

Thesis: Introduction

The focus of my thesis is the development and implementation of structure prediction methods (MembStruk) for a class of integral membrane proteins called the G-protein-coupled receptors (GPCRs). MembStruk involves several steps that are detailed and tested in the first couple chapters of this thesis. The validation of the most current form of this method is presented for bovine rhodopsin, which is the only GPCRs with a crystal structure available. The application of this method to an olfactory receptor (I7) and other peptide chemokine GPCRs (CCR5 and CXCR4) is detailed in the last chapters.

The cell membrane serves as a barrier between cells and the extracellular environment. The membrane is embedded with proteins and these membrane proteins vary in their density on the lipid bilayer, with some membranes containing as little as 18% of the membrane content (Myelin) all the way up to 75% (internal membrane of mitochondria). There are two kinds of proteins found on the lipid membrane: the first kind of membrane proteins are classified as peripheral having primarily only electrostatic and hydrogen-bond interactions with the membrane, the second kind are classified as integral membrane proteins and interact greatly with the hydrocarbon chains on the lipids. In fact, almost all integral membrane proteins span the lipid bilayer.

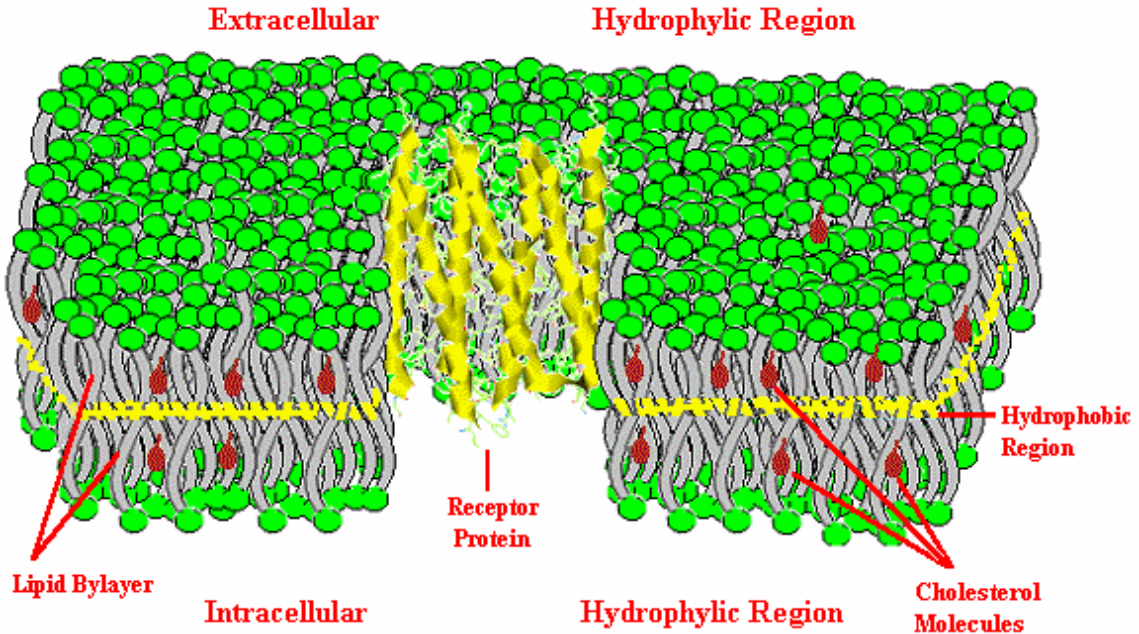


Figure 1 – Schematic diagram of a seven-helical GPCR in a bilipid layer.

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. The transmembrane (TM) receptors fall into four main types: GPCRs, ion-channel receptors that control ion flow, tyrosine kinase-linked receptors that activate tyrosine kinases, and tyrosine kinases that when activated lead to autophosphorylation. The protein superfamily of interest to most drug companies is the GPCRs.

GPCRs are 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, it 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 one of the most important targets 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, bovine rhodopsin (4-5). The sequence identity to rhodopsin is low for most GPCRs of interest (17 % for dopamine, 14 % for serotonin), making the use of homology modeling for obtaining reliable structures not a valid option (6).

With such sparse structural information about GPCRs, many pharmaceutical companies rely on chemical assays of large compound libraries to find a possible drug lead. However, using high throughput chemical screening (HTS) to simultaneously test thousands of compounds is expensive and often fails due to the poor quality of experiments being done on a large compound library, and frequently the size of the compound library is reduced drastically for these very reasons (7). 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 (15). 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”) (17, 19-20). Thus we believe that it is important to explore the viability of MembStruk for predicting structure of GPCRs.

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 MembStruk method (12-13), which uses no information from the high resolution crystal structure of rhodopsin or bacteriorhodopsin. The only information used comes from general principles, like the fact that hydrophobic residues prefer being next to the lipid membrane.

In my thesis I have detailed the various steps involved in the development of the MembStruk method that has been developed by myself, Trabanino, and Vaidehi to predict GPCR structures using no information from the crystal structure of bovine rhodopsin. We further predict function of GPCRs by using the first principles method HierDock for docking ligands to GPCRs. The predicted ligand binding sites and affinities are compared with the abundant experimental data available on ligand binding and mutagenesis for GPCRs. We have thus validated Membstruk method for various GPCRs by comparing the ligand binding sites predicted with the experimental mutagenesis data (9-13). This thesis is divided into three specific sections. The first section comprising chapters 2 and 3 is devoted to detailing the validation and development of the current methods used in MembStruk (version 4.30 and MembComp version 1.80). The second

section (Chapter 3) is giving a breakdown of the programs and GUI interface developed for the easy implementation of MembStruk. The last section (Chapters 4 and 5) contains applications of this method for the prediction of structure and function of specific GPCRs.

Chapter 1 contains a detailed description of the latest version of MembStruk version 4.30 and its validation against the bovine rhodopsin crystal structure. Several new methods were tested on the development of the MembStruk rhodopsin structure, and their implementation improves the accuracy of the binding site for cis-retinal. The comparison of the MembStruk binding site of cis-retinal and the crystal structure (1U19) was 1.37 Å RMS for the main chain atoms and 2.39 Å RMS for all atoms. This validation led to the release of version 4.30 that contains these new changes.

The new methods developed in Chapter 1 help to explain the conformational change that occurs with the binding of cis-retinal to rhodopsin. Studies of the helical energetics show the crucial importance of TMs 3 and 6 and the possible rotations of each between and active and inactive state. A new analysis of the data produced by TM2NDS (12), the step to predict the transmembrane regions of the GPCR sequence, that allows for a better prediction of the TM regions for rhodopsin is also detailed.

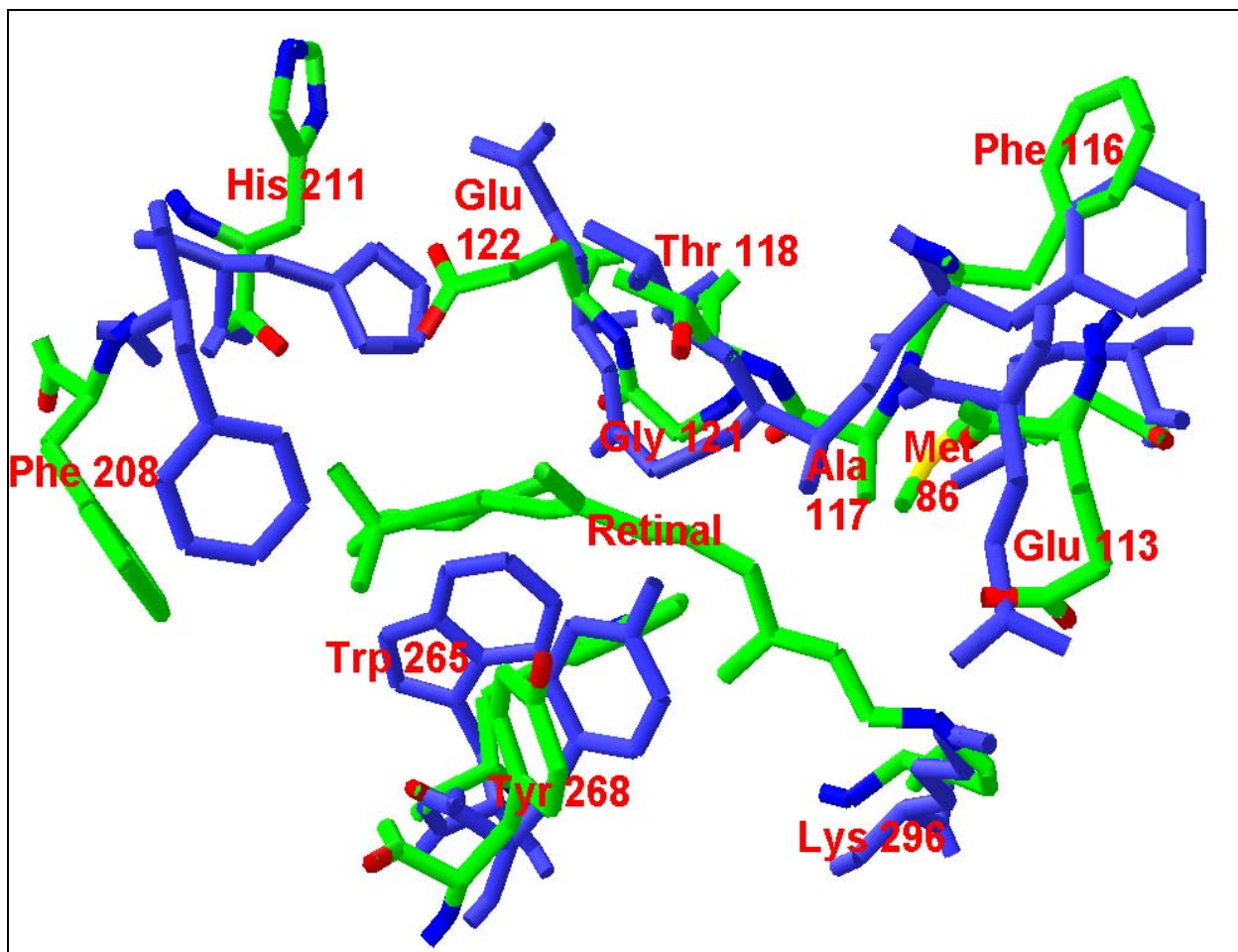


Figure 2 - 3.5 Angstrom view of MembStruk bovine rhodopsin GrovBox site aligned with the 1U19 crystal structure (1.37 main chain RMS).

Chapter 2 contains a detailed look at the step of the MembStruk method developed for optimizing the helical rotational orientation of the helices. The rotational orientation optimization of the transmembrane (TM) helices are embodied in the code called “MembComp”, which can also be used to compare rotational orientation of two TM proteins. MembComp was designed define key properties of membrane bound proteins to help compare protein structures in more detail than what is offered with RMS in coordinates. MembComp was used to show the similarity of the bacteriorhodopsin crystal structures and compare bacteriorhodopsin to bovine rhodopsin. Chapter 2 also

compares a homology model of D2 (from bacteriorhodopsin) to a bacteriorhodopsin crystal structure to demonstrate how little is changed in homology models.

Chapter 3 details all the scripts and programs that were written for the implementation of the various steps in the MembStruk method. MembStruk is comprised of 105 individual programs consisting of 50711 lines of code, which 65% of all code was written by myself and 31% having been written by Rene Trabanino (see Appendix MembStruk Methods). Each step in the methods was written as an independent set of code that is connected by a user GUI. This chapter will explain the GUI for MembStruk 4.10 and the current defaults used for starting a 3D protein structure.

Chapter 4 shows the application of the earlier version of the MembStruk versions 1.50 and 2.00 to build structures of I7 mouse and rat olfactory receptors. Comparison of these structures showed no large differences in their binding sites, yet previous literature had different ligand specificity for these two proteins for heptanal and octanal. The MembStruk structures built in a blind study showed no such differences in ligand affinity, and current experiments corroborated with this result.

Chapter 4 also details a method of analyzing moments of inertia for bound ligand structures to filter out false positives. The moments of inertia filter were found to be a natural indicator of the relative shape and size of the binding pocket. This correlates well with the observations that molecular length is critical for rat I7 olfactory receptor (14).

The last chapter describes in detail the application of MembStruk version 4.1 to the prediction of the structure and function of two chemokine receptors human CCR5 and CXCR4 using the MembStruk protocol. These two different proteins both bind to a similar ligand and offer a good study of similar binding sites among different proteins.

This demonstrates the utility of the MembStruk and HierDock methods on drug development in identifying leads that might not have the desired target specificity.

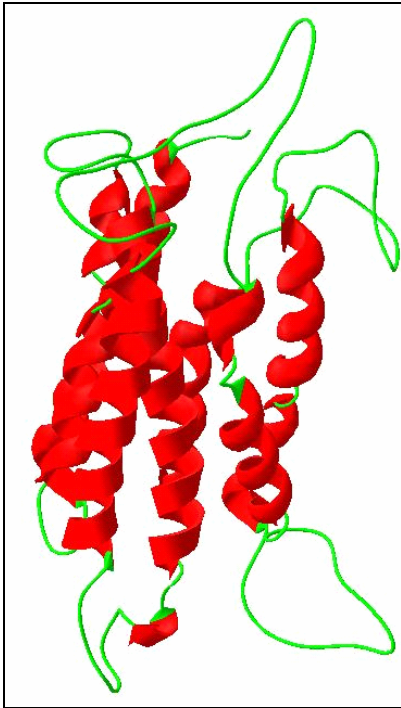


Figure 3 - MembStruk 4.10 structure of CCR5.

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. & 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) National Institute of Health. (2004) PA-04-068: Development of Assays for High Throughput Drug Screening. <http://grants.nih.gov/grants/guide/pa-files/PA-04-068.html>.
- 8) Schertler, G. F. X. (1998). Structure of rhodopsin. *Eye* 12, 504-510.
- 9) 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.
- 10) Hall, S E., Floriano, Wely B., Vaidehi, Nagarajan and William A. Goddard III (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.
- 11) 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.
- 12) 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

- 13) Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and William A. Goddard II, (2002). Prediction of structure and function of G protein-coupled receptors. *Proc. Nat. Acad. Sci.* Vol. 99 no. 20, 12622-12627.
- 14) Araneda, R. C., Kini, A. D., and Firestein, S. (2000). The molecular receptive range of an odorant receptor. *Nat. Neuroscience* 3(12):1248-1255.
- 15) John B., and Sali A., (2003), Comparative protein structure modeling by iterative alignment, model building and model assessment, *Nucleic Acids Research*, 31, 3982-3992.
- 16) 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, pp. 330–339, New York Academy of Sciences, New York.
- 17) Rost B., (1999). Twilight zone of protein sequence alignments *Protein Engineering*. 12, 85-94.
- 18) 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.
- 19) Chung S.Y., Subbiah S., (1996), A structural explanation for the twilight zone of protein sequence homology, *Structure* 4, 1123-1127.
- 20) 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.