TM regions are built first and optimized in their position, shape, and rotation. Then, using the helices as starting points, the more variable loop regions are built using homology modeling.

3.3.1 *Building the Canonical Helical Structure*

Using the sequences for the final predicted TM regions, right-handed canonical helices are built using extended side chain conformation for all the residues. Each helix

is then minimized in vacuum, and then these helices are fitted to the 7.5 angstrom electron density map of frog rhodopsin (27). The electron density map is only used for the initial placement of the tilt of the helices and the initial packing of the helices into a bundle.

3.3.2 *Rotation and Translation of Helices (hcenterTR)*

The seven hydrophobic centers calculated in step 3.2.4, are then fitted to plane which translates the helices relative to each other. The hydrophobic centers have been observed to lie in a plane for cases like the crystal structure of bovine rhodopsin and bacteriorhodopsin (14-16, 29). The translation of the helices are done along the principle moment of inertia axis (the length of the helical barrel). This places the hydrophobic centers so that they lie in the same plane that was occupied by the geometric centers on the frog rhodopsin electron density map.

Each helix will have a plane centered on its hydrophobic center with the z-axis along the smallest moment of inertia (along the length of the helix). The C-alpha atoms in the other six helices are projected onto this centered plane and the distance each projected atom on this plane is reduced to one. This forms a circle of radius 1 that contains all the C-alpha atoms, with the atoms from each helix connected to form an arc. The arc angle gaps in this circle are measured, and the largest one is taken to be the hydrophobic face angle (since this corresponds to the largest side of the helix that can face the lipid, see Chapter 2 - Figure 6). Now the C-alphas from the target helix are projected onto this plane, and the entire circle is scanned for the arc (equal to the face angle) that contains the most hydrophobic residues (using Eisenberg's scale). This effectively finds the largest hydrophobic side of the helix that will face the lipids. The

middle of this hydrophobic face is rotated to the middle of the largest gap found in the centered plane (See Chapter 2, Figure 2). The positions of the side chains will be optimized using SCWRL (30). The structure produced after this step is called the MembStruk template.

3.3.3 Optimization of Helical Bends and Kinks in the canonical helical structures (fixhelix)

The MembStruk template is a perfect canonical structure that needs to be relaxed into the correct helical shapes. Each individual helix is taken individually and run through a molecular dynamics simulation. First, each is minimized with hydrated ions, and then run through Cartesian molecular dynamics at: 300 Kelvin, for 200 picoseconds, time step of 2 fs, CMM level at 0, constant temperature and pressure (TVN), Nose-Hoover equations with a τ of 0.05. Another option is to use NEIMO dynamics using the same settings as above. Plus for either simulation, the simulation can be run using the standard charges or qeq neutral charges on every residue. After the simulation has ended, a snapshot of the structure is taken at every picosecond, and the structure with the lowest energy after 100 picoseconds is taken and minimized with hydrated ions. This snapshot file is now used as the final structure for that helix. Once all helices are finished running through dynamics, the entire structure is minimized with hydrated ions.

3.3.4 Coarse Grain Rotational Orientation Optimization of the TM Regions (hcenterTR-bgf)

The seven hydrophobic centers are fit to a plane using a three dimensional least squares approach (see Chapter 2, Section 3.2). This plane then has the hydrophobic centers and the hydrophobic moments of each helix projected onto the plane. Set zero

degrees rotation as the position of the line pointing away from the hydrophobic center of the helix and the center of the seven hydrophobic centers. The projected hydrophobic moments of each helix are then rotated to this zero position. The positions of the sidechains are then optimized using SCWRL (30).

Helix 3 is a special case since its center often is very near (within 1-2 angstroms) of the center of the protein. This means that zero degrees might change drastically with a slight change in the planar position of helix 3 (which is possible during the RBMD step). So a centered plane is created as in step 3.3.2, and the extracellular (top) 15 residues from the hydrophobic center are used to create a hydrophobic vector along with another vector created from the intracellular (bottom) 15 residues. The top vector is rotated to be inside the H2-H4 gap, while the bottom vector is rotated to be inside the H4-H5 gap. This is due to the tilt of helix 3 that one side at the extracellular region faces the lipid while the other side (~90 degrees) faces the lipid on the intracellular region. (See Chapter 2, Figure 7)

3.3.5 Fine Grain Rotation Optimization of TM Regions using all-atom FF Energy Minimization (rotmin)

Each helix is rotated plus and minus a specific set of rotational angles, within the entire protein, the side chains are optimized with SCWRL, the entire protein is minimized for 80 steps, and then minimized again (80 steps) with all but the target helix fixed to get the energy of the rotated helix. The energy minimization step default was set at 80, because it is sufficient to minimize the good rotations, without becoming too time consuming. This process generates three rotated structures: the positive angle rotation, the negative angle rotation and the zero angle rotation. The lowest energy rotation is

picked as the starting point for the next helix to be rotated. This is a systematic optimization procedure for finding a rotational energy minimum for the helices. The helices are rotated in the following order: 3, 2, 1, 7, 6, 5, 4. The rotational optimization begins with helix 3, since this is the helix that is located almost inside the TM barrel and hence optimization by hydrophobic moment is not sufficient for this helix. After all seven have been rotated, a new degree is chosen and all seven in the above order are rotated again. The order of the degrees to rotate is as follows: 5, 10, 15, 20, 25, 20, 15, 10, 5.

3.3.6 Fine Optimization of Helical Packing (RBMD)

Next, two layers of explicit lipid bilayers are added to the protein. This lipid consists of 52 molecules of dilauroylphosphatidylcholine lipid around the TM bundle of seven helices. For the lipids we used the DREIDING FF with QEq charges (31). The protein is minimized for 500 steps, and then rigid body molecular dynamics (RBMD) (300 Kelvin, for 50 picoseconds, CMM level at 4, TVN, Nose-Hoover equations with a tau of 0.05, 1 fs time step, hydrated ions, dielectric constant 8.0) is run in MPSim. The protein sidechains are then optimized in the lipid (SCWRL) and minimized (500 steps) again.

3.3.7 Generation of alternate Structures with different rotational orientations: Combinatorial Scanning of the Rotational Energy

GPCRs are very dynamic structures and are known to exist in more than one conformation, namely the active and inactive states even in the absence of an agonist (P.G. Strange, 2002, Trends in Pharmacological Science, 23(2), 89-95). Hence in this step, we generate alternate low energy TM barrel conformations by systematically varying the rotational orientation of each helix by 5 degrees and then optimizing the TM

barrel and calculating the total energy of the rotated helix. Each helix of the protein in lipid is rotated by +5 degrees, minimized (0.3 RMS, 500 steps) and then all but the target helix is fixed and minimized (0.3 RMS, 800 steps) to obtain an energy score. The number of inter-helical hydrogen bonds and salt bridges formed in each rotated conformation is counted as well. This is done repeatedly until the helix has been rotated a total of 360 degrees and a profile is made of the energies for each degree. Using this information, a set of rotations is chosen based on lowest energies for each helix. Next several structures are created (all possible combinations from the rotations chosen) minimized (0.3 RMS, 500 steps) and energies are obtained (fixing the lipids, 800 steps, 0.3 RMS). This information is then used to decide which combinatorial structure is best (more than one can be used and then all will be used in HierDock). A sample rotational scan plot is shown in Figure 2. This shows that the helix has 4 possible minima for rotations (0, 90, 165, -40). All these minima are included in the combinatorial analysis. Details of the procedure of rotation, axis of rotation are given in Chapter 2.

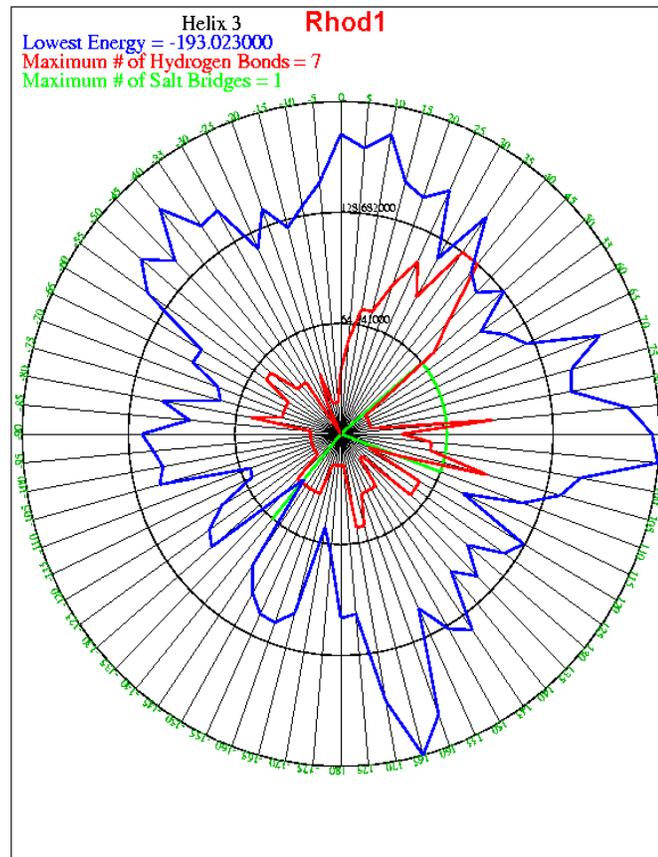


Figure 2 - Helical scan of New MembStruk structure of Bovine Rhodopsin.

3.3.8 Homology Modeling of Loops (*makeloops*)

The next step is to add the three extracellular and three intracellular loops for the best structure(s) chosen from the previous step. The program Modeller v7.0 (32-34) is used to add each of the loops individually. Then each loop will have its side chains optimized using SCRWL. The loop will be movable while the rest is fixed for minimization (0.3 RMS, 5000 steps, CMM level 4), and then the tops and bottoms of each helix (3 residues) connected to the loop will be minimized with the loops (0.3 RMS, 5000 steps, CMM level 4). Finally, the entire protein will be minimized (0.3 RMS, 5000 steps, CMM level 4). The loops can also be added simultaneously using Modeller, followed by the optimization procedure as above.

3.3.9 Closing of the EC2 Loop (*EC_LOOP_SIM*)(Optional)

The crystal structure of bovine rhodopsin has the second extracellular loop (EC2 loop) in a conformation that is covering the TM barrel. We call this a “closed loop” conformation for the EC2. We build the structures with both open conformation and closed conformation for EC2. The closing of EC2 loop is triggered by a disulfide bond formation between a Cys on TM3 and a cysteine on the EC2 loop. These two cysteines are conserved across many GPCRs though not all. Thus, this step of closing the loop is carried out by first making the disulfide bond and optimizing the loop in small steps towards the bond length of the disulfide linkage. The procedure is as follows: The loops added by Modeller are first optimized by conjugate gradient minimization while keeping the TM helices fixed. This step is to allow the option of forming selected disulfide linkages (generally there is at least one from the middle of the EC2 loop and the extracellular end of TM3, which is the case for bovine rhodopsin). The disulfide bond is formed and optimized with equilibrium distances lowered in decrements of 2 Å until the disulfide bond distance is reduced to 2 Å. Then the loop is optimized with the default equilibrium disulfide bond distance of 2.07 Å.

Subsequently, annealing MD is used to optimize the accommodation of EC2 loop inside the barrel. This involves 71 cycles of annealing where the loop atoms are heated from 50 K to 600 K and back to 50 K over a period of 4.6 ps. From each cycle the minimum potential energy structure is selected and minimized. During this process the rest of the atoms are kept fixed for the first 330 ps, and then all the atoms in the top half of the TM barrel are allowed to move for 100 ps. Lastly, a full atom conjugate gradient minimization of the protein is performed in vacuum using MPSim.

3.3.10 *Full Protein Relaxation (looprelax)*

The final looped protein is then minimized (5000 steps) with only the loops movable, and then the entire protein will be minimized (5000 steps). This final step is to prepare the final structure for HierDock by ensuring that the structure is in an energy minimum. This final structure is called the MembStruk structure and is now ready for HierDock.

3.4 *HierDock v2.5BS - Function Prediction for GPCRs*

Since there is only one crystal structure for GPCRs, we use ligand binding sites compared to experimental mutation studies to validate the predicted structures. The HierDock procedure (17, 28) is used to predict ligand binding sites and binding energies with no assumed knowledge of the ligand binding site(s). HierDock is a hierarchical docking strategy consisting of a coarse grain docking method to generate ligand/protein bound structures followed by fine grain all atom optimization of a certain number of these complexes using continuum solvation methods. The HierDock procedure has been validated for many globular and membrane proteins with known crystal structures (see Datta, Floriano, Huskey).

3.4.1 *HierDock ForceFields*

The cis-retinal ligand used in docking was described with the DREIDING FF (18), using charges from quantum mechanics (QM) calculations on the isolated ligand [ESP charges calculated using Jaguar (35)]. Some of the calculations were done in the vacuum (e.g., final optimization of receptor structure to approximate the low dielectric membrane environment). For structural optimization in the solvent (water), we used the Analytical Volume Generalized Born (AVGB) (36) approximation to Poisson-Boltzmann

(PB) continuum solvation (PB). All other calculations used the standard forcefields described in 3.3.1.

3.4.2 *Sphere Generation and Partitioning (GrowBox or Pass)*

For each ligand, the entire receptor is scanned for possible binding regions (except for the intracellular loops and the part in contact with the membrane). The void regions of the protein are mapped with spheres generated over the whole receptor using the Sphgen program in DOCK 4.0 (37). The protein is partitioned into areas to search by two methods: 1) The spheres are then partitioned into as many overlapping (2\AA) cubic docking boxes each with sides of 10\AA as will cover the entire protein (GrowBox, default), or 2) the use of Pass10 to find the centers of scanning regions in the receptor (44, optional). The dimensions of the scanning regions used by Pass10 are limited by the number of spheres allowed in each region (25 to 150 spheres per region). The HierDock calculations are then performed in each of these regions.

3.4.3 *Coarse Docking into all Potential Binding Regions*

Using Dock 4.0 (37), one thousand conformations are generated for each ligand to be run in HierDock. The flexible ligand docking option with torsion minimization, a non-distance dependent dielectric constant of 2.5, and a cutoff of 10\AA for energy evaluation will be used. The conformations will be ranked using the energy scoring function from DOCK 4.0. The best scoring 10% conformations for each ligand in each of the binding regions will be used as input for finer refinement of conformations and energy evaluation.

Next, a subset of the docked ligands are selected using a better FF and optimizing the structure of the ligand with protein fixed. The all atom DREIDING FF and Gasteiger charges for the ligand are used to calculate the energy of the ligand with protein fixed.

This is done for each of the 100 structures from the 10% found above. For each of these 100 minimized structures, the amount of the ligand surface that is buried in the receptor using Connolly's MS program from QCPE is calculated. For each region and each ligand, 10 best structures based on lowest energy and highest buried surface are then selected.

3.4.4 *Ranking of Coarse Docked Structures*

For each ligand, a possible protein binding site is determined by selecting the region with the best binding energy (below 200 kcal/mol) and a buried surface equal or greater than 90%. The final coordinates the ligand in this binding site are then used to define the "binding site" as all spheres (from Sphgen) of the protein structure in a 2.5Å radius from the ligand atoms. Often no ligand docked structure is found with a buried surface $\geq 90\%$ and binding energy ≤ 200 kcal/mol, so the binding regions are also ranked according to buried surface and binding energy.

All regions in the top (extracellular) half of the protein are searched for the top four regions in terms of highest buried surface. Starting at 90% buried surface and then subtracting 5% each time, looking for a buried surface limit that has at least four different regions that contain structures with buried surfaces equal or greater than that percentage. Once this buried surface limit is found, the 50 lowest energies in these regions with buried surfaces that meet the criteria are sorted. The regions are ranked according to the number of structures that are within this set of 50 (10 being the max). The region with the most structures in the top ten of this set is given one rank higher (to a possible 11). Regions can have equal rank. With multiple ligands, the regions are then ranked according to the summation of the ranks obtained with each ligand at that rank level or higher (see Chapter 5).

The top 3 regions (ranked accordingly) will have their sphere sets merged into one set. This merged set will then be reduced to keep only those spheres that are farther than 0.5 Å apart. Next, the merged sphere set is compared to the binding site found in the beginning to see if they match. Otherwise, both sites are used in the Fine Docking and the final structures are compared according to energy at the end of Hierdock.

3.4.5 *Fine Docking in the Selected Binding Site*

The ligands are then docked into the binding site, repeating step 3.4.3 of HierDock for just this site using 3000 conformations in Dock with 300 kept. Then 100 are chosen based on buried surface (> 75%) and energy (< 100 kcal/mol). These 100 are then minimized: 100 steps, with the protein fixed, and the top 5 are kept based on lowest energies.

To determine the best binding conformation of each ligand in the receptor, it is important to allow both the ligand and the receptor to optimize their conformations. Thus the 5 best scoring structures are now minimized allowing all atoms of both protein and ligand to be optimized: 100 steps.

3.4.6 *Scoring of the Fine Docked Structures*

Each structure will be scored using the equation (1) to find the binding energy:

$$BE = PE (\text{odorant in the receptor}) - PE (\text{odorant in water}) \quad (1).$$

Where BE is the binding energy of the odorant in the receptor, PE (odorant in receptor) is the potential energy of the odorant calculated with the OR fixed. PE (odorant in water) is the potential energy of the odorant in water calculated for the starting conformation of the ligand. The potential energy of the odorant in the receptor will be calculated using dreiding force field with the receptor atoms fixed. The PE of the odorant

in water will be calculated using the AVGB continuum solvation method (AVGB) (36):
solv_in_eps 1.11 and solv_out_eps 78.2.

3.4.7 Optimization of the Side Chains (SCREAM)

The best structure from step 3 (lowest BE) will now have its EC2 loop optimized using the method described in 3.3.9 with the ligand bound, if this has not already been done. Then the side chains within 5Å of the ligand will be optimized using SCREAM (V. W. T. Kam, N. Vaidehi, and W. A. Goddard 3rd, unpublished). This final structure will then have its binding energy calculated again as in step 3.4.8.

4.0 Results and Discussion

The function of a GPCR is to detect specific ligands and signal this information to the cell. Literature has indicated that this involves two distinct conformations of GPCRs, one active and one inactive, in equilibrium, even in the absence of ligands (2 38-39). There is a change in equilibrium when a ligand binds to the GPCR. Unfortunately, there is only the inactive form that is observed in the bovine rhodopsin crystal (4-5).

It has been postulated in a previous paper that the conformational change from active to inactive form is the closing of the EC-II loop on top of the protein (14). The GPCR is said to be in the active form when the EC-II loop is open, and the GPCR is said to be in the inactive form when the EC-II loop is closed. In this chapter using the method described in 3.3.7, we propose that not only does the EC-II loop move but also the TM region 3 by 90 degrees.

4.1 Comparison of MembStruk Rhodopsin Structures (Old and New)

The current MembStruk methods were used to develop a 3D structure of bovine rhodopsin (see Appendix A). In building this updated structure, a new method of

analyzing the hydrophobic graphs for TM prediction (see section 3.3.5) led to a better prediction of the TM regions involved in binding. This new prediction gives a significant improvement in TM3 and slight improvements in TMs 2 and 4, for an average error of ~3 residues per TM region (see below, compare to ~5 residue average error for the old structure).

| | |
|-------------------------|---|
| Old TM1: | PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH |
| New TM1: | PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH |
| <u>1HZX</u> (1U19) TM1: | <u>EPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH</u> |
| Old TM2: | PLNYILLNLAVADLFMVFGGFTTTLYTSLHG |
| New TM2: | PLNYILLNLAVADLFMVFGGFTTTLYTSLH |
| <u>1HZX</u> (1U19) TM2: | <u>TPLNYILLNLAVADLFMVFGGFTTTLYTSLHG</u> |
| Old TM3: | PTGCNLEGFFATLGGEIALWLSLVVLAIE |
| New TM3: | TGCNLEGFFATLGGEIALWLSLVVLAIERVYVVVCK |
| <u>1HZX</u> (1U19) TM3: | <u>FGPTGCNLEGFFATLGGEIALWLSLVVLAIERVYVVVCK</u> |
| Old TM4: | HAIMGVAFTWVMALACAAPPLVG |
| New TM4: | NHAIMGVAFTWVMALACAAPPLVG |
| <u>1HZX</u> (1U19) TM4: | <u>GENHAIMGVAFTWVMALACAAPPLV</u> |
| Old TM5: | ESFVIYMFVVHFIIPLIVIFFCYGQLVF |
| New TM5: | ESFVIYMFVVHFIIPLIVIFFCYGQLVF |
| <u>1HZX</u> (1U19) TM5: | <u>NNESFVIYMFVVHFIIPLIVIFFCYGQL</u> |
| Old TM6: | RMVIIMVIAFLICWLPYAGVAFYIFTH |
| New TM6: | RMVIIMVIAFLICWLPYAGVAFYIFTH |
| <u>1HZX</u> (1U19) TM6: | <u>EKEVTRMVIIMVIAFLICWLPYAGVAFYIFT</u> |
| Old TM7: | PIFMTIPAFFAKTSAVYNPVI |
| New TM7: | FMTIPAFFAKTSAVYNPVIYIMMNK |
| <u>1HZX</u> (1U19) TM7: | <u>GPIFMTIPAFFAKTSAVYNPVIYIMMN</u> |

This method is good at resolving the ends of TM regions between helices, giving a better prediction for TM 3. TM 7 looked to be the worst using this method compared to the older crystal structure (1HZX) TM regions, but the latest structure classifies “GPIFMTIPAFFAKTSAVYNPVIIMMN” as helical showing a better fit to the newer method. The difficulty with defining TM 7 is that the C-terminus is helical at the end of

TM7. However, the TM prediction for TM 7 compared to the new structure is good demonstrating the value of the summation analysis.

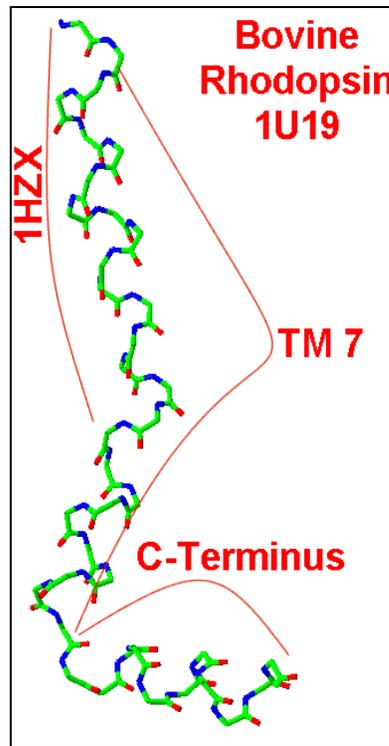


Figure 3 - TM 7 of the 1U19 crystal structure. Shown here are the helical regions that are a part of the TM region and part of the C-terminus. Differences are shown between 1HZX and 1U19 in defining the TM region.

The comparing the new structure to the latest crystal structure (1U19), the CRMS of the crystal TM regions to the new structure is 2.72 Å. This is better than the old structure that compares to the crystal TM regions with a 3.01 Å CRMS. The main chain RMS of the new structure to the crystal structure is 2.66 Å (2.87 Å for the old structure).

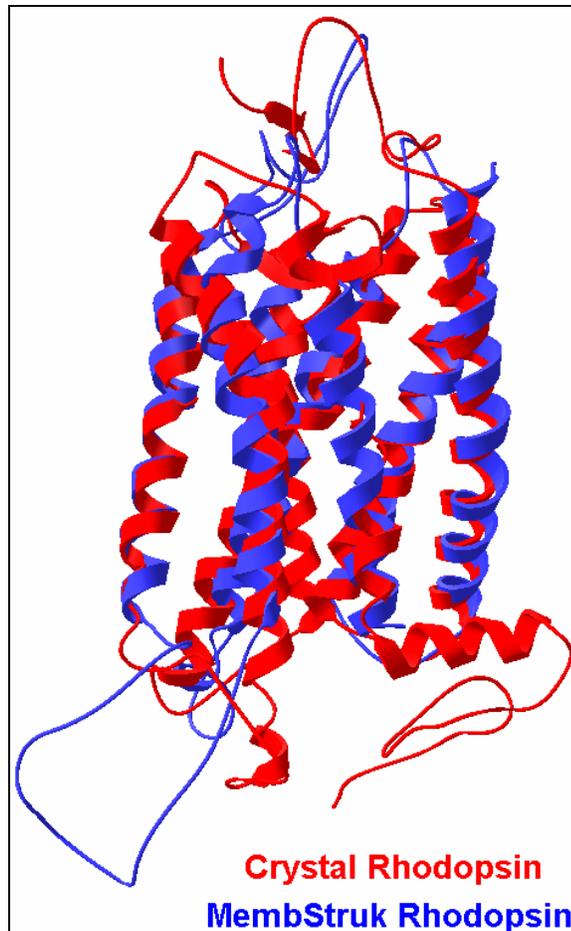


Figure 4 - Comparison of crystal Bovine Rhodopsin (1U19) (red) to the open MembStruk version 4.30 rhodopsin structure (blue).

4.2 Conformational Change from Open to Closed in the Rhodopsin Structure

Initial development of the rhodopsin MembStruk structure produced a structure with helix 3 rotated away from the crystal structure by 90 degrees (see Appendix A). The correct rotation was found to be a local minimum on the helical scanning graph (Section 3.3.7, Figure 2). The potential conformations for helix 3 were built and the addition of the EC-II loop was only possible with the correct conformation. Figure 4 shows that the location of the open conformation of cystine 110 cannot form a disulfide bridge in its position forcing an open loop conformation.

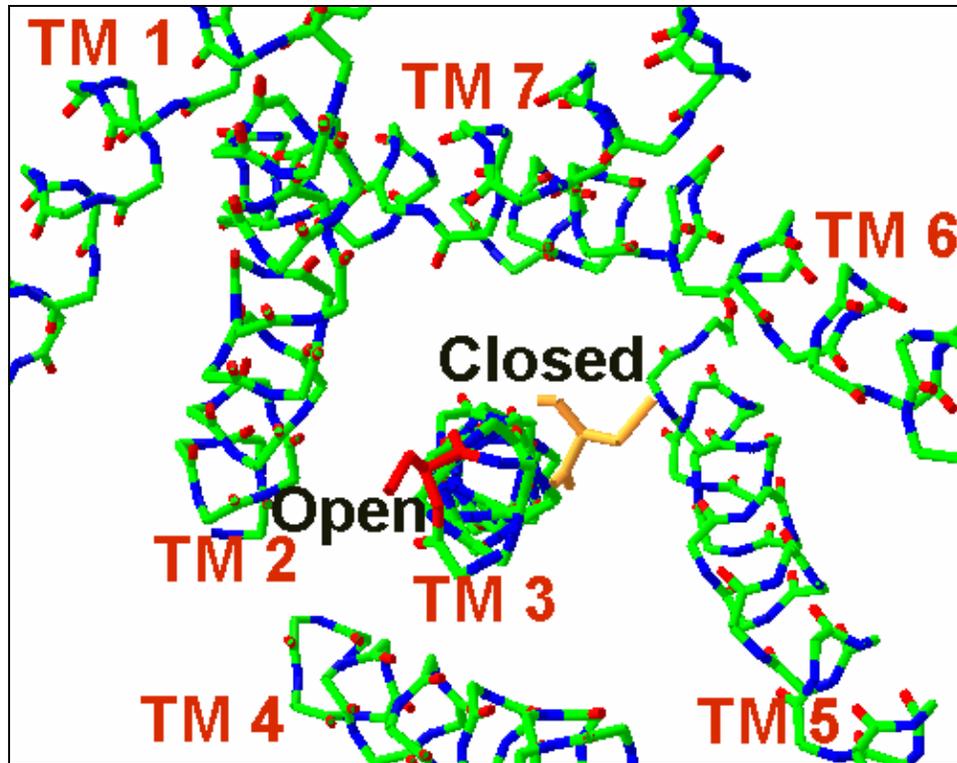


Figure 5 - The open and closed positions of Cys 110, show how the EC-II loop will fail to close over the protein in the open position.

In the open conformation, helix 3 also has Glu 122 forming a salt bridge with Lys 296 on TM 7. This salt bridge stabilizes the rotation of helix 3 making this an energetically favorable conformation. This indicates that as retinal enters the binding site it breaks this weak salt bridge between Glu 122 and Lys 296 to form a Schiff's bond to Lys 296 (14). Then helix 3 would be favored to rotate 90 degrees and form the disulfide bridge closing the EC-II loop into the closed conformation.

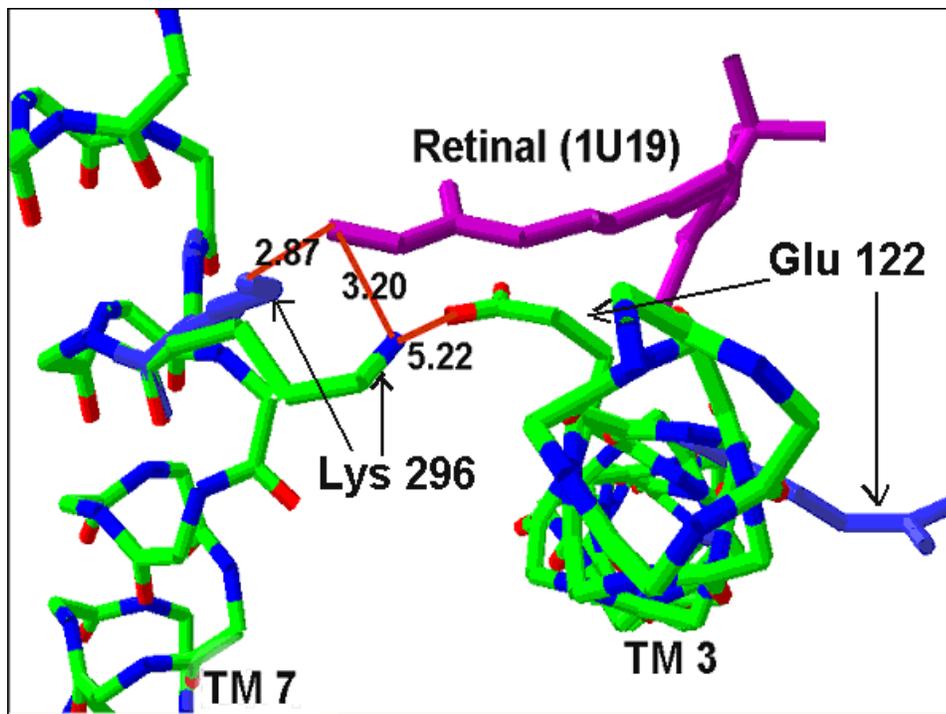


Figure 6 - Open and Closed (blue) positions of Glu 122 and Lys 296 with the aligned position of Retinal from 1U19. In the open conformation Glu 122 and Lys 296 are within 5.22 angstroms forming a weak salt bridge.

4.3 Scanning of the MembStruk Rhodopsin Structure

Two final structures were created for use in HierDock. The first structure was a structure in closed conformation with an open EC-II loop (open) and the second structure was a closed conformation with a closed EC-II loop (closed). The closing of the EC-II loop was done by modeling the EC-II loop from the closed “old” structure (14). The open conformation was not tested since Glu 122 effectively blocks the correct binding site. The two structures that we will refer to as open and closed were run through steps 3.4.2 to 3.4.4 of HierDock to locate the best binding location.

4.3.1 Scanning of the Open Structure

The open structure was run using both GrowBox and Pass (Section 3.4.4). Pass was able to find the correct binding site of cis-retinal (as compared to 1U19) and ranked

it as number 2. The top ranked region was found to be in the extracellular loops and so eliminated as a possible binding site. The second region (region 6) was located near the tail (the end of the long carbon chain of cis-retinal, Figure 6). The third region had its center located in the extracellular loops as well. The center of the fourth ranked region (region 4) was found to be next to the head (phenyl ring of cis-retinal).

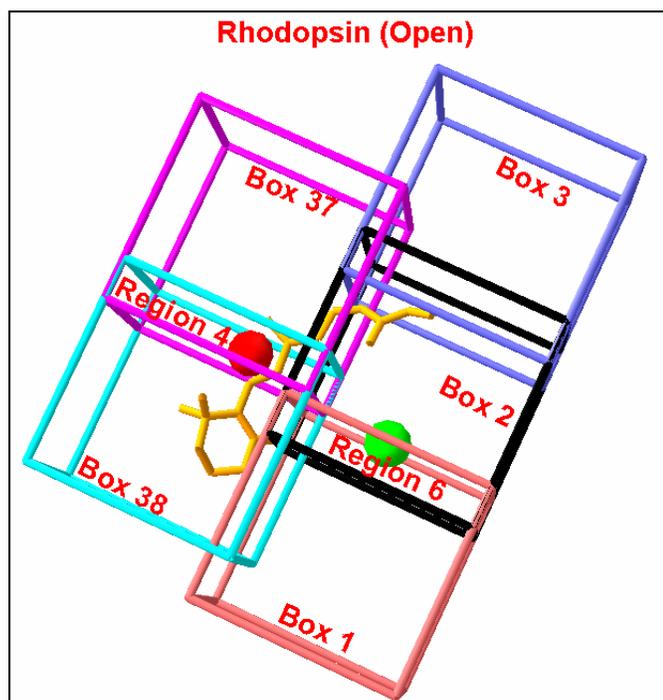


Figure 7 - Scanned binding site on the Rhodopsin Open Structure. The boxes are from GrowBox and the Regions are from Pass.

For Open-Pass10, the best 50 conformations are:

| | | | | | |
|----|----------|------------------|----|----------|------------------|
| 1 | Region 2 | Energy: -307.355 | 12 | Region 7 | Energy: -161.312 |
| 2 | Region 4 | Energy: -294.473 | 13 | Region 6 | Energy: -153.743 |
| 3 | Region 2 | Energy: -286.117 | 14 | Region 6 | Energy: -39.498 |
| 4 | Region 7 | Energy: -277.267 | 15 | Region 6 | Energy: 6.899 |
| 5 | Region 2 | Energy: -275.103 | 16 | Region 6 | Energy: 96.738 |
| 6 | Region 7 | Energy: -262.057 | 17 | Region 6 | Energy: 97.176 |
| 7 | Region 7 | Energy: -261.977 | 18 | Region 6 | Energy: 103.610 |
| 8 | Region 4 | Energy: -256.234 | 19 | Region 6 | Energy: 155.536 |
| 9 | Region 4 | Energy: -242.668 | 20 | Region 6 | Energy: 179.821 |
| 10 | Region 7 | Energy: -214.749 | 21 | Region 6 | Energy: 185.649 |
| 11 | Region 7 | Energy: -192.284 | 22 | Region 6 | Energy: 193.598 |

ret-QM 90% - 7 (6*) 6 (10) 2 (3) 4 (3)

The above shows the actual energies (kcal/mol) and ranking given to the top four regions found by Pass. Region 7 was ranked as first since it had the most conformations in the top 10 according to energy, but region 6 had all 10 conformations in the list and was not in the list and became the chosen binding site. The lowest energy conformation in region 6 was almost identical to the best conformation found from GrowBox.

GrowBox ranked the correct boxes as third, fourth, and fifth. The top two ranked boxes from GrowBox are near the correct location, but either translated too high or too low. These two regions had better energies as every one of the conformations was in a vertical position (parallel to the membrane). With a closed loop these conformations would not have been possible, and eliminating all vertical conformations the lowest energy conformation was found in Box 2 (conformation 1). This conformation is nearly identical to the conformation found in Pass.

For Open-GrowBox, the best 50 conformations are:

| | | | | | |
|----|--------|------------------|----|-------|-----------------|
| 1 | Box 3 | Energy: -296.295 | 16 | Box 2 | Energy: 25.448 |
| 2 | Box 3 | Energy: -294.300 | 17 | Box 1 | Energy: 31.357 |
| 3 | Box 3 | Energy: -289.092 | 18 | Box 1 | Energy: 45.393 |
| 4 | Box 37 | Energy: -288.099 | 19 | Box 1 | Energy: 52.212 |
| 5 | Box 3 | Energy: -278.476 | 20 | Box 2 | Energy: 75.867 |
| 6 | Box 3 | Energy: -275.580 | 21 | Box 2 | Energy: 84.020 |
| 7 | Box 35 | Energy: -249.194 | 22 | Box 2 | Energy: 90.565 |
| 8 | Box 38 | Energy: -188.038 | 23 | Box 2 | Energy: 109.125 |
| 9 | Box 1 | Energy: -88.772 | 24 | Box 1 | Energy: 109.722 |
| 10 | Box 2 | Energy: -48.316 | 25 | Box 1 | Energy: 127.173 |
| 11 | Box 1 | Energy: -22.120 | 26 | Box 1 | Energy: 146.356 |
| 12 | Box 38 | Energy: -8.838 | 27 | Box 2 | Energy: 148.490 |
| 13 | Box 2 | Energy: -3.075 | 28 | Box 1 | Energy: 202.079 |
| 14 | Box 2 | Energy: 6.068 | 29 | Box 2 | Energy: 245.387 |
| 15 | Box 1 | Energy: 10.430 | | | |

ret-QM 90% - 3 (5*) 1 (10) 2 (10) 38 (2) 37 (1) 35 (1)

The lowest energy structure from Box 2 (conformation 1) is the best structure and the conformation used to determine the nearest neighboring residues found in the binding site. This conformation correlates to the conformations found in Pass and from the scanning of the closed structure.

4.3.2 Scanning of the Closed Structure

For Closed-GrowBox, the best 50 conformations are:

| | | | | | | | |
|----|--------|---------|----------|----|--------|---------|----------|
| 1 | Box 19 | Energy: | -287.315 | 22 | Box 1 | Energy: | 102.011 |
| 2 | Box 19 | Energy: | -270.276 | 23 | Box 1 | Energy: | 102.827 |
| 3 | Box 19 | Energy: | -257.193 | 24 | Box 1 | Energy: | 108.813 |
| 4 | Box 19 | Energy: | -233.203 | 25 | Box 3 | Energy: | 114.573 |
| 5 | Box 19 | Energy: | -229.054 | 26 | Box 1 | Energy: | 117.904 |
| 6 | Box 19 | Energy: | -201.805 | 27 | Box 18 | Energy: | 122.511 |
| 7 | Box 18 | Energy: | -188.474 | 28 | Box 1 | Energy: | 124.959 |
| 8 | Box 19 | Energy: | -116.585 | 29 | Box 18 | Energy: | 129.820 |
| 9 | Box 19 | Energy: | -65.154 | 30 | Box 1 | Energy: | 143.853 |
| 10 | Box 2 | Energy: | -62.698 | 31 | Box 1 | Energy: | 155.683 |
| 11 | Box 2 | Energy: | -22.664 | 32 | Box 1 | Energy: | 158.384 |
| 12 | Box 1 | Energy: | -20.815 | 33 | Box 18 | Energy: | 168.997 |
| 13 | Box 2 | Energy: | -6.177 | 34 | Box 2 | Energy: | 176.909 |
| 14 | Box 1 | Energy: | 19.957 | 35 | Box 2 | Energy: | 228.386 |
| 15 | Box 18 | Energy: | 22.669 | 36 | Box 2 | Energy: | 246.194 |
| 16 | Box 18 | Energy: | 24.810 | 37 | Box 2 | Energy: | 268.651 |
| 17 | Box 18 | Energy: | 25.082 | 38 | Box 2 | Energy: | 320.461 |
| 18 | Box 18 | Energy: | 27.025 | 39 | Box 3 | Energy: | 506.077 |
| 19 | Box 18 | Energy: | 27.550 | 40 | Box 3 | Energy: | 547.073 |
| 20 | Box 2 | Energy: | 93.090 | 41 | Box 3 | Energy: | 1976.780 |
| 21 | Box 2 | Energy: | 96.077 | | | | |

ret_QM 90% - 19 (8*) 2 (10) 1 (10) 18 (9) 3 (4)

The Closed structure was scanned for potential binding sites using GrowBox. The closing of the EC-II loop did in fact bring the correct box to become ranked as first. The actual retinal crystal position was found to reside in the boxes ranked first, second, and fourth (boxes 2, 18-19). Conformation 3 of box 2 is in the exact same location and position as those found from scanning of the open structure. However, conformation 9 from box 19 is in an identical conformation as the crystal structure with exception of a translation of 6.83 Å at the heads of the retinal structures (see Figure 7) towards the extracellular region. This conformation however is close enough (3.4 Å) to form the Schiff's base and then annealing would help bring this conformation into the correct location. This indicated that even with the closed loop, the EC-II loop is still not closed enough to push this ligand down.

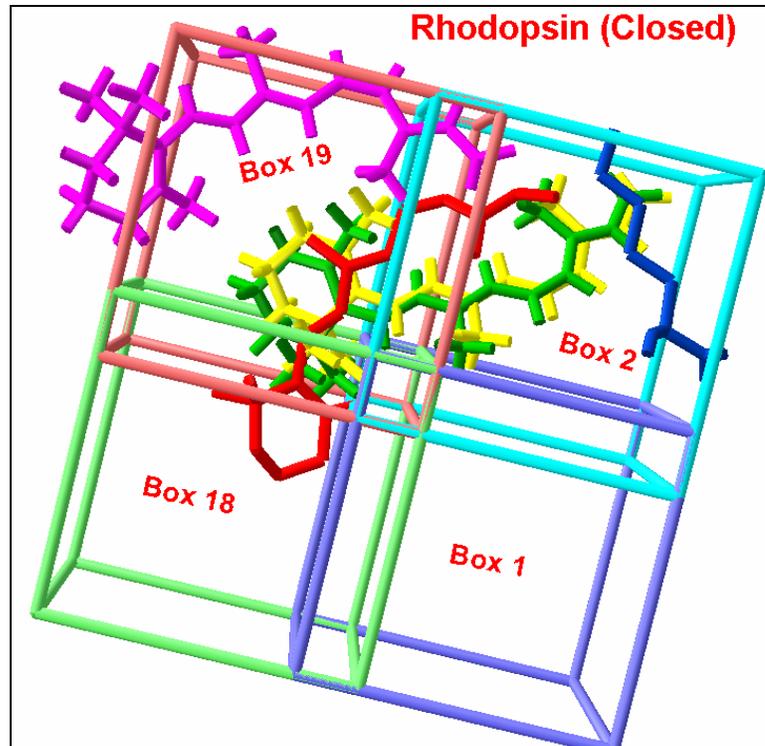


Figure 8 - Scanning of the Rhodopsin (Closed) structure. The purple retinal was the best ranked structure found in box 19, while the yellow and green retinal structures are from box 2 Open-GrowBox and Closed-GrowBox. The red retinal structure is the crystal structure.

4.4 Analysis of the Binding Site

The conformations from box 2 (the common structure found in all scanning results) was used to determine the residues involved in the binding site for cis-retinal. All residues within 3.5 angstroms (except for those found in the EC-II loop) of this conformation were selected and aligned to their crystal structure counterparts. Their alignment to the crystal structure (1U19) was 1.37 Å RMS for the main chain atoms and 2.39 Å RMS for all atoms (see figure 2 from the Introduction). The alignment to those residues in 5.0 Å is 2.19 RMS for the main chain atoms and 2.97 RMS for all atoms.

Comparison was done of the MembStruk Rhodopsin (Closed) structure's scanned binding site and the crystal structures binding site. The closed structure had 24 residues located in its binding site of which 12 (of a possible 25) matched those found in the

crystal structure's binding site. This is a good fit for a scanned binding site giving an excellent starting location for fine grain HierDock.

5.0 Conclusions

The rotation of helix 3 is important to the binding of retinal in helping to control the position of the EC-II loop with its cystine residue involved in the disulfide bridge. It also controls the size of the binding pocket with the movement of glu 122 forming a saltbridge with lys 296. The conformation showed up as a local energy minimum when tested with the new rotational analysis. The new methods for rotational analysis now allow for the sampling of possible local conformations that might be active or in-active forms.

The new structures produced showed a marked improvement over previous models done of bovine rhodopsin. This validation of the rhodopsin structure showed the importance of the EC-II loop for binding, and the accuracy for determination of the correct binding site from a MembStruk structure. This shows that functional prediction from MembStruk structures is possible and valuable information can be obtained for drug design.

6.0 References

- 1) Wallin, E. and von Heijne, G. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* 7, 1029-1038.
- 2) Schoneberg, T., Schulz, A. and Gudermann, T. (2002). The structural basis of G-protein-coupled receptor function and dysfunction in human diseases. *Rev. Phys. Biochem. Pharm.* 144, 145-227.
- 3) Pierce, K. L., Premont, R. T. and Lefkowitz, R. J. (2002). *Nat. Rev. Mol. Cell Biol.* 3, 639–650.
- 4) 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). Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* 289, 739-745.
- 5) Teller, D., Okada, T., Behnke, C., Palczewski, K. and Stenkamp, R. (2001). Advances in Determination of a High-Resolution Three-Dimensional Structure of Rhodopsin, A Model of G-Protein-Coupled Receptors (GPCRs). *Biochemistry* 40, 7761-7772.
- 6) Archer, E., Maigret, B., Escrieut, C., Pradayrol, L., and Fourmy, D. (2003). Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends Pharmacol. Sci.* 24, 36-40.
- 7) John B., and Sali A., (2003). Comparative protein structure modeling by iterative alignment, model building and model assessment, *Nucleic Acids Research*, 31, 3982-3992.
- 8) Strader, C.D., Fong, T. M., Tota, M.R., Underwood, D. and Dixon, R.A.F., (1994). Structure and Function of G-Protein Coupled Receptors. *Annu. Rev. Biochem.* 63, 101-132.
- 9) Parrill, A. L., Baker, D. L., Wang, D., Fischer, D. J., Bautista, D. L., van Brocklyn, J., Spiegel, S., and Tigyi, G. (2000). in *Lysophospholipids and Eicosanoids in Biology and Pathophysiology* (Goetzl, E. J., and Lynch, K. R., eds) Vol. 905, 330–339, New York Academy of Sciences, New York.
- 10) Rost B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering* 12, 85-94.
- 11) Chung S.Y., Subbiah S., (1996). A structural explanation for the twilight zone of protein sequence homology. *Structure* 4, 1123-1127.
- 12) Brenner S.E., Chothia C., Hubbard T.J.P., (1998). Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships, *Proc. Natl. Acad. Sci. USA*, 95, 6073-6078.
- 13) Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and Goddard III, William A. (2002). Prediction of structure and function of G protein-coupled receptors. *Proc. Nat. Acad. Sci.* Vol. 99(20),

12622-12627.

- 14) Trabanino R.J., Hall, S.E., Vaidehi N., Floriano W.B., and Goddard III, W.A. (2004). First Principles Predictions of the Structure and Function of G-Protein Coupled Receptors: Validation for Bovine Rhodopsin. *BioPhys. J.* 86, 1904-1921.
- 15) Freddolino, P. L., Kalani, M. S. Y., Vaidehi, N., Floriano, W. B., Hall, S. E., Trabanino, R. J., Kam, W. T. and Goddard III, W. A. (2004). Predicted 3D structure for the human beta2 adrenergic receptor and its binding site for agonists and antagonists. *Proc. Natl. Acad. Sci.* 101, 2736-2741.
- 16) Kalani, M. Y. S., Vaidehi, N., Hall, S. E., Trabanino, R. J., Freddolino, P. L., Kalani, M. A., Floriano, W. B., Kam, V. W. and Goddard III, W. A. (2004). The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists. *Proc. Natl. Acad. Sci.* 101, 3815-3820
- 17) Floriano W.B., Vaidehi, N., Singer M.S., Shepherd G.M., Goddard III, W.A. (2000). Molecular Mechanisms underlying differential odor responses of a mouse olfactory receptor. *P. Natl. Acad. Sci. USA* 97, 10712-10716.
- 18) Mayo, S. L., Olafson, B. D., and Goddard III, W. A. (1990). DREIDING—a generic force field for molecular simulations. *J. Phys. Chem.* 94, 8897–8909.
- 19) 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, I., 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). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102, 3586–3616.
- 20) Ding, H. Q., Karasawa, N., and Goddard W. A. (1992). Atomic level simulations on a million particles—the cell multipole method for Coulomb and London nonbond interactions. *Chem. Phys. Lett.* 97, 4309–4315.
- 21) Lim, K.-T., Brunett, S., Iotov, M., McClurg, R. B., Vaidehi, N., Dasgupta, S., Taylor, S., and Goddard III, W. A. (1997). Molecular dynamics for very large systems on massively parallel computers: the MPSim program. *J. Comput. Chem.* 18, 501–521.
- 22) Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. USA.* 8, 140–144.
- 23) Donnelly, D. (1993). Modeling alpha-helical transmembrane domains. *Biochem. Soc.* 21, 36–39.
- 24) Altschul, S. F., Gish, W., Miller, W., Myers, W. E., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

- 25) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- 26) Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). Clustal-W improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- 27) Schertler, G. F. X. (1998). Structure of rhodopsin. *Eye.* 12:504–510.
- 28) Floriano, W. B., Vaidehi, N., and Goddard III, W. A. (2004). Making sense of olfaction through molecular structure and function prediction of olfactory receptors. *Chem. Senses* 29, 269-290.
- 29) Hall, Spencer E., Floriano, Wely B., Vaidehi, Nagarajan, and Goddard III, William A. (2004). Predicted 3-D Structures for Mouse I7 and Rat I7 Olfactory Receptors and Comparison of Predicted Odor Recognition Profiles with Experiment *Chemical Senses* 29, 595–616.
- 30) Bower, M., Cohen, F. E., and Dunbrack, R. L., Jr. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* 267, 1268-1282.
- 31) Rappé, A.K. and Goddard III, W.A. (1991). Charge equilibration for molecular-dynamics simulations. *J. Phys. Chem.* 95, 3358-3363.
- 32) Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spacial restraints. *J. Mol. Biol.* 234, 779-815.
- 33) Fiser, A., Do, R. K., and Sali, A. (2000). Modeling of loops in protein structures. *Protein Science* 9, 1753-1773.
- 34) Marti-Renom, M. A., Stuart, A., Fiser, A., Sanchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biomol. Struct.* 29, 291-325
- 35) Jaguar v4.0, Schrodinger Inc. Portland, Oregon.
- 36) Zamanakos, G. (2001) Caltech Chemistry Thesis
- 37) Ewing, T.A. and Kuntz, I.D. (1997). *J Comput. Chem.* 18, 1175-1189.
- 38) Melia, T.J., Cowan, C.W., Angleson, J.K., Wensel, T.G. (1997) A comparison of the efficiency of G protein activation by ligand-free and light-activated forms of rhodopsin. *Biophys. J.* 73, 3182-3191.
- 39) Strange, P.G. (1998) Three-state and two-state models. *Trends Pharm. Sci.* 19, 85-86.
- 40) Jayasinghe, S., Hristova, K. and White, S. H. (2001). MPtopo: A database of membrane protein topology. *Protein Science* 10, 455-458.
- 41) Vaidehi, N., Jain, A., and Goddard III, W. A. (1996) Constant Temperature Constrained Molecular Dynamics: The Newton-Euler Inverse Mass Operator Method. *J. Phys. Chem.* 100, 10508.

- 42) Mathiowetz, A. M., Jain, A., Karasawa, N., and Goddard III, W. A. (1994). Protein Simulations using Techniques Suitable for Very Large Systems: the Cell Multipole Method for Nonbond Interactions and the Newton-Euler Inverse Mass Operator Method for Internal Coordinate Dynamics. *Proteins* 20, 227 CN 8921.
- 43) Vriend, G. (1990). WHAT IF- a molecular modeling and drug design program. *J. Mol. Graph.* 8, 52-56.
- 44) Brady Jr., G. P., and Stouten, P. F. W., (2000) Fast Prediction and Visualization of Protein Binding Pockets with PASS. *Journal of Computer-Aided Molecular Design.* 14, 383-401.