B

# Structural and Computational Approaches in Drug Design for G Protein–Coupled Receptors

**Babak Sokouti**
*Biotechnology Research Center, Tabriz University of Medical Sciences, Iran*

**W Bret Church**
*Faculty of Pharmacy, University of Sydney, Australia*

**Michael B Morris**
*Bosch Institute and Physiology, University of Sydney, Australia & Sydney Centre for Development and Regenerative Medicine, Kolling Institute of Medical Research, Australia & Royal North Shore Hospital, Australia*

**Siavoush Dastmalchi**
*Biotechnology Research Center, Tabriz University of Medical Sciences, Iran & School of Pharmacy, Tabriz University of Medical Sciences, Iran*

## INTRODUCTION

G protein-coupled receptors (GPCRs), encoded by about 5% of human genes represent the largest family of integral membrane proteins (IMPs), which in turn constitute about 30% of all human proteins (Herrick-Davis, Grinde, Lindsley, Cowan, & Mazurkiewicz, 2012). The GPCRs are the prime target for pharmaceutical intervention in many diseases. It is usually accepted that about 30 to 60% of marketed drugs target GPCRs (Dastmalchi, Church, & Morris, 2008) and worldwide annual sales exceed 50 billion Dollars every year (S. H. Xiao et al., 2008). This article presents current approaches in drug discovery by highlighting many computational efforts being spent to rationally develop ligands affecting GPCRs, such as application of bioinformatics in identification and classification, as well as the topology prediction of GPCR structures. Application of chemoinformatics and molecular modeling tools to the rational design of novel drugs targeting GPCRs was also discussed by providing examples of virtual ligand screening, and docking.

## BACKGROUND

The computational studies are dealt with different aspects of molecular properties of GPCRs, such as sequence-based identification and classification, sequence analyses, topology prediction and three-dimensional model generation.

Many distinct methodologies based on full-length sequence and motif-based search approaches, machine learning, and several alignment-free techniques have all been used successfully to identify and then classify GPCRs. GPCRTOP (Sokouti, Rezvan, Yuchdav, & Dastmalchi, 2014), GPCRHMM (Wistrand, Kall, & Sonnhammer, 2006), 7TMHMM (Moller, Vilo, & Croning, 2001), Pred-GPCR (Papasaikas, Bagos, Litou, Promponas, & Hamodrakas, 2004), and GPCRsClass (Bhasin & Raghava, 2005) are illustrative examples of many available methodologies being used for detection and classification of GPCRs. Topology prediction of GPCRs can be considered in the context of more general problem of IMP topology prediction. Consequently, methods such as TOP-PRED (von

DOI: 10.4018/978-1-4666-5888-2.ch046

Heijne, 1992), SOSUI (Hirokawa, Boon-Chieng, & Mitaku, 1998), HMMTOP (Tusnady & Simon, 2001) and TMHMM (Krogh, Larsson, von Heijne, & Sonnhammer, 2001) are used frequently for the purpose of topology prediction of GPCRs. In its simplest form, a GPCR topology prediction can be regarded as search for conserved seven hydrophobic stretches of residues in the sequence. Combination of the above mentioned methods was also used to improve discrimination of a particular family of GPCRs. According to latest classifications, GPCRs are grouped into six categories based on their shared sequence homologies, namely class A (Rhodoposin-like), class B (Secretin receptor family), class C (Metabotropic glutamate and pheromone), class D (Fungal mating pheromone receptors), class E (Cyclic AMP receptors) and class F (Frizzled and Smoothened) (X. Xiao, Wang, & Chou, 2011) (Figure 1). The importance of this superfamily of proteins, urged scientific community to setup specialized database servers to make the study of these proteins more accessible. GPCRDB (Vroling et al., 2010), IUPHAR (Sharman et al., 2012) and GPCR section of UniProt ("The Universal Protein Resource (UniProt) in 2010," 2009) are the well known publicly available databases. Structural information on GPCRs can be predicted at different levels ranging from detection of such proteins from their sequence information, prediction of locations and numbers of transmembrane segments within their sequences up to constructing their 3D model structures.

## Sequence Analysis of GPCRs, Identification and Classification

The prediction of sequence location and transmembrane orientation (collectively called topology) are the essential first steps in the analysis of GPCRs primary structures. The properties of the two-dimensional lipid bilayer impose a set of constraints on the folding of GPCRs, which can make predicting the location of TM segments relatively straightforward using a range of currently available methodologies (Rath et al., 2013).

## Exploration of Structures

In 2000, the first crystal structure of a GPCR, bovine rhodopsin, was published (Palczewski et al., 2000) and has been widely used as a template for homology modelling of other GPCRs (Latek, Pasznik, Carlomagno, & Filipek, 2013). That a single template would exist for all GPCRs seems unsupportable and so methods not entirely from the single template tradition became available. Up to now, more GPCR structures are known (i.e., near twenty GPCRs with known 3D structures belonging to class A) that their structures can be used for modelling of newly found GPCRs.

When little or no experimental structural information is available, the availability of computational methods means one can choose from a range of methods for predicting the structures, or aspects of the structural features of GPCRs.

The homology modelling of GPCRs has received great attention in part because of their importance as drug targets (Overington, Al-Lazikani, & Hopkins, 2006). Early homology modelling approaches for the Class A GPCR relied on using bacteriorhodopsin as the template (Nikiforovich, Galaktionov, Balodis, & Marshall, 2001) even though sequence identity is very low (~10%) (Cronet, Sander, & Vriend, 1993). Justification for this approach was based in part on the fact that both bacteriorhodopsin and bovine rhodopsin were activated by photons via a covalently bound retinal, that both proteins appeared to have 7 TM helices arranged sequentially with any crossover of helices prevented by the short length of the extra-membranous loops and that the N- and C-termini were placed on opposite sides of the membrane. Transposed sequence similarities were also considered to exist between some of the TMHs of bacteriorhodopsin and the GPCRs by some researchers (Taylor & Agarwal, 1993).

The PREDICT modeling method is a specialized GPCR modeling method based on observations of the topology of the crystal structures of both bovine rhodopsin and bacteriorhodopsin and sensitive to the membrane environment which has predicted the 3D structures of dopamine D2, neurokinin NK1, and neuropeptide Y Y1 receptors (Shacham et al., 2004).

The SWISS-MODEL (Parizot, 2002) modelling server in GPCR mode (Peitsch, Herzyk, Wells, & Hubbard, 1996)can also be used for 3D modelling of GPCR targets in three steps (Bordoli & Schwede, 2012).

*Ab initio* methods attempt to model structures from first principles and provide the computational framework for solving the protein folding problem. On the other hand, there are *de novo* methods, as applied to helical GPCRs, to be those which use knowledge-based approaches to predict specific structural properties of trans-