Homology Modeling of CCR5 – Verification by Molecular Dynamics and Docking Studies with Agonists and Antagonists

Dissertation submitted in partial fulfillment for the award of the degree of

M. Sc

in

Biotechnology and Bioinformatics

By

Merugu Ravi Raja Tejasvi

3561020005

Under the guidance of

Dr. Sreedhara R Voleti

Principal Research Scientist & Head of CADD

Institute of Life Sciences, Hyd-46.

DEPARTMENT OF BIOINFORMATICS

FACULTY OF SCIENCE AND HUMANITIES

SRM UNIVERSITY

KATTANKULATHUR

2012

DECLARATION BY THE CANDIDATE

This is to certify that the thesis entitled **"Homology modeling of CCR5 – verification by molecular dynamics and docking studies with agonists and antagonists",** submitted by me to SRM University, Chennai for the award of the degree of Master of Science [Biotechnology & Bioinformatics] is a bonafide record of research work carried out by me under the supervision of **Dr. Sreedhara R Voleti.** The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Chennai 603203

Date: Signature of the Candidate

ACKNOWLEDGEMENTS

It is indeed a pleasant task to thank the people who have contributed towards the successful completion of this project.

Firstly I take this opportunity to thank **Dr. M. Vijjulatha**, Department of Chemistry, Nizam College - Hyderabad for creating a wonderful opportunity to work in ILS, Hyderabad.

With profound respect, I would like to express my deep sense of gratitude to my guide **Dr. Sreedhara R Voleti**, for giving a wonderful opportunity to work in his sophisticated lab with nice ambience. For his dynamic and invaluable technical guidance, lab meetings, for the fun in extra-curricular activities and constant encouragement with a cool attitude.

I would like to thank the research associates Mr. Lalith, Mr. Giridhar and Mr. Ajith for their valuable suggestions and support and my co-project trainees Kishore, Lakshmi, Devika, Swetha, Anushka and Jyothsna for their cooperation and suggestions for the successful completion of the project.

My sincere thanks to **Mr.Gurunathan**, Head, Dept of Bioinformatics and **Mr.Balajee** Asst.Prof. for their valuable encouragement and providing the opportunity and facilities to work in a technical environment to successfully carry out the study.

I thank Andrej Sali, author of MODELLER software for providing the free academic license.

Last but not least, I would like to thank my parents for their co-operation and moral support.

M.RAVI RAJA TEJASVI

To

My Mom, Dad

&

Sir Isaac Newton (1642)

TABLE OF CONTENTS

Abstract:

CC chemokine receptor 5 (CCR5) ,one of the most familiar drug target as HIV entry inhibitor has been found to be used as target for treatment of prostate cancer and many other diseases in recent years. Recently, research on CCR5 as anticancer target has started and possibly succeeds in case of other diseases too. Modeling the 3-dimensional (3D) structure of CCR5 for rational drug discovery procedures, thus, is an interest for many scientists including us. A 3D model of CCR5 based on the human CXCR4 was built and optimized including loop refinement. The reliability of the model is assessed through molecular dynamics simulation and docking with three known agonists and antagonists. The docking results are correlated with data available from site directed mutagenesis. We believe this model would explain the structural basis for the activation, and inhibition processes of CCR5, and may provide insights for the design of new agonists and antagonists.

Introduction

1. INTRODUCTION:

1.1 Project Topic

The CC-chemokine receptor 5 (CCR5) is a member of CC-chemokine receptor family G protein-coupled receptors (GPCR) with the characteristic structure of a seven transmembranes. It regulates trafficking and effector functions of memory/effector Th1 cells, macrophages, NK cells, and immature dendritic cells. It is a receptor for the chemokines, and macrophage inflammatory proteins such as MIP-1α (CCL3) or MIP-1β (CCL4). Regulated upon ctivation Normal T-cell Expressed and Secreted (RANTES)/CCL5 and Monocyte Chemoattractant Protein-2 (MCP2)/CCL8 [1] .

It also serves as a CD4 co-receptor for R5-tropic Human Immunodeficiency Virus type-1 (HIV-1) entry into activated memory $CD4^+$ T-lymphocytes and monocyte-derived macrophages [2]. Recent investigations on the role of the CCL5/CCR5 in tumor growth and metastasis [3] and prostate cancer [4] have shown potential therapeutic importance. It is also implicated in the pathogenesis of many immune and inflammatory conditions, including multiple sclerosis, rheumatoid arthritis (RA), psoriasis and chronic hepatitis [5], along with age-related diseases like atherosclerosis, Alzheimer's [6-7] and control of parasite replication and acute cardiac inflammation following infection with *Trypanosoma cruzi* [8].

1.2 Objectives of project

The objective of this project is to understand the structural characterization of the GPCR Class-A and model the 3D structure of CC-Chemokine Receptor Type 5 (CCR5).

Most of the homology models of CCR5 that were modeled till date used either using Bovine Rhodopsin (PDB id: 1F88) and, recently, using β-Adrenergic receptor (PDB id: 2RH1) as the templates. In current work, CXC Chemokine Receptor Type 4 (CXCR4) not only is more homologous to the CCR5 than Rhodopsin or β-Adrenergic receptors, but also belongs to the same chemokine subfamily - thus serves the better choice as the template for performing homology modeling. We then validated the model with the stereochemistry of the protein, and, with docking studies of the known antagonists and agonists.

Literature Review

2. REVIEW OF LITERATURE

2.1 Introduction to receptor proteins

Proteins are not only the structural building blocks of life; they also act as signaling molecules and receptors to receive those signals. Receptor molecules are present as intrinsic, extrinsic and integral proteins.

In the field of biochemistry, a ***receptor*** is a molecule most often found on the surface of a cell, which receives chemical signals originating externally from the cell. Receptors are protein molecules, embedded in either the plasma membrane (cell surface receptors) or the cytoplasm or nucleus (nuclear receptors) of a cell, to which one or more specific kinds of signaling molecules may attach. A molecule which binds to a receptor is called a ***ligand***, and may be a peptide (short protein) or other small molecule, such as a neurotransmitter, a hormone, a pharmaceutical drug, or a toxin.

Ligands which merely block receptors without inducing any response are called ***antagonists***, and those which induce some biological response are called ***agonists***. Ligand-induced conformational changes in receptors result in functional responses (such as production in secondary messengers) within cellular environment.

2.1.1 Ternary Complex Models.

Three models of ternary complex formation were proposed [9,10,11]. In Panel-A (Figure-1) the receptor (R) interacts either with the ligand (L) or the G protein (G). The activated complex is generated by interaction with the remaining component resulting in LRG[9]. Panel-B illustrates the Extended Ternary Complex model as proposed by Samama *et al*.[34]. The receptor can either interact with the ligand to form LR or isomerize to form R*. R* can either interact with G or with L. The two possible complexes then interact with the remaining component to form LR*G. If R interacts with L prior to isomerization, then the conformation of R will result from interaction with L generating LR*. LR* then interacts with G to form the activated complex LR*G. Panel-C illustrates the method of receptor activation as proposed by Gether and Kobilka [11]. In this model, the receptor is in a neutral conformation that can interact with either agonists or inverse agonists. Agonists would shift the equilibrium towards the activated complex and inverse agonists would shift the

equilibrium towards in inactive complex. Partial agonists would stabilize an intermediate that could then change conformation to the activated complex. This model still allows for activation of the receptor without any ligand accounting for basal activity.

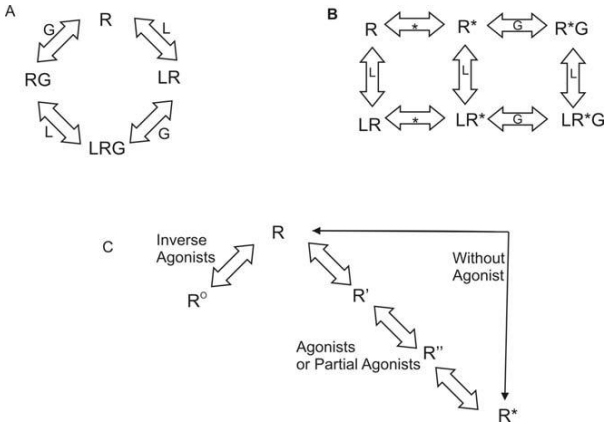


Figure 2.1: Ternary Complex Models Models proposed to demonstrate the equilibrium between Ligand, Receptor, G protein and LRG complex.

R - receptor

L - ligand

G - G protein

Courtesy: Abby L. Parrill, 2011

2.1.2 Five classes of Ligands:

Agonists: able to activate the receptor and result in a maximal biological response (ex: natural ligands)

Partial Agonists: not able to completely activate, causing responses which are partial compared to those of full agonists

Antagonists: bind to receptors resulting in receptor blockage, inhibiting the binding of agonists and inverse agonists.

Inverse agonists: reduce the activity of receptors by inhibiting their constitutive activity.

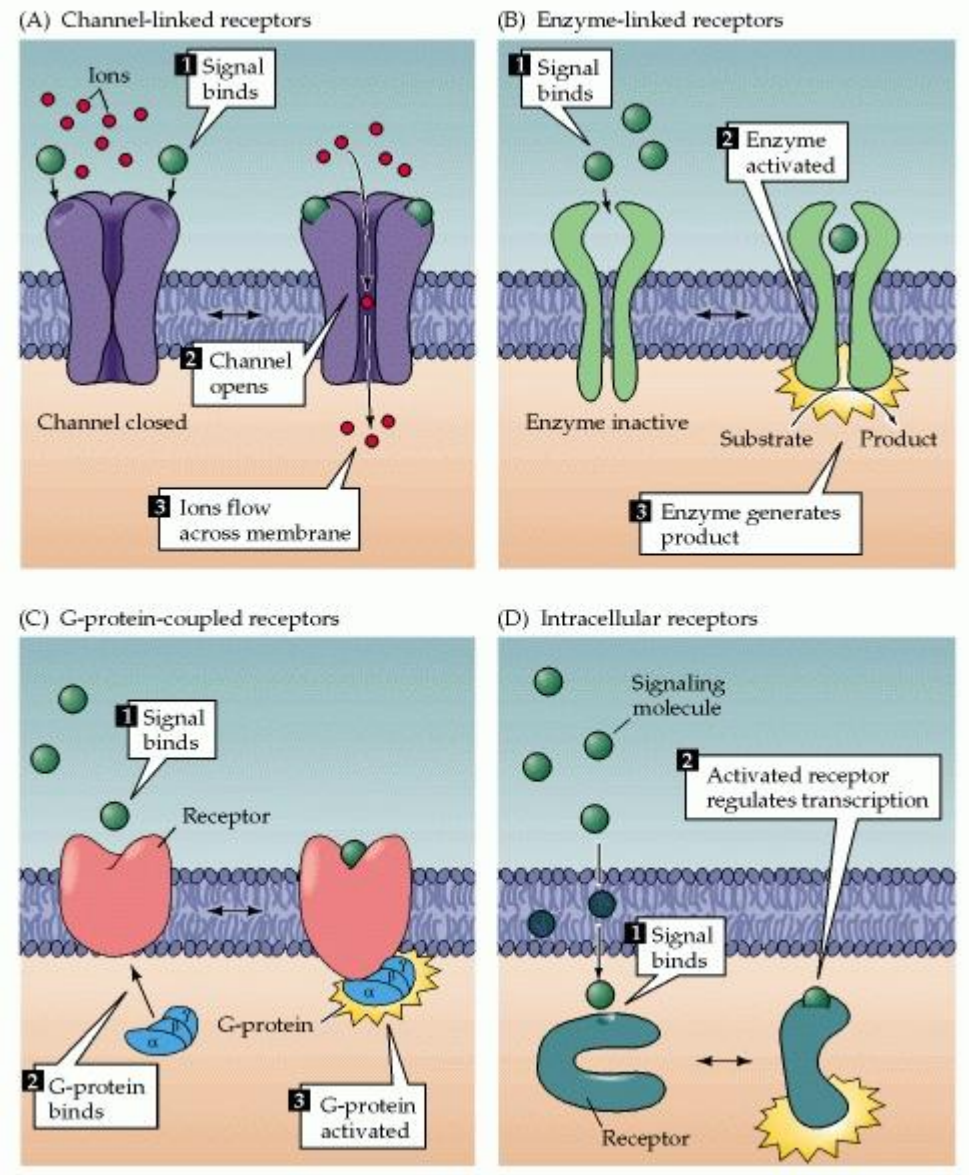
Neutral Antagonists: has no activity in the absence of an agonist or inverse agonist but can block the activity of either.

2.1.3 Classification of receptor proteins

Receptors can be grouped into following families defined by the mechanism used to transduce signal binding into a cellular response [12].

a. Channel-linked receptors (ligand-gated ion channels) have the receptor and transducing functions as part of the same protein molecule. Interaction of the chemical signal with the binding site of the receptor causes the opening or closing of an ion channel pore in another part of the same molecule.

b. **Enzyme-linked receptors:** The intracellular domain of such receptors is an enzyme whose catalytic activity is regulated by the binding of an extracellular signal. The great majority of these receptors are protein kinases and other receptors for growth factors.



Courtesy: D Purves et al., Neuroscience. 2nd edition

Figure 2.2: Categories of cellular receptors. Membrane-impermeate signaling molecules can bind to and activate either receptor types A, B, C. Membrane permeate signaling molecules receptors type D.

c. G-protein-coupled receptors regulates intracellular reactions by an indirect mechanism involving an intermediate transducing molecule, called the GTP-binding proteins (***G-proteins***). All receptors share the structural feature of crossing the plasma membrane seven times and hence also referred to as ***7-transmembrane receptors***.

d. Intracellular receptors are activated by cell-permeate or lipophilic signaling molecules. Many of these receptors lead to the activation of signaling cascades that produce new mRNA and protein within the target cell. Some intracellular receptors are located primarily in the cytoplasm, while others are in the nucleus. In either case, once these receptors are activated they can affect gene expression by altering DNA transcription.

2.2 G-Protein Coupled Receptors

G protein-coupled receptors (GPCRs) represent the largest super family of all receptors (approx. 2% of the human genome) and are the targets for nearly 50% of all currently used therapeutic drugs.GPCRs are membrane proteins that convey the majority (80%) of signal transduction across cell membranes. GPCRs are major targets for drug development today. Over the last 20 years, several hundred new drugs have been registered which are directed towards modulating more than 20 different GPCRs, and approximately 40% of the top 200 synthetic drugs act on GPCRs [13]. Many vital physiological events such as sensory perception, immune defense, cell communication, chemotaxis, and neurotransmission are mediated by GPCRs. They are activated by diverse ligands, which vary from single photons through ions, odorants, amino acids, fatty acids, neurotransmitters, peptides/polypeptides, to proteolytic enzymes, which cleave off receptor fragments to generate an activating ligand and adhesion molecules.

All GPCRs are located within the plasma membrane and have a common architecture consisting of seven-transmembrane (TM) domains embedded in the hydrophobic environment, connected by extracellular (ECL) and intracellular (ICL) loops present in the hydrophilic environment. The name GPCR came from the fact that these receptors interact with **G proteins (guanine nucleotide-binding proteins)** for the activation and deactivation of the signal which are discussed in later sections.

2.2.1 GPCR sequence numbering system

A number of numbering systems for easy identification of GPCRs have been suggested and used. Weinstein and Ballesteros [14] numbering is the most useful, but still has its own anomaly. The numbering is straight forward; amino acid residues in the TM domain are assigned two numbers (N1.N2)

N1 = TM number

N2 = number relative to the most conserved residue in this TM, which is assigned 50, numbers decreasing towards N-terminus and increasing towards C-terminus

Examples:

- P267 in TM6 of Rhodopsin is given number P6.50 or P6.50(267)
- E134/R135 in D/ERY sequence of TM3 in Rhodopsin has been given E3.49 (134)/R3.50 (135)

2.2.2 Classification of GPCRs:

Approximately 800 of these seven GPCRs have been identified of which over 300 are non-olfactory receptors .Subdivision on the basis of sequence homology allows the definition of rhodopsin, secretin, adhesion, glutamate and Frizzled receptor families. NC-IUPHAR[15] recognizes Classes A, B, and C, which equate to the rhodopsin, secretin, and glutamate receptor families

GPCRs are be classified in to three major classes/families.

- Class A
- Class B
- Class C
- Frizzled and Other 7TM proteins

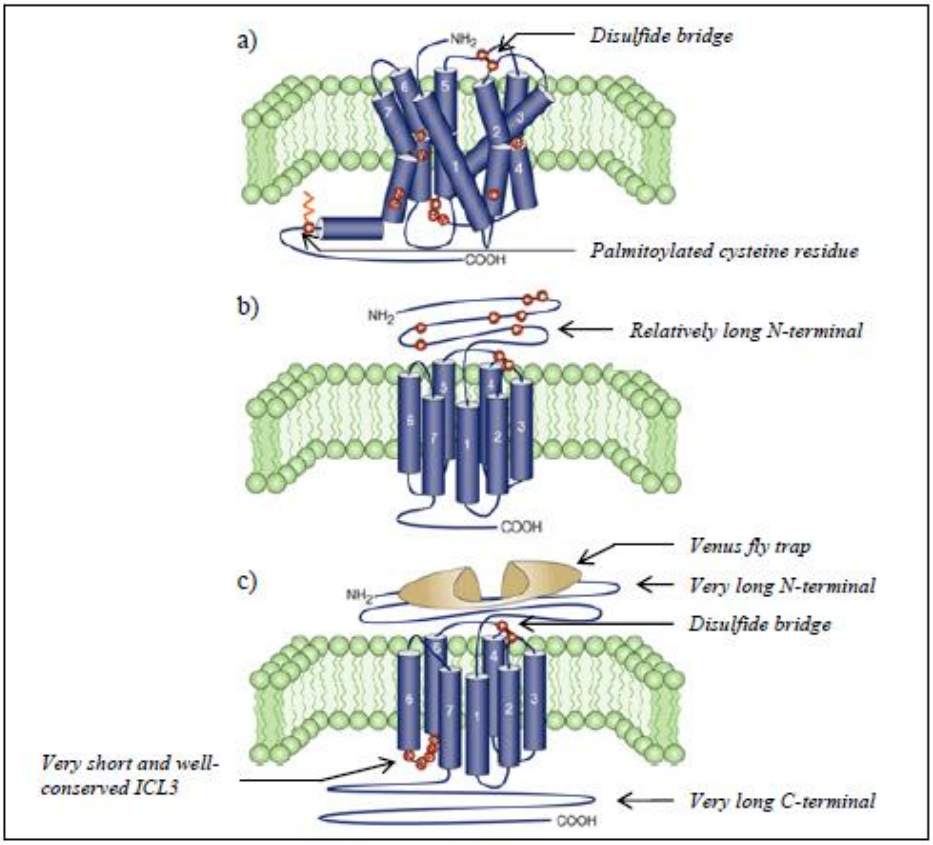
Family A:

Family A, the rhodopsin-like family, is the largest family of GPCRs comprising of 672 members, including 388 odorant receptors and accounts for nearly 85% of all GPCR genes [16]. The receptors of this family are characterized by short N-termini (without any conserved domains) and several highly conserved amino acids in the seven transmembrane bundles. Their structure is tilted with the disulfide bond that connects extracellular loops 1 and 2. They also have a palmitoylated cysteine residue at C-terminal.

Family B:

Family B, the Secretin family, is a small family of GPCRs consisting of 15 known hormone peptides receptors and 35 orphan receptors in humans. This family of receptors is

characterized by a relatively long N-terminal tail and a network of three conserved cysteine disulfide bridges which stabilize the N-terminal structure (**Figure 2.3**).



Courtesy: George et al., 2002.

Figure 2.3 Molecular structures of GPCRs. All GPCRs share the seven transmembrane α -helices which are connected by three extracellular loops (ECL1, ECL2 and ECL3) with an extracellular NH2 terminal and three intracellular loops (ICL1, ICL2 and ICL3) with an intracellular COOH-terminal. Red balls represent residues. **a) Family A GPCR.** The receptor is 'tilted' and 'kinked' with a disulfide bridge that connects ECL1 and ECL2 and a palmitoylated cysteine residue in the intracellular C-terminal domain. **b) Family B GPCR**. These receptors contain a relatively long N-terminal tail and a network of conserved disulfide bridges. **c) Family C GPCR.** Characterized by very large N- and C-terminal tails, a putative disulfide bridge connecting ECL1 and ECL2, together with a very short and well-conserved ICL3. The ligand-binding domain, located in the amino terminus is often described as being like a venus fly trap.

Family C:

Family C, forms the other small family of GPCRs in humans. This family contains 22 members including metabotropic glutamate receptors, γ-aminobutyric acid type B ($GABA_B$) receptors and calcium-sensing receptors and 10 orphan receptors (IUPHAR database, Accessed on 2012-03-16). The majority of Family C receptors are characterized by very large N- and C-terminal tails, a putative disulfide bridge connecting ECL1 and ECL2, together

with a 27 very short and well-conserved ICL3 (**Figure 2.3**). The ligand binding site is located in the N-terminal domain, which is often described as a venus fly trap. Most Family C GPCRs, except the $GABA_B$ receptors, contain a cysteine-rich domain with nine conserved cysteine residues, which links the venus fly trap to the seven transmembrane domains [17].

2.2.3 GPCR Trafficking

Trafficking is the mechanism by which a cell transports proteins to the appropriate positions in or out of the cell. GPCRs synthesized in the endoplasmic reticulum (ER), interact with endogenous chaperones, which assist in folding (or may retain incorrectly folded proteins) and are transferred to the plasma membrane(PM), where they exert their physiological functions. GPCRs are subjected to a stringent quality control mechanism at the endoplasmic reticulum, which ensures that only correctly folded proteins enter the secretory pathway. Because of this quality control system, point mutations that may result in the production of misfolded and disease-causing proteins, that are unable to reach their functional destinations in the cell. Functional rescue of misfolded mutant receptors by small nonpeptide molecules originally designed to serve as receptor antagonists has been demonstrated in two receptor proteins - gonadotropin-releasing hormone receptor and the vasopressin type-2 receptors. These small molecules called pharmacoperones [18] serve as molecular templates, promoting correct folding and allowing the mutants to pass the scrutiny of the cellular quality control system and be expressed at the cell surface membrane.

Although the steric character of the protein backbone restricts the spectrum of protein shapes that are recognized by the stringent quality control mechanisms, some features displayed by proteins including exposure of hydrophobic shapes, unpaired cysteines, or immature glycans have been identified as important in chaperone-protein association [19]. In fact, molecular chaperones possess the ability to recognize misfolded proteins by the exposure of hidden hydrophobic domains or particular motifs [20].

2.2.4 Internalization and desensitization of GPCRs

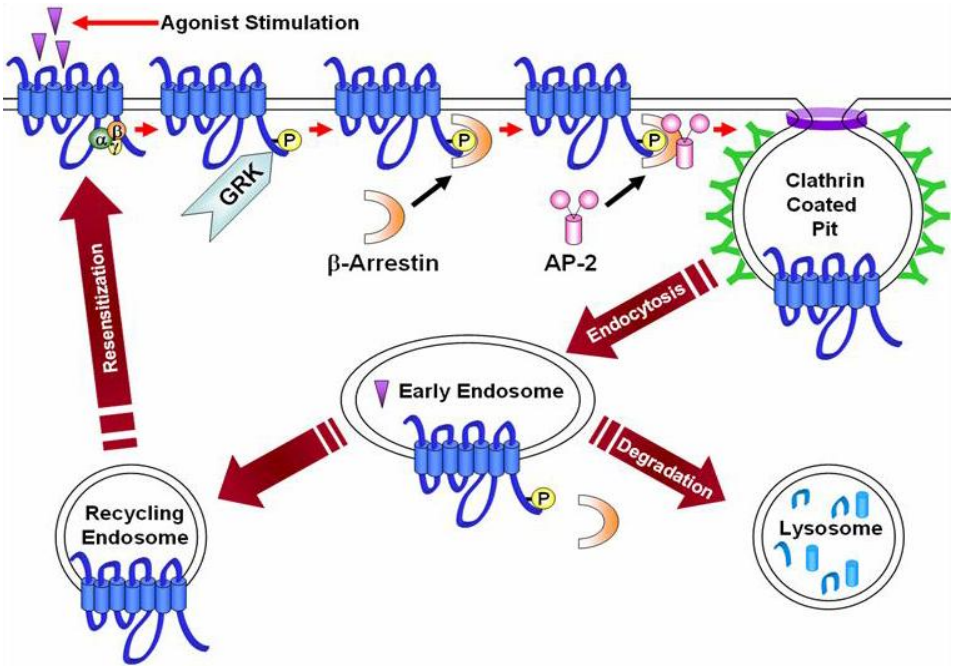
For most GPCRs, a loss of functional response, known as ***desensitization***, can occur when receptors are exposed to repeated or continuous agonist stimulation. Desensitization is cell specific and dependent upon both the expression and sub cellular localization of specific components that regulate the desensitization processes. Three general mechanisms are

associated with desensitization of GPCRs: 1) receptor phosphorylation; 2) receptor internalization or sequestration; and 3) receptor downregulation. Agonist-induced receptor phosphorylation often results in the most rapid desensitization. In this process, conformational changes occurring on agonist binding lead to phosphorylation of various serine or threonine residues of ICL3 and/or the C-terminal tail by GPCR kinases (GRKs) [21]. This facilitates β-arrestin binding and promotes receptor uncoupling from their cognate heterotrimeric G-proteins. Receptor desensitization can also occur in a GRK-independent manner through phosphorylation of different serine/threonine sites (also of ICL3 and the C-terminal tail) by secondary messenger-dependent protein kinases, e.g. PKA and PKC.

In addition to the role in GPCRs desensitization, β-arrestins promote receptor internalization involving Clathrin and adaptor protein-2 (AP-2) [22]. ***Internalization*** of components of the plasma membrane, associated ligands and fluid, is also called endocytosis, which is a fundamental process in eukaryotic cells. Once β-arrestin is bound to a GPCR, it undergoes a conformational change allowing it to serve as a scaffolding G-protein for an adaptor complex AP-2, which in turn recruits clathrin. If enough receptors in the local area recruit clathrin in this manner, they aggregate and the membrane buds inwardly as a result of interactions between the molecules of clathrin, in a process called ***opsonization***. Once the pit has been pinched off the plasma membrane due to the actions of amphiphysin and dynamin, it becomes an endocytic vesicle. Following release of the vesicles into the cytoplasm as clathrin-coated vesicles, the coated vesicles fuse with early endosomes. After targeting to the endosomal compartment, the adapter molecules and clathrin dissociate and GPCRs can be either rapidly dephosphorylated and recycled back to the plasma membrane or targeted to lysosomes for degradation (**Figure 2.4**). Thus, receptor trafficking is critical for regulation of the temporal and spatial aspects of GPCR signalling. Indeed, internalization controls the density of the receptor at the cell surface, signal termination and propagation as well as receptor resensitization [23].

At any point in this process, the β-arrestins may also recruit other proteins such as the non-receptor tyrosine kinase (nRTK), c-src, which may initiate activation of ERK1/2, or other mitogen-activated protein kinases (MAPKs) signalling through, for example, phosphorylation of the small GTP-ase, Ras, or recruit the proteins of the ERK cascade directly (i.e. Raf-1, MEK, ERK-1/2) at which point signalling is initiated due to their close proximity to one another. Another target of c-src is the dynamin molecules involved in

endocytosis. Dynamins polymerize around the neck of an incoming vesicle, and their phosphorylation by c-src provides the energy necessary for the conformational change allowing the final "pinching-off" from the membrane.



Courtesy: Strachan et al., 2009

Figure 2.4: Clathrin-dependent internalization of GPCRs. The current model for agonist-induced internalization of GPCRs is based primarily on the β2-adrenoceptor receptor and other GPCRs. Following activation, receptors are phosphorylated by kinases, resulting in rapid desensitisation due to G-protein uncoupling. G-protein uncoupling is further promoted by the binding of arrestins to the phosphorylated third intracellular loops and C-terminal tails of agonist-activated GPCRs. In addition, arrestins promote the targeting of desensitized receptors to clathrin coated pits for internalization by the interaction of the C-terminal portions of arrestin with both the clathrin heavy chain and the β2-adaptin subunit of AP-2. The dynamin induces neck formation of coated pits and their release into the cytoplasm as clathrin-coated vesicles. These coated vesicles fuse with early endosomes where the receptors may be dephosphorylated by specific phosphatases and recycled back to the plasma membrane fully resensitized or targeted to lysosomes for degradation

Although for most GPCRs, including the β2-adrenoceptor, internalization involves clathrin-dependent endocytosis pathways, some GPCRs, including the endothelin-A receptor, are internalized via clathrin-independent endocytosis pathways [24]. For example, endocytosis can occur through structures coated with the caveolin protein. However, this caveolae-mediate endocytosis is also dependent upon dynamin and responsible for endocytosis of some proteins that partition into cholesterol-rich membrane domains, especially in endothelial cells.

2.2.5 GPCR dimerisation and oligomerisation:

GPCRs can exist as dimers or higher oligomers for function. Until past few years GPCRs are thought to be existing in monomers only. However, a few literatures evidences show that they might be present as oligomeric arrays. Many of the key early studies have been reviewed by *Salahpour et al*[25]. The study on co-expression of two distinct, non-functional mutants of the angiotensin II AT1 receptor resulting in the appearance of a functional receptor (Monnot et al., 1996) gave support for the dimerisation of GPCR.

$GABA_B$ receptor heterodimerisation involves interactions between the coiled-coil domains within the C-terminal tails of the two $GABA_B$ receptors (Figure. 2.5). Coiled-coiled domains are present in many proteins and are well-characterized mediators of protein-protein interaction. However, this does not appear to be a general mechanism for GPCR oligomerisation. Although *Cvejic and Devi* report that a C-terminally truncated version of the dopioid receptor that lacks 15 residues cannot dimerise [26], this receptor does not have an obvious coiled-coil domain in this region akin to those in the $GABA_B$ receptor partners. The large extracellular extensions of class III GPCRs contain several cysteine residues, and disulphide cross-links can mediate (*Romano et al., 1996*) or certainly contribute to (*Romano et al., 2001*) dimerisation (Figure. 2.5). In the case of metabotropic glutamate receptor 1, removal of the extracellular domain prevents dimerisation. Mutation of Cys140 indicates that this residue contributes to dimerisation but it is not the only contact point for dimer formation [27,28]. Only class III GPCRs possesses the large extracellular domain. However, the N-terminus of the bradykinin B2 receptor has been reported to be involved in dimerisation of this class I GPCR [29]. Computational studies and direct experimental observation have indicated that dimerisation of other GPCRs may involve interfaces produced by transmembrane helices V and VI [30]. Two hypotheses exist for the mechanism of dimerisation [31] of class 1 GPCRs (Figure. 2.5).

1. The first of these is that domain swapping occurs between GPCR pairs. TMs I-V and VI-VII form essentially independent folding domains. GPCR sequences can be split between TM V and VI, and, when these fragments are co-expressed, functional receptors can be reconstituted.

2. A second model simply envisages that a GPCR dimer is formed by lateral packing of the individual polypeptides (Fig. 2.5). This model is supported by studies of mutants of the vasopressin V2 receptor. Interestingly, although unable to resolve these models, peptides

corresponding to TM regions V and VI of the β2-adrenoceptor and the dopamine D2 receptor interfere with dimerisation of their respective GPCRs.

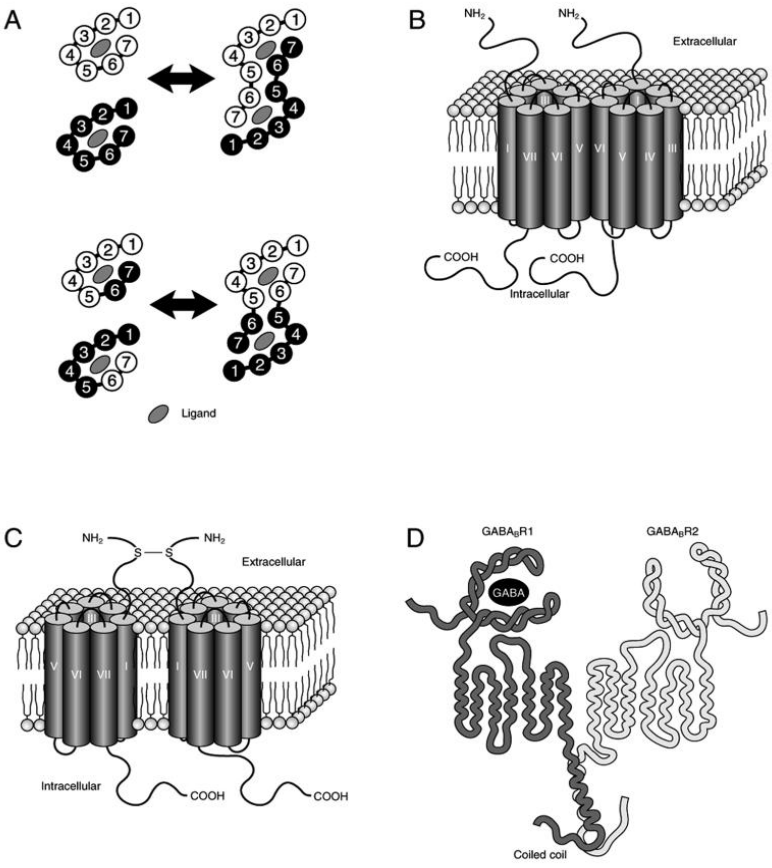
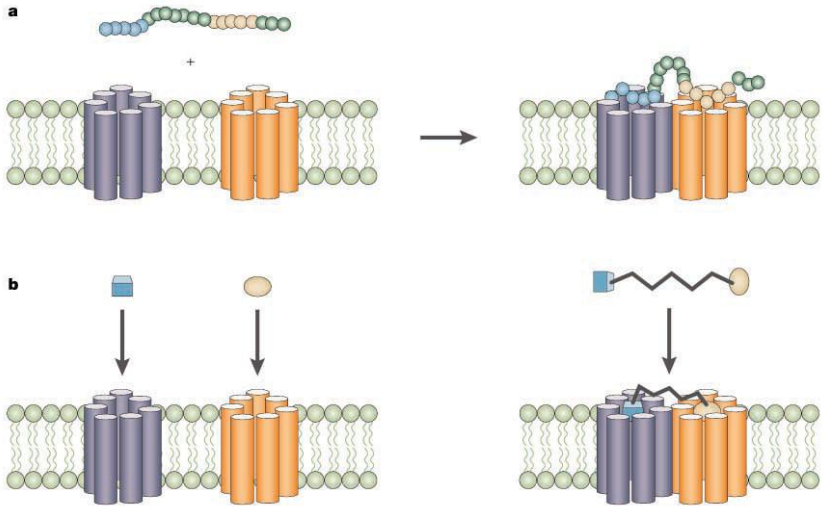


Figure 2.5: Models of GPCR dimerisation
(A)Domain swaps. Functional, ligand binding GPCRs can be reconstituted following expression of separated fragments of a GPCR comprising TM I-V and VI-VII. Because distinct ligand-binding properties can be produced when the two segments are derived from different GPCRs, various domain swapping models have been proposed.
(B)Lateral packing. For GPCRs that have ligands sufficiently large that all the binding determinants are unlikely to lie within the crevice formed by the architecture of the seven TM, lateral packing and/or microaggregation may be required to initate ligand-induced signal transduction. Evidence from peptide inhibition studies suggest that contacts may occur between TM V and VI (*Ng et al., 1996; Herbert et al., 1996*).
(C) Intramolecular disulphide bonds. Class III GPCRs have long extracellular N-terminal extensions that bind the ligand (*Kunishima et al., 2000*). These regions also contain cysteine residues that provide intramolecular disulphide bonds that contribute to dimer stabilization (*Romano et al., 2001*).
(D) Coiled-coils. Heterodimerisation between the GABABR1 and GABABR2 to produce a functional GABAB receptor requires the C-terminal tails. These regions are predicted to produce a coiled-coil interaction that is vital for membrane delivery of the proteins (*White et al., 1998*).

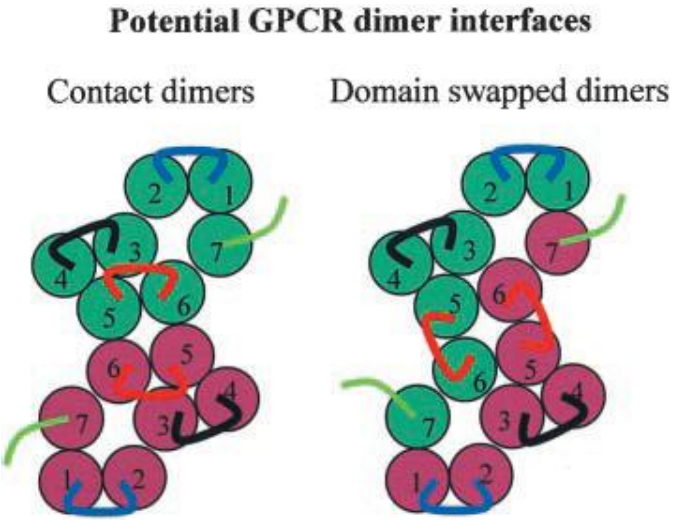
The dimerization of the receptor can also be induced by the attachment of the ligand. The procedure is depicted in the figure 2.6 and overview of changes in TMs in Figure 2.7



Courtesy: *Sandvig et al, 2008*

Figure 2.6: Bivalent and dimeric ligands

a.Schematic representation of ligand-induced heterodimerization by a bivalent peptide ligand. A bivalent peptide with two distinct receptor-recognition sites might bind two different receptors simultaneously and induce receptor heterodimerization. If the two receptors pre-exist in an oligomeric state, the ligand might enhance the heteromeric conformation by crosslinking two receptors.
b. Dimeric ligands generated from two monovalent ligands. Similar to a bivalent ligand, a dimeric ligand that has an appropriate linker might induce or enhance dimerization of two receptors. It is plausible that homodimeric and heterodimeric ligands might selectively bind homomeric and heteromeric receptors, respectively.



Courtesy: Gerda E. Breitwieser, 2004

Figure 2.7: Potential GPCR dimer interfaces. Contact dimers, in which the interface between GPCR monomers involves surface contact between helices of two independent monomers (indicated by color), and domain-swapped dimers, in which helices 6 and 7 are exchanged or "swapped" between GPCR monomers (indicated by color), are illustrated, viewed at the cytoplasmic face of the membrane. Note the origins and locations of the intracellular loops (i1 loop, blue; i2 loop, black; i3 loop, red) contributing to the putative G protein binding surface in the two models.

2.2.6. General mechanism of signaling by GPCR

If a receptor in an active state encounters a G protein, it may activate it. Some evidence suggests that receptors and G proteins are actually pre-coupled. For example, binding of G proteins to receptors affects the receptor's affinity for ligands. Activated G proteins are bound to GTP. Further signal transduction depends on the type of G protein. G proteins are classified based on activity as

1. $\mathbf{G_s}$ –stimulates adenylyl cyclase(AC)
2. $\mathbf{G_i}$ – inhibits adenylyl cyclase., activates K+ channel (GIRK)
3. $\mathbf{G_o}$ – ('o' for other) neither stimulates or inhibits AC, inhibits Ca^{2+} channels
4. $\mathbf{G_q}$ – stimulates/quenches PLCβ
5. $\mathbf{G_t}$ – (transducin), stimulates cGMP phosphodiesterase
6. $\mathbf{G_{12}}$ – RhoGEF

Currently, GPCRs are considered to utilize two primary types of transducers: ***G-proteins*** and ***β-arrestins***. Because β-arr's only have high affinity to the phosphorylated form of most GPCRs, the majority of signaling is ultimately dependent upon G-protein activation. However, the possibility for interaction does allow for G-protein independent signaling to occur.

A G-protein that activates cyclic-AMP formation within a cell is called a ***stimulatory G-protein***, designated G_s with alpha subunit $G_{s\alpha}$. G_s is activated, e.g., by receptors for the hormones epinephrine and glucagon. The β-adrenergic receptor is the GPCR for epinephrine. The α subunit of a G-protein (G_α) binds GTP, and can hydrolyze it to GDP + P_i. The α and γ subunits have covalently attached lipid anchors, that insert into the plasma membrane, binding a G-protein to the cytosolic surface of the plasma membrane. Adenylate Cyclase (AC) is a transmembrane protein, with cytosolic domains forming the catalytic site. The sequence of events by which a hormone activates cAMP signaling is summarized below and in the figure 2.8.

1. Initially the G-protein-a subunit has bound GDP, and the α, β, γ subunits are complexed together. $G_{\beta,\gamma}$ the complex of β & γ subunits, inhibits G_α.
2. *Hormone binding*, usually to an extracellular domain of a 7-helix receptor (GPCR), causes a conformational change in the receptor that is transmitted to a G-protein on

the cytosolic side of the membrane. The nucleotide-binding site on G_α becomes more accessible to the cytosol, where [GTP] is usually higher than [GDP]. G_α releases GDP and binds GTP. (GDP-GTP exchange)

3. Substitution of GTP for GDP causes another **conformational change** in G_α. G_α-GTP dissociates from the inhibitory bg subunit complex, and can now bind to and activate Adenylate Cyclase.
4. **Adenylate Cyclase, activated** by the stimulatory G_α-GTP, catalyzes synthesis of cAMP.
5. **Protein Kinase A** (cAMP-Dependent Protein Kinase) catalyzes transfer of phosphate from ATP to serine or threonine residues of various cellular proteins, altering their activity.

Turn off of the signal:

1. **G_α** hydrolyzes GTP to GDP + P_i (**GTPase**). The presence of GDP on G_α causes it to rebind to the inhibitory $G_{\beta\gamma}$ complex. Adenylate Cyclase is no longer activated.
2. **Phosphodiesterases** catalyze hydrolysis of cAMP to AMP.
3. **Receptor desensitization** varies with the hormone. In some cases, the activated receptor is phosphorylated via a G-protein Receptor Kinase. The phosphorylated receptor then may bind to a protein **β-arrestin**. β-arrestin promotes removal of the receptor from the membrane by clathrin-mediated endocytosis. β-Arrestin may also bind a cytosolic **phosphodiesterase**, bringing this enzyme close to where cAMP is being produced, contributing to signal turnoff.
4. **Protein Phosphatases** catalyze removal by hydrolysis of phosphates that were attached to proteins via Protein Kinase A.

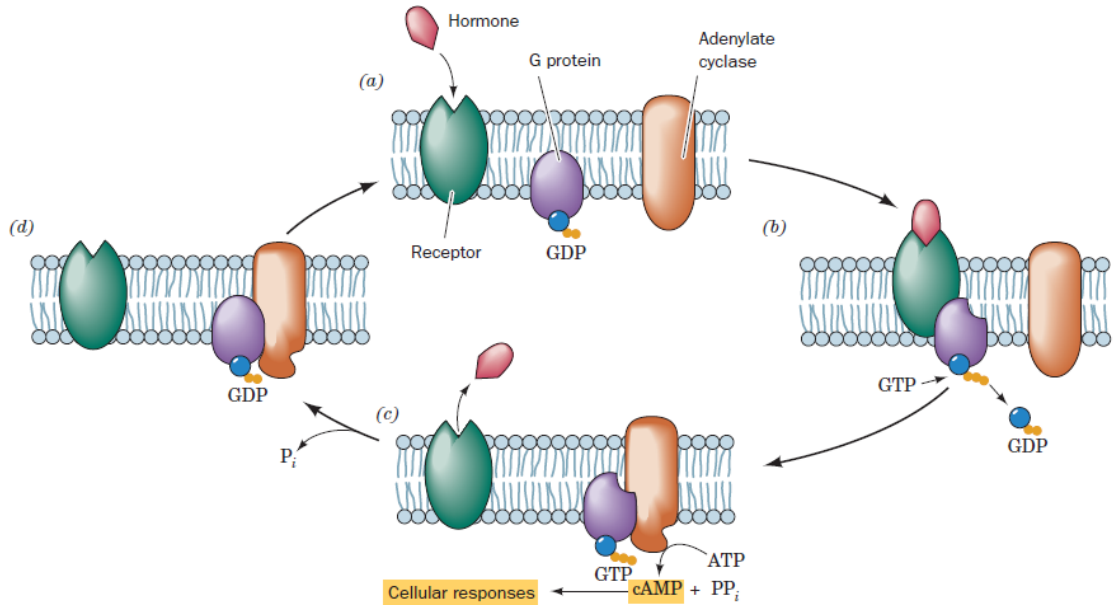
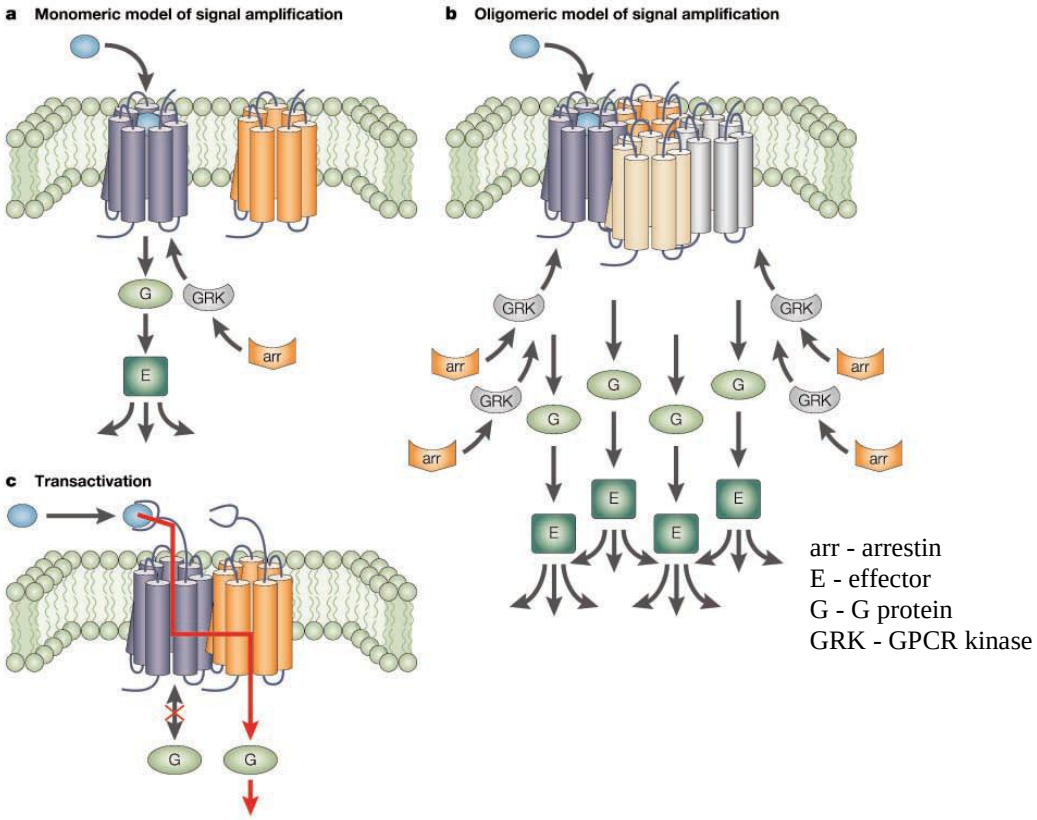


Figure 2.8: Activation/deactivation cycle for hormonally stimulated AC. (*a*) In the absence of hormone, heterotrimeric G protein binds GDP and AC is catalytically inactive. (*b*) The hormone receptor complex stimulates the G protein to exchange its bound GDP for GTP. (*c*) The G protein GTP complex, in turn, binds to and thereby activates AC to produce cAMP. (*d*) The eventual G protein-catalyzed hydrolysis of its bound GTP to GDP causes G protein to dissociate from and hence deactivate AC.

2.2.7 Transactivation of GPCRs: signal amplification at the receptor level

Traditionally, in models of G-protein-coupled receptor (GPCR) function, it has been assumed that a single agonist activates a single receptor and that signal amplification occurs at the level of the G protein or effector (Figure 2.9 panel a).However, investigations of the homo- and hetero- oligomerization of several GPCRs have shown that ligand binding to one receptor might activate neighboring receptors in the oligomeric complex (Figure 2.9 panel b). Similar results have been observed for GPCR heteromers, for which agonist binding to one receptor partner in the heteromer results in activation of the other receptor [32]. Interestingly, in the native γ-aminobutyric acid (GABA)B(1)– GABAB(2) heteromer, it seems that the GABAB(1) subunit is not capable of G-protein coupling, and the GABAB(2) subunit cannot bind ligand, indicating that heteromerization is required not only for proper transport in the GABAB receptor, but also for signalling, which seems to occur only through transactivation (Figure 2.9 panel c).

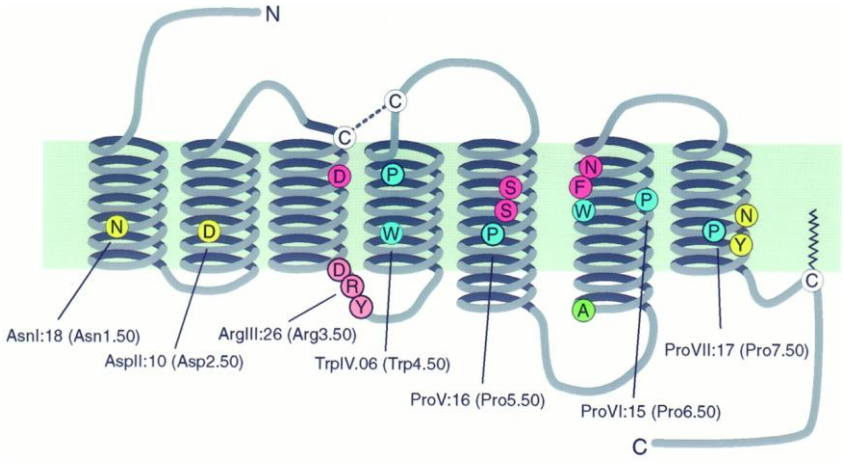


Courtesy: *Sandvig et al, 2008*

Figure 2.9: Transactivation of GPCRs

2.2.8 Structural Characterizations of GPCRs:

The first crystal structures that emerged after that of rhodopsin were from Brian Kobilka's laboratory[33]. Solving the β2-adrenergic crystal structure presented an enormous challenge in producing sufficient pure protein, stabilizing flexible domains such as ICL3, defining appropriate detergent/lipidic environments and crystallization conditions, and developing microdiffraction technology to obtain x-ray data from very small crystals. Kobilka and co-workers [33] first used an antibody to stabilize the ICL3 (which is highly mobile and compromises crystallization) and achieved a 3.4–3.7Å resolution, which revealed the TM domains very clearly. This was followed up with another crystal structure in which ICL3 was stabilized by substitution with lysozyme, a highly structured molecule, which improved the resolution to 2.4Å so that ECL structure could be visualized as well as elements of the ICLs [34]. The structure also shows membrane cholesterol interaction with TMs 2, 3, and 4, which is suggested to be the site at which small molecules can act allosterically at GPCRs to potentially activate or inactivate them [35].



Courtesy : Brian K. Kobilka,1998

Figure 2.10: The general structure characteristics of GPCR

The most conserved residue in each transmembrane segments is indicated both using the Schwartz nomenclature [36] and the Ballesteros-Weinstein nomenclature, in paranthesis. In the Schwartz nomenclature the most conserved residue in each helix had been given a generic number according to its position in the helix. In the Ballesteros-Weinstein nomenclature the most conserved residue in each helix had been given the number 50. A series of conserved tryptophans and prolines are indicated in blue. The kink that may be caused by ProVI:15 has been suggested to be critical for the conformational changes involved in receptor activation [37]. The almost invariable disulfide bridge between extracellular loops 2 and 3 and the conserved palmitoylation site in the C-terminal tail are indicated in white.

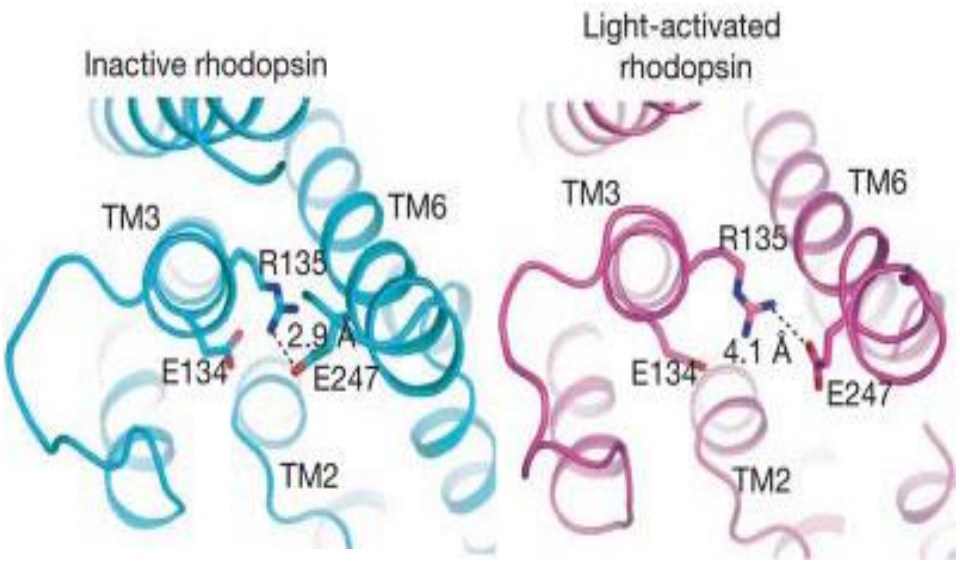
Sequence comparison between the different GPCRs revealed the existence of different receptor families sharing no sequence similarity. However, all these receptors have in common a central core domain constituted of seven transmembrane helices (TM-I through -VII) connected by three intracellular (i1, i2 and i3) and three extracellular (e1, e2 and e3) loops (Baldwin, 1993). Two cysteine residues (one in e1 and one in e2) which are conserved in most GPCRs, form a disulfide link which is probably important for the packing and for the stabilization of a restricted number of conformations of these seven TMs. Aside from sequence variations, GPCRs differ in the length and function of their N-terminal extracellular domain, their C-terminal intracellular domain and their intracellular loops. Each of these domains provides specific properties to these various receptor proteins, as discussed below.

The seven TM (7TM) regions constitute the core domain of these receptors, and a change in conformation of this domain is probably responsible for receptor activation. Mutagenesis and biochemical analysis with model GPCRs like rhodopsin revealed that the switch from the inactive to the active conformation is associated to a change in the relative orientation of TM-III and TMVI (with a rotation of TM-VI and a separation from TMIII), which unmasks G protein-binding sites (Farrens *et al*., 1996; Bourne, 1997; Javitch *et al*., 1997). In family A GPCR, one residue (Asp) in TM-II and a tripeptide (DRY or ERW) at the interface of TM-III and i2 are important for receptor activation (Oliviera *et al*., 1994; Scheer *et al*., 1996). Since these residues are not conserved in the other GPCR families, one may conclude that either the change in conformation of the core domain, or the molecular events leading to these changes are not conserved between members of these different families.

The change in conformation of the core domain generally affects the conformation of the i2 and i3 intracellular loop (which are directly linked to TM-III and TM-VI, respectively; see Figure 1B) that constitutes one of the key sites for G-protein recognition and activation (Spengler *et al*., 1993; Pin and Bockaert, 1995; Wess, 1997). In GnRH receptors, the i1 loop is essential for activation of Gs but not Gq (Arora *et al*., 1998). This perfectly illustrates the diversity of the solutions which have been selected during evolution to assume a good coupling to G proteins. It has been proposed that the C-terminal end of the G protein a-subunit binds in a pocket constituted by these intracellular loops in the various GPCR families

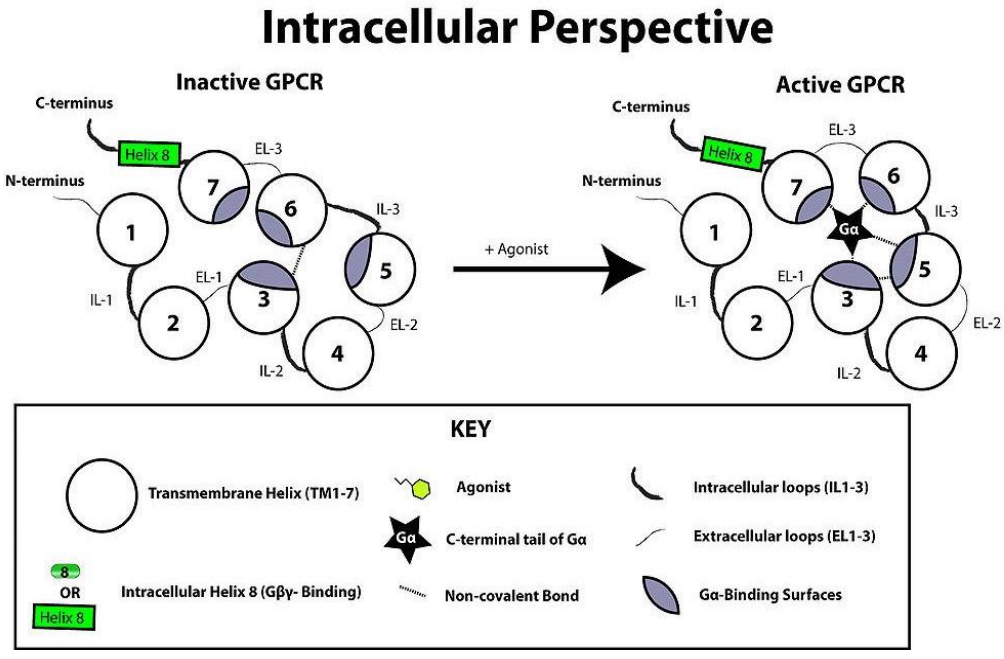
2.2.9 Comparison of active and inactive states

One of the big questions in GPCR function is how does receptor activation and inactivation occur? There have been many postulates over the years based on molecular modeling, biophysical data, and mutagenesis studies. The three-dimensional structures solved during the past year have contributed enormously to advancing our understanding of the molecular mechanisms involved in receptor activation. Figure 2.11 demonstrates what changes occur in the rhodopsin activation process. Here, opsin reflects the active state of rhodopsin; active state opsin bound to the C-terminal peptide of the G protein (transducin) ; and inactive rhodopsin. TM6, which is predicted to move, does so around 3 or 4Å in the active opsin crystal structures. This movement allows access for the G protein transducin to interact with TM6 and TM5 and induction of an α-helical-type structure as it associates with TM5, TM6, and TM7.



Courtesy: S. G. Rasmussen *et al, 2007*

Figure 2.11: Comparison of inactive and active rhodopsin conformations Presence of the ionic lock between Arg135 and the adjacent Glu134 in TM3 and Glu247 in TM6. In the light-activated structure the Arg135/Glu134 ionic bond is broken, and the distance of Glu247 from Arg135 has increased.



Courtesy: Repapetilto *et al*, 2010

Figure2.12: GPCR Conformational changes in Activation: Ligand binding disrupts an ionic lock between the E/DRY motif of TM-3 and acidic residues of TM6. As a result the GPCR reorganizes to allow activation of G-alpha proteins. The side perspective is a view from above and to the side of the GPCR as it is set in the plasma membrane The intracellular perspective shows the view looking up at the plasma membrane from inside the cell.

These molecular changes are brought about by the breaking of the ionic lock in Rhodopsin comprising ionic interactions in the conserved E/DR motif between Arg135 and the adjacent Glu134, and with a second Glu247 in TM VI (38). The recent crystal structures of active opsin and inactive rhodopsin clearly show that activation is accompanied by the disruption of the Arg135/Glu134 ionic bond (Fig. 2.10). This results in the formation of a new interaction of Arg135 with Tyr223 in TM5 (38).

2.3 CCR5 (CC-motif Chemokine Receptor Type 5):

CC-chemokine receptor type 5, also known as CCR5 or CD195, is a protein of 352 aminoacids in length and belongs to chemokines subfamily of class A GPCRs. It is predominantly expressed on T cells, macrophages, dendritic cells and microglia. It is likely that CCR5 plays a role in inflammatory responses to infection, though its exact role in normal immune function is unclear. It acts as a receptor for chemokines, which are most important regulators of leukocyte trafficking. The typical cellular response to stimulation of chemokine receptors is chemotaxis up or down a concentration gradient of the chemokine. The natural chemokine ligands that bind to this receptor (reviewed by Mueller & Strange, 2004) are RANTES (CCL5) and macrophage inflammatory proteins - MIP-1α (CCL3) and MIP-1β (CCL4).

2.3.1 Structural Characteristics

The C-C motif chemokine receptors type 5 is a polypeptide of 352 amino acids and belongs to chemokine subfamily of GPCR family A (GPCRDB). Although the 3D structure from any experimental studies is not present, a close member that belongs to same chemokine receptors subfamily - CXC4[40] structure was present that can be used for the structure prediction. It share approximately 32% amino acid identity, with CXCR4 although this is reduced to 20% when only the amino acids on the extracellular surface are. Both CCR5 and CXCR4 contain 352 amino acids and have three extracellular loops on the surface of the cell. The extracellular domain contains four cysteine residues (Figures 2.13), which form disulfide bonds between the Cys20 in N terminus and Cys269 on ECL3 and between the Cys101 on

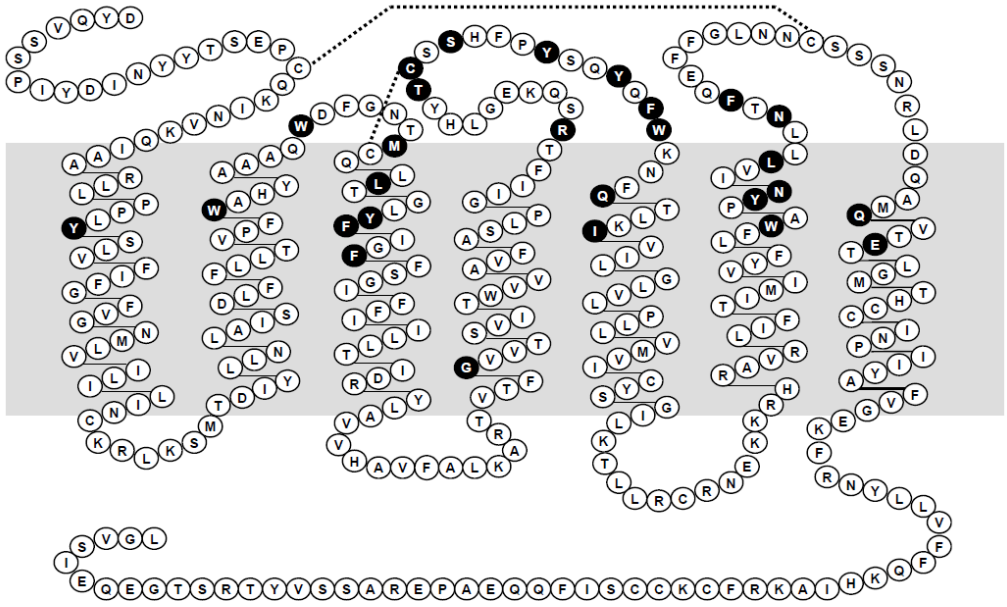


Figure 2.13: Snake plot of human CCR5 highliting the residues important for the binding of natural ligands and the antagonists and disulphide bonds between C20-C269 and C101-C178.

ECL1 and Cys178 on ECL2 (Berson *et al*., 1996). It is likely that mutations in one extracellular domain may indirectly affect the function of neighboring regions and that residues in all four extracellular regions can influence coreceptor usage (Horuk *et al*., 1994).

Most understandings about CCR5 are from the mutagenesis studies performed from the perspective of treating HIV i.e., studies are performed based on the loss/gain of the receptor binding affinity to the gp-120 protein(HIV protein that interacts with CCR5) and CCL3 (natural ligand). Mutagenesis studies by Javier Garcia-Perez *et al* [39] showed mutations of residues W86A in TM2, Y108A in TM3, Q194H in ECL2-TM5, I198A in TM5, Y251A/F/I in TM6, and E283A/Q has shown a considerable change in the binding affinity and the activity of the Maraviroc (MVC) and natural ligand CCL3, giving an insight that there residues are important for binding of the CCL3 and MVC. Furthermore, mutagenesis studies performed by Kenji Maeda et al,[40] has showed that Y37A in TM1, P84H in TM2, C101A in TM3, Y108A in TM3, F109A in TM3, G163R in TM4, C178A in ECL2, S180A in ECL2, K191N ECL2-TM5, T195P/S in TM5, K197A in TM5, I198A in TM5, W248A in TM6, Y251A in TM6, E283A in TM7 and M287E in TM7 considerably rescued the binding affinity of the antagonists. Of these mutations Y37A, C101A, Y108A, F109A, G163R, C178A T195P/S, W248A, Y251A, M287E drastically rescued the binding affinity (K_d) by more than fivefold. They also demonstrated that amino acid residues in the beta-hairpin structural motif of ECL2 are critical for HIV-1-elicited fusion and binding of the

spirodiketopiperazine-based inhibitors to CCR5. The direct ECL2-engaging property of the inhibitors likely produces an ECL2 conformation, which HIV-1 gp120 cannot bind to, but also prohibits HIV-1 from utilizing the inhibitor-bound CCR5 for cellular entry which is a mechanism of HIV-1's resistance to CCR5 inhibitors.

2.3.2 CCR5 as therapeutic target:

As the CCR5 is a receptor for the chemokines that actively regulate the various immune responses and regular cell signals , it is found to be therapeutically important in various diseases like HIV, prostate cancer, multiple sclerosis, rheumatoid arthritis (RA), psoriasis and chronic hepatitis, along with age-related diseases like atherosclerosis, Alzheimer's. The thereuptic importance of CCR5 first began with observation that Natural ligands of CCR5 can inhibit HIV-1 entry (Cocchi *et al*., 1995) that led to the identification of CCR5 as a co-receptor for HIV-1 (Alkhatib *et al*., 1996). Mutational analyses and the inhibition of chemokine and gp120 binding to CCR5 by monoclonal antibodies have identified ECL2 and a tyrosine-rich region within the N terminus of CCR5 as the major components involved in interactions with gp120 during HIV-1 entry (Dragic *et al*., 1998; Farzan *et al*., 1998; Farzan *et al*., 1999; Farzan *et al*., 2000; Farzan *et al*., 2002; Lee *et al*., 1999; Olson *et al*., 1999).

It also serves as a CD4 co-receptor for R5-tropic Human Immunodeficiency Virus type 1 (HIV-1) entry into activated memory $CD4^+$ T-lymphocytes and monocyte-derived macrophages. Recent investigations on the role of the CCL5/CCR5 in tumor growth and metastasis and prostate cancer have shown potential therapeutic importance. It is also implicated in the pathogenesis of many immune and inflammatory conditions, including multiple sclerosis, rheumatoid arthritis (RA), psoriasis and chronic hepatitis ,along with age-related diseases like atherosclerosis, Alzheimer's and control of parasite replication and acute cardiac inflammation following Infection with Trypanosoma cruzi .

CCR5 blockade along with rapamycin are effective in modulating transplant immunity and led to prolonged allograft survival in mice [41]. Expression of the inflammatory chemokine CCL5 (RANTES) by tumor cells is thought to correlate with the progression of several cancers. CCL5 was shown to induce breast cancer cell migration, mediated by the receptor CCR5. A CCR5 antagonist was demonstrated to inhibit experimental breast tumor growth [42]. Work done by Vaday et al [43] suggest that

inflammatory chemokines, such as CCL5, expressed by prostate cells may act directly on the growth and survival of Prostate Cancer (PCa) cells. Chemokine receptor antagonists may thus block autocrine mechanisms of PCa progression.

CCR5 is prominently expressed during both acute and chronic disease, suggesting a role in regulating leukocyte trafficking and accumulation within the heart following *T. cruzi* infection. Work by Jenny L Hardison[44] demonstrated that CCR5 contributes to the control of parasite replication and the development of a protective immune response during acute infection but does not ultimately participate in maintaining a chronic inflammatory response within the heart. Although CCR5 is thought to be implicated in the Rheumatoid Arthritis a study by van Kuijk et al [45] does not support the use of CCR5 blockade as a therapeutic strategy in patients with active RA. Placebo controlled trials by Marjan de Groot et al, [46] showed that CCR5 does not play a crucial role in the pathogenesis of psoriasis.

CCR5 is also expressed in the brain, where it can be crucial in determining the outcome in response to different conditions. CCR5 expression can be deleterious or protective in controlling the progression of certain infections in the CNS, but it is also emerging that it could play a role in non-infectious diseases. In particular, it appears that, in addition to modulating immune responses it has possible involvement in neuroprotective mechanisms [47]. In Alzheimer's disease (AD), there is an increase in presence of monocytes/macrophages and activated microglial cells in the brain [48]. Immunohistochemical studies show increased expression of chemokine receptor 5 (CCR5) on reactive microglia associated with amyloid deposits in AD, suggesting that CCR5 may play a role in the regulation of the immune response in AD. The secretion of chemokines is enhanced in diabetes and there by CCR5 on monocytes receives the signal and enhances the recruitment to renal tissues and differentiation to macrophages. This assembly of macrophages in the glomerulus may play a key role in development or progression diabetic nephropathy. Blockade of chemokines and chemokine receptors in kidney diseases, with a special focus on the therapeutic potential and possible adverse effects of anti-chemokine strategies in renal inflammation are reviewed by *Vielhauer Vet al.[49]*

CCR5 is An effective immune response to hepatitis C virus (HCV) infection requires efficient recruitment and activation of inflammatory cells to the liver, the site of infection. In particular, the interaction between chemokines CCL3 (MIP-1α), CCL4 (MIP-1β) and CCL5 (RANTES) and their receptor, In case of the hepatitis C virus (HCV), CCR5 may be critical

in regulating T cell functions by mediating recruitment, polarization, activation and differentiation of antiviral type 1 cytokine secreting T-helper and cytotoxic T-cells [50]. This is the therapeutic application where we target the disease by enhance the activity of CCR5 (agonistic fashion).

Although there is a wide scope of targeting CCR5 for treating many disease, lack of complete information on the mechanism is a limiting factor for the studies. It is successful to some extent in case of HIV as entry inhibitor target. Recent works has been started that are focusing on prostate cancer. Agonist induced internalization which is a pseudo form of receptor inhibition limits the knowledge of antagonistic and agonistic roles in signal controlling.

2.3.3 Current Antagonists and Agonists

Lot of research in CCR5 designing antagonists are done focusing on treating the HIV. Earlier efforts to block the gp120 and CCR5 interaction employed chemical modifications of the known CCR5 ligands, especially RANTES (Regulated on Activation, Normal T-Cells Expressed and Secreted). These compounds suffered, however, from poor oral bioavailability and the induction of intracellular signaling cascades upon CCR5 binding [51]. The most promising drugs in this class, however, are nonpeptidic, orally bioavailable, small-molecule antagonists of CCR5. These small molecules are noncompetitive allosteric antagonists of CCR5. As allosteric inhibitors, these antagonists do not have to induce conformational changes in CCR5 per se, but may alter the binding affinity of gp120 by perturbing the equilibrium between different CCR5 conformations that are naturally present. The CCR5 antagonists have been given generic names with the suffix ***"-viroc,"*** an abbreviation for "viral receptor occupancy."

Compound	**Status**
Maraviroc	Approved
Vicriviroc (SCH-D)	Phase III
INCB 9471	Phase IIa
SCH-C	Phase I
TAK-652	Phase I
Aplaviroc	Discontinued from Phase I (Liver Toxicity)

AD101	Preclinical only
TAK-779	Preclinical only
TBR-652	Phase IIb
INCB009471	phase IIb
SP-01A	phase IIa
Anibamine	Preclinical only

Table 2.1: Current antagonists status [52,53]

Most of the molecules that went through clinical trials showed some counter effects like liver toxicity or counter effects with the other diseases or lack oral bioavailability (incase of aplaviroc) [54] . Among all the inhibitors of CCR5, Anibamine is the first and the only natural product that has been identified as a CCR5 antagonist with high affinity. Clearly, anibamine possesses a novel structural skeleton compared with all other known CCR5 antagonists.

Currently, Maraviroc originally designated UK-427857 is the only nonpeptidic, small molecule human immunodeficiency virus type 1 (HIV-1) entry inhibitor that has received full US FDA approval on August 6, 2007 for use in treatment experienced patients infected with only R5-viruses [52]. The agonists and antagonists in current strategy are:

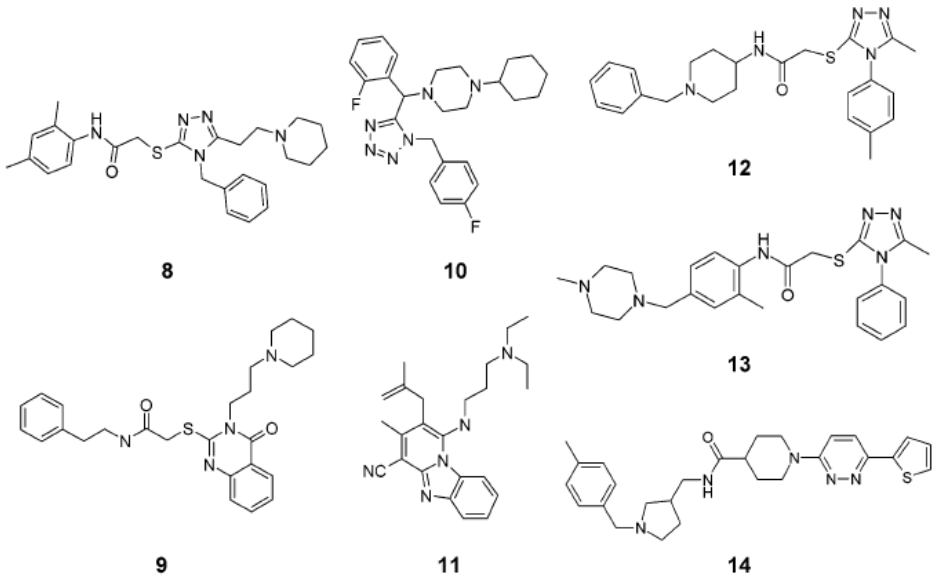


Figure 2.14: Structures of known agonists of CCR5

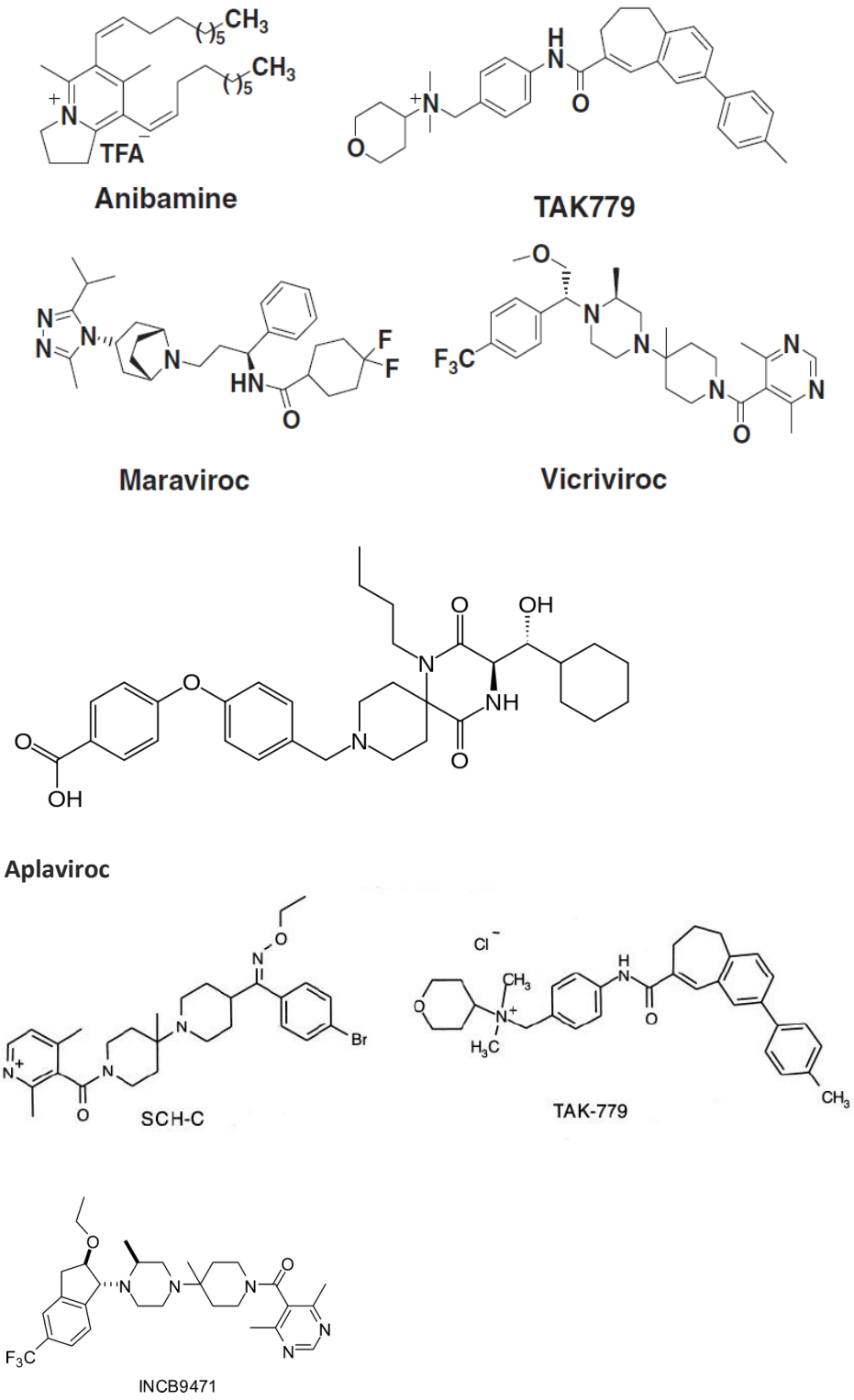


Figure 2.15: Structures of antagonists

Identifying GPCR agonists by structure-based *in silico* screening is more challenging, because GPCRs are expected to undergo significant conformational changes upon activation. By searching for antagonists of the chemokine CCR5 receptor as an attempt to derive anti-

HIV agents, Garcia-Perez et al., [53] reported unexpected identification of nonpeptide CCR5 agonists (Figure 2.14). Although handful of antagonists are currently in clinical trials, each have their own negative points, that made researchers to come up with more potent molecules with good efficiency. The current antagonists in clinical trials are shown in the Figure 2.15.

Computational Methods

3.1 Homology Modeling:

With the recent efforts in the crystallization of these proteins, homology modeling approaches are becoming widely used as a method for obtaining quantitative and qualitative information for structure-based drug design as well as the interpretation of experimental data [54]. Recently determined GPCR structures provide opportunities for advancements in GPCR modeling. Four important conclusions are given in the review on GPCR homology modeling by *Niv MY et al.*,[55] :

(i) Multi-template models may produce better structures than single-template models, although inferior models may also be generated by multi-template approaches, warranting the development and application of improved model assessment methods;

(ii) Cautious incorporation of knowledge-based constraints can improve the quality of models and docking;

(iii) Molecular dynamics simulations account for structural features not observed in X-ray structures and may refine docking poses.

(iv) While progress in de novo methods for long loop prediction is ongoing, loopless models provide a practical alternative for docking and virtual screening applications.

3.1.1 Backbone Modeling with Modeller

Homology modeling was proved fruitful for CCR5 in case of many antagonists that were reported till now. This gives confidence in using the homology model for determining the 3D structure of the CCR5. Homology model was built with Modeller [56] which is one of the most successful tools for comparative modeling. The amino acid sequence (FASTA sequence) Of the Human CCR5 protein was taken from the UniProtKB database [57], accession no. P51681. Sequence homology searches were carried out using BLAST [58] web tool (Basic Local Alignment Search Tool) against protein data bank [59] (PDB). From the given sequence homologous proteins CXCR4 (PDBid: 3ODU) was selected as the template which has the sequence identity of 34% and E-value of 7e-37. CXCR4 is selected as template for 2 reasons: 1) Maximum identity among the known structures and it belongs to the same sub family chemokine receptors. Though CXCR4 has the maximum identity, the N-terminus and C-terminus are lacking in the crystal structure of CXCR4, for this reason we have used the N-terminus and C-terminus of the Bovine Rhodopsin (PDBid: 1F88) for modeling the complete structure through the complete Transmembranes(TMs) are built only based on the

CXCR4. We have also included the NMR determined structure of N-terminus of CCR5 (PDBid: 2L87) determined by Anglister *J et al.*,[60]. The overall model was built with the templates 3ODU, 2L87 and 1F88, where 3ODU is used to model the TMs only and N-terminus of 1F88 and 2L87 were used to model the N-terminus region of the CCR5 and C-terminus of 1F88 was used to model the C-terminus of the CCR5.

Modeller requires 2 to files build homology model. They are:

1. Alignment file (*.ali)
2. Script file that contains the instructions; what to perform (*.py)

Multiple sequence alignment was done in Multiple Sequence Viewer modules [61] in Schrodinger-2011 Suite keeping the conserved residues aligned and obtain maximum identity. The alignment used for the model is shown in the Figure (3.1). All the conserved residues in the Class A are constrained to be aligned like residue N in TM1, DRY motif in ICL2. The complete alignment is shown in the Figure 3.2.

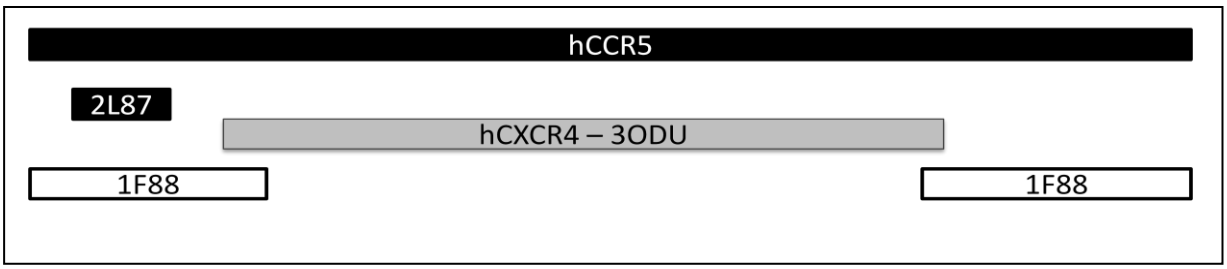


Figure 3.1: Picture showing the overall alignment used for modeling the final model

Coming to the script used for the model generation, first we have used two classes for model generation one at a time: 1) ***automodel()*** - used for modeling without any loop refinement, 2) ***autoloop()*** - which performs loop refinement automatically along with normal *automodel* modeling. The *autoloop* class for the model generation did not gave expected results, whereas *automodel* class gave good results with disulphide bonds between C101(TM2) and C178(ECL2) without giving any constraints of disulphide bond. With this satisfactory results, from the automodel class, we have added additional constraint of disulphide bond between C20(N-terminus) and C269(ECL3). This modified *automodel* method gave good result with expected 2 disulphide bonds.

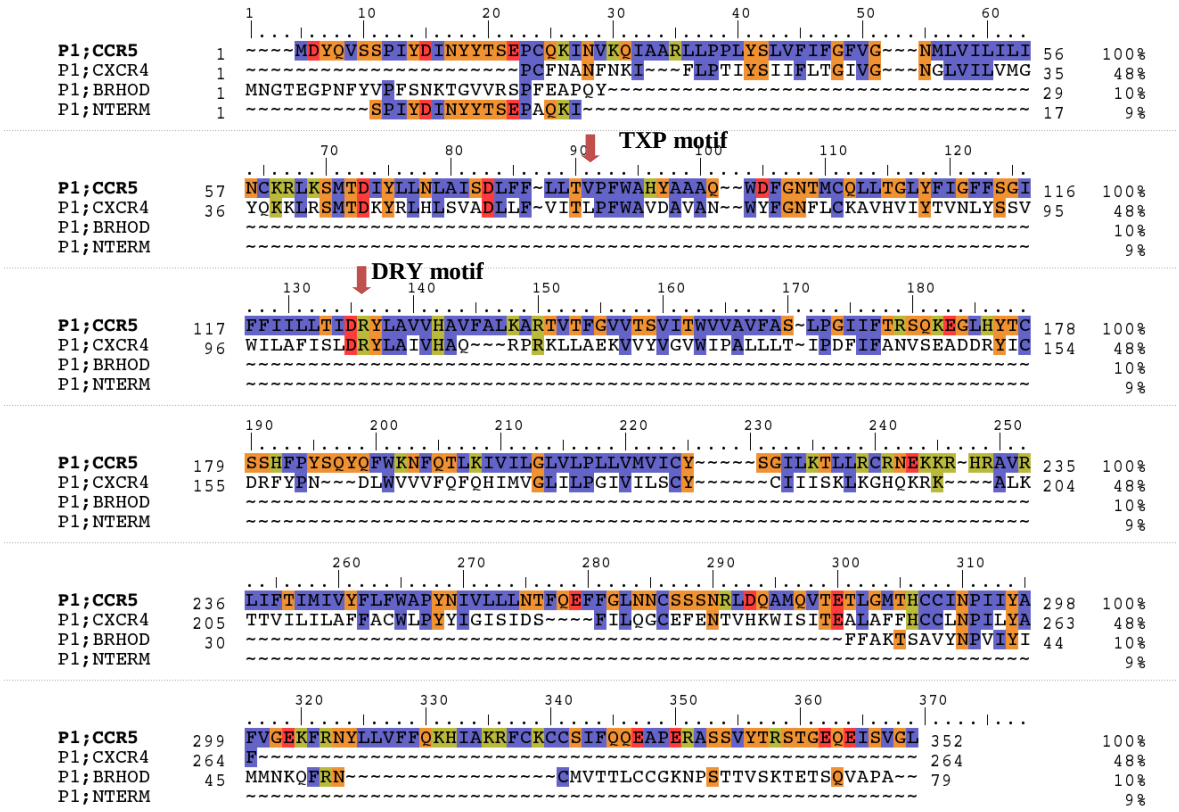


Figure 3.2: Sequence alignment used for modeling

In the automodel class model assessing methods DOPE (Discrete Optimized Protein Energy) score [62] and GA341 score [63] were included to access the final models. GA341 is is sufficiently sensitive to distinguish between two 'very good' models. Any good model will give a score very close to 1 (e.g. running GA341 on most PDB structures will give a 1.0 score). It is used only to discard 'bad' models (e.g. those with score less than 0.6). GA341 score above 0.7 indicate a correct fold with more than 35% atoms super imposable within 3.5A of their correct positions (Practical bioinformatics by Janusz M. Bujnick, pp: 86). As all the models shows 0.998 to 1 GA341 score, DOPE may be more sensitive in determining the good score based on the least possible score. DOPE scoring is based on an improved reference state that corresponds to non-interacting atoms in a homogeneous sphere with the radius dependent on a sample native structure; it thus accounts for the finite and spherical shape of the native structures. The normalized DOPE scores near to 0 and negative(-1) are considered to be native state structures.

A total of 1000 models were generated, which were filtered to 50 models based on the DOPE score. The models with least normalized DOPE score and G3A1 score nearly 1 are generally considered to be the best models, we observed there is not much variation in the

DOPE score of the models, so we have given a cutoff values of -42,000 for DOPE score and 0.60 for normalized dope score. After the cutoff, we have manually selected the filtered models for the β-hairpin structure in the ECL2 and disulphide bonds which are the characteristic features of the chemokine family.

Filename	molpdf	DOPE_score	GA341_score	Normalized_DOPE
n-n-c.B99990958.pdb	1933.65	-43186.57	1.00	0.50
n-n-c.B99990985.pdb	1962.70	-43137.79	1.00	0.51
n-n-c.B99990962.pdb	2004.21	-43103.41	1.00	0.52
n-n-c.B99990987.pdb	1905.43	-43103.09	1.00	0.52
n-n-c.B99990955.pdb	1897.22	-43045.96	1.00	0.53
n-n-c.B99990937.pdb	1972.97	-43031.00	1.00	0.53

Table 3.1 Table showing the DOPE, G3A1 and Normalized DOPE score of the top 5 molecules generated by the Modeller from the 1000 models

3.1.2 Loop optimization with Modeller

After the manual selection we have checked the selected models with the PROCHECK [64] server. Two models with most favorable stereochemistry from the Procheck results are taken. All the unfavorable residues are present in the loop regions, so we have subjected the models to loop optimization in modeler using the *MyLoopModel()* class. The models were the rechecked with the PROCHECK and loop optimization until the model stereochemistry was good enough. The best model among the optimized models was selected as the final model. Then, the model was prepared for further studies with the Protein preparation wizard in Schrodinger Suite[65] by adding the Hydrogens(modeller doesn't give a model with H) and fixing the side chain amides and bumps, and minimizing the protein with Conjugate Gradient method until a gradient of 0.05 Kcal/(mol*A) is reached . The results of the PROCHECK are given in the Section 4.1.

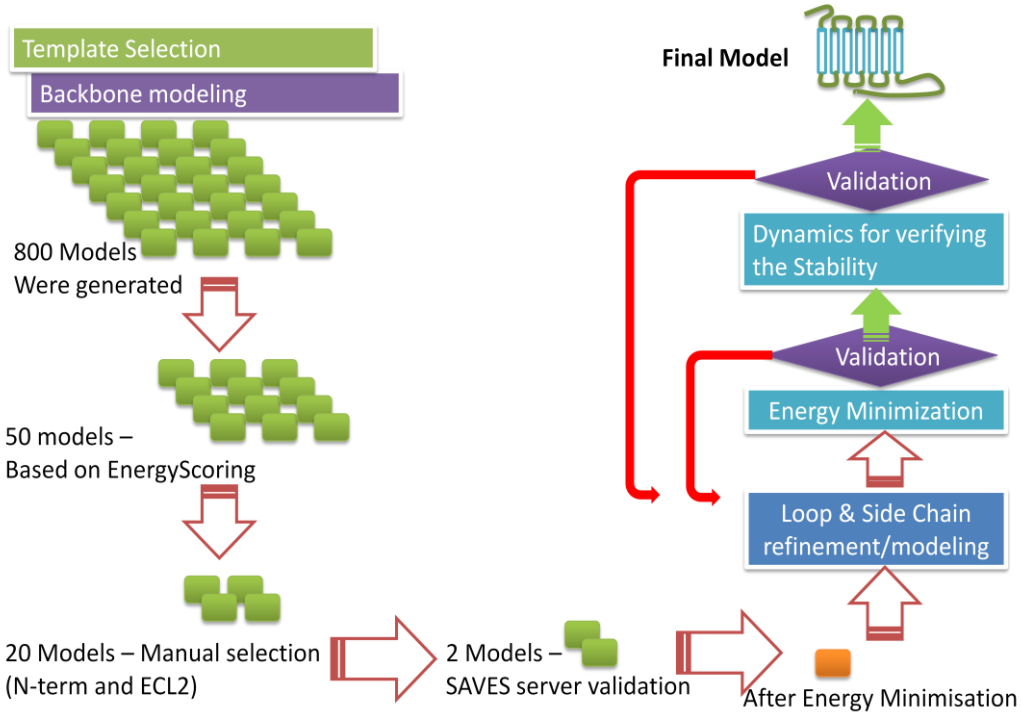


Figure 3.3: Protocol used to build the model

3.1.3 Molecular Dynamics with Gromacs

The final model was then subjected to Molecular Dynamics (MD) simulation of 20 ns in vacuum to check the stability and integrity of the TMs and loops of the model. MD was performed with the GROMACS [66].

The steps involved in simulation with GROMACS are as follows:

1. Generate topologies for the Protein

2. Define a box to simulate the box with in (solvate if necessary-optional).

3. Minimize the energy. (Optional)

4 Equilibration (NVT/NPT based on your choice)

5 Production MD simulation

Further information on GROMACS is provided in the Chapter 4.

3.1.3.1. Generate topologies for the Protein

Gromacs provides topology generation feature with the tool ***g_pdb2gmx.*** It uses a simple pdb file to generate the topology of the protein and *.gro (gromacs format of structure information).

```
#g_pdb2gmx -f protein.pdb -o protein.gro –water none -ignh
```

This will prompt for the force field you want to use for topology generation, I have chosen GROMOS96 43A1 force field for topology generation.

This will generate a toplogy file topol.top which cantains the toplogy information of the protein.

3.1.3.2 Define a box for simulate.

Since we are performing a gas phase simulation, we don't need to solvate the box. All we need to do is define a box to simulate the protein. Gromacs provide ***g_editconf*** to define the box.

```
#g_editconf –f protein.gro –o box.gro –c –d 1.0 –bt cubic
```

The flags –c is used to center the molecule in box, and –bt represents the shape of the box and –d tells the tool to keep the box edge at least 1 nm from the protein, to make sure protein is in the box in all axis.

3.1.3.3 Minimize the energy.

Minimization of the protein was generally done to minimize the energy of the system with the current force fields being used. Energy minimization of the system was performed with 100000 steps of steepest descent minimization with a step size of 1fs, till it reached the maximum force of 1KJ/mol/nm. The script (minim.mdp) file used is given in the appendices. We use two tools provided by the gromacs for this ***g_grommp*** and ***g_mdrun****;* g_grommp prepares the files required for the simulation where as g_mdrun do the simulation.

```
#g_grommp  -f minim.mdp -c box.gro –p topol.top –o em.tpr
```

Once the files required for the simulation are prepared ie., em.tpr and mdout.mdp (final script) we can run the minimization with g_mdrun. Script used is given in Script2 in Appendices.

#g_mdrun –v –deffnm em

Flag –v is used for verbose messaging and deffnm is used to name the default name em to all output files.

3.1.3.4 Production MD

The final production run was done with the script (md.mdp with all standard conditions. Since there are no boundaries and the pressure is not applicable as the system is not limited to any boundary. Since the pbc is not used and no positional constraints are being used, the forces used for the simulation may convert the potential/kinetic energy into the rotational energy, so we used **comm-mode**(comm-mode=angular) in the script to remove the center of mass motion removal in all axes allowing the protein to vibrate but not rotate around its center of mass which causes the unfolding of the protein. We have disabled the neighbor list updating frequency since there is only one entity (protein) in the system and we don't need to update the neighbor list (nstlist=0). The overall temperature of the protein system was kept constant at 300 K with a Berendsen thermostat [67]. All the outputs like trajectory, energy were calculated for every 2ps. The integration of the equations of motion was performed by using a leap frog algorithm (md integrator) with a time step of 2 fs till 20ns (100000000 steps). A cutoff of 1 nm was implemented for the Lennard–Jones and Coulombic interactions. All the covalent bonds were constrained with the LINCS algorithm [68]. Each nanosecond of simulation required approximately three 2.5 CPU hrs on Intel Xeon CPU (2.4 GHz processor).

#g_grommp -f md.mdp -c em.gro –p topol.top –o md.tpr

#g_mdrun –v –deffnm md –nt 4

Flag –nt is used to describe number of threads of CPU it can run on. Since pbc is absent and nstlist is 0, complete domain decomposition is not possible and hence we need to manually give the number of domains the work can be split into, in my case its 4 based on the lincs-order variable being 4. Script used is given in Script2 in Appendices.

3.1.3.5 Post MD Analysis:

Three graphs were plotted against the 1) Total Energy (KJ/mol) vs. Time (ps), 2) Root Mean Square Deviation (RMSD) of the protein (nm) vs. time (ps) and 3) Root Mean Square Fluctuations(RMSF) vs. atoms and 4) Energy gradient vs time.

Graphs 1 and 2 from above list shows the region on the simulation which are stable in the conformation based on the energy of the system (protein) and the deviation from the initial structure. The RMSD graph when comes in parallel to the axis states that the all the consecutive structures are having same spatial arrangement with minute vibrations ie., the deviation from the initial structure, this gives a possibility that the atoms in the structure have attained a stable positions where there is not much deviation among the consecutive time steps. The energy graph also when reached a stable parallel line to the X-axis states that there is no much variations in the energy of the system. The graphs were given in the Chapter 1V, Figure. The graphs shown a established energies after 11ns and hence we have taken an average structure of the 11ns to 16ns which is the most stable region of energies. The average model was taken as the pre-final model, and subject to the protein preparation module in Sybyl. In this step, the side chain amides and side chain bumps are treated and short minimization of 1000 iteration to minimize the energy increased due to the side chain treatment .Thus a final model was obtained for the further studies.

3.2 Validation of Model with Docking Studies

The final validation of the model was done in two steps:

1. PROCHECK
2. Docking with reported Agonists and Antagonists

The model thus obtained was validated with PROCHECK server, 87% of residues are in most favored, 13% are in additionally favored and 0.3% are in disallowed regions , and interestingly this resides is a part of C-terminus and was ignored.

Further, we used the final model for the docking studies with 3 known Antagonists and Agonists to verify the ability of the model in predicting the binding modes. All the docking studies were done using Glide in Schrodinger Suite 2011 [69].

The model thus obtained was the prepared for further analysis using the protein preparation tool in Schrodinger Suite. A grid is defined with the centroid of the residues Y108, K191, Y251, and E283. All these residues were reported to be important for the binding of the antagonists and other natural ligands [70] .The receptor scaling factor of 0.8 was used that gives protein some flexibility unlike the rigid molecule and the all the receptor hydroxyl groups in the grid were allowed to rotate.

All the ligands were minimized with the Ligprep [71] module of the Schrodinger Suite. All the possible conformations at pH 7+/- 2 are generated for the docking studies. All the conformations generated for the antagonists and agonists are docked into the grid defined by the active site residues.

Antagonists and agonists used for model validation with docking studies are:

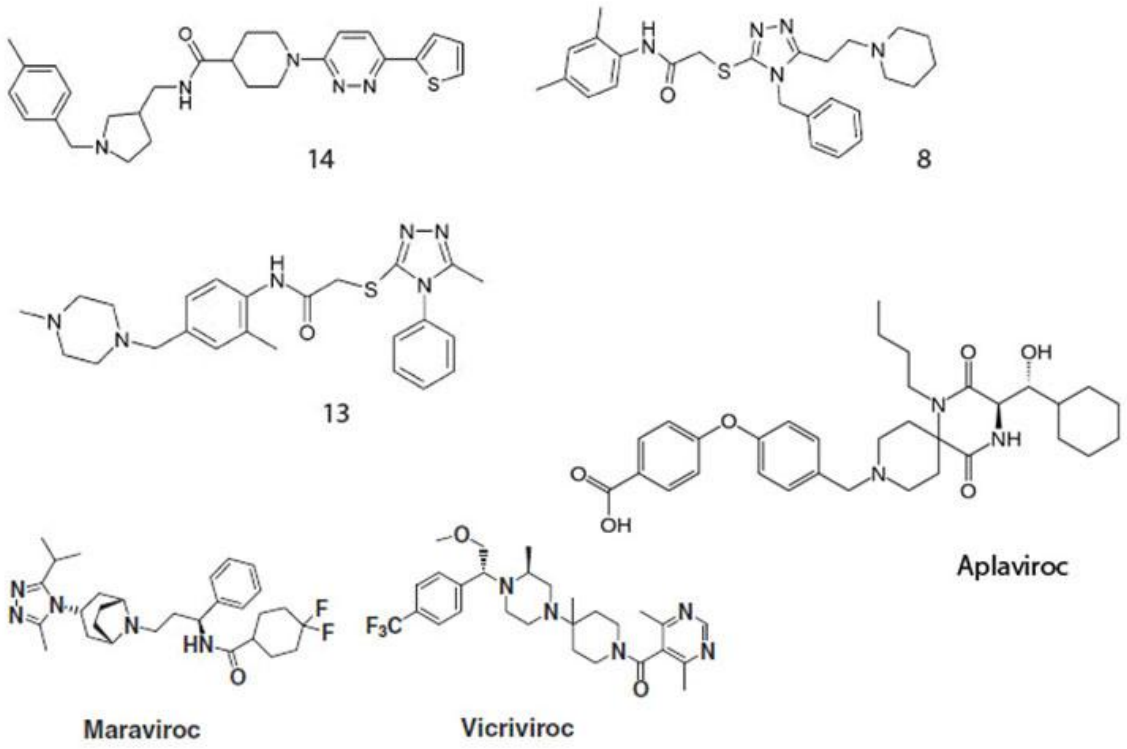


Figure: 3.4: Agonists (8, 13, 14) and Antagonists (Aplaviroc, vicriviroc, and maraviroc) used in the present study.

The prediction of the docked poses of the Antagonists and Agonists gave an assurance that this model was effective in predicting the binding modes of the newly designed molecules and can be used for further studies like virtual screening and other structure based approaches.

Results & Discussion

4.1 Homology Modeling

The model that was obtained by modeller was subjected to iterations of loop modeling and minimizations and validated with the PROCHECK in each step until a model with good score in PROCHECK was obtained. The relative positions of the Class A are conserved with the

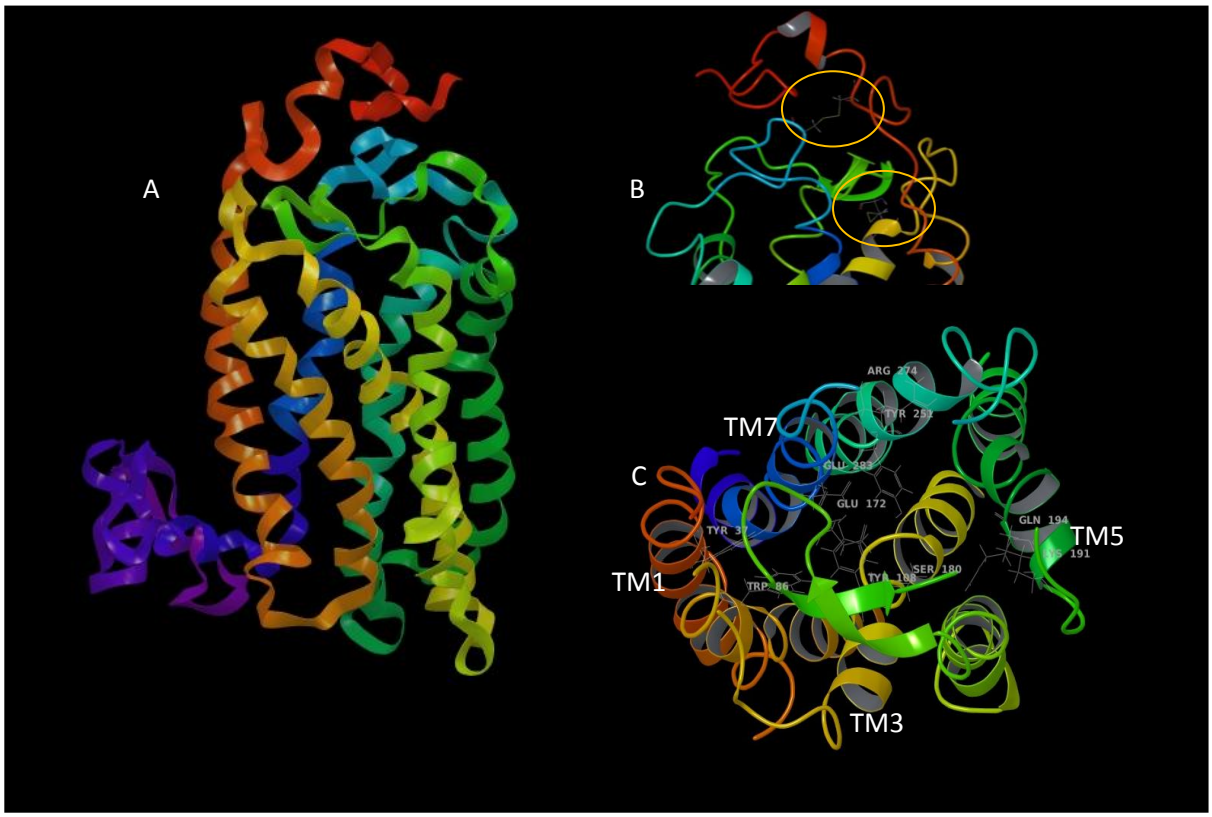


Figure 4.1: Structure of final CCR5 model. Red color represents N-terminus and Blue color represents the C-terminus. (A) Final structure of the CCR5 (B) Showing the disulphide bonds, and (C) Top view of the model showing the ECL2 and the arrangement of TMs along with positions of active site residues.

The model has got the expected hairpin loop in ECL2 and the expected disulphide bonds between C20-C269, C178-C101. The top view of the TMs shows a perfect architecture of the typical GPCR with TM3 going diagonally to the TM5. Active site identification by Sitemap analysis showed expected binding pocket in mid of the TMs with a volume of 1280A^3. All the sidechains of active site are exposed to the center of the cavity allowing the interactions with the incoming agonists and antagonists.

4.1.1Evaluation of model:

All the homology models(1000 models) generated were filtered to single structure with filters like DOPE score and the presence of the hairpin loop in the ECL2 with 2 ant parallel β sheets and presence of the conserved disulphide bonds between TM2-ECL2 and N-term-ECL3. The procedure of filtering is discussed in Chapter 3.1. The quality of final model

obtained from MD was evaluated by PROCHECK and RAMPAGE. The result of the RAMPAGE server is shown in the figures 4.2.

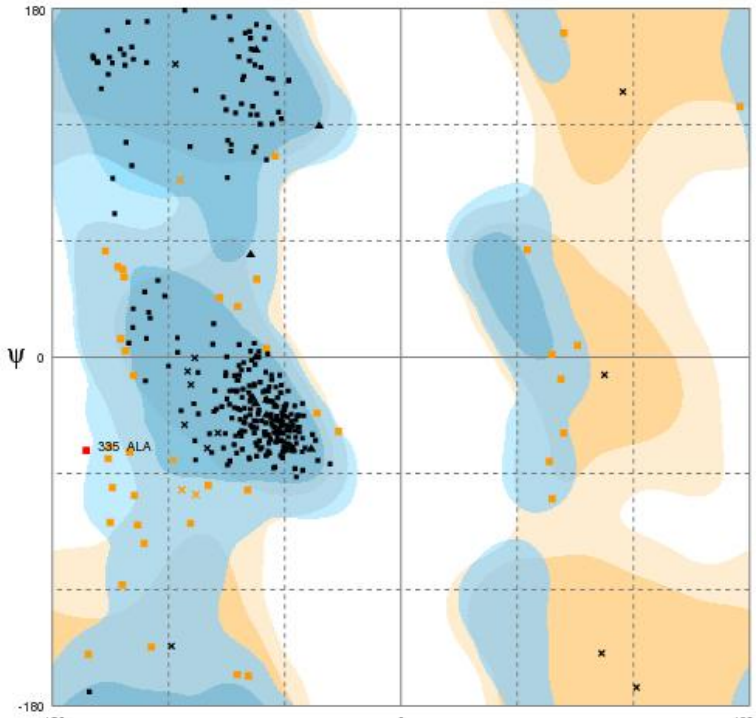


Figure 4.2: Result of RAMPAGE showing the positions of the residues in the Ramachandran plot

The PROCHECK analyses provide an idea of the stereo chemical quality of the protein structure. It highlights regions of the protein which appear to have unusual geometry and provide an overall assessment of the structure as a whole. RAMPAGE indicates the percentage of residues located in the favored regions of the Ramachandran plot. This plot gives the main chain conformations as pairs of ψ and φ dihedral angles for each residue in the protein.

Note	**Main-chain params: 6 better 0 inside 0 worse** [PostScript] • [PDF] • [JPG]
Note	**Side-chain params: 5 better 0 inside 0 worse** [PostScript] • [PDF] • [JPG]
Error	*** Residue properties: Max.deviation: 5.8 Bad contacts: 0 *** **Bond len/angle: 5.2 Morris et al class: 1 1 2 G-factors** **Dihedrals: -0.38 Covalent: -0.12 Overall: -0.25** [PostScript] • [PDF] **Images: 1 2 3 4**
Note	**G-factors Dihedrals: -0.38 Covalent: -0.12 Overall: -0.25** [PostScript] • [PDF] • [JPG]
Note	**M/c bond lengths:100.0% within limits 0.0% highlighted** [PostScript] • [PDF] **Images: 1 2**
Note	**M/c bond angles: 89.3% within limits 10.7% highlighted** [PostScript] • [PDF] • [JPG]
Warning	**+ Planar groups: 98.4% within limits 1.6% highlighted** [PostScript] • [PDF] • [JPG]

Figure 4.3: The **PROCHECK** results showing the most of the stereochemistry of the protein is error free.

The stereo-chemical parameters showed that more than 84.4% of the residues of the model has the (ψ, φ) dihedral angles in the favored regions of the Ramachandran plot as expected for a good model and among the rest 15.6% residues, 12.3 % are in allowed regions and 0.3% is in disallowed region. The results of PROCHECK are depicted in the figure 4.3 indicating no bad contacts and major steric clashes in the protein. The results shows most of the parameters are in the safe zone indicating the model a good one , and the error in the Residue property is due to the C terminus part of the model , and hence ignored.

4.1.2 Molecular Dynamics:

Molecular dynamics simulation of Protein in vacuum for the 20 ns was good in showing the stability of the protein and stable arrangement of TMs. In the simulation the protein attained a constant energy after 11ns with. The energy of the model was fluctuating around the 500KJ/mol and the RMSD of the protein atoms were stabilized after an RMSD of 0.25nm from the initial structure is reached. Thus, we can conclude that there are not many variation in the protein structure after the simulation of 11ns and the stability in the energy with constant fluctuations also confirms that there is no much variation in the energy of the system and hence has attained a point where the structure is stabilized in terms of RMSD of the atoms of protein and energy of the protein. We have taken the average structure of the 11ns to 16ns which seemed to be most stabilized region of the simulation. There is no much variation in the initial structure of the MD and final model after MD. There is almost very minute amount of variations one could expect as the RMSD of the protein from initial structure from the initial structure is 0.23 nm from the initial structure (Figure 4.5). The graphs obtained from the results of the GROMACS are shown in the Figures 4.4 and 4.5. There is an energy gradient of 500KJ/mol in the simulation from 11ns to 16ns.

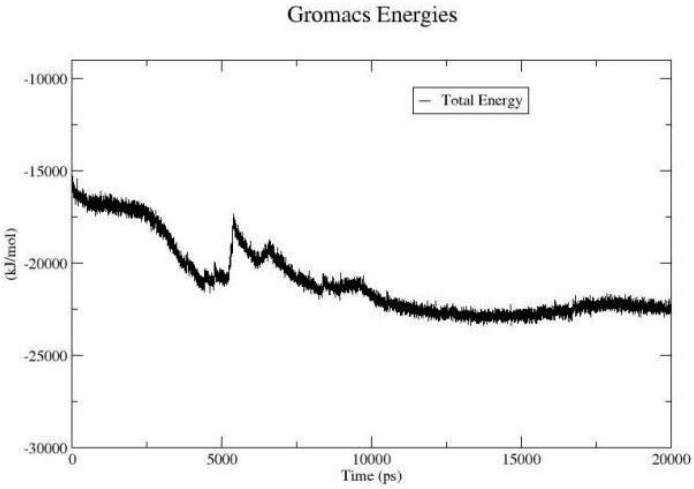


Figure 4.4: Energy vs. time graph showing the change in energy during the 20ns simulation.

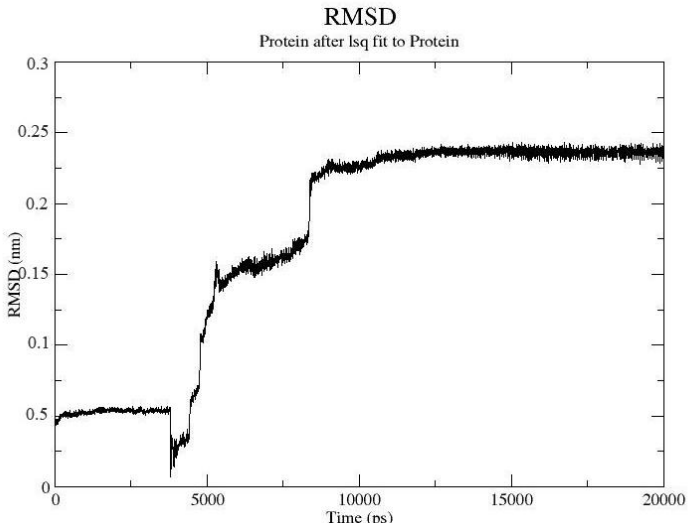


Figure 4.5: RMSD vs. time graph showing change in the structure during the simulation

4.2 Docking Studies:

The docking of agonists and antagonists are performed using Glide with SP (standard precision) method. All the agonists and antagonists are binding in the active site ie., in the center of the cavity formed by the TMs and has shown the expected. E283 which is expected to form ionic lock interactions are reproduced. There is good correlation between the docking scores and IC_{50} values of the molecules indicating a possibility that prediction with the model could be good. Since all the agonists used for the study are very large molecules with more than 2 rotatable bonds and there is huge degree of freedom for the conformations. Though the interacting residues shown in the table are with single residues,

these molecules are showing H bonding with the other residues around but keeping the residue given in the table constant.

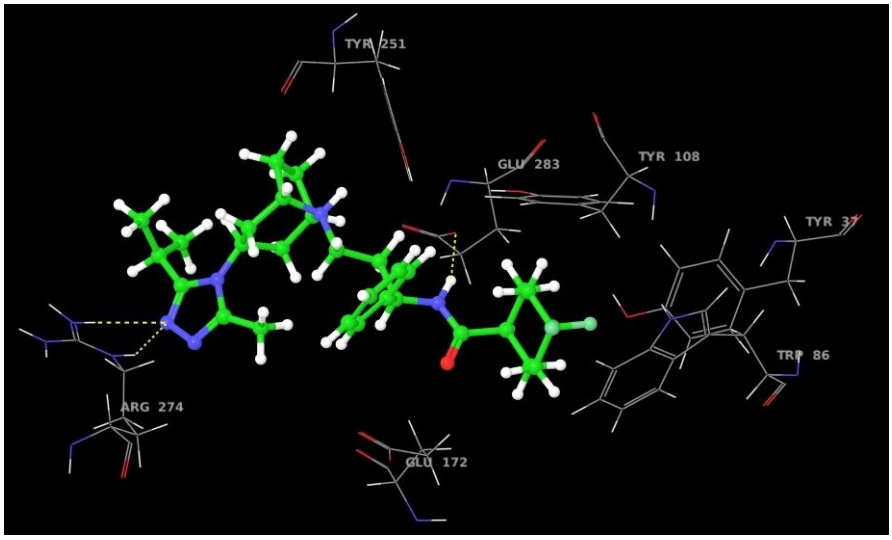
Name	Most interacting residues	Gscore	IC50
Aplaviroc	Trp 86,Tyr 108, Glu 283, Tyr 251	-7.32	0.28 nM
Viciviroc	Glu 283 , Arg274	-7.55	3.2 nM
Maroviroc	Ser 108, Glu 283 , Arg 274	-6.79	5.3 nM
Ag 14	Glu 283,Arg 274	-7.13	5.8 µM
Ag 13	Glu 283	-6.91	14.1 µM
Ag 8	Glu 283 ,Tyr 105	-6.27	17 µM

Table: 4.1: Docking score along with IC_{50} values

The Docking poses of the Agonists and Antagonists used for the study are:

Figue 4.6: Docked poses of the Agonists and Antagonists

Figure 4.6.a: Docked pose of Maraviroc showing interactions with R274 and E283



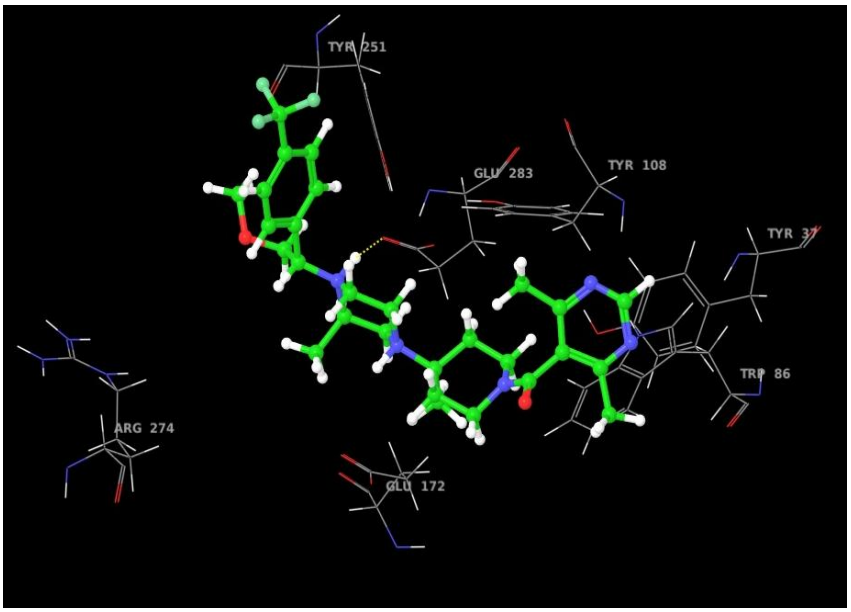


Figure 4.6.b: Docked pose of Vicriviroc showing interactions with E283 .

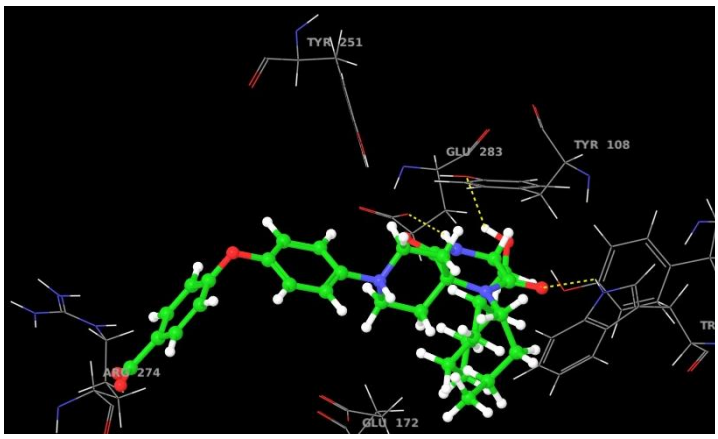


Figure 4.6.c: Docked pose of Aplaviroc: showing interactions with E283, W86

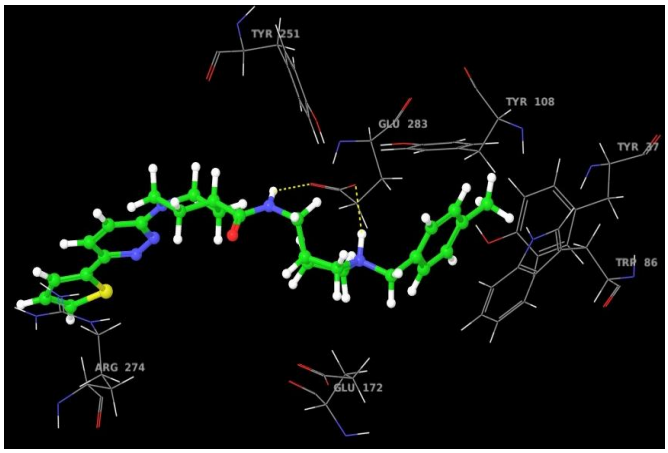


Figure 4.6.b: Docked pose of Ag13 showing interactions with E283

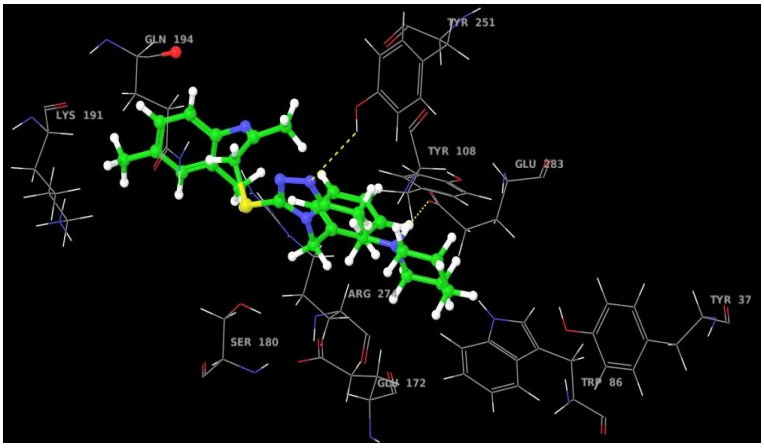


Figure 4.6.b: Docked pose of Ag8 showing interactions with Y251.

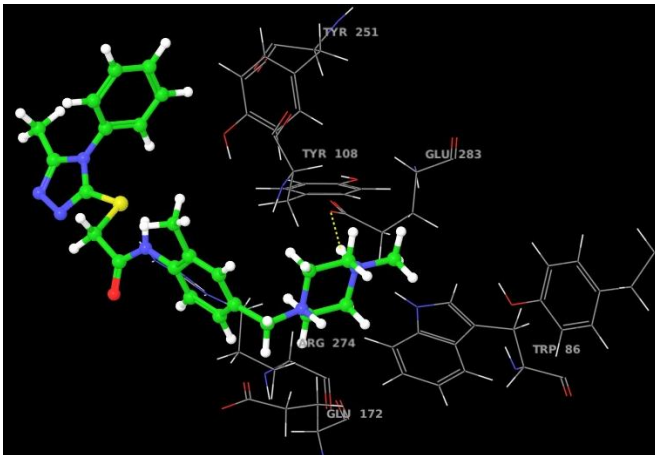


Figure 4.6.b: Docked pose of Ag13 showing interactions with E283.

Figure 4.7: Diagram showing the docked site of all the conformations all ligands.

All the residues are interacting with the E283 and are stabilized by the H-bonding on other end of the ligand molecule with R274 in molecules (Vicriviroc, Maraviroc, and Ag14). All the molecules are binding in the expected site interacting with the key residues. Figure 4.7 shows the position of docked ligands.

4.2.1 Correlation with the Site Directed Mutagenesis

Site mutagenesis studies done by Kellenberger et al.,[39]has showed that mutations D11A,Y37A, Y108A,S180T, I198A ,E283A has drastically reduced the binding affinity of the natural ligands (MIP-1α and MIP-1β and RANTES), where as Y108A, Y251A, E283A has reduced the binding affinity of the HIV-gp120/CD4. These observations gave us clue that these residues are interacting with the natural ligands and gp120/CD4 complex which when mutated lost the side chains that are interacting with the ligands.

Conclusions

5. CONCLUSIONS:

A 3D model of the human CCR5 was built from the human CXCR4 X-ray structure (PDBid: 3ODU). The resulting model was verified for the analysis of stereochemistry of the protein with the PROCHECK. A good model with good scores in PROCHECK is obtained with iterations of loop refinement. Further, the model is validated by 20 ns MD conformational analysis in vacuum which showed its consistency in the arrangement of the TMs. To completely exploit this model, and to give further insights into the molecular basis of the initiation and development of new agonist and antagonists, we docked three antagonist and three agonists into an average structure from the stable energy region of energy vs time graph (11ns to16ns). These interactions and binding poses of the agonists and antagonists are nicely correlated with the available site directed mutagenesis studies evidence and hence provided a consistent structural model for agonist and antagonist binding.

The results of our combined modelling, dynamics, and docking study provide structural insights into CCR5 interactions with the agonists and antagonists, which may be useful in the rational design of new molecules as agonists and antagnoists.

References

REFERNCES:

[1] Parmentier et al., M. Blood(1999) 94, 1899-1905

[2] Berger, EA et al., Science. (1996) 272, 1955-1958.

[3] Adit Ben-Baruch et al., The CCL5/CCR5 Axis in Cancer CHEMOKINE RECEPTORS IN CANCER Cancer Drug Discovery and Development. (2009)109-130.

[4] Lawrence DM et al., Expression of CCL5 (RANTES) and CCR5 in prostate cancer. Prostate. (2006) ;66(2):124-34.

[5] Luster AD et al., Chemokines and their receptors: drug targets in immunity and inflammation. Annu. Rev. Pharmacol. Toxicol.(2008), 48, 171-197.

[6] A P Davenport et al., Chemokine receptors as therapeutic targets in atherosclerosis: pharmacological characterisation of the human CCR5 knock-in mouse. Heart (2011);97(20):e7

[7] Candore G et al., CCR5 receptor: biologic and genetic implications in age-related diseases. Ann N Y Acad Sci. (2007) Apr; 1100:162-72.

[8] Jerry E. Manning et al., The CC Chemokine Receptor 5 Is Important in Control of Parasite Replication and Acute Cardiac Inflammation following Infection with Trypanosoma cruzi .Infect. Immun. (2006) 74(1)135-143

[9] Lefkowitz, R.J et al., A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled beta-adrenergic receptor. J. Biol. Chem.(1980) 255, 7108-7117.

[10] Lefkowitz, R.J et al., A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model. J. Biol. Chem. (1993) 268, 4625-4636.

[11]. Gether, U.; Kobilka, B.K. G protein-coupled receptors. II. Mechanism of agonist activation. J. Biol. Chem.(1998) 273, 17979-17982.

[12] Purves D, Augustine GJ, Fitzpatrick D, et al., Neuroscience. 2nd edition. Sunderland (MA): Sinauer Associates; 2001. Receptor Types.

[13] Behan, D.P et al., The Use of Constitutively Active GPCRs in Drug Discovery and Functional Genomics. Nature Reviews, Drug Discovery 1 (2002) 599-608.

[14] Ballesteros, J. & Weinstein, H. (1995) Methods Neurosci. 25, 366-428.

[15] Harmar AJ et al.,; NC-IUPHAR. IUPHAR-DB: new receptors and tools for easy searching and visualization of pharmacological data. Nucleic Acids Res. (2011) Jan;39(Database issue):D534-8.

[16] Claire L. Newton et al., The Year In G Protein-Coupled Receptor Research. Mol Endocrinol, January (2010) 24(1):261–274

[17] Hans Bräuner-Osborne et al., .Molecular Pharmacology of Promiscuous Seven Transmembrane Receptors Sensing Organic Nutrients. Molecular Pharmacology September 2009 76(3) 453-465.

[18] Bouvier M et al., Pharmacological chaperones: potential treatment for conformational diseases. Trends Endocrinol Metab (2004) 15: 222-228.

[19] Helenius A et al., ER quality control: towards an understanding at the molecular level. Curr Opin Cell Biol (2001)13: 431-437.

[20] Limbird LE et al., Membrane trafficking of G protein-coupled receptors. Annu Rev Pharmacol Toxicol (2004) 44: 559-609.

[21] Peng H, et al. Salmeterol stimulation dissociates beta2-adrenergic receptor phosphorylation and internalization. Am J Respir Cell Mol Biol. (2007);36:254–261.

[22] Harvey T. McMahon ., Mechanisms of Endocytosis. Annual Review in Biochemistry. (2009) 78, 857-902.

[23] J. Trejo et al., Ubiquitination differentially regulates clathrin-dependent internalization of protease-activated receptor-1. J. Cell Biol. (2007) 177:905-916

[24] A. Helenius et al., Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. Nat. Cell Biol(2001). 3:473–483.

[25] Michel Bouvier et al., DIMERIZATION: An Emerging Concept for G Protein–Coupled Receptor Ontogeny and Function. Annu. Rev. Pharmacol. Toxicol.(2002)42:409–35

[26] Clauser E et al. Cloning and functional characterization of a novel mas-related gene, modulating intracellular angiotensin II actions. Mol. Endocrinol. (1991) 5:1477–1487

[27] Jingami H et al., Cryptic Dimer Interface and Domain Organization of the Extracellular Region of Metabotropic Glutamate Receptor Subtype 1. J Biol Chem. (2000);275(36):28144-51.

[28] Hauschild BC et al., Cys-140 is critical for metabotropic glutamate receptor-1 dimerization. J Biol Chem.(2000) Nov 3; 275(44):34245-51.

[29] Quitterer, U et al., Involvement of the amino terminus of the B(2) receptor in agonist-induced receptor dimerization. J. Biol. Chem. (1999) 274, 26079-26084.

[30] Reynolds CA et al., Dimerization and domain swapping in G-proteincoupled receptors: a computational study. Neuropsychopharmacology (2000) 23, S60-S77.

[31] Graeme Milligan et al., Oligomerisation of G-protein-coupled receptors Journal of Cell Science(2001) 114, 1265-1271

[32] Robbins, M. J. et al,. GABA(B2) is essential for G-protein coupling of the GABA(B) receptor heterodimer. J. Neurosci. 21, 8043–8052 (2001).

[33] Kobilka BK et al., Crystal structure of the human β2 adrenergic G-protein-coupled receptor. Nature (2007) 450: 383–387.

[34] Stevens RC et al., High-resolution crystal structure of an engineered human β2-adrenergicGprotein-coupled receptor. Science(2007) 318:1258–1265.

[35] Kobilka BK et al.Structural insights into G-protein-coupled receptor activation. Curr Opin Struct Biol (2008)18:734–740

[36] Schwartz TW. Locating ligand-binding sites in 7TM receptors by protein engineering. Curr Opin Biotechnol. (1994)5(4):434-44.

[37] Kobilka BK et al., Agonists induce conformational changes in transmembrane domains III and VI of the beta2 adrenoceptor. EMBO J. (1997) 16(22):6737-47

[38] Ernst OP et al. Crystal structure of opsin in its Gprotein- interacting conformation. Nature (2008) 455:497–502

[39] Garcia-Perez J, Rueda P, Alcami J, Rognan D, Arenzana-Seisdedos F, Lagane B, Kellenberger E. Allosteric model of maraviroc binding to CC chemokine receptor 5 (CCR5).J Biol Chem. (2011) 23; 286(38):33409-21.

[40] Mitsuya H et al., Involvement of the second extracellular loop and transmembrane residues of CCR5 in inhibitor binding and HIV-1 fusion: insights into the mechanism of allosteric inhibition. J Mol Biol. (2008)12;381(4):956-74

[41] Joshi S et al., CCR5 as target for HIV-1 gene therapy Curr Gene Ther. 2008 Aug;8(4):264-72..

[42] Lawrence DM et al., Expression of CCL5 (RANTES) and CCR5 in prostate cancer. Prostate. 2006 Feb 1;66(2):124-34.

[43] Diana M. Lawrence et al., Expression of CCL5 (RANTES) and CCR5 in prostate cancer . (2006)66(2) 124–134

[44] Jerry E. Manning et al., The CC Chemokine Receptor 5 Is Important in Control of Parasite Replication and Acute Cardiac Inflammation following Infection with Trypanosoma cruzi . Infect. Immun. (2006)74,1135-143.

[45] Tak PP et al., CCR5 blockade in rheumatoid arthritis: a randomised, double-blind, placebo-controlled clinical trial.Ann Rheum Dis. 2010 Nov;69(11):2013-6.

[46] Menno A. de Rie et la., Expression of the chemokine receptor CCR5 in psoriasis and results of a randomized placebo controlled trial with a CCR5 inhibitor. Arch Dermatol Res. 2007 September; 299(7): 305–313.

[47] Krause KH et al., Prog Neurobiol. The chemokine receptor CCR5 in the central nervous system 2011 Feb;93(2):297-311. Epub 2010 Dec 14.

[48] Vijay K. Kalra et al. Mechanism of amyloid peptide induced CCR5 expression in monocytes and its inhibition by siRNA for Egr-1. Am J Physiol Cell Physiol (August 2005) 289(2) C264-C276.

[49] Anders HJ et al. Chemokines and chemokine receptors as therapeutic targets in chronic kidney disease. Front Biosci (Schol Ed). 2009 Jun 1;1:1-12.

[50] Ulrich Spengler et al.CC-chemokine receptor 5 (CCR5) in hepatitis C—at the crossroads of the antiviral immune response? J. Antimicrob. Chemother. (2004) 53 (6): 895-898.

[51] Pohlmann S et al. Prospects of HIV-1 entry inhibitors as novel therapeutics. Rev Med Virol. (2004)14(4):255-270.

[52] Juan C Bandres et al. Maraviroc: a CCR5-receptor antagonist for the treatment of HIV-1 infection. Clinical Therapeutics (2008) 30(7) 1228-1250

[53] Kellenberger E et al. Allosteric model of maraviroc binding to CC chemokine receptor 5 (CCR5).J Biol Chem. 2011 Sep 23;286(38):33409-21. Epub 2011 Jul 20.

[54] Sexton PM et al. Homology modeling of GPCRs.Methods Mol Biol.(2009) 552:97-113.

[55] Niv MY et al. Homology modeling of G-protein-coupled receptors with X-ray structures on the rise. Curr Opin Drug Discov Devel. 2010 May;13(3):317-25.

[56] Sali A et al. Modeller: generation and refinement of homology-based protein structure models.Methods Enzymol. (2003)374:461-91.

[57] Magrane M. and the UniProt consortium. UniProt Knowledgebase: a hub of integrated protein data. Database, 2011: bar009 (2011).

[58] Thomas L. Madden et al. NCBI BLAST: a better web interface. Nucl. Acids Res. (2008) 36 (suppl 2): W5-W9.

[59] P.E.Bourne et al.The Protein Data Bank. Nucleic Acids Research, (2000)28. 235-242.

[60] Anglister, J et al.,The Conformation and Orientation of a 27-Residue CCR5 Peptide in a Ternary Complex with HIV-1 gp120 and a CD4-Mimic Peptide.Journal: (2011) J.Mol.Biol. 410: 778-797.

[61] Multiple Sequence Veiwer , Schrödinger, LLC, New York, NY, 2011.

[62] Andrej Sali et al. Statistical potential for assessment and prediction of protein structures. Protein Sci. (2006) 15: 2507-2524.

[63] Melo F et al., Statistical potentials for fold assessment Protein Sci. (2002)11, 430-448.

[64] RA Laskowski et al.,. PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst. (1993), 26, 283-291.

[65] Schrödinger, Inc., New York, NY, 2011.

[66] Berendsen et al., GROMACS: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. (1995)91, 43–56.

[67] Berendsen HJC , J Chem Phys (1984)81:3684–3690

[68] Hess B et al., LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem. (1997)8,1463–1472.

[69] Glide, version 5.7, Schrödinger, Inc., New York, NY, 2011.

[70] Didier Rognan et al. Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening. J. Med. Chem. (2007), 50, 1294-1303

[71] Ligprep, version 2.4, Schrödinger, Inc., New York, NY, 2011.

Appendices

List of Figures:

List of Table

List of Abbreviations:

AC – adenyl cyclase

CCR5 – C-C chemokine receptor type 5 / Chemokine (C-C motif) ligand 5

MIP - Macrophage Inflammatory Protein

NK - Natural killer

RANTES - Regulated upon Activation Normal T-cell Expressed and Secreted

CCL - CC chemokine Ligand

MCP - Monocyte Chemo attractant Protein

CD4 - Cluster of differentiation 4

RA - Rheumatoid arthritis

CXCR - C-X-C motif chemokine receptor

PDB - Protein Data Bank

GPCR - G protein-coupled receptors

G- G-protein

L- Ligand

GTP - Guanosine-5'-triphosphate

TM – Transmembrane

AC - adenyl cyclase

ECL -extracellular loop

ICL - intracellular loop

GRK - GPCR kinases

PKA – Protein Kinase A

PKC – Protein Kinase C

GABA - gamma-amino butyric acid

DOPE – Discrete Optimized potential Energy

Scripts used

Script 1:

model.py – used to build the models in modeller along with DOPE, GA341 score assessment

```
# Homology modeling with multiple templates
from modeller import *              # Load standard Modeller classes
from modeller.automodel import *    # Load the automodel class

# Redefine the special_patches routine to include the additional disulfides
# (this routine is empty by default):
class MyModel(automodel):
    def special_patches(self, aln):
        # A disulfide between residues 8 and 45:
        self.patch(residue_type='DISU', residues=(self.residues['20'],
                                                  self.residues['269']))
        self.patch(residue_type='DISU', residues=(self.residues['101'],
                                                  self.residues['178']))


    def special_restraints(self, aln):
        rsr = self.restraints
        at = self.atoms
#       Add some restraints from a file:
#       rsr.append(file='my_rsrs1.rsr')
#       Residues 32 through 38 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('32:A',
'38:A')))


log.verbose()    # request verbose output
env = environ()  # create a new MODELLER environment to build this model in

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'
#env.io.atom_files_directory = ['.', '../atom_files']


a = automodel(env,
              alnfile  = 'align-multiple.ali', # alignment filename
              knowns   = ('BRHOD','CXCR4','NTERM'),    # codes of the
templates
              sequence = ('CCR5'),
              assess_methods=(assess.DOPE,
                              assess.GA341,
                              assess.normalized_dope))

# code of the target
a.starting_model= 900                # index of the first model
a.ending_model  = 1000               # index of the last model
               # (determines how many models to calculate)
```

```
a.make()                                # do the actual homology modeling


# Get a list of all successfully built models from a.outputs
ok_models = [x for x in a.outputs if x['failure'] is None]

# Rank the models by DOPE score
key = 'DOPE score'
if sys.version_info[:2] == (2,3):
    # Python 2.3's sort doesn't have a 'key' argument
    ok_models.sort(lambda a,b: cmp(a[key], b[key]))
else:
    ok_models.sort(key=lambda a: a[key])

# Get top model
m = ok_models[0]
print("Top model: %s (DOPE score %.3f)" % (m['name'], m[key]))
```

Script 2:

minim.mdp – used for minimization in gromacs

```
; Parameters describing what to do, when to stop and what to save
integrator        = steep     ; Algorithm (steep = steepest descent
minimization)
emtol             = 0.05      ; Stop minimization when the maximum force <
1000.0 kJ/mol/nm
emstep            = 1         ; Energy step size
nsteps            = 500000    ; Maximum number of (minimization) steps to
perform
continuation      = yes       ; Restarting
; Parameters describing how to find the neighbors of each atom and how to
calculate the interactions
nstlist           = 1         ; Frequency to update the neighbor list and
long range forces
ns_type           = simple    ; Method to determine neighbor list (simple,
grid)
rlist             = 1.0       ; Cut-off for making neighbor list (short
range forces)
coulombtype       = PME       ; Treatment of long range electrostatic
interactions
rcoulomb          = 1.0       ; Short-range electrostatic cut-off
rvdw              = 1.0       ; Short-range Van der Waals cut-off
pbc               = no        ; Periodic Boundary Conditions (yes/no)
```

Script 3:

md.mdp: used for production md in gromacs

```
 title            = CCR5 Protein Stability Test
; Run parameters
integrator        = md              ; leap-frog integrator
nsteps            = 10000000        ; 20 * 500000 = 20000 ps, 20 ns
dt                = 0.002           ; 2 fs
comm_mode         = angular         ; Remove center of mass translation and
rotation around the center of mass
```

```
; Output control
nstxout           = 1000           ; save coordinates every 2 ps
nstvout           = 1000           ; save velocities every 2 ps
nstxtcout         = 1000           ; xtc compressed trajectory output
every 2 ps
nstenergy         = 1000           ; save energies every 2 ps
nstlog            = 1000           ; update log file every 2 ps
; Bond parameters
continuation      = yes            ; Restarting after EM
constraint_algorithm = lincs       ; holonomic constraints
constraints       = all-bonds      ; all bonds (even heavy atom-H bonds)
constrained
lincs_iter        = 1              ; accuracy of LINCS
lincs_order       = 4              ; also related to accuracy
; Neighborsearching
ns_type           = simple         ; search neighboring simple
nstlist           = 0              ; 0 fs
rlist             = 0              ; short-range neighborlist cutoff(nm)
rcoulomb          = 0              ; short-range electrostatic cutoff (nm)
rvdw              = 0              ; short-range van der Waals cutoff (nm)
; Electrostatics
coulombtype       = Cut-off        ; Cut-off electrostatics
pme_order         = 4              ; cubic interpolation
fourierspacing    = 0.16           ; grid spacing for FFT
; Temperature coupling is on
tcoupl            = V-rescale      ; modified Berendsen thermostat
tc-grps           = Protein        ; two coupling groups - more accurate
tau_t             = 0.1            ; time constant, in ps
ref_t             = 300            ; reference temperature, one for each
group, in K
; Pressure coupling is on
pcoupl            = no             ; pressure coupling off
; Periodic boundary conditions
pbc               = no             ; No PBC
; Dispersion correction
DispCorr          = No             ; account for cut-off vdW scheme
```