Chapter 1

Introduction

G-protein coupled receptors (GPCRs) make up 4% of the human genome, comprising one of the largest and most studied superfamilies of membrane proteins.[1–4] These proteins integrate in the cell membrane and mediate communication with the cell by responding to an extracellular ligand and activating a guanine nucleotide-binding protein (G-protein) at the intracellular side of the membrane.[5–7] The ligand may be a hormone, neurotransmitter, drug, odorant, or covalently bound molecule that responds to light. Once in an active state, the associated G-protein initiates a second messenger cascade, a signaling mechanism first characterized in 1958.[8] The processes mediated by GPCRs range from sensory recognition to immune responses to system regulation, and are not yet fully understood.[9,10] Their ubiquity and ligand-based mode of activation make them attractive drug targets: more than 50% of the current top selling drugs target GPCRs.[11–15] Of the 266 human proteins targeted by approved drugs, 26% of them are Class A GPCRs.[16] Small molecule drug development still focuses heavily on this superfamily for new advances in medicine.

With accurate binding site models of important GPCRs, it may be possible to

design new potent and selective drugs as well as improve the selectivity of existing drugs. Early work[17] in GPCR study confirmed the seven transmembrane (TM) helix structure that is conserved throughout the superfamily. Ligands bind primarily to the TM bundle rather than to the flexible extracellular loops,[18] and the conserved structure continues to drive understanding of GPCR structure and binding sites.

GPCRs are divided into three classes, each sharing a series of conserved residues and interactions. Class A, or rhodopsin-like receptors feature the conserved seven-TM structure, bind small ligands inside this TM core, and are the primary GPCR targets for drug development.[14,16,19] Conserved residues include a cluster of charged residues in TMs 1, 2, and 7 that interact via interhelical salt bridges and are important for stabilization of the receptor[20] and ion regulation,[21] a WXPFF motif in TM6 responsible for TM6 shape and implicated in both ligand binding[22] and receptor activation,[23,24] an NPXXY motif at the intracellular side of TM7, and a three-residue D- or ERY motif at the intracellular side of TM3 that can interact with TMs 5 or 6 to stabilize the inactive receptor.[25,26] In general, the universally conserved receptors contribute to stabilization of either an active or inactive state, while residues only conserved within a subfamily are responsible for subtype selective binding.

These highly conserved residues form the basis of the Ballesteros-Weinstein numbering system,[27] used throughout this work as a way of describing residues in a particular GPCR so that comparisons with related receptors can be made easily. This numbering scheme names a conserved residue in each TM as X.50, where

X is the TM helix number, and residues around it are numbered in sequence. A residue of interest will be denoted with its three-letter code and its number in the sequence, with the Ballesteros-Weinstein designation in superscript. When referring to a conserved residue without reference to a particular receptor, only the three-letter code and Ballesteros-Weinstein designation are used.

The increasing availability of GPCR X-ray crystal structures has dramatically advanced understanding of GPCR structure and ligand binding sites. The first structure available was bovine rhodopsin, first published in 2001.[28–36] It verified previously determined information about general structure, the role of highly conserved residues, and the ligand binding site, but it also provided a starting point for further structure predictions through homology modeling. The conserved 7-TM structure allowed this distantly related receptor to provide insight into human receptors in different systems, but the low sequence identity between bovine rhodopsin and many receptors of interest cast doubt on the validity of the resulting models. This changed again in 2007 with the publication of the human β2 adrenergic receptor (β2),[37–40] then in 2008 with the turkey β1 (β1) adrenergic receptor[41] and human adenosine A_{2a}.[42] An engineered mutant of another GPCR, BLT1, has been reported[43] and may result in yet another crystal structure. These structures revitalized the homology model as a valid avenue of structure prediction, both because of the similarity between the two related structures and the new availability of crystal templates expected to be more similar to receptors of interest.

Many important GPCRs are constitutively active, or self-activating, adding an-

other avenue of exploration.[44] GPCRs exist in equilibrium between active and inactive states, and ligand binding may stabilize one state or preserve the equilibrium by preventing other ligands from binding.[45–47] In general, ligands that stabilize an active or inactive state are known as agonists or inverse agonists; many known antagonists are now understood to be allosteric antagonists that simply block the binding site but do not affect the activation state.[48] As a result of this dynamic behavior, ligand studies alone do not completely illuminate the mechanism of activation. As structure identification has progressed, due to advances both in methods and availability of X-ray structures, activation has become the new frontier of GPCR research.

The bovine rhodopsin crystal structure contains a salt bridge between TMs 3 and 6,[28,31,49] connecting the highly conserved ERY motif at the intracellular end of TM3 with a conserved asparagine at the intracellular end of TM6. These residues are conserved throughout the family, and although this interaction is not observed in subsequent crystal structures there is evidence that the interaction is conserved in the native protein.[25] When considering the absence of this interaction in the available adrenergic crystal structures, it is important to consider the necessary modifications for crystallization may have disrupted this interaction. Experimental[25,50] and theoretical evidence show that this salt bridge creates an "ionic lock," stabilizing TMs 3 and 6 in their inactive positions. This lock may break during activation, but MD simulations imply the lock breaks late in the activation process, after some receptor reorganization takes place. One available crystal structure shows

the conformation of an active receptor.[32] This structure shows the ionic lock broken, but also that the intracellular end of TM5 is extended beyond the membrane into the cytosol. The changes observed between the active and inactive conformations of rhodopsin are useful as a starting point for further study into activation, but without an understanding of the intermediate steps the complete mechanism remains unclear.

1.1 Adrenergic Receptors

Adrenergic receptors are a family of nine GPCRs that mediate the stress response to endogenous catecholamine agonists epinephrine and norepinephrine (Figure 1.1). Much of the seminal work in GPCR research, including two of the five currently available crystal structures, has targeted this family of receptors. They are found throughout the sympathetic nervous system, and regulate functions such as blood pressure and heart rate at rest and in response to stress. The β receptors' role in regulating blood pressure drew initial attention to the family as drugs were developed to fight hypertension, then as the α subtypes were discovered and characterized it was possible to create drugs that targeted a specific part of the sympathetic nervous system.

β adrenergic receptors are perhaps the best studied subtype, as they include targets for widely prescribed drugs for asthma and blood pressure. β1 antagonists are standard anti-hypertensive and cardiovascular disease treatments,,[51] including propranolol, the first selective β blocking drug,.[52] β2 agonists including

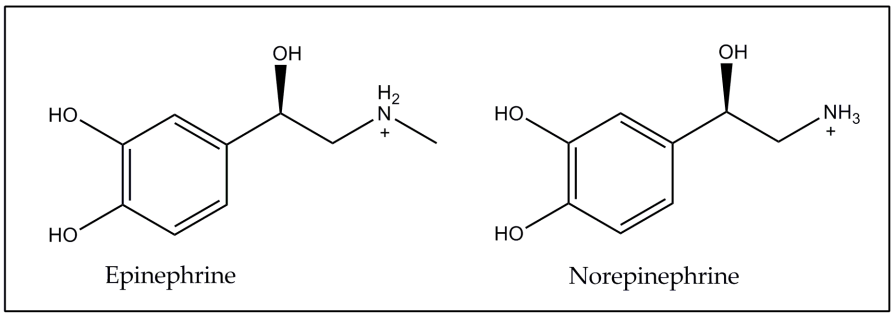

Figure 1.1: Adrenergic receptors respond to the endogenous catecholamine agonists epinephrine and norepinephrine.

salbutamol and formoterol are widely prescribed for asthma.[53] Cross-reaction of these drugs can have adverse effects for patients who suffer from both asthma and hypertension, so the utility of selective drugs for this subtype is immediate and widespread.

These receptors stimulate the production of cAMP through the activation of adenylyl cyclase and are found in the heart, lungs, and adipose tissue. Much of the structure-activity relationship work done for the adrenergic family has been on β2,[54–57] and currently the only human adrenergic crystal structure available is for β2.[37–40] As many binding site residues are conserved, this structure offers an excellent foundation for homology modeling, as discussed in Chapter 4.

The α1 receptors are located in vascular smooth muscle, the digestive tract, the liver, and central nervous system. The three subtypes are designated α1a, α1b, and α1d (the α1c designation was later discovered to be identical to α1a).[58] They stimulate production of inositol triphosphate and Ca^{++} [59] and regulate vasoconstriction in a variety of systems. α-blockers are sometimes prescribed as alternative antihypertensives instead of the traditional β-blockers.

Unlike the other two subtypes, the α2 receptors are inhibitory, decreasing the intracellular concentration of second messenger and decreasing the release of norepinephrine by inhibiting adenylyl cyclase.[60,61] These receptors are involved in blood pressure,[62] development, the startle reflex, and locomotion, among other processes. They are implicated in cognitive functions and working memory.[63] The canonical α2 antagonist, yohimbine, is an aphrodesiac, but also shows effects similar to α1 and β agonists, reflecting its role in the adrenergic feedback loop.[48,64]

All nine adrenergic receptors respond to the endogenous ligands epinephrine and norepinephrine, but their effects differ greatly depending on their associated G-protein and distribution throughout the body. In addition to the Class A GPCR conserved motifs, the adrenergic receptors share a series of conserved residues responsible for binding the endogenous agonists. A TM3 aspartic acid recognizes the protonated amine in nearly all adrenergic ligands, both agonists and antagonists. TM5 contains two absolutely conserved serine residues and a third serine conserved through all adrenergic receptors except α1a (and shifted one position in α2a and α2c). The Class A WXPFF motif participates in agonist binding.[22] TM7 contains a conserved tyrosine residue that may interact with the protonated amine on a ligand to form a ligand-mediated salt bridge between TMs 3 and 7 (some examples shown in Chapter 4). These conserved residues create an adrenergic pharmacophore for the family, and subtype selectivity depends on the binding site residues around these conserved motifs.

1.2 Computational Advances in GPCR Research

Theoretical predictions and models offer atom-level understanding and fine control of systems that are otherwise difficult to explore scientifically. Through structure prediction and molecular dynamics (MD) simulations, theory can illuminate details of both static structures as well as dynamic systems, providing insight important to drug development and understanding of GPCR activation. Homology models are often used to visualize results of experimental data, as well as provide a starting point for such studies. *Ab intio* methods that do not rely directly on an X-ray crystal structure are used to visualize systems that cannot be accurately modeled using existing structural data. Finally, molecular dynamics provides a way of monitoring the motions and activation mechanisms of GPCR systems that "snapshot" structural data cannot yet achieve.

1.2.1 Homology Modeling

The conserved seven-TM helix structure of GPCRs offers a useful starting point in structure prediction for GPCRs with no direct structural data. Once the crystal structure of bovine rhodopsin was available, direct homology modeling was widely used to create models of GPCR structures of interest as well as binding site models for drug development and screening for off-target effects. Rhodopsin's low sequence identity to human receptors and receptor families of interest complicated the structure prediction process, however, and homology models based on rhodopsin often needed heavy modifications.[65–67] The publication of β2 and

adenosine A_{2A} structures offered human template receptors to work with and turkey β1 provided a comparison to a well-studied, highly targeted family of receptors. While ligand-based homology modeling can work well,[12] it does not provide a useful method for systems that have yet to be extensively studied, or that have no known endogenous ligand. Chapter 4 explores the utility of simple homology modeling in generating structures for the entire human adrenergic family based on the two available, highly similar adrenergic X-ray crystal structures.

1.2.2 *Ab initio* Structure Prediction

For systems with less mutation and SAR data, and those that are only distantly related to available crystal structures, methods must be developed to predict their structures without direct use of the high-resolution crystal structures homology models rely upon. The PREDICT method[68] is one such method starting from the amino acid sequence of the receptor of interest, and Fanelli *et al.* developed a method for the prediction of the α1b receptor.[69]

The conserved structure of all GPCRs that inspired the homology modeling approach may also be a starting point for *ab initio* structure prediction. By considering the hydrophobicity of a region of amino acids as well as their tendency towards αhelical structure, it is possible to determine what parts of a GPCR sequence will form the TM helices and which will form the loops using only the amino acid sequence and thermodynamic information. Developed in the Goddard group, the Membstruk[70] and later MembSCREAM[71] protocols focus on optimization of these

predicted TM bundles, as most residues involved in ligand binding are located in the TM core.

1.2.3 Molecular Dynamics

Molecular dynamics (MD) provide a method to observe a GPCR interacting with its environment. While X-ray crystal structures offer high-resolution snapshots of the GPCR in an environment optimized for crystallization rather than native behavior, force field calculations can approximate protein structure and motion in a more native environment. MD also offers the opportunity to model sections of and interactions in the protein unavailable via crystallization.

This approach has been applied extensively to bovine rhodopsin to determine the regions and interactions important to activation.[72–76] These involve a variety of approaches, including constrained dynamics to test specific interactions, observation of the isomerization of retinal from *cis* to *trans* initiating the transition to the activated state, and they observe a range of changes including changes to the orientations of TMs 5 and 6. α1b has been studied with molecular dynamics based on a predicted structure,[26,69] and recently the crystal structure of β2 has been studied for 600 ns.[77] The potential for insight into activation from MD is profound, but in order to sample the entire millisecond timescale of GPCR activation it is necessary to incorporate some information from experimental studies of activation.

1.3 Subject of this Thesis

This thesis explores each of the major theoretical efforts in GPCR research: *ab initio* structure prediction, nanosecond (ns) timescale dynamics to elucidate activation mechanism, and homology modeling to obtain an entire family of structures based on closely related crystal structures. It explores two methods of validation for predicted structures: docking to verify the integrity of the ligand binding site and verification that residues implicated in binding are involved in the predicted structure, and comparison to stabilizing mutation data. I focus on the adrenergic family, utilizing the wealth of experimental data available to develop and test robust methods, leading to a greater understanding of the adrenergic receptors as well as methods that may be applied with confidence to other systems with less experimental data.

The thesis first addresses β1 turkey structure prediction directly from the amino acid sequence without assistance from crystal data, and validates the structure using stabilizing mutation data, before the crystal structure was available. This served as a test case for many of the recent developments in *ab initio* structure prediction. The next section starts with the β1 crystal data to initiate 10 ns of molecular dynamics on the apo protein, followed by ligand binding and a subsequent 10 ns of equilibration in the presence of ligand. These simulations also include a full intracellular loop, a region of the protein not currently available in the β1 crystal structure. Finally, the thesis turns towards homology models of the entire human adrenergic family, exploring the utility of available crystal structures for obtaining

3D models of highly similar receptors. Several of these models are validated with docked antagonists.

The central idea of this thesis is that theory and experiment can and must work in concert, with the findings from one propelling advances in the other in the mutual pursuit of knowledge. The methods developed in the course of this work are applied to systems with a great deal of experimental knowledge, but may be applied to those that have been less thoroughly characterized. Over the course of these explorations, new subtleties in adrenergic structure have been illuminated, and may drive further exploration into selective binding and the activation mechanism of these and other receptors.