UNIVERSITAT AUTÒNOMA DE BARCELONA

TESI DOCTORAL

# Influence of Ser and Thr residues in the geometry of transmembrane helices: implications on the structure and function of G protein-coupled receptors

Memòria presentada per Xavier Deupí i Corral per optar al grau de Doctor en Bioquímica i Biología Molecular

Tesi Doctoral realitzada al Laboratori de Medicina Computacional, Institut de Neurociències, sota la direcció de Leonardo Pardo Carrasco

Tesi adscrita al Departament de Bioquímica i Biología Molecular, Universitat Autónoma de Barcelona

# Agraïments

Hi ha moltes maneres de rebre ajuda, algunes directes i d'altres més o menys indirectes, i totes elles són importants. Així que entre unes i altres, hi ha molta gent a la qual estic agraït per la seva ajuda en l'elaboració d'aquesta tesi.

El suport més gran és el de la meva família. Ells, per moltes raons, són els principals responsables de que hagi arribat fins aquí.

Mama, gràcies per l'amor, la felicitat, la pau i la confiança dels que sempre m'has envoltat. La teva fe en el que podia arribar a fer i arribar a ser m'han portat fins aquí. Mai t'ho podré arribar a agrair prou.

Josep, gràcies per tota la teva ajuda en els meus estudis, però molt especialment, per tot el que has fet per nosaltres, i en particular per la meva mare.

Mati i Kike, gràcies per donar-me una infantesa tant feliç i per ser els millors germans que ningú podria desitjar.

La principal font d'ajuda de caire més pràctic ha estat, evidentment, els companys de l'Autònoma.

Leo, gràcies per la confiança que has dipositat en mi i per buscar el millor per la gent que treballem amb tu; per provar sempre de fer les coses una mica millor, per deixar-nos marge d'iniciativa i empenye'ns així a superar-nos.

Merche, gràcies per estar sempre disposada a ajudar, per la disposició a aprendre i a ensenyar-nos coses noves; per la teva actitud positiva, que dóna ànima al grup.

Gràcies a l'esquerra nacionalista del Departament, representada en aquest cas pel flamant regidor de Manlleu i per la qui va estar a punt de convertir-se en la Primera Dama de Tagamanent. David, gràcies pel teu inesgotable bon humor i per la tota la teva ajuda en la complicada transició de químic orgànic nou vingut a projecte de biòleg molecular. Honorabilitat i servei!. Mireia, gràcies pels bons moments que hem compartit, per estar sempre disposada a donar un cop de ma en el que faci falta; per les teves històries, les teves cabòries i la teva alegria, que fan que el Departament no sigui només el lloc de feina. I, finalment, gràcies Joffre per tota la teva ajuda en temes científics i lúdics; per les ganes i l'entusiasme amb que et prens tots els projectes que comencem, els acabem o no; i, sobretot, per les bones estones que passem cada dia.

Al Departament hi ha més gent a la qual vull donar les gràcies, en particular al Jesús, per tenir sempre una estona per tothom, ja sigui per xerrar o per donar-te un consell; a la Teresa i Ceci, per fer-nos la feina una mica més fàcil, i pel seu bon humor; al Javier i al Fernando

per la seva ajuda; i a la gent que ha anat passant durant aquests anys, i que han fet els dies més entretinguts: el David i el Jose, el Manu, la Indiana, la Gema, ...

A més, hi ha un grup de persones amb les quals he tingut la sort de treballar durant la tesi, de les quals he après molt.

Cédric, Cédric... what can I say... I've been lucky to work with you: I've learned a lot and I've had a room in SF, what else can I ask for? Seriously, you've done a lot for me and, moreover, we've had fun... Merci beaucoup, mon ami!

Joan, parlar amb tu sempre és molt bo per l'autoestima i l'optimisme. Gràcies per ser una font d'idees inesgotable, per compartir-les amb nosaltres, i per ser un excel·lent amfitrió!

Joost "dj Flowerpot"... I hope we can work, go out and have a laugh together for many years. You have to come again!

Hi ha també aquella zona nebulosa on la feina i el terreny personal es barregen. En aquesta franja de penombra, brilla amb llum pròpia el Carles. Carles, fer el TFC amb tu va ser el millor que em va poder passar. Quan només tenia una idea més o menys romàntica del que volia dir investigar, tu em vas ensenyar el que és fer recerca meticulosa i de qualitat (suposo... ;-) ), i em vas posar "en el bon camí", per dir-ho d'alguna manera. A més, ets una de les persones amb les que més em puc arribar a riure. Realment, coneixe't ha estat una sort. Gràcies per tot.

Per sort, puc dir que tinc algú que ho engloba tot i que dóna sentit al que faig. Laura, gràcies pel teu suport, pel teu amor, per tot el que em dónes i pel que em fas sentir... per ser com ets.

I found it hard.
It was hard to find.
Oh, well, whatever,
nevermind...

Kurt Cobain

# Contents

# Chapter 1
# Introduction

## *1.1. Bioinformatics: a different approach to biology*

Decades of biochemical experiments have produced a wealth of biological data. As an example, the 2003 edition of the Federation of American Societies for Experimental Biology annual meeting (Experimental Biology 2003, San Diego (CA), April 11-15, 2003) hosted around 7500 contributions (1). Moreover, the recent genome sequencing projects are generating a vast quantity of biological information. For instance, on the May 1st 2003 version of the GOLD (Genomes OnLine Database. http://wit.integratedgenomics.com/GOLD/) there are 717 genome projects reported, with 139 published complete genomes and 578 ongoing projects.

The application of computers for the analysis and interpretation of this great amount of data is defined as *Bioinformatics*. This discipline is, by nature, an interdisciplinary field, which brings together biologists, computer scientists, chemists, physicists and mathematicians. Due to its success in the analysis of genomic and proteomic data, bioinformatics has become a basic tool essential for management of information in modern biology and medicine.

Bioinformatics is a very broad discipline that has been divided in three main applications (2):

a) connection between all different forms of data gathered by new high-throughput techniques such as systematic sequencing, expression arrays, yeast two-hybrid and high throughput screenings. This application of bioinformatics is referred as *data mining*, and is considered the mainstream bioinformatics.

b) tool to study specific biological problems, through its integration in the daily work of molecular biologists, using the growing set of methods available in the World Wide Web.

c) management and representation of highly complex biological information, as the simulation of cellular systems or molecular interaction networks, and to the study of fundamental biological questions, such as protein folding or mechanisms of protein

function. The contribution of bioinformatics to these areas is related to the development of concepts in theoretical molecular biology.

This thesis is framed in this later aspect of bioinformatics, the study of particular biological systems as the source of inspiration that guides the formation of general ideas from specific cases to general principles. In particular, the theoretical study of the influence of certain amino acids on the structure and dynamics of the secondary structure elements of proteins has been applied to homology modelling of G protein-coupled receptors and to the study of their mechanisms of activation.

## *1.2. Protein structure*

### 1.2.1. Protein structure and function

Protein synthesis is carried out in the ribosomes (3). In this process, known as translation, the amino acids are incorporated into the protein according to the mRNA template. The polypeptide chain must then fold into the appropriate three-dimensional conformation, which results from interactions between the side chains of their constituent amino acids. All the information required for a protein to adopt the correct three-dimensional conformation is provided by its amino acid sequence. This was established by Christian Anfinsen's experiments demonstrating that denatured RNase can spontaneously refold in vitro to its active conformation (4). However, it has been shown that in the cell the folding of proteins is mediated by the activity of chaperones, which catalyze protein folding by assisting the self-assembly process, without conveying any additional information for the proper folding (3).

So then, there is a close relationship between the three-dimensional folding of a protein and its function. At the same time, this folding is encoded in the primary sequence of the protein. So, although there is a great leap between the simple list of amino acids (or nucleotides in the gene) and how the protein works, the information about the function is buried in that one-dimensional list.

### 1.2.2. From primary to secondary structure

The amino acid sequence of a protein's polypeptide chain is called its primary structure. Different regions of the sequence form local regular secondary structures, such as α-helices or β-strands. Packing such structural elements into one or several compact globular domains forms the tertiary structure. The final protein may contain several polypeptide chains arranged in a quaternary structure. By formation of such tertiary and quaternary structure amino acids far apart in the sequence are brought close together in the three dimensions to form a functional region, an active site. Although the complexity and irregularity of the structure of proteins,

there are regular features the most important of which is their secondary structure.

But, how are these secondary structures formed? Although the process of protein folding is not yet fully understood, it has been shown that the main driving force for folding water-soluble globular protein molecules is to pack hydrophobic side chains into the interior of the molecule, thus creating a hydrophobic core and a hydrophilic surface (3). However, the main chain, which is highly polar, must also fold into the interior. These main chain polar groups must be neutralized by hydrogen bond formation in a hydrophobic environment. This problem is solved by the formation of regular secondary structures within the interior of the protein. Such secondary structure is usually one of two types: α-helices or β-sheets. Both types are characterized by having the main chain NH and CO groups participating in hydrogen bonds to each other.

Although membrane proteins were considered "inside-out" proteins, in the sense that they would have a core of polar residues surrounded by hydrophobic residues facing the apolar core of the membrane, that is, the reverse of the characteristics found in water-soluble proteins, it has been shown that this is not the case. Neither polar nor hydrophobic residues show any bias for the exterior or the interior of a transmembrane domain. Packing of α-helices in membranes is better understood by maximization of van der Waal's forces, rather than by a general segregation of hydrophobicities driven by lipid exclusion (5-7).

### 1.2.3. Elements of secondary structure: structure of α-helices

α-helices as structural elements of proteins were first described by Linus Pauling in 1951, who predicted its structure on the basis of accurate geometrical parameters derived from the results of crystallographic analyses of the structures of a range of small molecules (8). This prediction was completely verified from the high resolution structure of myoglobin (9).

α-helices in proteins are found when a stretch of consecutive residues all have the φ and ψ backbone dihedral angle pair approximately -60º and –50º, corresponding to the allowed region in the bottom left quadrant of the Ramachandran plot (3) (Fig. 1.1). The α-helix has 3.6 residues per turn with hydrogen bonds between C=O of residue *i* and NH of residue *i-4* (Fig. 1.2). Thus, all NH and CO groups are joined with hydrogen bonds except the first NH groups and the last CO groups at the ends of the α-helix. As a consequence, the ends of α-helices are polar. α-helices vary considerably in length in globular proteins ranging from four of five amino acids to over 40 residues, with an average length of around 10 residues, corresponding to three turns. Since the rise per residue of an α-helix is 1.5 Å along the helical axis, this corresponds to about 15 Å from one end to the other of an average α-helix. The α-helices observed in proteins are almost always right-handed. All the hydrogen bonds in an α-helix point in the same direction so the peptide units are aligned in the same orientation along the helical axis. Since a peptide unit has a dipole moment arising from the different polarity of NH and CO groups, these dipole moments are also aligned along the helical axis. The overall effect is a significant net dipole for the α-helix that

gives a partial positive charge at the amino end and a partial negative charge at the carboxyl end of the α-helix.

### 1.2.4. Proline residues in α-helices

Proline is unique among the natural aminoacids, as its side chain is bonded to the preceding peptide bond nitrogen forming a pyrrolidine ring (10) (Fig. 1.3). This ring imposes rigid constraints on the φ backbone dihedral angle, which is fixed at the preferred α-helical value, -60º. The ψ angle is more flexible but hindered, with two preferred regions, corresponding to the values of φ=-55º and +145º (11). These regions correspond to the *cis*/*trans* conformations of the peptide bond. It is worth to note that Pro is the only residue for which the *cis* peptide bond in energetically accessible (10).

In α-helices the backbone dihedral angles of Pro are shifted to φ = –61º and ψ = –35º (10). The more negative values of φ help to reduce the steric clash between the $C_\alpha$ hydrogen of the preceding residue and the carbonyl carbon of the proline, while the shift to more positive ψ value helps to reduce the steric conflict between the $C_\alpha$ of the preceding residue and the carbonyl oxygen of the proline. In addition, the bulky pyrrolidine ring restricts the available conformational space of adjacent residues in an α-helix. The φ backbone dihedral angle of the residue preceding proline presents higher values than those two positions before the Pro residue. At the same time the corresponding ψ values tend to show an inverse correlation: the ψ angle in the residue which precedes the Pro changes to more negative values, to minimize the bad contacts between $C_\beta$ of the residue before the proline and $C_\delta$ of the proline, which helps to accommodate the proline ring (Fig. 1.4). These changes in the angles of residue at the *i-1* position result in an upward movement of the proline ring away from the carbonyl at the *i-4* position, while the shifts involving residue at the *i-2* position causes a lateral clockwise movement away from it, resulting in a global kink (*Pro-kink*) of the helix in that point. The angle by which helices bend due to the presence of Pro has been estimated to be around 20º (12-16). Moreover, the presence of the pyrrolidine ring induces a local opening in the α-helix at the Pro-kink turn (Fig. 1.5).

Another helix destabilizing feature is that the Pro nitrogen lacks a proton, as it is covalently bonding the Pro side chain. This way this amide nitrogen cannot participate in a hydrogen bond to a neighbouring carbonyl group. In helical segments, this situation frees up the carbonyl group four residues away for participation in a hydrogen bond interaction. In addition the backbone hydrogen bond between the *i-3* carbonyl and the *i+1* amide is also disrupted, due to the presence of the bulky pyrrolidine ring (11) (Fig. 1.3). Furthermore, the oxygen atom of the C=O of the residue preceding Pro is more electronegative than those preceding other amino acids, with the result that this carbonyl group has an enhanced tendency to accept and form strong hydrogen bonds (17).

Although prolines disrupt the structure of α-helices, they are frequently found in the in the putatively α-helical transmembrane segments of multi-spanning integral membrane proteins that function as receptor subunits or transporters (11;17;18).

The effects of prolines in a transmembrane helix (TMH) can be categorized as structural and/or dynamic (17). Structural roles include purely static effects such as those arising from a kinked helix. For instance, proline residues serve to generate bent helices that may suit specific packing motifs. Moreover, this putative structural role may be related to electronic effects that stem from increased local polarity. Dynamic roles describe the participation of Pro in conformational change.

In view of these exceptional properties it is not surprising that proline tends to be a conserved residue among homologous proteins and plays a role in protein structure and function. For instance, it has been suggested that some prolines may be actively involved in the regulation of transmembrane proteins (19-24).

### 1.2.5. Ser and Thr residues in α-helices

Ser, Thr and Cys residues have short polar side chains with a strong hydrogen bond potential. In these residues, the *gauche* – (*g*-), *gauche* + (*g*+) and *trans* (*t*) staggered side chain conformations are strongly preferred relative to the eclipsed conformations (25). In the context of an α-helix, these conformations have different properties. For instance, *g*- and *g*+ conformation of Ser and Thr are capable to hydrogen bond the backbone carbonyls at positions *i-3* or *i-4* in the previous turn of the helix (Fig. 1.6), while a Cys residue in its *g*+ conformation can hydrogen bond the carbonyl at the *i-4* position. However, the *g*- conformation is forbidden for the Cys residue, due to the steric clash between the $S_\gamma$ atom and the carbonyl at the *i-3* position. On the other hand, *t* conformations are not capable to establish intrahelical hydrogen bonds (Fig. 1.6). Moreover, the methyl group at the gamma position of a helical Thr residue in the *t* conformation would clash with the backbone carbonyl at the *i-3* position, which prevents this conformation in the context of a helix. Thus, in transmemebrane helices, where the hydrogen bond capability of these polar residues must be somehow satisfied, *g*- and *g*+ conformations of Ser and Thr and *g*+ of Cys are more common that *t* conformations.

These properties of Ser and Thr residues may help to explain that they are regularly found at TMH, although not so well conserved as the strongly polar residues. For instance, patterns of multiple Ser and Thr residues at the helix interface have been shown to be important in mediating and stabilizing the interactions between TMH through a series of side chain hydrogen bonds (26;27). Ser and Thr can also have an important role in membrane protein folding, stabilizing the formation of helical oligomers (28). In addition, when involved in a hydrogen bond with nearby backbone carbonyls, Ser and Thr have been shown to promote a bending in TMHs, possibly facilitating conformational changes in membrane proteins (29).

## *1.3. G Protein-Coupled Receptors (GPCRs)*

### 1.3.1. The GPCR family

Genome sequencing projects have identified the GPCR superfamily as one of the largest classes of proteins in mammalian genomes (30). For instance, preliminary analysis of the human genome has revealed up to 600 GPCRs (31;32). In addition, GPCRs are scored as the most common family in the human proteome at the Proteome Analysis Database of the European Bioinformatics Institute, with more than 800 sequences (see http://www.ebi.ac.uk/proteome/HUMAN/interpro/top15f.html). GPCRs mediate the transmembrane signal transduction in response to ligands of a great structural diversity, as hormones, peptides, nucleotides, amino acids, neurotransmitters, Ca2+ or light (33), with distinct binding modes (34). Moreover, GPCRs are the principal signal transducers for the senses of sight and smell (35). In addition, GPCRs are important drug targets as evidenced by the fact that approximately 50% of marketed drugs treat diseases by targeting GPCRs (36).

Based on specific sequence homologies, GPCRs can be divided into several subfamilies (37;38) Recently, using phylogenetic analyses of the human genome, these receptors have been classified into five main familes: glutamate, rhodopsin, adhesion, frizzled/taste2, and secretin. The rhodopsin family, which is the largest, is subdivided into four main groups (105) (Fig. 1.7). The number of families could increase, as a large number of genes encoding new GPCRs for which no natural ligand are known, the so-called orphan receptors, are being identified through the on-going genome sequence projects.

GPCRs are able to regulate the activity of heterotrimeric guanine-nucleotide exchange factors, the G proteins (33). G proteins are composed of three subunits: $\alpha$, $\beta$ and $\gamma$, and the different G proteins are named after their $\alpha$ subunits. According to their sequence homology, the $G_\alpha$ subunits can be classified into four main subtypes: $\alpha_s$, $\alpha_{i/o}$, $\alpha_{q11}$ and $\alpha_{12}$ (39). In the inactive state of the receptor, the G protein interacts with its cytoplasmatic region. Studies of point mutations, receptor chimeras and partially deleted receptors indicate that the sites for G protein coupling and activation are mainly in the third intracellular loop, but also in the second intracellular loop and in the C-terminal domain (40). The activation of the receptor induces conformational changes in the $\alpha$ subunit of the G protein, which lead to the release of GDP and the binding of GTP. Subsequently, the G protein dissociates into the alpha and beta-gamma subunits, which can modulate several cellular signalling pathways. For instance, $G_i$ inhibits adenylyl cyclase and activates phopholipase C and potassium channels, $G_o$ inhibits calcium channels and stimulates potassium channels and phospholipase C and $G_s$ stimulates adenylyl cyclase and calcium channels (40). It has also been suggested that GPCRs could act using other signalling cascades (33). A whole GPCR family can couple to one particular G protein subtype (for instance, all opioid receptors couple to $G_{i/o}$), but different members of one family can also couple to different G protein subtypes (for instance, the muscarinic receptors couple to $G_{i/o}$ or $G_q$ (40)). It is also possible that some receptors exert their function via several G protein

subtypes. For instance, the TSH receptor activates $G_s$, $G_q$, $G_i$ and $G_{12}$ (41). This promiscuity depends not only on the properties of the receptor, but also on the relative amounts of ligand, receptors and G proteins (42).

### 1.3.2. Mechanism of activation of GPCRs

Activation of GPCRs is commonly discussed in terms of the extended ternary complex model, which proposes that the receptor exists in equilibrium between two states: the inactive state (R) and the active state (R*) (43) (Fig. 1.8). GPCRs are kept in its inactive conformation through a network of constraining intramolecular interactions (33;35), and spontaneous release of some of these interactions can lead to the co-existence of different conformations of the receptor (44). Some of these conformations will correspond to active forms of the receptor, which can bind the G protein with high affinity and, thus, induce its activation. The active forms of the receptor can be stabilized by the binding of full or partial agonists and by the interaction with the G protein through specific interactions. GPCRs ligands are classified according its influence on this equilibrium: agonists stabilize R*, inverse agonists stabilize R, while antagonists don't change the equilibrium rate.

Some GPCRs present a significant basal signal in the absence of agonist, a so-called constitutive activity, which is caused by a small proportion of the receptor population in the R* active state (43;45;46). Studies on the $\beta_2$-adrenergic receptor (47) show that constitutively activated mutants of this receptor present a higher degree of conformational flexibility, which is attributed to the disruption of stabilizing conformational constraints of the receptor.

However, this simple two-state model may not be able to explain all properties of GPCR activation (48). For instance, recent studies point to the existence of different states of the native form of the receptor (44). It has also been shown that different ligands can stabilize distinct conformations of the active state (49). In summary, the process of activation is a highly complex process, which may involve multiple conformational states of the receptor, the ligand and the G protein. The different conformations of the receptor will differ in the arrangement of the transmembrane bundle. For instance, extensive mutagnesis and biochemical experiments in rhodopsin as well as a variety of other GPCRs suggest that during the process of signal transduction specific interactions between the agonist and the receptor are translated into changes in the network of interactions of specific residues which lead to a rearrangement of the transmembrane bundle, specially in the relative orientations of TMH3 and TMH6 (50;51). These movements of TMHs are then thought to affect the conformation of G protein-interacting intracellular regions of the receptor and thus uncover previously masked G protein-binding sites on the second, third and fourth cytoplasmatic loops, leading to activation of the G protein (52-56). In this way, the signal progresses from the extracellular side through the transmembrane bundle to the intracellular domain.

### 1.3.3. Three-dimensional structure of GPCRs

The low natural abundance of GPCRs and the difficulty in producing and purifying recombinant protein has hindered obtaining high-resolution structures through X-ray crystallographic studies. Only one member of the family, bovine rhodopsin, has been crystallized so far (57). From the analysis of this single structure, three structural models of bovine rhodopsin are available at the Protein Data Bank (PDB), at 2.8 Å (PDB identifiers 1F88 and 1HZX) and at 2.6 Å (PDB identifier 1L9H) (58). These high resolution structures have altered deeply the study of GPCRs, yielding the first detailed mechanisms of how rhodopsin propagates its signal (59;60). The structure of GPCRs consists in a single polypeptide unit that forms seven membrane-spanning domains, plus an eight helix which lies approximately parallel to the plane of the membrane and perpendicular to the helical bundle, with an extracellular N-terminus and a cytoplasmic C-terminus. The seven TMHs are joined through three cytoplasmatic and three extracellular loops (Fig. 1.9). While small molecular weight ligands bind within the core formed by the TMHs, binding sites for peptides and protein agonists include the N-terminus and extracellular loops joining the transmembrane domains (Fig. 1.10).

Many models of GPCRs have been published over the last years. Low resolution structures of rhodopsin, obtained from electron diffraction of two-dimensional crystals, combined with data from biochemical and biophysical experiments, as site-directed mutagenesis, crosslinking, spin labelling and scanning accessibility, lead to a three-dimensional model (61) which predicted successfully the arrangement and relative orientation of the transmembrane domains (around 3 Å of root mean square deviation for approximately 200 alpha carbons of the transmembrane bundle) (59). As this model included explicitly the patterns of conservation of other GPCRs, its high accuracy suggests that the transmembane bundle structure is conserved among GPCRs. By contrast, the loops and termini are more divergent in amino acid sequence and probably in three-dimensional structure. It is expected that extracellular domains will be highly variable among GPCRs because they are involved in ligand recognition. On the other hand, cytoplasmatic loops, involved in the interaction with the G protein, would present a higher degree of conservation, as there only exist a few G protein subtypes.

From August 2000, the availability of the three-dimensional structure of bovine rhodopsin (57) allows the use of homology modelling techniques for the building of three-dimensional models of GPCRs. Homology modelling is a computer-assisted method used to obtain a three-dimensional model of a protein with an unknown structure. It is based on the use of the structure of rhodopsin as a template and on the amino acid sequence alignment between rhodopsin and the target GPCR.

But there are a number of reasons that make rhodopsin a peculiar GPCR:

- it is a light receptor that has an inverse agonist as a ligand
- it possesses a low degree of homology with other GPCRs

In addition, there is wide diversity within the GPCR family. This way, the straightforward use of the rhodopsin template will not be sufficient to accurately predict the structure of any GPCR.

### 1.3.4. GPCRs dimerization

Models that describe the interaction of GPCRs with G proteins are generally based on the assumption that these receptors exist as monomers. However, there is growing experimental evidence that GPCRs can exist as complexes. For instance, a number of Class C GPCRs have been shown to form dimers in the plasma membrane, as the calcium-sensing receptor (62), the $GABA_B$ receptor (63-65), and the metabotropic glutamate receptors (66). The strongest support for heterodimerization in GPCRs comes precisely from the studies on these $GABA_B$ receptors. It has been found that co-expression of $GABA_{B(1a)}$ and $GABA_{B(2)}$ is explicitly required for proper trafficking to the cell surface (67), and that a physical interaction is needed to generate functional receptors (63-65;68;69). Experimental results suggest that the binding of GABA to $GABA_{B(1a)}$ causes $GABA_{B(2)}$ to bind to and activate the G-protein (70;71). In addition, many class A rhodopsin-like receptors, including the dopamine D2 receptor (72-78), the $\beta_2$-adrenergic receptor (79;80), the sphingosine-1-phosphate receptor (81) or the oxytocin and vasopressin V1a and V2 receptors (82) have been also proposed to form dimers. Heterodimerization in Class A GPCRs has been reported either in closely related receptors, for instance, between *kappa* and *delta* opioid peptide receptors (83) or between *mu* and *delta* opioid peptide receptors (84;85), or distantly related receptors, as between $\alpha_2$-adrenoreceptors and muscarinic acetylcholine type 3 receptors (86), dopamine type 2 receptors and somatostatin subtype 5 receptors (87), $\beta_2$-adrenoreceptors and *delta* opioid peptide receptors (88;89), $\beta_2$-adrenoreceptors and *kappa* opioid peptide receptors (88) or angiotensin type 1 receptors and bradykinin type 2 receptors (90). Interestingly, this last interaction has been proposed to be involved in regulation of blood pressure *in vivo*. Finally, members belonging to different families have also been shown to be physically associated, for instance adenosine type 1 receptors, from the Class A, with metabotropic glutamate type 1 receptors, from the Class C (91).

The structural mechanism for cross-talk between receptors in a dimeric complex is not known yet. Two hypotheses have been proposed for GPCR dimerization: domain swapping and contact dimerization (92;93). In the context of contact dimerization, different types of interactions have been proposed between the units of the dimer. For instance, these interactions may involve the extracellular domain, as in calcium-sensing receptors (94), or in metabotropic glutamate type 1 receptors. In this case, the crystallographic structure of the extracellular domain show that it appears as a dimer, and a conserved cysteine residue plays an important role in the dimerization (95). In other dimers, as between $\beta_2$-adrenoreceptors and dopamine type 2 receptors (80), hydrophobic, hydrogen bond and/or ionic

interactions between regions of the TM bundle are proposed to stabilize the association. In some particular cases, the dimer interface has been identified. For instance, the extracellular end of the fourth transmembrane (TM4) segment in the dopamine D2 receptor (96), or the GxxxG motif in TM6 in the $\beta_2$-adrenergic receptor (80). Finally, interactions between the cytoplasmatic COOH-terminal region have also been proposed for $GABA_B$ receptors (63-65;68;69). It is also possible that in some cases interaction between multiple domains is required for dimerization. Computational studies have proposed a number of different potential interfaces (92;97;98), but these have not yet been experimentally verified.

Very recently, closely packed rhodopsin dimers organized in arrays have been observed in native membranes using infrared-laser atomic-force microscopy (99) (Fig. 1.11b). The same authors have proposed a semi-empirical model of the higher-order structure of rhodopsin where the intradimeric contacts involve helices IV and V, the formation of rhodopsin dimer rows is facilitated by contacts between helices I and II and the cytoplasmic loop connecting helices V and VI, and contacts between rows are on the extracellular side and involve helix I. In addition, a theoretical model of the transducin-rhodopsin dimer complex is proposed (Fig 1.11a) (100).

In conclusion, considerable evidence suggests that GPCRs form homomeric and heteromeric dimers *in vivo*, and that dimerization of some GPCRs leads to important changes in receptor function (101, 102). Dimerization might be required in some families to produce active receptors, while in other cases, it might be associated with different signalling routes that those induced by monomers. However, it is not clear if dimerization is a mechanism common to all GPCRs.

## *1.4. Objectives*

In this thesis, bioinformatic tools will be used to study the influence of certain aminoacids on the geometry of α-helices. In Chapter 3.1, the influence of Ser and Thr residues on the structure of TMHs is studied through the analysis of structural databases. In Chapter 3.2 this study is focused to Pro-containing α-helices, and their dynamic behaviour is studied using molecular dynamics simulations. In addition, the putative structural or functional role of Ser/Thr and Pro helical motifs is studied through the analysis of sequence databases. In Chapter 3.3 some of our hypothesis are investigated through an additional molecular dynamics study, and they are experimentally tested in collaboration with the Institut de Recherche Interdisciplinaire en Biologie Humaine et Nucléaire (IRIBHN), Université Libre de Bruxelles, by the group of Prof. G. Vassart and Prof. M. Parmentier. Finally, in Chapter 3.4, our structural hypotheses are used to build a three-dimensional model of the chemokine CCR5 receptor. This model is also experimentally tested in the IRIBHN.

In summary, this thesis analyzes the effects of single Ser/Thr residues and Ser/Thr and Pro combinations on the three-dimensional structure of GPCRs, using a set of bioinformatic tools as protein sequence analysis, protein structure database search and analysis and molecular dynamics simulations. Our theoretical results are used to design experiments to gain insight in the three-dimensional structure of GPCRs.

## References

1. Experimental Biology 2003 * Translating the Genome * April 11-15 * San Diego, CA. 17[3], 532-568. 2003 Mar 01.
2. Valencia A 2002 Bioinformatics: biology by other means. Bioinformatics 18:1551-2
3. Voet D, Voet J 1995 Biochemistry, ed. J. Wiley & Sons, New York
4. Anfinsen CB 1973 Principles that govern the folding of protein chains. Science 181:223-30
5. Stevens TJ, Arkin IT 1999 Are membrane proteins "inside-out" proteins? Proteins 36:135-43
6. Rees DC, Eisenberg D 2000 Turning a reference inside-out: commentary on an article by Stevens and Arkin entitled: "Are membrane proteins 'inside-out' proteins?" (Proteins 1999;36:135-143). Proteins 38:121-2
7. Stevens TJ, Arkin IT 2000 Turning an opinion inside-out: Rees and Eisenberg's commentary (Proteins 2000;38:121-122) on "Are membrane proteins 'inside-out' proteins?" (Proteins 1999;36:135-143).

Proteins 40:463-4

8. Pauling L, Corey RB, Branson HR 1951 The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. PNAS 37:235-240
9. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC 1958 A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181:662-666
10. MacArthur MW, Thornton JM. Influence of proline residues on protein conformation. J 218. 218[2. 2], 397-412, 397-412. 1991 Mar 20 Mol Biol 1991 Mar 20.
11. von Heijne G 1991 Proline kinks in transmembrane alpha-helices. J Mol Biol 218:499-503
12. Barlow DJ, Thornton JM 1988 Helix geometry in proteins. Journal of Molecular Biology 201:601-619
13. Sankararamakrishnan R, Vishveshwara S. 1990 Conformational studies on peptides with proline in the right-handed alpha-helical region. Biopolymers 30: 287-98.
14. Woolfson DN, Williams DH. 1990 The influence of proline residues on alpha-helical structure. FEBS 277, 185-8.
15. Yun RH, Anderson A, Hermans J. 1991 Proline in alpha-helix: stability and conformation studied by dynamics simulation. Proteins 10. 219-28.
16. Sankararamakrishnan R, Vishveshwara S 1992 Geometry of proline-containing alpha-helices in proteins. Int J Pept Protein Res 39:356-63
17. Williams KA, Deber CM. 1991 Proline residues in transmembrane helices: structural or dynamic role? Biochemistry 30[37], 8919-23.
18. Sansom MS. 1992 Proline residues in transmembrane helices of channel and transport proteins: a molecular modelling study. Protein 5, 53-60.
19. Hong S, Ryu KS, Oh MS, Ji I, Ji TH. 1997 Roles of transmembrane prolines and proline-induced kinks of the lutropin/choriogonadotropin receptor. J Biol Chem 272, 4166-71.
20. Ri Y, Ballesteros JA, Abrams CK, Oh S, Verselis VK, Weinstein H, Bargiello TA 1999 The role of a conserved proline residue in mediating conformational changes associated with voltage gating of Cx32 gap junctions. Biophys J 76:2887-98
21. Tieleman DP, Shrivastava IH, Ulmschneider MR, Sansom MS. 2001 Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. Proteins 44, 63-72.
22. Govaerts C, Blanpain C, Deupi X, Ballet S, Ballesteros JA, Wodak SJ, Vassart G, Pardo L, Parmentier M 2001 The txp motif in the second transmembrane helix of ccr5. a structural determinant of chemokine-induced activation. J Biol Chem 276:13217-25
23. Stitham J, Martin KA, Hwa J 2002 The critical role of transmembrane

prolines in human prostacyclin receptor activation. Mol Pharmacol 61:1202-10

24. Cordes FS, Bright JN, Sansom MS. Proline-induced distortions of transmembrane helices. J 323. 323[5. 5], 951-60, 951-60. 2002 Nov 8 Mol Biol 2002 Nov 8.
25. Gray TM, Matthews BW 1984 Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. Journal of Molecular Biology 175:75-81
26. Dawson JP, Weinger JS, Engelman DM 2002 Motifs of serine and threonine can drive association of transmembrane helices. J Mol Biol 316:799-805
27. Eilers M, Patel AB, Liu W, Smith SO 2002 Comparison of helix interactions in membrane and soluble alpha-bundle proteins. Biophys J 82:2720-36
28. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM 2000 Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nat Struct Biol 7:154-60
29. Ballesteros JA, Deupi X, Olivella M, Haaksma EE, Pardo L 2000 Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. Biophysical Journal 79:2754-2760
30. Takeda S, Kadowaki S, Haga T, Takaesu H, Mitaku S 2002 Identification of G protein-coupled receptor genes from the human genome sequence. FEBS Lett 520:97-101
31. International Human Genome Sequencing Consortium 2001 Initial sequencing and analysis of the human genome. Nature 409:860-921
32. Venter JC, *et al.* 2001 The sequence of the human genome. Science 291:1304-51
33. Gether U 2000 Uncovering molecular mechanisms involved in activation of G protein- coupled receptors. Endocr Rev 21:90-113
34. Ji TH, Grossmann M, Ji I 1998 G protein-coupled receptors. I. Diversity of receptor-ligand interactions. Journal of Biological Chemistry 273:17299-17302
35. Gether U, Kobilka BK 1998 G protein-coupled receptors. II. Mechanism of agonist activation. Journal of Biological Chemistry 273:17979-17982
36. Drews J 2000 Drug discovery: a historical perspective. Science 287:1960-4
37. Parnot C, Miserey-Lenkei S, Bardin S, Corvol P, Clauser E 2002 Lessons from constitutively active mutants of G protein-coupled receptors. Trends Endocrinol Metab 13:336-43
38. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G 2003 GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 31:294-7

39. Simon MI, Strathmann MP, Gautam N 1991 Diversity of G proteins in signal transduction. Science 252:802-8

40. Horn F, van der Wenden EM, Oliveira L, IJzerman AP, Vriend G 2000 Receptors coupling to G proteins: is there a signal behind the sequence? Proteins 41:448-59

41. Laugwitz KL, Allgeier A, Offermanns S, Spicher K, Van Sande J, Dumont JE, Schultz G 1996 The human thyrotropin receptor: a heptahelical receptor capable of stimulating members of all four G protein families. Proc Natl Acad Sci U S A 93:116-20

42. Oliveira L, Paiva AC, Vriend G 1999 A low resolution model for the interaction of G proteins with G protein- coupled receptors. Protein Eng 12:1087-95

43. Samama P, Cotecchia S, Costa T, Lefkowitz RJ 1993 A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model. Journal of Biological Chemistry 268:4625-4636

44. Peleg G, Ghanouni P, Kobilka BK, Zare RN. 2001 Single-molecule spectroscopy of the beta(2) adrenergic receptor: observation of conformational substates in a membrane protein. Proc Natl Acad Sci U S A 98, 8469-74

45. Kjelsberg MA, Cotecchia S, Ostrowski J, Caron MG, Lefkowitz RJ 1992 Constitutive activation of the alpha 1B-adrenergic receptor by all amino acid substitutions at a single site. Evidence for a region which constrains receptor activation. J Biol Chem 267:1430-3

46. Scheer A, Fanelli F, Costa T, De Benedetti PG, Cotecchia S 1996 Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. EMBO J 15:3566-78

47. Gether U, Ballesteros JA, Seifert R, Sanders-Bush E, Weinstein H, Kobilka BK 1997 Structural instability of a constitutively active G protein-coupled receptor. Agonist-independent activation due to conformational flexibility. J Biol Chem 272:2587-90

48. Kenakin T 1997 Agonist-specific receptor conformations. Trends Pharmacol Sci 18:416-7

49. Ghanouni P, Gryczynski Z, Steenhuis JJ, Lee TW, Farrens DL, Lakowicz JR, Kobilka BK. 2001 Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta 2 adrenergic receptor. J Biol Chem 276, 24433-6.

50. Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG 1996 Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. Science 274:768-770

51. Altenbach C, Yang K, Farrens DL, Farahbakhsh ZT, Khorana HG, Hubbell WL 1996 Structural features and light-dependent changes in the cytoplasmic interhelical E-F loop region of rhodopsin: a site-directed spin- labeling study. Biochemistry 35:12470-8

52. Franke RR, Konig B, Sakmar TP, Khorana HG, Hofmann KP 1990 Rhodopsin mutants that bind but fail to activate transducin. Science 250:123-5

53. Ernst OP, Meyer CK, Marin EP, Henklein P, Fu WY, Sakmar TP, Hofmann KP 2000 Mutation of the fourth cytoplasmic loop of rhodopsin affects binding of transducin and peptides derived from the carboxyl-terminal sequences of transducin alpha and gamma subunits. J Biol Chem 275:1937-43

54. Marin EP, Krishna AG, Zvyaga TA, Isele J, Siebert F, Sakmar TP 2000 The amino terminus of the fourth cytoplasmic loop of rhodopsin modulates rhodopsin-transducin interaction. J Biol Chem 275:1930-6

55. Cai K, Itoh Y, Khorana HG 2001 Mapping of contact sites in complex formation between transducin and light-activated rhodopsin by covalent crosslinking: use of a photoactivatable reagent. Proc Natl Acad Sci U S A 98:4877-82

56. Itoh Y, Cai K, Khorana HG 2001 Mapping of contact sites in complex formation between light-activated rhodopsin and transducin by covalent crosslinking: use of a chemically preactivated reagent. Proc Natl Acad Sci U S A 98:4883-7

57. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M 2000 Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289:739-45

58. Mirzadegan T, Benko G, Filipek S, Palczewski K 2003 Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. Biochemistry 42:2759-67

59. Meng EC, Bourne HR 2001 Receptor activation: what does the rhodopsin structure tell us? Trends Pharmacol Sci 22:587-93

60. Okada T, Ernst OP, Palczewski K, Hofmann KP 2001 Activation of rhodopsin: new insights from structural and biochemical studies. Trends Biochem Sci 26:318-24

61. Baldwin JM, Schertler GF, Unger VM 1997 An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. Journal of Molecular Biology 272:144-164

62. Bai M, Trivedi S, Brown EM 1998 Dimerization of the extracellular calcium-sensing receptor (CaR) on the cell surface of CaR-transfected HEK293 cells. J Biol Chem 273:23605-10

63. White JH, Wise A, Main MJ, Green A, Fraser NJ, Disney GH, Barnes AA, Emson P, Foord SM, Marshall FH 1998 Heterodimerization is required for the formation of a functional GABA(B) receptor. Nature 396:679-82

64. Jones KA, Borowsky B, Tamm JA, Craig DA, Durkin MM, Dai M, Yao WJ, Johnson M, Gunwaldsen C, Huang LY, Tang C, Shen Q, Salon JA, Morse K, Laz T, Smith KE, Nagarathnam D, Noble SA, Branchek TA, Gerald C 1998 GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. Nature 396:674-9

65. Kaupmann K, Malitschek B, Schuler V, Heid J, Froestl W, Beck P, Mosbacher J, Bischoff S, Kulik A, Shigemoto R, Karschin A, Bettler B 1998 GABA(B)-receptor subtypes assemble into functional heteromeric complexes. Nature 396:683-7

66. Romano C, Yang WL, O'Malley KL 1996 Metabotropic glutamate receptor 5 is a disulfide-linked dimer. J Biol Chem 271:28612-6

67. Margeta-Mitrovic M, Jan YN, Jan LY 2000 A trafficking checkpoint controls GABA(B) receptor heterodimerization. Neuron 27:97-106

68. Kuner R, Kohr G, Grunewald S, Eisenhardt G, Bach A, Kornau HC 1999 Role of heteromer formation in GABAB receptor function. Science 283:74-7

69. Sullivan R, Chateauneuf A, Coulombe N, Kolakowski LF Jr, Johnson MP, Hebert TE, Ethier N, Belley M, Metters K, Abramovitz M, O'Neill GP, Ng GY 2000 Coexpression of full-length gamma-aminobutyric acid(B) (GABA(B)) receptors with truncated receptors and metabotropic glutamate receptor 4 supports the GABA(B) heterodimer as the functional receptor. J Pharmacol Exp Ther 293:460-7

70. Havlickova M, Prezeau L, Duthey B, Bettler B, Pin JP, Blahos J 2002 The intracellular loops of the GB2 subunit are crucial for G-protein coupling of the heteromeric gamma-aminobutyrate B receptor. Mol Pharmacol 62:343-50

71. Duthey B, Caudron S, Perroy J, Bettler B, Fagni L, Pin JP, Prezeau L 2002 A single subunit (GB2) is required for G-protein activation by the heterodimeric GABA(B) receptor. J Biol Chem 277:3236-41

72. Armstrong D, Strange PG 2001 Dopamine D2 receptor dimer formation: evidence from ligand binding. J Biol Chem 276:22621-9

73. Lee SP, O'Dowd BF, Ng GY, Varghese G, Akil H, Mansour A, Nguyen T, George SR 2000 Inhibition of cell surface expression by mutant receptors demonstrates that D2 dopamine receptors exist as oligomers in the cell. Mol Pharmacol 58:120-8

74. Ng GY, O'Dowd BF, Lee SP, Chung HT, Brann MR, Seeman P, George SR 1996 Dopamine D2 receptor dimers and receptor-blocking peptides. Biochem Biophys Res Commun 227:200-4

75. Zawarynski P, Tallerico T, Seeman P, Lee SP, O'Dowd BF, George SR 1998 Dopamine D2 receptor dimers in human and rat brain. FEBS Lett 441:383-6

76. Rios CD, Jordan BA, Gomes I, Devi LA 2001 G-protein-coupled receptor dimerization: modulation of receptor function. Pharmacol Ther 92:71-87

77. George SR, O'Dowd BF, Lee SP 2002 G-protein-coupled receptor oligomerization and its potential for drug discovery. Nat Rev Drug Discov 1:808-20

78. Angers S, Salahpour A, Bouvier M 2002 Dimerization: an emerging concept for G protein-coupled receptor ontogeny and function. Annu Rev Pharmacol Toxicol 42:409-35

79. Angers S, Salahpour A, Joly E, Hilairet S, Chelsky D, Dennis M, Bouvier M 2000 Detection of beta 2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET). Proc Natl Acad Sci U S A 97:3684-9

80. Hebert TE, Moffett S, Morello JP, Loisel TP, Bichet DG, Barret C, Bouvier M 1996 A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. J Biol Chem 271:16384-92

81. Van Brocklyn JR, Behbahani B, Lee NH 2002 Homodimerization and heterodimerization of S1P/EDG sphingosine-1- phosphate receptors. Biochim Biophys Acta 1582:89-93

82. Terrillon S, Cheng LL, Stoev S, Mouillac B, Barberis C, Manning M, Durroux T 2002 Synthesis and characterization of fluorescent antagonists and agonists for human oxytocin and vasopressin V(1)(a) receptors. J Med Chem 45:2579-88

83. Jordan BA, Devi LA 1999 G-protein-coupled receptor heterodimerization modulates receptor function. Nature 399:697-700

84. George SR, Fan T, Xie Z, Tse R, Tam V, Varghese G, O'Dowd BF 2000 Oligomerization of mu- and delta-opioid receptors. Generation of novel functional properties. J Biol Chem 275:26128-35

85. Gomes I, Jordan BA, Gupta A, Trapaidze N, Nagy V, Devi LA 2000 Heterodimerization of mu and delta opioid receptors: A role in opiate synergy. J Neurosci 20:RC110

86. Maggio R, Vogel Z, Wess J 1993 Coexpression studies with mutant muscarinic/adrenergic receptors provide evidence for intermolecular "cross-talk" between G-protein- linked receptors. Proc Natl Acad Sci U S A 90:3103-7

87. Rocheville M, Lange DC, Kumar U, Patel SC, Patel RC, Patel YC 2000 Receptors for dopamine and somatostatin: formation of hetero-oligomers with enhanced functional activity. Science 288:154-7

88. Jordan BA, Trapaidze N, Gomes I, Nivarthi R, Devi LA 2001 Oligomerization of opioid receptors with beta 2-adrenergic receptors: a role in trafficking and mitogen-activated protein kinase activation. Proc Natl Acad Sci U S A 98:343-8

89. McVey M, Ramsay D, Kellett E, Rees S, Wilson S, Pope AJ, Milligan G 2001 Monitoring receptor oligomerization using time-resolved fluorescence resonance energy transfer and bioluminescence resonance energy transfer. The human delta -opioid receptor displays constitutive oligomerization at the cell surface, which is not regulated by receptor occupancy. J Biol Chem 276:14092-9

90. AbdAlla S, Lother H, Quitterer U 2000 AT1-receptor heterodimers show enhanced G-protein activation and altered receptor sequestration. Nature 407:94-8

91. Ciruela F, Escriche M, Burgueno J, Angulo E, Casado V, Soloviev MM, Canela EI, Mallol J, Chan WY, Lluis C, McIlhinney RA, Franco R 2001 Metabotropic glutamate 1alpha and adenosine A1 receptors assemble

into functionally interacting complexes. J Biol Chem 276:18345-51

92. Gouldson PR, Snell CR, Bywater RP, Higgs C, Reynolds CA 1998 Domain swapping in G-protein coupled receptor dimers. Protein Eng 11:1181-93

93. Gouldson PR, Higgs C, Smith RE, Dean MK, Gkoutos GV, Reynolds CA 2000 Dimerization and domain swapping in G-protein-coupled receptors: a computational study. Neuropsychopharmacology 23:S60-77

94. Zhang Z, Sun S, Quinn SJ, Brown EM, Bai M 2001 The extracellular calcium-sensing receptor dimerizes through multiple types of intermolecular interactions. J Biol Chem 276:5316-22

95. Kunishima N, Shimada Y, Tsuji Y, Sato T, Yamamoto M, Kumasaka T, Nakanishi S, Jingami H, Morikawa K 2000 Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. Nature 407:971-7

96. Guo W, Shi L, Javitch JA 2003 The fourth transmembrane segment forms the interface of the dopamine d2 receptor homodimer. J Biol Chem 278:4385-8

97. Filizola M, Weinstein H 2002 Structural models for dimerization of G-protein coupled receptors: the opioid receptor homodimers. Biopolymers 66:317-25

98. Filizola M, Olmea O, Weinstein H 2002 Prediction of heterodimerization interfaces of G-protein coupled receptors with a new subtractive correlated mutation method. Protein Eng 15:881-5

99. Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K 2003 Atomic-force microscopy: Rhodopsin dimers in native disc membranes. Nature 421:127-8

100. Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A 2003 Organization of the G Protein-coupled Receptors Rhodopsin and Opsin in Native Membranes. J Biol Chem (in press).

101. Devi LA 2001 Heterodimerization of G-protein-coupled receptors: pharmacology, signaling and trafficking. Trends Pharmacol Sci 22:532-7

102. Gazi L, Lopez-Gimenez JF, Strange PG. 2002 Formation of oligomers by G protein-coupled receptors. Curr Opin Drug Discov 5, 756-63.

103. Kraulis J 1991 MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. J. Appl. Cryst. 24:946-950

104. Merritt EA, Bacon DJ 1997 Raster3D: Photorealistic Molecular Graphics. Methods in Enzymology 277:505-524

105. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB 2003 The g-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63:1256-72

# Chapter 2
# Methods

This section is aimed to review briefly a few general aspects of some of the bioinformatics tools used within this thesis. The particular methodology used in each specific case study is detailed in the 'Results' chapter under each section of 'Methods'.

## *2.1. Homology modelling of GPCRs*

In homology modelling techniques, a protein sequence with unknown three-dimensional structure is mapped against a known structure of one or several similar proteins. Therefore the known structure is used as a template for modelling the unknown structure of the target protein. This method is based in assuming that the template and target proteins will have reasonable structural similarity as long as they have similar origin and function (1;2).

In homology modelling of GPCRs, the three-dimensional structure of bovine rhodopsin (3) is used as a template. Only the structure of the transmembrane bundle is likely to be conserved in the Class A of GPCRs, so only the coordinates of the TMHs, as in the rhodopsin crystal, have been used in the homology modelling procedure. Therefore, the sequence of the transmembrane region of the target receptor is superimposed in this template. The superimposition is done using the amino acid sequence alignment between rhodopsin and the target GPCR (4), and using as anchors the position of the highly conserved motifs in each TMH (Fig. 2.1).

Following this procedure, a model for the three dimensional structure of the backbone is built. Then, the sidechains of the conserved residues between rhodopsin and the target receptor are left in the same conformation that in the crystal, while the non-conserved sidechains are placed according a backbone dependent library of side chain conformations (5). During the placement of the sidechains each rotamer is chosen so it is not clashing with nearby sidechains. However, this automatized procedure can lead to some non-optimal interactions. Therefore, this initial model is first checked manually in order to be sure that some key interactions are kept (for instance, the interaction between the conserved Asp/Glu$^{3.49}$ and Arg$^{3.50}$). Each individual receptor will have a set of interactions proposed in base of

biochemical or biophysical studies, which can be satisfied in this initial model by adjusting manually the side chain conformations.

Then, the positions of this side chains are optimized using an energy minimization protocol, which will lead the system to the nearest energy minimum, basically through the optimization of van der Waals and electrostatic interactions. Around 500-1000 steps of minimization will usually suffice to reach a local energy minimum.

With this protocol we have built an initial model of the target receptor, which consists in the rhodopsin backbone with an initial configuration for the side chains. It is important to note that, if this initial model is used as a starting point in a molecular dynamics study, many of the side chain conformations are likely to be kept, so it is important to check this initial structure in some detail.

## *2.2. Principles of molecular dynamics*

Dynamic properties of the transmembrane α-helices have been studied using molecular dynamics (MD) simulations. MD is a technique to compute the equilibrium and transport properties of a classical many-body system (6;7). The potential energy of the system could, in principle, be solved from the electronic structure of the atoms. However, in the case of complex molecular systems, the Born-Oppenheimer approximation is applied, *i.e.*, electronic motions are included implicitly in the parameterization and the potential energy is calculated only as a function of nuclear positions. Then, the motion of the constituent particles is considered to obey the laws of classical mechanics. In practice, Newton's equations of motion are solved for a model system until it is equilibrated, *i.e.*, its properties no longer change with time. This technique is used to study large systems, as nucleic acids or proteins, with tenths of thousands of atoms, far above the capabilities of current computer implementation of *ab-initio* or semi-empiric methods.

A MD algorithm comprises the following steps (7):

a) read in the parameters (initial temperature, number and characteristics of the particles which compose the system, …)

b) set initial positions and velocities of all the particles. Velocities are assigned from a Maxwell-Boltzmann distribution at the initial temperature of the system. If this temperature is 0 K, the velocities will be calculated from the forces instead

c) compute forces on all particles. If there are no external influences, each atom moves according to the influence of the rest of the atoms of the system. The force on each atom is calculated as the negative gradient of the potential energy function

$$\boldsymbol{F}_i = -\frac{\partial E}{\partial r_i} \quad i = 1,\mathrm{K}\ , N$$

where $\boldsymbol{F}_i$ is the force on the atom $i^{th}$ and N=number of atoms.

In order to calculate the energy of the system we need a potential energy surface as a function of certain coordinates of the system. This potential energy surface is calculated using molecular mechanics. In this procedure, the potential energy on each atom is split in the contribution of bonded and non-bonded atoms (8) (Fig. 2.2).

$$\mathrm{E} = \mathrm{E}_{\mathrm{bonded}} + \mathrm{E}_{\mathrm{non\text{-}bonded}}$$

However, the exact term describing the potential energy may vary with each particular system. For instance, extra terms accounting for the planarity of improper angles or specific constraints can be used.

The *bonded* term includes atoms connected by one (bond term), two (angle term) or three (dihedral term) consecutive bonds:

$$\mathrm{E}_{\mathrm{bonded}} = \mathrm{E}_{\mathrm{bond}} + \mathrm{E}_{\mathrm{angle}} + \mathrm{E}_{\mathrm{dihedral}}$$

The bond and angle term are described as harmonic oscillators:

$$E_{bond} = \sum_{bonds} k_b (b - b_0)^2$$

where $b$ is the distance between atoms, $b_0$ is the equilibrium bond length and $k_b$ is a force constant.

$$E_{angle} = \sum_{angles} k_\theta (\theta - \theta_0)^2$$

where $\theta$ is the angle between atoms, $\theta_0$ is the angle of equilibrium and $k_\theta$ is a force constant.

The dihedral term is represented as a cosine expansion:

$$E_{dihedral} = \sum_{dihedrals} k_\phi (1 + \cos(n(\phi - \phi_0))$$

where $\phi$ is the value of the dihedral angle, $\phi_0$ is the equilibrium value, $k\phi$ is a force constant affecting the barrier height and $n$ the multiplicity (number of minima in the function as the bond is rotated 360º).

The *non-bonded* term of the potential energy includes electrostatic and van der Waals interactions between pairs of atoms.

$$\mathrm{E}_{\mathrm{non\text{-}bonded}} = \mathrm{E}_{\mathrm{electrostatic}} + \mathrm{E}_{\mathrm{vdw}}$$

Electrostatics is described using a Coulombic term:

$$E_{electrostatics} = \sum_{i<j} \frac{q_i q_j}{\varepsilon r_{ij}}$$

where $q_i$ and $q_j$ are the point charges, $r_{ij}$ is the distance and ε is the dielectric constant of the medium.

The van der Waals interaction, which accounts for the forces of repulsion and dispersion between atoms, is represented using a Lennard-Jones term:

$$E_{vdw} = \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)$$

where $A_{ij}$ and $B_{ij}$ are the strength parameters for repulsion and dispersion, respectively, which are related to the 'hardness' of the atoms, the depth of the energy well and the respective van der Waals radii of the particles.

The calculation of the pairwise interactions in the *non-bonded* term of the potential energy is the most time-consuming part of MD algorithms, as the time needed for evaluation of the forces scales as $N^2$, where N is the number of atoms of the system (7). However, there are techniques to speed up the evaluation of forces. For instance, in order to limit the number of pairs, cut-off distances, usually combined with smoothing functions, can be defined for both electrostatic and van der Waals terms. As van der Waals interactions are of a short-range nature, the use of a cut-off is a good approximation. But electrostatics is, by nature, a long-range interaction and the use of a cut-off can lead to inaccuracies (9;10). On the other hand, when periodic boundary conditions have been defined for the system, the Ewald summations method can be used. This method takes into account explicitly the all the electrostatic interactions in the infinite array of periodic replicas of the central simulation cell, so the electrostatic term is, in principle, exact. The periodicity and symmetry of the system are used to express the electrostatic potential as two summations that are convergent (7). This method is computationally expensive, but the gain in accuracy in the calculation of the energy is worth it. The Ewald method is usually implemented using the Particle Mesh Ewald method (11;12), where the sums are calculated via fast Fourier transforms..

As seen in the above formulaes, potential energy is determined by a number of parameters ($b_0$, $k_b$, $\theta_b$, $k_\theta$, ...), which are determined experimentally (for instance bond parameters can be estimated from IR, Raman and microwave spectra, or from electronic and X-ray diffraction patterns) or using quantum mechanics calculations (for instance for the calculation of electrostatic charges).

d) When all forces between particles have been computed, we can integrate Newton's equations of motion. Positions and velocities of the atoms are allowed to evolve according to the Newtonian equation of motion.

$$-\frac{\partial E}{\partial \vec{r}_i} = m_i \frac{\partial^2 \mathbf{r}_i(t)}{\partial t^2}, \quad i = 1,\mathrm{K}\,,N$$

where $\mathbf{r}_i(t)$ is the position of the atom $i^{th}$ at the time $t$.

The differential equations of motion are integrated numerically using an algorithm according to which the particle positions and velocities are updated. The integration algorithms are usually based in Taylor

expansions, as the Verlet algorithm, which is fast and exhibits a little long-term energy drift.

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \frac{\mathbf{F}_i(t)}{m}(\Delta t)^2 + O\left[(\Delta t)^4\right]$$

The estimate of the new position contains an error that is of order $\Delta t^4$. The Verlet algorithm does not use the velocity to compute the new position, but it is derived from knowledge of the trajectory using a central distance formula.

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{2\Delta t} + O\left[(\Delta t)^2\right]$$

which is only accurate to order $\Delta t^2$.

Thus, using this algorithm, velocity and position of each particle after a certain time step can be calculated from the last velocity and position.

c) and d) steps are repeated until the system has been simulated during the desired length of time.

If we assume that the *ergodic hypothesis* applies to the system that we are studying, many properties of the system, as energy, heat capacity, temperature or pressure, can be calculated at each time step of the simulation. The *ergodic hypothesis* states that if we wish to compute the average of a function of the coordinates and momenta of a many-particles system, the mean value of the property is equivalent to the average of this property during time. So, to measure a property in a MD simulation we must be able to express it as a function of the positions and momenta of the particles in the system, and study the evolution of the system during a long enough time

For instance, the temperature of the system can be calculated from the velocities, using statistical mechanics:

$$T(t) = \frac{1}{(3N - n)k_B}\sum_{i=1}^{N} m_i \left|v_i(t)\right|^2$$

where $N$ is the number of atoms, $(3N-n)$ is the number of degrees of freedom, $k_B$ is the Boltzmann constant and $v_i(t)$ is the velocity of the atom $i$ at the time $t$.

The MD simulations using the methane box environment (see Chapter 2.3) are carried out in an NVT ensemble, which is characterized by a constant number of particles, constant volume and constant temperature. Temperature is kept constant coupling the system to a heat bath using the Berendsen algorithm, which re-scales the velocities at each simulation step. On the other hand, periodic boundary conditions are needed to carry out a MD simulation in constant volume. When the system is solvated using an explicit solvent, it is necessary to define a box which contains all the atoms of the system. This box, which is usually a parallelepiped, is copied in each

direction, so the origin cell is surrounded by copies of itself simulating an infinite system. While the motions are computed only for the central box, the calculation of pairwise interactions may involve atoms of the surrounding boxes. A particular particle in different boxes is under the same forces, so it moves in the same way and, if a particle leaves the unit box, it will enter by the opposite side. The solvent box has to be large enough to avoid interactions between solutes in different boxes. Moreover the use of periodic boundary conditions permits the reduction of abnormal behaviour of the particles at the boundaries. Finally, if pressure has to kept constant, the positions are re-scaled using a factor related with the isotermic compressibility.

The integration time step, $\Delta t$ in the above expressions for the calculation of positions and velocities, is the increment of time between each new force calculation over each particle. It has to be short enough to make the correction tems in the Taylor expansions as small as possible, thus ensuring the validity of numeric integrations. In addition, a short time step avoids large force variations between two consecutive steps, which would lead to high velocities and a system blow-up. However, the time step affects inversely the real time needed to perform the simulation, as if it is decreased, the number of force calculations will increase. Moreover, if has to be large enough to effectively sample the phase space. The fastest motions determine the size of the time step in the system, which are the bond vibrations. The SHAKE algorithm (13) removes this bond stretching freedom, allowing a larger time steps to be used. Typical values for the time step are in the order of femtoseconds ($10^{-15}$ s). In our simulations we use a 2 fs time step

Molecular dynamics simulations have some inherent drawbacks. Due to the small scales of time used in this technique (~ ns) simulations of certain processes are not possible:

a) those which need long simulation times, such as the encounter between a ligand and its receptor

b) fast but relatively infrequent processes, such as the rotation of an aminoacid side chain

c) processes which involve the creation or disruption of covalent bonds

## 2.3. The methane box as a model for the membrane core

As we are studying only the transmembrane region of membrane receptors, the environment of our model system corresponds to the core of the cell membrane. This apolar environment has been mimicked using a box of methane molecules with a density about half of the membrane core. This is the maximum density which could be reached to have stable MD simulations, and corresponds to the density of methane molecules at its equilibrium distance (14). This approximation was chosen because it leads

to relatively small systems (~20000 atoms) and because of the relatively easy setup and equilibration process of the methane box. We have shown that this environment, when combined with the use of the Ewald method to compute the electrostatic term of the energies, leads to a correct description of the geometrical properties of trasmembrane helices (14). It has to be noted that this environment forces the use of constant volume algorithms when periodic boundary conditions are required.

In order to prepare the system for MD simulations, the following procedure is followed. First, the initial model of the transmembrane region of the receptor is solvated in a cubic box of methane molecules, with a size of ~60×36×36 Å (around 1600 methane molecules).

This system is subjected to a pre-equilibration protocol consisting in:

- energy minimization of the solvent.
- equilibration of the solvent with 50 ps of molecular dynamics. The solvent is heated from 0K to 300K in 35 ps, and then it is simulated for further 15 ps at this temperature
- energy minimization of the whole system.

Although the methane box environment has shown its utility in the study of transmembrane region of membrane receptors, it has some important inherent drawbacks:

- we are neglecting the influence of the phospholipid polar heads and, probably more important, of the water, in the zone of transmembrane regions near to the extracellular or cytoplasmic interface
- because of the nature of the methane box approximation, we are not taking into account the loops, N-terminal and C-terminal regions of the receptor. Although the modelling of these regions is far less straightforward than the TM bundle, and the models including them may be highly speculative, these regions have to be taken into account in order to model the structure or function of the whole receptor with more reliability.

For this reason we have recently started to use models of explicit lipidic bilayers using POPC as model phospholipid. Using ~90 POPC molecules and ~10000 water molecules we have a system size of ~70×60×80 Å, with a total of ~45000 atoms. Although these systems are larger, they can take advantage of special fast routines used for the simulation of water molecules. The modelling of the loops, N-terminal and C-terminal regions of the receptor will be tailor-made for each GPCR studied.

## References


1. Sali A 1995 Modeling mutations and homologous proteins. Curr Opin Biotechnol 6:437-51

2. Orengo CA, Todd AE, Thornton JM 1999 From protein structure to function. Curr Opin Struct Biol 9:374-82

3. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M 2000 Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289:739-45

4. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G 2003 GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 31:294-7

5. Bower MJ, Cohen FE, Dunbrack RL Jr 1997 Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. J Mol Biol 267:1268-82

6. Allen MP, Tildesley DJ 1987 Computer simulation of liquids., ed. Clarendon Press, Oxford

7. Frenkel D, Smit B 1996 Understanding Molecular Simulations from Algorithms to Applications, ed. Academic Press,

8. AMBER 7, University of California, San Francisco. Case DA, Pearlman DA, Caldwell JW, Cheatham TE, Wang W.S. , Ross WS, Simmerling TA, Darden TA, Merz KM, Stanton RV, Cheng AL, Vicent JJ, Crowley M, Tsui H, Gohlke RJ, Radmer Y, Duan J, Pitera I, Massova GL, Seibel GL, Singh UC, Weiner PK, Kollman PA. 2002. University of California San Francisco.

9. Alper HE, Bassolino-Klimas D, Stouch TR 1993 The limiting behaviour of water hydrating a phospholipid monolayer: A computer simulation study. J. Chem. Phys 99:5547-5559

10. Alper HE, Bassolino-Klimas D, Stouch TR 1993 Computer simulation of a phospholipid monolayer-water system: the influence of long range forces on water structure and dynamics. J. Chem. Phys. 98:9798-9807

11. Darden T.A., York D, Pedersen P 1993 Particle Mesh Ewald: An N log(N) method for Ewald sums in large systems. J. Chem.Phys. 98:10089-10092

12. Essmann U, Perera. L., Berkowitz L 1995 A Smooth Particle Mesh Ewald Method. J. Chem. Phys. 103:8577-8593

13. Ryckaert J, Ciccotti G, Berendsen H 1977 Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23:327-341

14. Olivella M, Deupi X, Govaerts C, Pardo L 2002 Influence of the Environment in the Conformation of alpha-Helices Studied by Protein Database Search and Molecular Dynamics Simulations. Biophys J 82:3207-13

# Chapter 3
# Results

## *3.1. Serine and threonine residues bend α–helices in the χ1 = g– conformation*

Juan A. Ballesteros[*#], Xavier Deupi[*], Mireia Olivella[*], Eric E. J. Haaksma[†], Leonardo Pardo[*]

[*]Laboratori de Medicina Computacional. Unitat de Bioestadística. Facultat de Medicina. Universitat Autònoma de Barcelona. 08193 Bellaterra, Spain.

[#]Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, NY 10029, U.S.A.

[†]Department of Chemistry, Boehringer Ingelheim Pharma KG, 88400 Biberach a.d. Riss, Germany.

## Abstract

The relationship between the Ser, Thr, and Cys side–chain conformation ($\chi_1$ = *g*–, *t*, *g*+) and the main chain conformation ($\phi$ and $\psi$ angles) has been studied in a selection of protein structures that contain α–helices. The statistical results show that the *g*– conformation of both Ser and Thr residues decreases their ϕ angles and increases their $\psi$ angles relative to Ala, used as a control. The additional hydrogen bond formed between the $O_\gamma$ atom of Ser and Thr and the *i–3* or *i–4* peptide carbonyl oxygen induces or stabilizes a bending angle in the helix 3–4° larger than for Ala. This is of particular significance for membrane proteins. Incorporation of this small bending angle in the transmembrane α–helix at one side of the cell membrane results in a significant displacement of the residues located at the other side of the membrane. We hypothesize that local alterations of the rotamer configurations of these Ser and Thr residues may result in significant conformational changes across transmembrane helices, and thus participate in the molecular mechanisms underlying transmembrane signalling. This finding has provided the structural basis to understand the experimentally observed influence of Ser residues on the conformational equilibrium between inactive and active states of the receptor, in the neurotransmitter subfamily of G protein–coupled receptors.

## 3.1.1. Introduction

Wide ranges of biologically active substances, such as neurotransmitters, elicit their action through signal transduction pathways that involve membrane proteins like *G*–protein coupled receptors (GPCRs). The membrane–bound domain of GPCRs adopts the conformation of a bundle of seven transmembrane helices (TMH) (1;2). Pharmacological and mutagenesis studies (see (3) for a review) have shown that neurotransmitters bind, at the extracellular side of the membrane, with their protonated amine to the conserved $Asp^{3.32}$ (nomenclature of (25)), in TMH 3. Similarly identified (3) are a series of conserved Ser residues ($Ser^{5.43}$ and $Ser^{5.46}$), in TMH 5, which act as hydrogen–bonding sites for the hydroxyl groups present in the chemical structure of many neurotransmitters. The molecular function of constitutively active receptors (5;6) and transgenic mice with receptor over expression (7) provides direct evidence that GPCRs exist in equilibrium between inactive and active states. Spectroscopic studies (see (8) for a review) have suggested the movement of TMH 3 and TMH 6 during the formation of the active form of the receptor. Moreover, it has recently been shown that the Ser residues in TMH 5 do not only provide a docking site for the agonist, but also control the equilibrium of the receptor between both conformational states (9). Deletion of these –*OH* groups from the $\beta_2$–adrenergic receptor (Ala replacement of $Ser^{5.43}$ and $Ser^{5.46}$) decreases the constitutive activity of the receptor (9). Therefore, the side chain of Ser has a significant effect on the conformation of the helix and in consequence of the receptor.

A pioneer survey of protein α–helices in 'hydrophilic' and 'hydrophobic' environments revealed that additional hydrogen bonds between the peptide carbonyl oxygen to a solvent molecule produce a significant change in the main chain torsion $\phi$ and $\psi$ angles and in the curvature of the helix (10). It has also been shown that Ser, Thr, and Cys residues might form an intrahelical hydrogen bond between the $O_\gamma$ (or $S_\gamma$) atom and the *i–3* or *i–4* carbonyl oxygen (11). This hydrogen bond interaction between side–chain and main chain atoms is feasible in the $\chi_1$ = *gauche*– (*g*–) or $\chi_1$ = *gauche*+ (*g*+) conformation (12). It does not occur in the $\chi_1$ = *trans* (*t*) conformation. We aim to explore the possibility that this intrahelical hydrogen bond of the polar side chain of Ser, Thr, and Cys could change the conformation of the α–helix. This would provide the structural basis to understand the experimentally observed influence of Ser on the conformational equilibrium between inactive and active states of the receptor (9). We have analyzed the relationship between the Ser, Thr, and Cys side chain conformation (the torsion $\chi_1$ angle) and the main chain conformation (the torsion $\phi$ and $\psi$ angles and bend angle) in two independent samples of protein structures. First, in the four available helix bundle membrane protein structures: bacteriorhodopsin (13), cytochrome c oxidase (4), the photosyntethic reaction center (14), and the potassium channel (15). Second, in a selection of soluble proteins that contains α–helices.

## 3.1.2. Methods

### *Membrane protein structures*

The atomic coordinates of *H. Halobium* bacteriorhodopsin (PDB access number 2brd, 3.5 Å resolution), bovine cytochrome c oxidase (1occ, 2.8 Å), *R. Sphaeroides* photosyntethic reaction center (1aij, 2.2 Å), and *S. Lividans* potassium channel (1bl8, 3.2 Å) were obtained from the Brookhaven Protein DataBank (16). The coordinates of the residues corresponding to transmembrane helices 1–7 of 2brd; 2–3, 7, 9, 12, 14–15, 19–20, 23, 28–30, 32–35, 41, 54, 59–60, and 63–66 of 1occ; 2, 5–6, 11, 13, 17, 22–23, 28, 31–32, and 34 of 1aij; and 1 and 3 of 1bl8, in the HELIX annotation of the PDB files, were extracted for analysis. This results in a total of 45 TMHs. These TMHs were split into amino acid stretches of 12 residues long with Ala (standard α–helix used as control), Cys, Ser, or Thr at the 8$^{th}$ position. Stretches with Pro residues in the sequence were removed from the database. The side–chain conformation of Ser, Thr, and Cys was categorized into *g*– (0°<$\chi_1$<120°), *t* (120°<$\chi_1$<240°), or *g*+ (240°<$\chi_1$<360°) depending on the value of the torsional $\chi_1$ angle. The following distribution of residues and conformations were observed: Ala (48), Cys (4; *g*+: 4, *t*: 0, *g*–: 0), Ser (34; *g*+: 16, *t*: 5, *g*–: 13), and Thr (41; *g*+: 32, *t*: 0, *g*–: 9).

### *Soluble protein structures*

Iditis 3.1 (Oxford Molecular) was used for the selection of protein structures in the Brookhaven Protein DataBank (16). The chosen α–helices possess: *i)* a resolution of 2.0 Å or better; *ii)* 12 residues length with either Ala, Cys, Ser, or Thr at the 8$^{th}$ position; and *iii)* no Pro residues in the sequence. If two α–helical segments have more than 80% sequence identity (if 10 or more than 10 residues out of 12 are identical) only the structure with best resolution was considered. This systematic search provided the following distribution of residues and conformations: Ala (730), Cys (66; *g*+: 46, *t*: 20, *g*–: 0), Ser (245; *g*+: 129, *t*: 74, *g*–: 42), and Thr (247; *g*+: 211, *t*: 2, *g*–: 34).

### *Statistical analysis*

The torsion angles of the backbone of the residues at positions 8, populated by Ala, Cys, Ser, or Thr ($\phi_i$ and $\psi_i$); 7 ($\phi_{i-1}$ and $\psi_{i-1}$); 6 ($\phi_{i-2}$ and $\psi_{i-2}$); 5 ($\phi_{i-3}$ and $\psi_{i-3}$); and 4 ($\phi_{i-4}$ and $\psi_{i-4}$) were calculated for statistical analysis with SAS 6.11 (SAS Institute, Cary, NC). One–way analysis of variance plus a posteriori two–sided Dunnett's T tests was employed for contrasting the calculated torsion angles in Ser, Thr, and Cys residues in the *g*–, *t*, and *g*+ rotamer conformations with the control Ala in both membrane and soluble proteins. No statistical differences were observed in the torsion $\phi_{i-1}$, $\phi_{i-2}$, $\phi_{i-3}$, $\phi_{i-4}$, $\psi_{i-1}$, $\psi_{i-2}$, $\psi_{i-3}$, and $\psi_{i-4}$ angles in both membrane and soluble proteins. The only exceptions (2 out of 112 comparisons) were found in $\phi_{i-3}$ in Cys/*g*+ and $\psi_{i-4}$ in Thr/*g*+ for soluble proteins (results not shown). These two exceptions were not further considered because of the lack of consistency among residues, conformational classes, or protein type.

The bend angle of the amino acid stretches of 12 residues long was calculated from the two axes that minimize the distance to the main chain atoms of residues 1–4 and 9–12 (17). One–way analysis of variance plus a posteriori one–sided Dunnett's T tests was employed to contrast if the bend angle of Ser, Thr, and Cys residues in the *g*–, *t*, and *g*+ rotamer conformations is greater than the control Ala in the sample of soluble proteins.

The $\chi^2$ distribution was employed to compare the frequencies of residues and conformations in membrane and soluble proteins.

## 3.1.3. Results

Table 3.1.1 summarizes the means and standard deviations for the backbone $\phi_i$ and $\psi_i$ dihedral angles of α–helices containing Ala (standard α–helix used as control) and Ser, Thr, and Cys residues in the three possible rotamer conformations: *g*–, *t*, and *g*+. The histograms in Fig. 3.1.1 depict the mean values and the lines extending from the bar represent the standard deviation of $\phi_i$ (panels a and b) and $\psi_i$ (panels c and d) dihedral angles. The results are presented for membrane (panels a and c) and soluble (panels b and d) proteins. The difference in degrees ($\Delta$) relative to the control Ala (black solid bar in Fig. 3.1.1) is also shown in Table 3.1.1.

### *The g– conformation*

The *g*– conformation significantly decreases $\phi_i$ ($\Delta$ of –4.3°) and increases $\psi_i$ ($\Delta$ of 10.1°), relative to Ala, in membrane proteins (Table 3.1.1). Moreover, the effect caused by both Ser and Thr is similar in magnitude. Ser/*g*– decreases $\phi_i$ in –3.6° and increases $\psi_i$ in 11.1°, whereas Thr/*g*– decreases $\phi_i$ in –5.4° and increases $\psi_i$ in 8.7° (Table 3.1.1 and Figs. 3.1.1a and 3.1.1c). However, these differences relative to Ala, calculated independently for Ser/*g*– and Thr/*g*–, are significant from a statistical point of view only in $\psi_i$. The lack of statistical significance of $\phi_i$ is attributed to the smaller number of points in the split Ser/*g*– (13 structures) and Thr/*g*– (9 structures) categories than in the total *g*– (22 structures) category (Table 3.1.1). Thus, in order to reinforce this finding of the influence of the *g*– conformation in both $\phi_i$ and $\psi_i$ angles, we have undertaken a similar analysis in soluble proteins for which larger number of high resolution structures are available (see methods). The *g*– conformation of Ser and Thr residues in soluble proteins has a statistically significant effect in both $\phi_i$ and $\psi_i$ (Table 3.1.1). Notably, the magnitude and direction of the effect is the same as observed in membrane proteins. The *g*– conformation of Ser and Thr decreases $\phi_i$ ($\Delta$ of –3.4° and –6.5°, respectively) and increases $\psi_i$ (7.7° and 6.2°), relative to Ala (Table 3.1.1 and Figs. 3.1.1b and 3.1.1d). Similar behaviour in $\phi_i$ and $\psi_i$ cannot be observed in the Cys residue since no experimental data is available in either membrane or soluble proteins. The *g*– conformation of Cys is totally forbidden because of the steric clash between the $S_\gamma$ atom and the carbonyl oxygen of residue *i–3* (12).

The conformation of the α–helix, driven by the *g*– conformation of Ser or Thr is illustrated in Fig. 3.1.2. Fig. 3.1.2a shows the conformation of a poly-Ala

α–helix (red) and a poly-Ala α–helix with a single Ser or Thr (blue) residue in between. The location of either Ser or Thr in the α–helix is shown throughout the $C_\alpha$–$C_\beta$ bond. The helices were constructed with the average $\phi_i$ and $\psi_i$ angles reported in Table 3.1.1 for Ala (–60.9° and –44.4°) and the *g*– conformation (–65.2° and –34.3°) in membrane proteins. Clearly, the *g*– conformation induces a bending angle in the helix (see below). Incorporation of this bending angle at one side of the cell membrane results in a significant displacement of the residues located at the other side of the membrane. The magnitude of the relocation might be estimated from the models depicted in Fig. 3.1.2a. Thus, the distance between the alpha–carbon positions, in the straight helix (red) and the bent helix (blue), is 3.3 Å for an amino acid located 15 residues away from Ser or Thr.

Figs. 3.1.2b and 3.1.2c show the crystal structure of helix 32, which contains $Thr^{277}$ and $Thr^{279}$ in *g*–, from the photosyntethic reaction center and helix 1, which contains $Thr^{33}$ also in *g*–, from the potassium channel, respectively. To emphasize the structural consequences of the *g*– conformation, the transmembrane α–helices were superimposed to an ideal α–helix (red). The backbone atoms of the amino acids from *i* (the residue in *g*–) to *i–4* are shown as balls and sticks, whereas tube ribbons represent the rest of the backbone atoms (Fig. 3.1.2c). Remarkably, the presence of these polar residues in the *g*– conformation modifies the direction of the α–helix. The additional intrahelical hydrogen bond formed between the side chain $OH_\gamma$ of Ser or Thr and the *i–3* or *i–4* peptide carbonyl oxygen of the preceding turn seems to produce this effect (10). Fig. 3.1.2c also shows a detailed view of this hydrogen bond network in the potassium channel. The average hydrogen bond $O_\gamma \cdots O_{i-3}$ and $O_\gamma \cdots O_{i-4}$ distances (see broken lines in Fig. 3.1.2c) are 3.4 Å and 3.5 Å in membrane proteins and 3.1 Å and 3.5 Å in soluble proteins, respectively. Thus, the $O_\gamma$ atom is located between $O_{i-3}$ and $O_{i-4}$, closer in average to $O_{i-3}$. However, the small difference between the $O_\gamma \cdots O_{i-3}$ and $O_\gamma \cdots O_{i-4}$ distances and the absence of the $H_\gamma$ atom in the crystal structures does not allow identifying to which carbonyl oxygen the $OH_\gamma$ side chain preferentially hydrogen bonds.

### *The g+ conformation*

The *g*+ conformation is the most abundant rotamer conformation in both membrane and soluble proteins (Table 3.1.1). Thus, the statistical contrasts between Ala and *g*+ possess higher statistical power than between Ala and *g*– or *t*. Despite this fact, the *g*+ conformation produces a statistically significant change only in $\psi_i$ of Thr in soluble proteins ($\Delta$ of –3.0°, Table 3.1.1 and Fig. 3.1.2). The lack of consistency of this variation among protein type and the other residues (Ser and Cys) does not led us to conclude that an α–helix with Ser, Thr, or Cys in the *g*+ conformation leads to a different conformation than an α–helix with Ala.

### *The t conformation*

The hydrogen bonding capacity of either Ser, Thr, or Cys must be satisfied, in a hydrophobic environment like the cell membrane, by the hydrogen

bond interaction, in either the *g*+ or *g*– conformation, with the carbonyl oxygen in the preceding turn of the helix (11). Thus, only 5 residues in the *t* rotamer conformation are found in membrane proteins. This lack of structures prevents the statistical analysis on membrane proteins. The *t* conformation produces in soluble proteins a statistically significant change in $\psi_i$, without modifying $\phi_i$ (Table 3.1.1 and Figs. 3.1.2b and 3.1.2d). Thus, both Cys and Ser residues in the *t* conformation decrease, relative to Ala, $\psi_i$ by –5.0° and –2.8°. No statistical differences are obtained for Thr because only 2 cases are found in the analysis. The steric clash between the methyl group and the carbonyl oxygen of residue *i–3* (11) explains the lack of Thr residues in this conformation. The conformation of the α–helix caused by Ser in *t* conformation (green, $\phi_i$ of –62.7° and $\psi_i$ of –44.0°), compared with the *g*– conformation (blue, –66.5° and –33.5°) and the ideal poly-Ala (red, –63.1° and –41.2°) are illustrated in Fig. 3.1.2d. The fact that $\phi_i$ does not change and the smaller change in $\psi_i$ produced by the *t* conformation, relative to the *g*– conformation, is reflected in the reported structures. The α–helix with Ser in *t* (green) is comparable to Ala α–helix (red). However, it is important to note that the obtained changes in $\psi_i$, in *g*– and *t* conformations, occur in opposite directions (increases in *g*– and decreases in *t*, relative to Ala), which result in a bend of the helices pointing toward different positions in space (Fig. 3.1.2d).

### *Bend angle*

Bend angles of the helices are calculated from the two axes that minimize the distance to the main chain atoms of the residues at the beginning and the end of the helix (17). Thus, only 4 residues (12 atoms) at the beginning and the end of the helix are employed in the calculation of the axes. Therefore, a small variation in the undersized number of main chain atoms results in an intermediate variation in the helical axis and a large variation in the calculated bend angle. This effect is very noticeable in membrane proteins because of the low-resolution structural information available and the limited number of them. Therefore the analysis of bend angle is presented only for soluble proteins. Figure 3.1.3 and Table 3.1.2 shows the means and standard deviations for the bending angle calculated from high–resolution crystallographic structures. Notably, the *g*– conformation significantly increases the bend angle ($\Delta$ of 3.8°), relative to Ala. No statistical differences are observed for the *g*+ ($\Delta$ of 0.5°) or *t* ($\Delta$ of –0.4°) conformations. The observed statistical significance for the *g*– conformation is not preserved when the analysis is independently done for Ser/*g*– and Thr/*g*– despite the magnitude of the differences continues similar to the *g*– category: Ser/*g*– increases the bend angle 4.3° and Thr/*g*– 3.2° relative to Ala. The smaller number of points in the Ser/*g*– and Thr/*g*– categories seems responsible for this lack of significance.

### 3.1.4. Discussion

The ability of all naturally occurring amino acids to form a turn when placed in the middle of a transmembrane helix has recently been measured (18). The observed rank order for turn–stabilizing tendencies are Asn = Arg = Pro (1.7) > Asp = Glu = His = Lys = Gln = (1.6) > Gly (1.3) > Ser = Trp (0.7) > Cys = Ile = Tyr (0.6) > Ala = Met = Val (0.5) > Leu = Phe = Thr (0.4). Clearly, there are two sets of residues with either high (≥ 1.3) or low (≤ 0.7) turn propensity. Charged or polar residues induce a turn (≥ 1.3), whereas hydrophobic residues plus Ser, Thr and Cys remain α–helical (≤ 0.7). Moreover, statistical analysis of transmembrane sequences has shown that the most frequent amino acids are Leu > Ile > Val > Ala > Phe > Gly > Ser > Thr (19). These amino acids comprise more than two–thirds of the total. Thus, Ser and Thr are regularly found in transmembrane segments. Consistent with these findings, the ratio of Ala:Ser:Thr:Cys residues found in the present survey of protein α–helices is 12 : 8.5 : 10.2 : 1 in membrane proteins and 11.1 :3.7 : 3.7 : 1 in soluble proteins. Ser and Thr residues occur almost as often as Ala in membrane proteins and three times less in soluble proteins. In addition, the ratio of $g^+$ : $g^-$ for Ser and Thr residues are 1.2 : 1 and 3.5 : 1 in membrane proteins and 3.1 : 1 and 6.2 : 1 in soluble proteins, respectively. There is a noticeable increase of the population of $g^-$ conformation if the α–helix is embedded in a hydrophobic environment like the cell membrane. Notably, Ser possesses as many side chains in $g^-$ as in $g^+$ in membrane proteins. These findings suggest a structural role of Ser and Thr residues in transmembrane segments. We have shown that the presence of Ser and Thr residues adopting the $g^-$ conformation correlates with a significant bending of the α-helix at this locus. Therefore, we hypothesize that local alterations of the rotamer configurations of these Ser and Thr residues may result in significant conformational changes across transmembrane α-helices, and thus participate in the molecular mechanisms underlying transmembrane signalling.

It should be noted that the statistical correlation found between Ser and Thr adopting the g- conformation and helix bending does not clarify whether the Ser/Thr side chain induces or stabilizes the observed helix bending. However, we would favour the causal relationship between side chain to main chain hydrogen bonding and helix bending, following the argument put forward by Blundell et al. (10). The authors compared the 180° angle of a linear *NH···O=C* α-helical backbone hydrogen bond that occurs in a straight helix, with the 120° of the same angle in a bifurcated *(NH, HOH)···O=C* hydrogen bonding when a water molecule also hydrogen bonds the backbone carbonyl. This difference in the hydrogen bonding angle would explain the characteristic bending observed in high resolution α-helical structures, where the water exposed face is bent (120°) relative to the more straight (180°) buried face of the helix (10). For the case of the Ser and Thr side chains, the side chain hydroxyl moiety may play a similar role as the water hydroxyl, inducing a similar bifurcated *(NH, OH)···O=C* hydrogen bond with an angle of 120° that would, by itself, induce a local bend in the α-helix.

We suggest that Ser$^{5.43}$ and Ser$^{5.46}$ in the $\beta_2$–adrenergic receptor, which provide the docking site for the agonist (see above), adopt the *g*– conformation, in the absence of the extracellular ligand. Possibly, Ala replacement of Ser$^{5.43}$ and Ser$^{5.46}$ by site–directed mutagenesis changes the conformation of helix 5, from the bent helix (Ser/*g*– in blue, see Fig. 3.1.2d) to the straight helix (Ala in red). This would explain the influence of these Ser residues in helix 5 on the conformational equilibrium between inactive and active states of the receptor (9). Moreover, substitution of two Ser residues, located three residues apart and thus in the same face of the helix, augments the magnitude of the relocation of helix 5 by Ala substitution.

Finally, we would like to remark the structural consequences derived from the hydrogen bond formation between the neurotransmitters and the Ser residues in helix 5. Ser must adopt the *t* conformation, if it acts as hydrogen bond donor, in the process of hydrogen bonding to the hydroxyl moieties of the ligand. Thus, ligand binding might require the conformational transition of Ser from the *g*– (see α–helix in blue in Fig. 3.1.2d) to the *t* (green) conformation. This process of rotation around $\chi_1$, from *g*– to *t*, induces a change in the direction of the helix toward different positions in space (see above and Fig. 3.1.2d).

It is important to note that Ala replacement of Ser$^{5.43}$ and Ser$^{5.46}$ (conformational transition from Ser/*g*– in blue to Ala in red, see Fig. 3.1.2d) decreases the levels of intracellular cAMP (9). In contrast, ligand binding to Ser$^{5.43}$ and Ser$^{5.46}$ (conformational transition from Ser/*g*– in blue to Ser/*t* in green, see Fig. 3.1.2d) increases the levels of intracellular cAMP (9). This opposite effect cannot merely be understood from these reported conformational changes of helix 5. Thus, ligand binding might trigger more complex processes that finally lead to the active form of the receptor. It has been suggested that agonists of the $\beta_2$–adrenergic receptor also induce conformational changes in transmembrane domains 3 and 6 (20). Moreover, the ligand might produce unfavourable changes (21) in the receptor binding site that triggers the significant change in the conformational properties of the receptors that are transmitted to the intracellular site (22).

This statistical analysis on the influence of Ser and Thr residues to the curvature of α–helices has provided the structural basis to understand the mechanism by which the Ser residues in helix 5 in the neurotransmitter family of GPCR control the equilibrium between inactive and active states of the receptor. Because our findings are based on general principles of protein structure, it is conceivable that Ser and Thr residues on α-helices of other integral membrane proteins, such as gap junctions (23), may also participate in the conformational changes underlying transmembrane signalling.

## Acknowledgments

This work was supported in part by grants from CICYT (SAF99–073) and Fundació La Marató TV3 (0014/97). Computer facilities were provided by the Centre de Computació i Comunicacions de Catalunya.

## References


1. Baldwin JM, Schertler GF, Unger VM 1997 An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. Journal of Molecular Biology 272:144-164
2. Unger VM, Hargrave PA, Baldwin JM, Schertler GF 1997 Arrangement of rhodopsin transmembrane alpha-helices. Nature 389:203-206
3. van Rhee AM, Jacobson KA 1996 Molecular Architecture of G Protein-Coupled Receptors. Drug Development research 37:1-38
4. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S 1996 The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. Science 272:1136-44
5. Lefkowitz RJ, Cotecchia S, Samama P, Costa T 1993 Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins. Trends Pharmacol Sci 14:303-7
6. Samama P, Cotecchia S, Costa T, Lefkowitz RJ 1993 A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model. Journal of Biological Chemistry 268:4625-4636
7. Bond RA, Leff P, Johnson TD, Milano CA, Rockman HA, McMinn TR, Apparsundaram S, Hyek MF, Kenakin TP, Allen LF 1995 Physiological effects of inverse agonists in transgenic mice with myocardial overexpression of the beta 2-adrenoceptor. Nature 374:272-276
8. Gether U, Kobilka BK 1998 G protein-coupled receptors. II. Mechanism of agonist activation. Journal of Biological Chemistry 273:17979-17982
9. Ambrosio C, Molinari P, Cotecchia S, Costa T 2000 Catechol-binding serines of beta(2)-adrenergic receptors control the equilibrium between active and inactive receptor states. Molecular Pharmacology 57:198-210
10. Blundell T, Barlow D, Borkakoti N, Thornton J 1983 Solvent-induced distortions and the curvature of alpha-helices. Nature 306:281-283
11. Gray TM, Matthews BW 1984 Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. Journal of Molecular Biology 175:75-81
12. McGregor MJ, Islam SA, Sternberg MJ 1987 Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. J Mol Biol 198:295-310
13. Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R 1996 Electron-crystallographic refinement of the structure of bacteriorhodopsin. J Mol Biol 259:393-421
14. Stowell MH, McPhillips TM, Rees DC, Soltis SM, Abresch E, Feher G 1997 Light-induced structural changes in photosynthetic reaction

center: implications for mechanism of electron-proton transfer. Science 276:812-6

15. Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R 1998 The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science 280:69-77

16. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M 1977 The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112:535-42

17. Chou KC, Némethy G, Scheraga HA 1984 Energetic Approach to the packing of a-helices. 2. General treatment of nonequivalent and nonregular helices. J. Amer. Chem. Soc. 106:3161–3170

18. Monne M, Hermansson M, von Heijne G 1999 A turn propensity scale for transmembrane helices. J Mol Biol 288:141-5

19. Senes A, Gerstein M, Engelman DM 2000 Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. J Mol Biol 296:921-36

20. Gether U, Lin S, Ghanouni P, Ballesteros JA, Weinstein H, Kobilka BK 1997 Agonists induce conformational changes in transmembrane domains III and VI of the beta2 adrenoceptor. EMBO Journal 16:6737-6747

21. Gether U, Ballesteros JA, Seifert R, Sanders-Bush E, Weinstein H, Kobilka BK 1997 Structural instability of a constitutively active G protein-coupled receptor. Agonist-independent activation due to conformational flexibility. J Biol Chem 272:2587-90

22. Pardo L, Campillo M, Giraldo J 1997 The effect of the molecular mechanism of G protein-coupled receptor activation on the process of signal transduction. European Journal of Pharmacology 335:73-87

23. Ri Y, Ballesteros JA, Abrams CK, Oh S, Verselis VK, Weinstein H, Bargiello TA 1999 The role of a conserved proline residue in mediating conformational changes associated with voltage gating of Cx32 gap junctions. Biophys J 76:2887-98

24. Kraulis, J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. J. Appl. Cryst. 24:946 –950.

25. Ballesteros, J. A. and Weinstein, H. (1995). Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. Methods in Neurosciences 25, 366-428.

## 3.2. Ser and Thr residues modulate the conformation of Pro-kinked transmembrane α-helices

Xavier Deupi[§], Mireia Olivella[§], Cedric Govaerts[#$], Juan Antonio Ballesteros[†], Mercedes Campillo[§], and Leonardo Pardo[§*]

[§]Laboratori de Medicina Computacional, Unitat de Bioestadística and Institut de Neurociències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain; [#]Institut de Recherche Interdisciplinaire en Biologie Humaine et Nucléaire, Université Libre de Bruxelles, Campus Erasme, 808 route de Lennik, B-1070 Bruxelles, Belgium; and [†]Novasite Pharmaceuticals, Inc., San Diego, California

[$]present address: Cellular & Molecular Pharmacology, University of California, San Francisco, 600 16th Street, San Francisco, CA 94143-2240

*Corresponding author. E-mail address: Leonardo.Pardo@uab.es

## Abstract

The conformation of Pro-kinked transmembrane helices containing Ser or Thr residues, in both *g*+ and *g*- rotamers, has been studied by molecular dynamics simulations in a hydrophobic environment. Analysis of the simulations shows that (S/T)P and (S/T)AP motifs in the *g*+ rotamer and the TAP and PAA(S/T) motifs in the *g*- rotamer induce a bending angle in the helix that is ~5-12º larger than a control Pro-kinked helix. The additional hydrogen bond formed between the OγH group of Ser/Thr and the peptide carbonyl oxygen in the preceding turn of the helix, together with the additional flexibility provided by the adjacent Pro, seems to cause the observed effect. Notably, the formation of the hydrogen bond between the OγH group of Ser and the carbonyl oxygen at position j-3, as observed in SP and SAP motifs in the *g*- rotamer, induces a bending angle that is ~15-30º larger and alters the direction of the helix. This change in direction is attributed to the closing of the helix at the Pro-kink, as illustrated by the increase of the twist angle relative to the control Pro-kinked helix. In contrast, (S/T)AAP and PA(S/T) motifs, in both *g*+ and *g*-, and PAA(S/T) in *g*+ rotamers decrease (~3-6º) the bending angle of the helix by either reducing the steric clash between the pyrrolidine ring of Pro and the helical backbone, or by adding a constrain in the form of a hydrogen bond in the curved in face of the helix. The large number of occurrences of Ser/Thr residues in combination with Pro found in both a database of transmembrane helices and a database of rhodopsin-like G protein-coupled receptors suggest that membrane proteins incorporate these motifs for functional purposes.

Keywords: transmembrane helices, proline-kink, serine, threonine, molecular dynamics, membrane proteins, G protein-coupled receptors.

## 3.2.1. Introduction

Membrane receptors and channels are integral membrane proteins composed of several transmembrane (TM) α-helices that assemble through tertiary or quaternary structures to form a bundle that crosses the lipid bilayer (1;2). Biological function of these proteins involve conformational rearrangement of this transmembrane bundle. For example, it is thought that activation of G protein-coupled receptors (GPCRs) require rigid-body motions of several if not all TM helices (3). Such conformational changes require local flexibility in α-helices, which can, for instance, be provided by proline residues within the helix (4). Although Pro has the least helix-forming tendency (5) and the highest turn–stabilizing tendency in the membrane (6) of all naturally occurring amino acids, Pro residues are normally observed in TM helices (7) where they usually induce a significant distorsion named Pro-kink (8). The break is produced in order to avoid a steric clash between the pyrrolidine ring of Pro (at position i) and the carbonyl oxygen of the residue in the preceding turn (position i-4) (9) leading to a bending of the helical structure (10). Furthermore, the absence of the backbone N–H group in Pro prevents the formation of the hydrogen bond with the backbone carbonyl oxygen of residue i-4. The role of the N-H group is partly fulfilled by the $C_\delta H$ atoms of Pro, which hydrogen bond a backbone carbonyl oxygen usually located three, four, or five residues preceding the Pro (9). This type of $C_\delta$–H•••O=C interaction can help to stabilize the local break. Moreover, due to the distorsions of the Pro-kink, the $N_{i+1}$-H•••O=$C_{i-3}$ hydrogen bond is frequently disrupted, and the $N_{i-1}$-H•••O=$C_{i-5}$ hydrogen bond is sometimes replaced by the $N_{i-1}$-H•••O=$C_{i-4}$ hydrogen bond (9). Therefore, in the context of an α-helix, Pro not only induces a break in the helix, but it also destabilizes the hydrogen bond network that normally stabilizes the secondary structure. As a result, Pro introduces a flexible point in the α-helix which could be of functional importance for membrane proteins (11).

We and others have shown that, in transmembrane proteins, specific sequence motifs preceding the Pro can provide structural properties for α-helices that are significantly different than those of a regular Pro-kink (12;13). These motifs include Ser and Thr residues in the vicinity of the Pro, which induce a structural modification of the Pro-kink through hydrogen bonds of their polar sidechains. Experimental studies involving site-directed mutagenesis of the residues forming these motifs have demonstrated the functional importance of the Ser and Thr residues in Pro-containing α-helices (12-14).

Therefore the question arises as whether combinations of Ser/Thr and Pro are generically used as structural modulators in integral membrane proteins. To address this issue we have systematically surveyed sequences of TM helices to analyze the frequencies of such combinations. Moreover, the conformation of Pro-kinked TM helices containing Ser or Thr residues was assessed by molecular dynamics (MD) simulations in a hydrophobic environment. Our study indicates that membrane proteins, and GPCRs in particular, favor combined motifs of Pro and Ser or Thr in their TM helices

and that these motifs have specific structural properties. This suggests that transmembrane proteins have evolved these specific motifs for functional purposes.

## 3.2.2. Results

### Statistical analysis of the presence of Ser/Thr residues in Pro-kinked transmembrane helices

Table 3.2.1 lists the observed number of occurrences of the (S/T)xxP, (S/T)xP, (S/T)P, P(S/T), Px(S/T), and Pxx(S/T) patterns in a non-homologous database of sequences of TM helices (7). In order to find over-represented and under-represented patterns we calculated the expected number of occurrences with the TMSTAT formalism that accounts for both the relative frequency of the amino acids and length of the sequence (7) (see http://bioinfo.mbb.yale.edu/tmstat/). We used a *p*-value cut-off of 0.10 to select statistically significant over-represented (odds ratio greater than 1) and under-represented (odds ratio less than 1) patterns. Both TxP (odds ratio of 1.10, *p*-value of 0.07) and SxP (1.12, 0.02) motifs are over-represented pairs, suggesting that (S/T)xP is a common pattern in TM helices. In contrast, the SP (0.87, 0.01), PS (0.84, 0.001), and PxxS (0.84, 0.004) are under-represented pairs.

### Statistical analysis of the presence of Ser/Thr residues in Pro-kinked transmembrane helices of the rhodopsin-like family of GPCRs

GPCRs, or 7TM receptors, are one of the largest family of membrane proteins with more than 800 human sequences (15). These receptors have been classified into five main familes: glutamate, rhodopsin, adhesion, frizzled/taste2, and secretin (16). The members of the rhodopsin family (the largest subgroup) and bovine rhodopsin, for which the structure is known (1), share a large number of conserved sequence patterns in the TM segments. This allows for a straightforward identification of TM helices and provides a large set to survey the association of Ser/Thr and Pro residues. Table 3.2.2 shows the positions, in the generalized numbering scheme of Ballesteros & Weinstein (17), and the number of occurrences of Pro residues within the TM helices of the 1948 GPCR sequences denoted as Class A Rhodopsin like in GPCRDB (18), as of September 2002 release (6.1). We found a total of 9715 Pro residues, providing a very large statistical sample of Pro-kinked TM helices. Whereas Pro are found at many locations throughout the family, a typical TM helix does not contain more than one Pro-kink (except in TM 4). Partly or highly conserved prolines can be found in TM 1 (positions 1.36 and 1.48), TM 2 (2.58, 2.59 and 2.60), TM 4 (4.59 and 4.60) TM 5 (highly conserved 5.50), TM 6 (highly conserved at position 6.50 and occasionally at 6.59) and TM 7 (7.38, 7.45, 7.46, and 7.50 which belongs to the highly conserved NPxxY motif). Table 3.2.2 also shows the probability of finding a Ser or Thr residue in the direct vicinity of the Pro, namely at positions i-3 (S/T)xxP, i-2 (S/T)xP, i-1 (S/T)P, i+1 P(S/T), i+2 Px(S/T), and i+3 Pxx(S/T) relative to Pro.

The statistical analysis shows that 52% of the Pro-containing TM helices contain either Ser or Thr in the vicinity of the proline. The most common combinations are the (S/T)xP motifs, found in 16% of all Pro-kinks in GPCRs (9% for TxP, 7% for SxP). Notably, the distribution of TxP and SxP is not uniform over the different Pro-kinks, as some specific locations, such as 2.58, 4.59 and 7.38, show much higher frequencies than the others. Moreover, Ser and Thr are not equally found. For example, $Pro^{2.58}$ predominantly associates with a Thr (forming a $TxP^{2.58}$ motif) while $Pro^{4.59}$ displays four times more $SxP^{4.59}$ than $TxP^{4.59}$ motifs. This indicates that Ser and Thr are not equivalent and already suggests, as their hydrogen bonding propensities are similar, that the methyl group differentiating their sidechains may have a defined structural or functional role (see below). Additionally, SP motifs (2%) are observed at lower frequency than TP motifs (6%) further suggesting different roles for these two side chains. Finally, the frequency of Ser/Thr residues at positions preceding Pro (35% of the Pro-kinks) is larger than at positions following Pro (16% of the Pro-kinks).

## Conformation of Pro-kinked transmembrane helices containing Ser or Thr residues at positions from i-3 to i+3 relative to Pro

The strong association of Ser/Thr residues with Pro in TM helices, described above, suggests a possible structural role. Due to the paucity of membrane protein structures, it is not possible to investigate such structural role by surveying TM helices in known crystal structures. We have previously shown that structural characteristics of TM helices (including Pro-containing helices) can be accurately reproduced by MD simulations (19). Thus, we have used a similar approach here to assess the conformation of model peptides containing the (S/T)AAP, (S/T)AP, (S/T)P, P(S/T), PA(S/T), and PAA(S/T) patterns, with Ser and Thr side chains built in both $\chi_1$=*gauche*+ (*g*+) or *gauche*- (*g*-) conformation (see Methods). In α-helices, the side chains of Ser and Thr adopt primarily the *g*+ or *g*- where the hydrogen bonding capacity of their $O_\gamma H$ side chain can be satisfied by interacting with the backbone carbonyl oxygen in the preceding turn of the helix (20). In contrast, the *trans* (*t*) conformation is not stable because it can not form such hydrogen bond as it points the $O_\gamma H$ group away from the backbone. The control of these simulations was the simulation of a Pro-containing polyAla α–helix (see Methods). We monitored the distorsions caused by these patterns in the α-helix by the hydrogen bond interactions in the Pro-kink turn (Figures 3.2.1 and 3.2.2), the bending angle (Table 3.2.3), and the unit twist angles (Figure 3.2.3) (see Methods). The results of the simulations are presented separately for each motif. The notation employed in the manuscript assigns the i position to Pro and the j position to Ser/Thr, and all other amino acids are named relative to these reference points. Moreover, the distribution, in percentage, of the intra-helical hydrogen bond between the $O_\gamma H$ side chain of Ser/Thr and the carbonyl oxygen at positions j-4 and j-3 as observed during the MD simulations, is also shown in Table 3.2.3.

### *Pro*

The bending angle of 20° we have obtained by MD simulations (Table 3.2.3) coincides with the reported bending angle of 21° measured from a database of Pro-containing TM helices (10), whose structures are known at a resolution of 4 Å or better. Figure 3.2.3 shows the evolution of the unit twist angles along the α-helix, from turns (i-11, i-8) to (i+4, i+7), as observed during the MD simulations. The helical distorsions induced by the Pro residue, in the control simulation, is reflected at the level of twist angles from aproximately turn (i-5, i-2) up to (i-1, i+2). Figure 3.2.1a shows a detailed view of the Pro-kink. Both $C_\delta H$ atoms of Pro act as hydrogen bond donors in the hydrogen bond interaction with the carbonyl oxygen at position i-4 (9).

### *(S/T)P*

#### *gauche+*

The side chain of both Ser and Thr residues always interact with the backbone carbonyl oxygen at position j-4 (Table 3.2.3). As a result, these helices are strongly bent, with average bending angles of 31° for SP and 27° for TP (~7-11° larger than the 20° of the reference Pro-kinked α-helix) (Table 3.2.3). Morover, there is an increase of 13° in SP and 4° in TP in the twist angle (i.e. a closing of the helix) at the (i-4, i-1) (Figure 3.2.3a). Figures 3.2.1b and 3.2.1c show a detailed view of the hydrogen bond pattern. The $O_\gamma H$ side chain of Ser/Thr interacts, in addition to the $N_{i-1}$-H amide, with the carbonyl oxygen at position i-5. The intrahelical $N_{i+1}$–H•••O=$C_{i-3}$ and $N_{i-1}$–H•••O=$C_{i-5}$ hydrogen bonds are preserved. It is important to note that the larger bending angle observed for SP are due to a small number of conformations visited during the MD simulation. These conformations present a bending angle of 35º due to a peculiar hydrogen bond network in which the $C_\delta H$ atoms of Pro acts as a hydrogen bond donor in the interaction with both carbonyl oxygens at positions i-3 and i-4, the $O_\gamma H$ side chain of Ser interacts with the carbonyl oxygen at position j-4, the intrahelical $N_{i+1}$–H•••O=$C_{i-3}$ is disrupted; and the $N_{i-1}$–H amide is halfway between O=$C_{i-4}$ and O=$C_{i-5}$ (not shown).

#### *gauche-*

Notably, this rotamer differenciates Ser and Thr. While Ser can form a hydrogen bond with either the carbonyl oxygen at position j-3 (33%, see Table 3.2.3) or j-4 (67%), Thr is almost only hydrogen bonding the carbonyl oxygen at position j-4 (92%) due to steric restriction of the additional methyl group (see discussion). When the $O_\gamma H$ side chain of Ser/Thr interacts with the carbonyl oxygen at position j-4, there is little or no structural effect on the Pro-kink, as measured by the bending angle (20° for SP or 21° for TP versus 20° for the reference Pro-kink) or the twist angle profile (Figure 3.2.3a) of the helix. Each $C_\delta H$ group of Pro acts as a hydrogen donor, either with the O=$C_{i-4}$ carbonyl oxygen or the $O_\gamma$ atom of Ser/Thr (Figures 3.2.1d and 3.2.1e). The intrahelical $N_{i+1}$–H•••O=$C_{i-3}$ and $N_{i-1}$–H•••O=$C_{i-5}$ hydrogen bonds are preserved.

In contrast, when the $O_\gamma H$ side chain of Ser interacts with the carbonyl oxygen at position j-3 ($O_\gamma$-H•••O=$C_{j-3}$), there are significant changes in the hydrogen bond network of the Pro-kink turn, inducing a strong structural effect. The $C_\delta H$ atoms of Pro interact with both carbonyl oxygens at positions i-3 and i-4, causing the intrahelical $N_{i+1}$–H•••O=$C_{i-3}$ to be disrupted, and the $N_{i-1}$–H amide to relocate between O=$C_{i-4}$ and O=$C_{i-5}$ (Figure 3.2.1f). This induces an increase in the bending angle (50° versus 20° of Pro, see Table 3.2.3) and a closing of the helix at the Pro-kink, as reflected by an increase of the twist angle at the (i-4, i-1) turn ( 121° versus 101° for the reference Pro-kink simulation, see Figure 3.2.3a). In order to corroborate these results, we performed additional MD simulations on the SP motif, with the side chain built in the *g-* conformation, and restraining the intra-helical hydrogen bond of the $H_\gamma$ atom of Ser to the carbonyl oxygen at positions j-3 and j-4. The analysis of these independent restrained MD trajectories confirms our initial conclusions obtained from unrestrained MD trajectories. The hydrogen bond patterns are similar to the ones depicted in Figures 3.2.1e and 3.2.1f: the bending angles are 16° and 48° and the twist angles at the (i-4, i-1) turn are 99° and 120°, for the SP motif with the $O_\gamma$–H•••O=$C_{j-4}$ and $O_\gamma$–H•••O=$C_{j-3}$ hydrogen bond, respectively.

### *(S/T)AP*

#### *gauche+*

TAP and SAP motifs induce an important bending of the α-helix (27° and 30° respectively, ~7-10° more than a regular Pro-kink). However, in contrast to (S/T)P in *g+*, there are no large changes in the unit twist angle profile relative to the reference Pro-kinked α-helix (Figure 3.2.3b). These effects are due to the specific hydrogen bond network, in the Pro-kink turn, as depicted in Figures 3.2.1g and 3.2.1h. The hydroxyl moiety of both Ser and Thr interacts, in addition to the $N_{i-2}$-H amide, with the carbonyl oxygen at the i-6 position (j-4 to Ser/Thr) (Table 3.2.5). Moreover, the $C_\delta H$ groups of Pro act as a hydrogen bond donors in the hydrogen bond interaction with both lone pairs of the carbonyl oxygen at i-4. The intrahelical $N_{i+1}$–H•••O=$C_{i-3}$ and $N_{i-1}$–H•••O=$C_{i-5}$ hydrogen bonds are preserved.

#### *gauche-*

As for (S/T)P, the *g-* rotamer shows significant differences between SAP and TAP motifs. The $O_\gamma H$ side chain of Thr hydrogen bonds preferably the carbonyl oxygen at the j-4 position (88%, Table 3.2.3) and only rarely to j-3, due to steric hindrance between the methyl moiety of the sidechain and the carbonyl oxygen of residue j-4 (see discussion). In contrast, the hydroxyl group of Ser hydrogen bonds almost exclusively the carbonyl oxygen at position j-3 (i-5 to Pro) (98%, Table 3.2.3). This leads to a very different hydrogen bond pattern and conformation of the Pro-kinked α-helix (Figures 3.2.1i and 3.2.1j). When both the $O_\gamma H$ of Thr and the $N_{i-2}$-H amide are interacting with the i-6 carbonyl oxygen, the $C_\delta H$ group of Pro act as a hydrogen bond donor in the hydrogen bond interaction with the carbonyl oxygen at position i-4. The $N_{i+1}$–H•••O=$C_{i-3}$ and $N_{i-1}$–H•••O=$C_{i-5}$ hydrogen bonds are preserved. In this case, the average bend angle of the α-helix is 27° (~7° larger than the reference Pro-kinked α-helix, Table 3.2.3); and the

unit twist angle profile is not strongly modified relative to the reference system (Figure 3.2.3b). In other words, the TAP motifs presents a similar structural effect in both *g+* and *g-* rotamers.

However, in the SAP *g-* motif, the $O_\gamma H$ of the Ser residue is interacting with the carbonyl oxygen at position j-3 (i-5 relative to the Pro), and while the $C_\delta H$ group of Pro keeps interacting with the carbonyl oxygen at position i-4 and the $N_{i+1}$–H•••O=$C_{i-3}$ hydrogen bond is preserved, the $N_{i-1}$–H•••O=$C_{i-5}$ hydrogen bond is disrupted (the $N_{i-1}$–H group seems to interact with the $O_\gamma$ atom of Ser) and the $N_{i-2}$–H amide reorients between O=$C_{i-5}$ and O=$C_{i-6}$ (Figure 3.2.1i). This results in an average bend angle of 35º (~15º larger than the reference Pro-kinked α-helix, Table 3.2.3), and an increase in the twist angle at the (i-5, i-2) turn of 17º relative to the Pro-kinked helix (Figure 3.2.3b).

### *(S/T)AAP*

#### *gauche+*

The effect of (S/T)AAP motifs in the helix is radically different from (S/T)P or (S/T)AP. SAAP (16º) and TAAP (15º) significantly *decrease* the bending angle of the helix relative to the reference Pro-kink (20º, Table 3.2.3). The decrease in bending is due to specific hydrogen bond networks which reduce the steric clash between the pyrrolidine ring of Pro and the O=$C_{i-4}$ carbonyl (see discussion). Figures 3.2.1k and 3.2.1l show how, in both cases, the $N_{i-3}$–H amide (N–H group of Ser/Thr) is located between the O=$C_{i-6}$ and O=$C_{i-7}$ carbonyl oxygens. Adaptation of the standard intrahelical $N_{i-3}$–H•••O=$C_{i-7}$ hydrogen bond towards the $N_{i-3}$–H•••O=$C_{i-6}$ hydrogen bond, modifies the orientation of the O=$C_{i-4}$ carbonyl. Notably, this relocation of the O=$C_{i-4}$ carbonyl allows the $C_\gamma H$ group of Pro to interact with the carbonyl oxygen instead of the standard $C_\delta H$ atoms.

#### *gauche-*

There is also a straightening of the helix for both SAAP (14º) and TAAP (14º) relative to the standard Pro-kink (20º, Table 3.2.3). As in the *g+* conformation, both Ser and Thr residues establish preferably the hydrogen bond with the backbone carbonyl at position j-4 (i-7 to Pro) (Table 3.2.3). The hydrogen bond network at the Pro-kink turn is also very similar: the $N_{i-3}$–H amide is located between the O=$C_{i-6}$ and O=$C_{i-7}$ carbonyl oxygens, the relocation of O=$C_{i-4}$ reduces the steric clash with the Pro side chain, and the $C_\delta H$ and $C_\gamma H$ groups of Pro interact with the O=$C_{i-4}$ carbonyl oxygen (Figures 3.2.1m and 3.2.1n)

### *P(S/T)*

#### *gauche+, gauche-*

PS and PT motifs in both *g+* and *g-* rotamers show no significant changes in either the bending angle (Table 3.2.3) or the twist profile (Figure 3.2.3d) relative to the control simulation. The side chain of both Ser and Thr residues forms the hydrogen bond, in addition to the $N_{i+1}$–H amide, with the carbonyl oxygen at position j-4 (i-3 to the Pro) (Table 3.2.3). The $C_\delta H$

group of Pro acts as a hydrogen bond donor in the hydrogen bond interaction with the carbonyl oxygen at the i-4 position (Figures 3.2.2a-3.2.2d).

### *PA(S/T)*

#### *gauche+*

The average bending angles of the α-helices for PAS (16º) and PAT (14º) are slightly decreased relative to the reference Pro-kink (Table 3.2.3), while the twist angle profile remains unchanged (Figure 3.2.3e). The fact that the steric clash between the $C_{\delta}H$ atoms of Pro and the $O=C_{i-4}$ carbonyl group and the hydrogen bond between the $O_{\gamma}H$ side chain of Ser/Thr and the $O=C_{i-2}$ carbonyl group, are in opposite sides of the α-helix could be responsible for this effect (see discussion). While the $C_{\delta}H\bullet\bullet\bullet O=C_{i-4}$ clash tends to bend the helix towards the opposite side of the Pro residue, the additional $O_{\gamma}H\bullet\bullet\bullet O=C_{i-2}$ hydrogen bond partly counters the bending of the helix in this direction. The detailed view of the hydrogen bond network depicted in Figures 3.2.2e and 3.2.2f shows that the hydroxyl group of either Ser or Thr side chains is interacting with carbonyl oxygen at the j-4 position (i-2 to the Pro) (Table 3.2.3).

#### *gauche-*

Ser and Thr side chains behave differently in this *g-* conformation. While Ser can form a hydrogen bond with the carbonyl oxygen at position j-3 (35%, Figure 3.2.2i) or at position j-4 (65%, Figure 3.2.2g), the methyl group of Thr forces the $O_{\gamma}H$ side chain to hydrogen bond the carbonyl oxygen at position j-4 (93%, Figure 3.2.2h) (Table 3.2.3). All these conformations, despite these variations in the hydrogen bond pattern, reduce the bending angle of the helix relative to the standard Pro-kink (14º or 15º for PAS hydrogen bonding the j-4 or the j-3 carbonyl oxygen, respectively; and 16º for PAT, Table 3.2.3). Similarly to the *g+* conformation, the formation of either the $O_{\gamma}H\bullet\bullet\bullet O=C_{i-2}$ or $O_{\gamma}H\bullet\bullet\bullet O=C_{i-1}$ hydrogen bond in the reverse side of the helix, from where Pro is located, impedes its bending. There are no significant changes in the twist angles profile (Figure 3.2.3e).

### *PAA(S/T)*

#### *gauche+*

These motifs also reduce the bending of the helix relative to the control (14º for PAAS and 16º for PAAT, Table 3.2.3). The side chain of both Ser and Thr residues is hydrogen bonding the $O=C_{i-1}$ carbonyl oxygen ($O=C_{j-4}$ relative to Ser/Thr) (Table 3.2.3) (Figures 3.2.2j and 3.2.2k). As previously observed for the PAS motif in the *g-* rotamer conformation, the formation of this additional $O_{\square}H\bullet\bullet\bullet O=C_{i-1}$ hydrogen bond impedes the bending of the helix. There are no significant changes in the twist angles profile (Figure 3.2.3f).

#### *gauche-*

PAAS and PAAT motifs in *g-* are the only patterns of all Ser/Thr residues following Pro in which the bending of the helix increases relative to the

control Pro-kinked helix (see Table 3.2.3). Moreover, the Thr side chain in PAAT is the only case of all patterns that contain Thr in which the OγH side chain is able to interact, in a significant amount of conformations (26%, Table 3.2.3), with the $O=C_{j-3}$ carbonyl ($O=C_i$ relative to Pro). In contrast to the other motifs, the formation of this hydrogen bond is feasible due to an increase of 6º in the unit twist angle (i.e. a closing of the helix) at the (i-1, i+2) helical turn (see Figure 3.2.3f), which relocates the position of the $O=C_{i-1}$ and $O=C_i$ carbonyls. In all cases, the $C_\delta H$ groups of Pro hydrogen bond the $O=C_{i-4}$ carbonyl, and the intrahelical $N_{i+1}–H•••O=C_{i-3}$ and $N_{i-1}–H•••O=C_{i-5}$ hydrogen bonds are preserved (Figures 3.2.2l-3.2.2o), leading to similar bending angles (30º for PAAS hydrogen bonding either $O=C_{j-4}$ or $O=C_{j-3}$; and 25º or 32º for PAAT hydrogen bonding $O=C_{j-4}$ or $O=C_{j-3}$, respectively). Although the side chain of Ser/Thr in both *g*+ and *g*- can interact with the same $O=C_{i-1}$ carbonyl, the effect on the bending of the helix is opposite, decreasing in *g*+ and increasing in *g*- (Table 3.2.3). While in the *g*+ conformation the $O_\gamma H•••O=C_{i-1}$ hydrogen bond is in the opposite face of the helix that the steric $C_\delta H•••O=C_{i-4}$ clash (Figures 3.2.2j and 3.2.2k, see discussion), in the *g*- conformation both interactions occur in the same face (Figures 3.2.2l and 3.2.2n).

Due to the different behavior of the PAA(S/T) motifs in g-, these results were further investigated by additional MD simulations restraining the intra-helical hydrogen bond of the Hγ atom of Ser/Thr to the carbonyl oxygen at positions j-3 and j-4. Analysis of these new trajectories leads to bend angles similar to those previously obtained (28º or 31º for PAAS hydrogen bonding O=Cj-4 or O=Cj-3, respectively; and 27º or 29º for PAAT hydrogen bonding O=Cj-4 or O=Cj-3, respectively), thus confirming our results.

## 3.2.3. Discussion

Signalling membrane proteins achieve structural diversity within a constraining environment. Specific TM helices must have included specific structural motifs, adapted to the functional role of each of these proteins. Deciphering the structure-function relationship in membrane proteins will therefore require an understanding of the structural motifs that govern the conformational diversity. Besides the important role of Pro residues (4), Ser and Thr residues also induce helical distorsions of possible functional importance (21). In the present study we have shown that combinations of these residues can produce dramatic effects on the conformation of TM α-helices, while their unusual distribution in TM segments further suggest a possible functional role as structural adapters.

### *The (S/T)P and (S/T)AP motifs in the gauche+ rotamer and the TAP and PAA(S/T) motifs in the gauche- rotamer increase the bending of Pro-kinked transmembrane helices*

Figure 3.2.4a shows the conformation of Pro-kinked (control, yellow), SP (red) and SAP (dark green) in the *g*+ rotamer, and TAP (light green) and PAAS (purple) in the *g*- rotamer TM α-helices. The OγH group of Ser/Thr is

hydrogen bonding the peptide carbonyl oxygen at position j-4, relative to Ser/Thr, in all the cases. The helix conformation of TP and TAP in *g*+ and PAAT in *g*- are similar to SP and SAP in *g*+ and PAAS in *g*-, respectively, and are omitted for clarity. Clearly, all these motifs and rotamer conformations induce a bending angle in the helix that is ~7-12º larger than the control Pro-kinked helix (see Results and Table 3.2.3). Incorporation into the α-helix of these modified conformations at one side of the cell membrane results in a significant displacement of the residues located at the other side of the membrane. The magnitude of the relocation might be estimated from the models depicted in Figure 3.2.4a. Thus, the distances between the alpha–carbon positions between the control Pro-kinked helix (yellow) and the other modified helices are in the 2-4 Å range for an amino acid located 10 residues away from Pro. The additional hydrogen bond formed between the OγH group of Ser or Thr and the backbone carbonyl oxygen in the preceding turn of the helix, together with the additional flexibility provided by the adjacent Pro, seems to cause the observed conformation of the helices.

### *The (S/T)P motifs in the gauche- rotamer do not modify the bending of Pro-kinked transmembrane helices*

SP and TP motifs in the *g*- rotamer can also hydrogen bond the i-5 carbonyl oxygen (j–4 relative to Ser/Thr), but do not change the bending of the Pro-kinked TM helix (see Results and Table 3.2.3). In contrast to other motifs and conformations, in this *g*- rotamer the Oγ atom of Ser or Thr hydrogen bonds the $C_\delta H$ group of Pro (see Figures 3.2.1d and 3.2.1e). This new intrahelical hydrogen bond adds an additional constrain which counterbalances the increase of bending angle caused by the additional hydrogen bond formed between the Ser/Thr side chain and the i-5 peptide carbonyl oxygen. The $C_\delta H \bullet\bullet\bullet O_\gamma$ hydrogen bond is only feasible when Ser/Thr is preceding Pro and Ser/Thr are in the *g*- rotamer. In the *g*+ rotamer the lone pairs of the Oγ atom of Ser/Thr are pointing away from the helix (see Figures 3.2.1b and 3.2.1c).

### *The SP and SAP motifs in the gauche- rotamer interacting with the carbonyl oxygen at position i-3 increase the bending and alter the direction of Pro-kinked transmembrane helices*

SP (33%) and SAP (98%) motifs in the *g*- rotamer can hydrogen bond the j-3 carbonyl oxygen relative to Ser (Table 3.2.3). The formation of this interaction, rather than the standard with the j-4 carbonyl oxygen relative to Ser, induces a striking modification in the conformation of the helix, as illustrated in Figures 3.2.4b and 3.2.4c. Notably, the bend angle of the helix increased significantly from 20º for Pro to 50º for SP or 35º for SAP α-helices (Table 3.2.3). Moreover, the helices point to a completely different direction in space (Figure 3.2.4c): towards the center for Pro (control, yellow), left for SAP (light green), or right for SP (orange) α-helices. This change in direction is attributed to the closing of the helix at the Pro-kink, as illustrated by the increase, relative to the Pro-kinked helix, of the twist angle at the (i-4, i-1) turn for SP (20º), and at the (i-5, i-2) turn for SAP (17º)

(see Figure 3.2.3). It is probable that the closing of the helix at different turns, (i-4, i-1) for SP or (i-5, i-2) for SAP, makes both helices to point towards opposite directions in space.

### *The methyl group of Thr prevents the interaction of OγH with the carbonyl oxygen at position j-3*

Figure 3.2.5 shows the evolution of the distance between the Hγ atom of Ser and the oxygen of the $O=C_{j-4}$ (solid line) or $O=C_{j-3}$ (broken line) carbonyl, relative to Ser, and the values of $\chi_1$ of Ser obtained during the production run of the MD simulations of SP (left panel) and SAP (right panel) in the *g*- rotamer. There is a clear correlation between the type of side chain interaction and $\chi_1$. When $O_\gamma H$ is hydrogen bonding $O_{j-4}$ (from ~850ps to 1500ps in SP) $\chi_1$ stays in the ~35°-55° range (average value of 46º for SP). The value of $\chi_1$ goes up to the ~55°-75° range (average values of 68º for SP and 65º for SAP) if $O_\gamma H$ is hydrogen bonding $O_{j-3}$ (from 500ps to ~850ps in SP and from ~550ps to 1500ps in SAP). Thus, the formation of the hydrogen bond with the j-3 carbonyl oxygen, relative to Ser, requires an increase of the $\chi_1$ dihedral. These values of $\chi_1$ are unfavorable for Thr because of the steric hindrance between its methyl group and the $O=C_{j-4}$ carbonyl. Thus, Thr residues rarely (see Table 3.2.3) form the interaction with the $O=C_{j-3}$ carbonyl. In these conformations local re-arrangement of the backbone is necessary to accommodate the methyl group (results not shown). The side chain of Thr in the PAAT motif interacts with the $O=C_{j-3}$ carbonyl ($O=C_i$ relative to Pro) in a larger number of structures (26%) than the other Thr-containing motifs (Table 3.2.3). The closing of the helix at the (i-1, i+2) helical turn (see Figure 3.2.3f) relocates the position of the $O=C_{i-1}$ (the $O=C_{j-4}$ carbonyl that would clash with the methyl group of Thr) and $O=C_i$ (interacting with the $O_\gamma H$ side chain of Thr) carbonyls, thus facilitating the accomodation of the methyl group.

### *Molecular mechanisms of decreasing the bending of Pro-kinked transmembrane helices*

(S/T)AAP and PA(S/T) motifs, in both *g*+ and *g*-, and PAA(S/T) in *g*+ rotamers decrease the bending of Pro-kinked TM helices (see Table 3.2.3). However, the mechanisms by which these motifs modify the conformation of the helix differ.

(S/T)AAP motifs moderate the steric clash between the pyrrolidine ring and the $O=C_{i-4}$ carbonyl oxygen by modifying the conformation of the $N_{i-3}$-H amide group. The additional hydrogen bond formed between $O_\gamma H$ of $Ser/Thr_{i-3}$ and the $O=C_{i-7}$ carbonyl oxygen (j-4 relative to Ser/Thr) modifies the intrahelical $N_{i-3}$–H•••$O=C_{i-7}$ hydrogen bond. Adjustment of the $N_{i-3}$-H amide group (see bottom arrow in Figure 3.2.6a) alters the orientation of the $O=C_{i-4}$ carbonyl (see top arrow in Figure 3.2.6a), due to the planarity of the $OC_{i-4}$-$N_{i-3}H$ peptide bond (in green), away from the helix; reducing the steric clash betwen the $C_\delta H$ atoms of Pro and the $O=C_{i-4}$ carbonyl; and decreasing the bending angle of the helix (Table 3.2.3).

PA(S/T), in both *g*+ and *g*-, and PAA(S/T) in *g*+ rotamers obstruct the bending of the helix by adding a constrain in the form of a hydrogen bond in the opposite side of the Pro residue, the curved in face of the helix. Thus, the formation of either the $O_\gamma H \bullet\bullet\bullet O{=}C_{i-2}$ or $O_\gamma H \bullet\bullet\bullet O{=}C_{i-1}$ hydrogen bond by $Ser/Thr_{i+2}$ or $Ser/Thr_{i+3}$ side chain (Figure 3.2.6b) in the other side of the helix where the steric $C_\delta H \bullet\bullet\bullet O{=}C_{i-4}$ clash occurs, attenuates the bending of the helix. However, PAA(S/T) in the *g*- rotamer also interact with the $O{=}C_{i-1}$ carbonyl, but in this case there is an increase of the bending angle (Table 3.2.3). In this *g*- conformation the formation of the $O_\gamma H \bullet\bullet\bullet O{=}C_{i-1}$ hydrogen bond by the $Ser/Thr_{i+3}$ side chain occurs on the same side as the steric clash (compare Figures 3.2.2j and 3.2.2k with 3.2.2l and 3.2.2n), hence reinforcing the deformation.

P(S/T) motifs in both *g*+ and *g*- rotamers do not modify the conformation of the helix relative to the standard Pro-kink. In this case the additional hydrogen bond between the $Ser/Thr_{i+1}$ side chain occur with the $O{=}C_{i-3}$ carbonyl, which is neither at the same nor at the opposite face of the helix than the Pro side chain. We suggest that this particular hydrogen bond network maintains the conformation of the Pro-kinked helix unchanged.

### *Influence of the Ser/Thr side chain rotamer in mechanisms of signal transduction*

Our simulations suggest that Ser and Thr could act as molecular switches in signal transduction. While in TM helices these sidechains are normally in either *g*+ or *g*- conformation, interaction with an external partner (i.e. bound ligand) could induce a change towards the *trans* rotamer. Our results indicate that such rearrangement of the Ser or Thr can lead to conformational modification of the whole helix, with possible functional importance. Similarly, structural changes could be induced by a rotamer change from a *g*+ to *g*- conformation (and vice versa). In order to illustrate this mechanism, Figure 3.2.7a shows the conformation of SAP in the *g*+ rotamer (dark green), *t* rotamer (yellow, taken as a regular Pro-kinked α-helix), and *g*- rotamer interacting with the carbonyl oxygen at position j-3 relative to Ser (light green); and Figure 3.2.7b shows the conformation of SP in the *g*+ rotamer (red) and *g*- rotamer interacting with the carbonyl oxygen at position j-4 (dark orange) and j-3 (light orange). Clearly, variation of the side chain rotamer (Figure 3.2.7a), or changing the hydrogen bond from j-4 to j-3 or vice versa (Figure 3.2.7b), covers a wide structural range in the α-helix. Similarly, rotation of Ser/Thr side chain in the PAA(S/T) motif from *g*+ to *g*- would modify the bend angle of the helix from ~15º to ~28º (see Table 3.2.3).

### *Functional relevance of Ser and Thr as Pro-kinked modulators of transmembrane α-helices*

The statistical analysis of the presence of these motifs in the rhodopsin-like family of GPCRs (Table 3.2.2) shows that, among all the occurrences found in the search, the predominant sequences are TxP (9%), SxP (7%), TP (6%), SxxP (6%), TxxP (5%). Thus, there is a clear tendency of Ser/Thr residues to be found at positions preceding Pro. Among all the possibilities of Ser/Thr

preceding Pro, only the SP motif is not often observed in the database (2%). This might be attributed to the extreme conformation this motif achieves in the *g*- rotamer interacting with the carbonyl oxygen at position j–3.

Clearly, (S/T)xP motifs are highly observed in both a non-homologous database of sequences of TM helices and a database of sequences of rhodopsin-like GPCRs. The functional role of Ser/Thr residues in the highly conserved Thr-x-Pro$^{2.58}$ motif of the chemokine receptors (TM 2) was probed by site-directed mutagenesis and functional assay of the CCR5 receptor (13;22). While mutation of Thr$^{2.56}$ to Ser, Cys, Ala, or Val does not affect chemokine binding, it strongly influences the functional response. The functional impairment is highly dependant on the specific residue substituted for the Thr, and the rank order parallels the structural deformation of the α-helix (13).

The second TM segment of connexin32 contains the TP$^{87}$ motif. Mutations of Thr$^{86}$ to Ser, Ala, Cys, Val, Asn, or Leu shift the conductance-voltage relation of wild-type, such that the mutated channels close at smaller transjunctional voltages (12). It was proposed that the hydrogen bonding potential of Thr$^{86}$ mediates the conformational changes between open and closed channel states (12).

Bacteriorhodopsin contains TP$^{91}$ in TM helix C. Mutation of Thr$^{90}$ to Ala alters the proton pumping efficiency, and suggests that Thr$^{90}$ has an important structural role in the proton pumping mechanism (14). Figure 3.2.8 shows the superimposition of helix C as observed in the crystal structure of bacteriorhodopsin (PDB entry 1AP9) that contains the TP$^{91}$ motif in the *g*+ rotamer (red), and the computationally determined conformations of a Pro-kinked (yellow) and TP in the *g*+ rotamer (orange) polyAla α-helices. Clearly, helix C of bacteriorhodopsin exhibits a larger bending than a regular Pro-kinked helix that matches the computationally determined conformation of TP in the *g*+ rotamer. The bending angle of 30° obtained around the Pro-kink of helix C is in good agreement with the average value of 27° obtained in computer simulations (see Table 3.2.3).

## 3.2.4. Conclusions

Our simulations on model peptides offer molecular explanations for the high association of Ser and Thr residues with Pro in TM helices. Through specific action of their polar side chains, these residues modulate the structural deformations caused by the pyrrolidine ring of Pro. Such modulations cover a wide structural range, with, for instance, a variation of the bending angle from 14° to 30° by simply rotating the side chain of Ser in the PAAS motif from *g*+ to *g*-. These significant conformational changes could obviously have important functional roles, allowing or preventing specific interactions to take place. Therefore, combinations of Ser and Thr residues with Pro appear as a possible mechanism for structural adaptation of membrane proteins. Of course, our single helix simulations cannot take into account helix-helix interactions, which most likely have important structural effects as well. Nevertheless, our findings show how structurally (and hence functionally) important polar residues such as Ser and Thr can modulate the conformation of Pro-kinked TM helices, and therefore provide

a structural basis for further understanding the function of membrane proteins.

### 3.2.5. Methods

MD simulations were performed on the model peptides Ace-$(Ala)_{10}$-x-$(Ala)_{11}$-NHMe, where x is either AAAP, AASP, AATP, ASAP, ATAP, SAAP, or TAAP; and Ace-$(Ala)_{13}$-x-$(Ala)_{8}$-NHMe, where x is either PSAA, PTAA, PASA, PATA, PAAS, or PAAT. These model peptides were built in the standard Pro-kinked α-helix conformation (Table 3.2.2 in (19)). The side chains of Ser or Thr were built on both the $\chi_1$= *g*+ and *g*- rotamer conformations.

These initial structures were placed in a rectangular box (~62 Å x 55 Å x 55 Å in size) containing methane molecules (~4200 molecules in addition to the TM α-helix) to mimic the membrane environment, at a density approaching half of hydrocarbons in lipid bilayer. This is due to the different equilibrium distance between carbons in the methane box and in the polycarbon chain of the lipid (19). The peptide-methane systems were subjected to 500 iterations of energy minimisation, and then heated to 300K in 15ps. This was followed by an equilibration period (15-500ps), and a production run (500-1500ps). The simulations were carried out at constant volume and constant temperature (300 K), with the latter maintained through coupling to a heat bath. The particle mesh Ewald method was employed to compute electrostatic interactions (23). Structures were collected for analysis every 10ps during the last 1000ps of simulation (100 structures). Bending angles of these structures were calculated as the angle between the axes computed as the least square lines through the backbone atoms (N, $C_\alpha$, C) of the residues before (from 2 to 11) and after (from 16 to 24) the Pro-kink, using the InsightII software (Accelrys, San Diego). One–way analysis of variance plus a posteriori two–sided Dunnett's T tests was employed, with the SPSS 11 program (SPSS Inc. Chicago, Illinois), to contrast if the bend angle of the model peptides differs from the control. It is important to remark that this statistical test can only be applied to independent samples. In order to achieve this prerequisite, structures were collected for analysis every 10ps so that a given structure is not related to the previous and succeeding structures (Lyapunov instability) (24). The unit twist angles of these structures were calculated, for each set of four contiguous $C_\alpha$ atoms along the α-helix, using the program HELANAL (25). Representative structures for each trajectory were selected by automatically clustering the collected geometries into conformationally-related subfamilies with the program NMRCLUST (26). The MD simulations were ran with the Sander module of AMBER 5 (27), using an all-atom force field (28), the SHAKE bond constraints on all bonds, and a 2 fs integration time step.

## Acknowledgments

This work has been supported by the Ministerio de Ciencia y Tecnología (SAF2002-01509 and SAF2003-02730), and the Improving Human Potential

of the European Community (HPRI-CT-1999-00071). Computer facilities were provided by the Centre de Computació i Comunicacions de Catalunya.

## References

1. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. L., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289, 739-745.
2. Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T. et al (2003). X-ray structure of a voltage-dependent K(+) channel. *Nature*. 423, 33-41.
3. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L., and Khorana, H. G. (1996). Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* 274, 768-770.
4. Sansom, M. S. P. and Weinstein, H. (2000). Hinges, swivels and switches: the role of prolines in signalling via transmembrane α-helices. *Trends Pharmacol. Sci* 21, 445-451.
5. O'Neil, K. T. and DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250, 646–651.
6. Monne, M., Hermansson, M., and von Heijne, G. (1999). A Turn Propensity Scale for Transmembrane Helices. *J. Mol. Biol.* 288, 141-145.
7. Senes, A., Gerstein, M., and Engelman, D. M. (2000). Statistical Analysis of Amino Acid Patterns in Transmembrane Helices: The GxxxG Motif Occurs Frequently and in Association with □-branched Residues at Neighboring Positions. *J. Mol. Biol.* 296, 921-936.
8. Von Heijne, G. (1991). Proline kinks in transmembrane alpha-helices. *J. Mol. Biol.* 218, 499-503.
9. Chakrabarti, P. & Chakrabarti, S. (1998). C--H...O hydrogen bond involving proline residues in alpha-helices. *J. Mol. Biol.* 284, 867-873.
10. Cordes, F. S., Bright, J. N. & Sansom, M. S. (2002). Proline-induced distortions of transmembrane helices. *J. Mol. Biol.* 323, 951-960.
11. Gether, U. (2000). Uncovering molecular mechanisms involved in activation of G protein- coupled receptors. *Endocr. Rev.* 21, 90-113.
12. Ri, Y., Ballesteros, J. A., Abrams, C. K., Oh, S., Verselis, V. K., Weinstein, H., and Bargiello, T. A. (1999). The Role of a Conserved Proline Residue in Mediating Conformational Changes Associated with Voltage Gating of Cx32 Gap Junctions. *Biophys. J.* 76, 2887-2898.
13. Govaerts, C., Blanpain, C., Deupi, X., Ballet, S., Ballesteros, J. A., Wodak, S. J., Vassart, G., Pardo, L., and Parmentier, M. (2001). The TxP motif in the second transmembrane helix of CCR5: a structural

determinant in chemokine-induced activation. *J. Biol. Chem.* 276, 13217-13225.

14. Peralvarez, A., Barnadas, R., Sabes, M., Querol, E. & Padros, E. (2001). Thr90 is a key residue of the bacteriorhodopsin proton pumping mechanism. *FEBS Lett.* 508, 399-402.
15. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. et al (2001). The sequence of the human genome. *Science.* 291, 1304-1351.
16. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. (2003). The G protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol.* 63, 1256-1272.
17. Ballesteros, J. A. and Weinstein, H. (1995). Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. *Methods in Neurosciences* 25, 366-428.
18. Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F. E. & Vriend, G. (2003). GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.* 31, 294-297.
19. Olivella, M., Deupi, X., Govaerts, C. & Pardo, L. (2002). Influence of the Environment in the Conformation of alpha-Helices Studied by Protein Database Search and Molecular Dynamics Simulations. *Biophys J.* 82, 3207-3213.
20. Gray, T. M. and Matthews, B. W. (1984). Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.* 175, 75-81.
21. Ballesteros, J. A., Deupi, X., Olivella, M., Haaksma, E. E. J., and Pardo, L. (2000). Serine and Threonine Residues Bend α-Helices in the χ1=g- conformation. *Biophys. J.* 79, 2754-2760.
22. Govaerts, C., Bondue, A., Springael, J. Y., Olivella, M., Deupi, X., Le Poul, E. et al (2003). Activation of CCR5 by chemokines involves an aromatic cluster between transmembrane helices 2 and 3. *J. Biol. Chem.* 278, 1892-1903.
23. Darden, T. A., York, D., and Pedersen, L. (1993). Particle Mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089-10092.
24. Frenkel, D. and Smit, B. (1996). Understanding Molecular Simulation. From Algorithms to Applications. San Diego, Academic Press.
25. Bansal, M., Kumar, S. & Velavan, R. (2000). HELANAL: a program to characterize helix geometry in proteins. *J. Biomol. Struct. Dyn.* 17, 811-819.
26. Kelley, L. A., Gardner, S. P., and Sutcliffe, M. J. (1996). An Automated Approach For Clustering An Ensemble Of NMR-Derived Protein Structures Into Conformationally-Related Subfamilies. *Protein Eng.* 9, 1063-1065.

27. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vicent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1997). AMBER 5, University of California, San Francisco.

28. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr., K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Amer. Chem. Soc.* 117, 5179-5197.

29. Kraulis, J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Cryst.* 24, 946-950.

30. Merritt, E. A. and Bacon, D. J. (1997). Raster3D: Photorealistic Molecular Graphics. *Methods in Enzymology* 277, 505-524.

## *3.3. The TxP motif in the second transmembrane helix of CCR5: a structural determinant in chemokine-induced activation.*

Cédric Govaerts[1], Cédric Blanpain[1], Xavier Deupi[2], Sébastien Ballet[1], Juan A. Ballesteros[4], Shoshana Wodak[3], Gilbert Vassart[1], Leonardo Pardo[2] and Marc Parmentier[1].

[1]IRIBHN, Université Libre de Bruxelles, Campus Erasme, 808 route de Lennik, B-1070 Bruxelles, Belgium.

[2]Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

[3]SCMB, Université Libre de Bruxelles, CP 160/16, Avenue F. Roosevelt, 1050 Bruxelles, Belgium.

[4]Novasite Pharmaceuticals, Inc., San Diego, California

## Abstract

CCR5 is a G-protein coupled receptor activated by the chemokines RANTES, MIP-1α, MIP-1β and MCP-2, and is the main co-receptor for the macrophage-tropic HIV strains. We have identified a sequence motif (Thr-x-Pro, TxP) in the second transmembrane helix of chemokine receptors and investigated its role by theoretical and experimental approaches. Molecular dynamics simulations of model α-helices in a non-polar environment show that a TxP motif strongly bends α-helices, due to the coordinated action of the proline, which kinks the helix, and of the threonine, which further accentuates this structural deformation. Site-directed mutagenesis of these residues in the CCR5 receptor allowed us to probe the consequences of these structural findings in the context of the whole receptor. The P84$^{2.58}$A mutation leads to decreased binding affinity for chemokines and nearly abolished the functional response of the receptor. In contrast, mutation of T82$^{2.56}$ into Val, Ala, Cys, or Ser did not affect chemokine binding. However, the functional response was strongly dependent on the nature of the substituted side chain. The rank order of impairment of receptor activation is P84A > T82V > T82A > T82C > T82S. This ranking of impairment parallels the bending of the α-helix observed in the molecular simulation study.

### 3.3.1. Introduction

Chemokine receptors are currently one of the most extensively studied sub-families of the G-protein coupled receptors (GPCRs). This is due to their key role in the immune response, where they act as attractors and stimulators of specific leukocyte populations (1), and their essential role in HIV infectioon. In particular, the chemokine receptor CCR5 is the main co-receptor for macrophage-tropic HIV-1 strains, which are responsible for disease transmission and predominate during the asymptomatic phase of the disease (2;3). It hence appears as one of the crucial targets for developing new therapeutic strategies against HIV.

CCR5 is activated by the chemokines RANTES, MIP-1α, MIP-1β and MCP-2 and binds a natural chemokine antagonist, MCP-3 (4). Chemokines are small globular proteins, 60-100 residues long, comprising a well structured domain, and a flexible $NH_2$ terminus, with a Cys-Cys or Cys-x-Cys motif (x being a variable residue) marking the limit between the two parts (1).

The mechanisms by which chemokines bind their receptors and induce activation are currently unclear. Mutagenesis studies of chemokines suggest that a major role in binding is played by receptor interactions with their compact domain, while the flexible $NH_2$ terminus is required mainly for receptor activation (5;6). $NH_2$-terminally truncated chemokines usually bind their receptors with wild-type affinities, but elicit a severely impaired functional response (6;7).

On the receptor side, several studies have shown that its extracellular domains play an essential role in chemokine binding (8;9). In particular, the $NH_2$-terminal domain of the receptor was shown to be mandatory for chemokine binding, with several charged and aromatic residues playing a crucial role (10;11). On the other hand, most of the ligand specificity has been shown to be encoded in the second extracellular loop of CCR5 (12).

Clearly belonging to the rhodopsin-like family of GPCRs, chemokine receptors share all of the highly conserved sequence motifs characteristic of this family. The overwhelming majority of these sequence motifs are located in the transmembrane region, suggesting the conservation of a common fold for this region throughout the entire rhodopsin-like family. The existence of a conserved fold may in turn imply similar mechanisms in receptor activation involving the membrane-embedded portion of the proteins.

A detailed atomic model representing this common fold has at long last become available with the determination of the high resolution X-ray structure of bovine rhodopsin (13). This structure confirms the well documented seven-transmembrane α-helix bundle topology, proposed on the basis of lower resolution structural studies (14). It furthermore provides a detailed atomic picture of the interactions between the transmembrane helices, particularly those involving the conserved GPCR sequence motifs.

It has been established that for many receptors, which are activated by small ligands like neurotransmitters, agonist binding and subsequent triggering of activation involves a water-accessible pocket centrally located within the transmembrane helix bundle. This pocket corresponds roughly to the retinal binding site in rhodopsin (15-17). The strong similarity of the transmembrane regions of chemokine receptors to those of other rhodopsin-

like GPCRs suggests that these proteins undergo ligand-induced activation processes, involving analogous conformational changes. Some of these changes have been monitored for various GPCRs, using different techniques (for a review, see (17)). In particular, transmembrane helix 6 (TMH6) was reported to rotate its cytosolic end away from TMH3 in several receptors (18-24).

This crucial rigid-body motion of a part of TMH6 is thought to be enabled by the presence of a highly conserved proline in the middle of the helix (25), which introduces a local break in the helix structure. Such break, denoted a proline kink (Pro-kink), is likely to impart the backbone flexibility (16;26-28) required for the conformational change associated with the activation process. Mutations of the conserved Pro in TMH6 in several receptors were indeed shown to produce phenotypes ranging from impaired or severely diminished expression of the receptor (29) to reduced functional coupling (30) or even constitutive activation (31). Mutations of conserved prolines in other helices, notably TMH5 (32) and TMH7 (33;34) were also found to cause significant perturbations. For instance, proline mutations in the conserved NPxxY motif in TMH7 often produce particularly strong phenotypes, including impaired activity (33;35;36).

However, although some of the structural rearrangements associated with activation are likely to be conserved throughout the rhodopsin-like receptor family, the extraordinary diversity of ligand types, ranging from small size neurotransmitters to large glycoprotein hormones (17), suggests that receptor sub-families have presumably evolved specific binding modes with activation mechanisms probably requiring somewhat different structural adaptations.

This study investigates such sub-family specific properties in the chemokine receptors. All chemokine receptors are shown here to share a proline in TMH2. Analysis of their aligned sequences also reveals the presence of a conserved threonine residue two positions upstream of this Pro forming the TxP motif. Considering that threonine residues have been observed to induce small distortions in α-helices (37;38), we hypothesize here that the conjunction of these two conserved and structurally relevant residues in chemokine receptors might constitute a key motif required for proper receptor function. The influence of Thr in Pro-containing helices had been identified earlier for other integral membrane proteins (39).

This hypothesis is investigated using an approach that combines theoretical and experimental procedures. The theoretical procedures are aimed at characterizing the effect of the TxP motif on the intrinsic conformational properties of the transmembrane helix, and on its putative interactions with other helices in the bundle. To this end we perform molecular dynamics (MD) simulations of an isolated polyalanine helix comprising a TxP motif and several variants thereof in which Thr is replaced by other residues, Ser, Cys, Val and Ala respectively. In addition, using the three-dimensional structure of rhodopsin as a template, the structural role of this motif in the context of the seven-helix bundle is assesed. This allows to formulate hypotheses on how the conformational states of the TxP motif-containing helix might act to produce structural changes in the helix bundle.

The experimental procedures involve site-directed mutagenesis, in which the Thr of the TxP motif of TMH2 in CCR5 is replaced by the same residues as in the simulation studies, and where Ala is substituted for the Pro in order to abrogate the Pro-kink. The different CCR5 mutants are then tested in order to determine their ligand binding and activation properties.

Our results reveal a significant correspondence between the modulating effect on the Pro-kink angle and helix conformational flexibility by Thr *versus* other residues in the TxP motif and the activation properties measured experimentally for the corresponding mutants in CCR5. The implications of these findings for chemokine-induced activation are discussed.

## 3.3.2. Methods

### *Numbering scheme of GPCRs*

In this work, we use a general numbering scheme to identify residues in the transmembrane segments of different receptors (28). Each residue is numbered according to the helix (helix 1 through 7) in which it is located and according to the position relative to the most conserved residue in that helix, arbitrarily assigned to 50. For instance $Pro^{2.58}$ is the proline in the transmembrane helix 2 (TMH2), eight residues following the highly conserved aspartic acid $Asp^{2.50}$.

### *Survey of helices containing a TxP motif in known protein structures*

Since stable structural motifs are likely to recur in proteins of known structures (40), the Protein Data Bank (41) was searched for α-helical segments featuring a TxP, SxP or CxP motif, with no other Thr, Ser or Cys within the Pro-kink, and no other Pro anywhere in the segment. Since measuring the bending angle requires at least one helical turn prior to the Pro-kink and a helical turn following it, we selected helical segments at least 12 residues. We performed our search in a nonredundant set of protein structures with resolutions of 3 Å or better, identifying 16 helical segments (Protein Data Bank numbers: 1AR1, 1B7E, 1B94, 1BDB, 1BPO, 1FCB, 1FIY, 1FVK, 1OCC, 1PJC, 1REQ, 1RVE, 1TCO, 1VHB, 2AK3, and 2GST). Detailed analysis of the structures showed that most of these helices are exposed to solvent, with water molecules often interacting with the backbone at the level of the Pro-kink. Since such interactions are not likely to occur in a membrane-embedded helix, structures displaying these interactions were rejected, finally yielding only seven structures (see Table 3.3.1). The bending angle of these structures is defined as the angle between the axes computed as the least square lines through the backbone atoms (*N, $C_\alpha$, C*) of the α-helical part before and after the Pro-kink (using the InsightII software, MSI, San Diego).

### *Molecular dynamics simulations of transmembrane helices*

To study the conformational properties of an α-helix containing a TxP motif, we performed molecular dynamic simulations of the model peptides Ace–$Ala_{11}$–xxx–$Ala_{11}$–NMe, where xxx is either AAP, TAP, SAP, CAP, or VAP. These 25-residue peptides were built in the standard α-helical conformation ($\phi,\psi$=58º, -47º).

In a hydrophobic environment, the side chains of Ser, Thr and Cys are most likely to form hydrogen bonds with other polar groups of the protein or of the ligand, whenever present. Surveys of known protein structures (42;43) show that in α-helices, these side chains hydrogen bond primarily the carbonyl group in the preceding turn of the helix (residue *i-4* or *i-3*).

For Thr and Ser such bonds can form only in the *g+* or *g-* side chain conformations, whereas for Cys, they can be formed only in the *g+* conformation. The *g-* conformation of Cys is energetically unfavourable because of the steric clash between the $S_\gamma$ atom and the carbonyl oxygen of residue *i*–3 (43). For these side chains, the *t* cannot form such hydrogen bond as it points the *OH* group away from the backbone. The model peptides were hence built with the Thr and Ser side chains in either *g+* or g- , and with the Cys side chain in *g+*. The hydrophobic Val side chain was built in the *t* conformation.

Starting structures were placed in a rectangular box (59 Å x 37 Å x 38 Å) containing methane molecules at a density approaching half that of hydrocarbons in lipid bilayer, in order to mimic the plasma membrane environment. The peptide-methane systems were subjected to 500 iterations of energy minimisation, and then heated to 300 K in 15 ps. This was followed by an equilibration period (15-500 ps), and a production run (500-1500 ps). The simulations were carried out at constant volume and constant temperature (300 K), with the latter maintained through coupling to a heat bath. The particle mesh Ewald method was employed to compute electrostatic interactions (44). Structures were collected for analysis every 10 ps during the last 1000 ps of simulation. The molecular dynamics simulations were run with the Sander module of AMBER5 (45), using an all-atom force field (46), the SHAKE bond constraints on all bonds, and a 2 fs integration time step.

Bending of the Pro containing peptiedes was measured as described avobe using backbone atoms of helical segments comprising residues 2-11 (before the Pro) and 16-24 (after the Pro). One–way analysis of the variance plus *a posteriori* one–sided Dunnett's *t* tests were performed to contrast if the bend angle of the helices containing the TAP, SAP, CAP, and VAP motifs are greater than the bend angle of the helix containing the AAP motif, taken as reference. To choose representative structures for each trajectory, the structures saved during the production run were clustered on the basis of their relative backbone root mean square deviation using the NMRCLUST program (47) with a cut-off of 3 Å.

### *CCR5 mutants*

Plasmids encoding the CCR5 mutants studied here were constructed by site-directed mutagenesis using the Quickchange method (Stratagene). Following sequencing of the constructs, the mutated coding sequences were sub-cloned into the bicistronic expression vector pEFIN3 as previously described for generation of stable cell lines (12). All constructs were verified by sequencing prior to transfection.

### *Expression of mutant receptors in CHO-K1 cells*

CHO-K1 cells were cultured in Ham's F-12 medium supplemented with 10% fetal calf serum (Life Technologies), 100 units/ml penicillin and 100 μg/ml streptomycin (Life Technologies). Constructs encoding wild-type or mutant CCR5 in the pEFIN3 bicistronic vector were transfected using Fugene 6 (Boerhinger Mannheim) in a CHO-K1 cell line expressing an apoaequorin variant targeted to mitochondria (48). Selection of transfected cells was made for 14 days with 400 μg/ml G418 (Life Technologies) and 250 μg/ml zeocin (Invitrogen, for maintenance of the apoaequorin encoding plasmid), and the population of mixed cell clones expressing wild-type or mutant receptors was used for binding and functional studies. Cell surface expression of the receptor variants was measured by flow cytometry using monoclonal antibodies recognizing different CCR5 epitopes: 2D7 (phycoerythrin-conjugated, Pharmingen), MC-1, MC-4, MC-5 and MC-6 (kindly provided by Mathias Mack, Medizische Poliklinik, Ludwig-Maximilians, University of Munich, Munich, Germany) were detected by anti-mouse IgG phycoerythrin-coupled secondary antibody (Sigma).

### *[125I]-RANTES binding assays*

CHO-K1 cells expressing wild-type or mutant CCR5 were collected from plates with $Ca^{2+}$ and $Mg^{2+}$-free phosphate-buffered saline supplemented with 5 mM EDTA, gently pelleted for 2 min at 1000 × g and resuspended in binding buffer (50 mM Hepes pH 7.4, 1 mM $CaCl_2$, 5 mM $MgCl_2$, 0.5% bovine serum albumin). Competition binding assays were performed in Minisorb tubes (Nunc), with 40,000 cells in a final volume of 0.1 ml. The mixture contained 0.05 nM [$^{125}$I]-RANTES (2000 Ci/mmol, Amersham Pharmacia Biotech) as tracer, and variable concentrations of competitors (R&D Systems). Total binding was measured in the absence of competitor, and non-specific binding was measured with a 100-fold excess of unlabelled ligand. Samples were incubated for 90 minutes at 27° C, and then bound tracer was separated by filtration through GF/B filters pre-soaked in 0.5% polyethylenimine (Sigma) for [$^{125}$I]-RANTES. Filters were counted in a $\beta$-scintillation counter. Binding parameters were determined with the Prism software (GraphPad Software) using nonlinear regression applied to a one-site competition model.

### *Functional assays*

Functional response to chemokines was analysed by measuring the luminescence of aequorin as described (49;50). Cells were collected from plates with $Ca^{2+}$ and $Mg^{2+}$-free Dulbecco's modified Eagle's medium supplemented with 5 mM EDTA. They were then pelleted for 2 min at 1000 × g, resuspended in Dulbecco's modified Eagle's medium at a density of 5 x $10^6$ cells/ml and incubated for 2 h in the dark in the presence of 5 $\mu$M coelenterazine H (Molecular Probes Inc., Eugene, OR). Cells were diluted 5-fold before use. Agonists in a volume of 50 $\mu$l of Dulbecco's modified Eagle's medium were added to 50 $\mu$l of cell suspension (50,000 cells) and luminescence was measured for 30 s in a Berthold luminometer.

## 3.3.3. Results

### *A conserved TxP motif in TMH2 of chemokine receptors.*

Multiple sequence alignments of the second transmembrane helix of 55 mammalian chemokine receptors were performed. Fig. 3.3.1 shows alignment of the human and mouse sequences, together with TMH2 of bovine rhodopsin. Inspection of the aligned sequences reveals a highly conserved TxP sequence motif in TMH2, where x represents a variable hydrophobic residue. Pro, at position $84^{2.58}$ (84 is the residue number in the CCR5 sequence and *2.58* is the corresponding number in general numbering scheme), is completely conserved across all chemokine receptors. The Thr residue is also highly conserved, present in 47 sequences out of 55, while Ser is found in 4 receptors. The last 4 receptors have an Ile or Leu in *2.56*.

A survey of the 1200 rhodopsin-like GPCRs present in the G Protein-Coupled Receptors Database (51) reveals that this motif is also found at the equivalent positioon in the sequences of about 50 non-chemokine receptors, comprising esentially peptidergic, such as angiotensin and opioid, receptors.

Fig. 3.3.1 also shows that the primary structure of bovine rhodopsin TMH2 is highly similar to that of chemokine receptors from its $NH_2$ terminus (cytosolic border) to the TxP motif, but strongly diverges between the TxP motif and its COOH terminus (beginning of ECL1). This suggests a structural and functional conservation in the cytosolic half of this transmembrane segment.

### *Influence of the TxP motif on the conformation of a transmembrane helix.*

To asses the influence of the TxP motif on the conformation of TMH2, complementary approaches were used. A first approach consisted in surveying known protein structures in the Protein Data Bank for α-helices containing either a TxP, SxP or CxP motif, using the criteria specified under

"Experimental Procedures". This resulted in identifying only seven structures, all of which displayed a strong bend, with angles ranging from 25° to almost 50° as shown in Table 3.3.1. It is noteworthy that the reported average bend angle of α-helices containing Pro is about 25° (26). The helices identified here thus seem to be as strongly bent as, if not more strongly bent than, the average Pro-kink, but the very small number of observations precludes drawing a reliable conclussion. A second approach was therefore undertaken. This involved performing molecular dynamics simulations on polyalanine helices, 25 residues long, embedded in a non-polar solvent, and containing respectively the AAP, TAP, SAP, CAP, and VAP motifs, respectively, in their midst. Table 3.3.2 lists the average helix kink angles in conformations along the MD trajectories. Representative structures from the different trajectories are displayed in Fig. 3.3.2, A and B.

We find that the presence of single proline (AAP) or of the VAP motifs produces helix bend angles around 20°. For the simulation of AAP containing peptide, our results agree with those of earlier simulation studies on Pro-containing polyalanine (52).

Significantly larger bending angles (27°-35°) are observed when Thr, Ser or Cys are introduced two positions before the Pro. Detailed analysis of the conformations in the trajectories show, as expected, that the side chain hydroxyl (or SH) groups of these residues form hydrogen bonds with the carbonyl group of residue *i-4*, in a significant proportion of the conformations (85-100 %). Such hydrogen bonds are formed with Thr in the *g+* or *g-* conformations, Ser in *g+* and Cys in *g+*.

On the other hand, Ser in *g-* achieves and maintains hydrogen bonding with the carbonyl of residue *i-3* position. This seemingly minor alteration in the hydrogen bonding pattern appears however to induce a dramatic modification in the conformation of the helix, as illustrated in Fig. 3.3.2B. Not only is the helix kink angle increased significantly, but the COOH-terminal moiety of the helix points to a completely different direction in space.

The results of our simulation analysis hence suggest that in a nonpolar environment, the nature of the residue located at position *i-2* relative to the proline modulates the magnitude and direction of the Pro-kink through the formation of an hydrogen bond between the side chain and the backbone carbonyl group at positions *i-3* or *i-4*. In particular, the presence of Thr, Ser, and Cys side chains at position *i-2* relative to the Pro increases the average helix bend angle by about 10°, whereas that of Val does not.

### *Accommodating a kinked TMH2 helix in the receptor three-dimensional structure*

To obtain a rough idea on the possible consequences that the presence of TxP motif might have on the structure of the receptor, and more particularly on the TM region, we perform a molecular modeling exercise using the three-dimensional structure of rhodopsin as the template. As shown in the alignment (Fig. 3.3.1), the sequence of TMH2 is strongly conserved (about 50% sequence identity) between chemokine receptors and rhodopsin between the cytosolic border and the TxP motif. This leaves no ambiguity in

aligning the rhodopsin and CCR5 sequences in this region and allowd us to readily position representative structures from the simulations of the AAP and TAP containing model peptides into the TM bundle of rhodopsin. In particular, the backbone atoms of the two helical turns preceding the Pro-kink in our model peptides are superimposed on those of the two turns preceding the equivalent residue (in CCR5, the Pro-kink starts at *2.54*, four residue before Pro$^{2.58}$) in rhodopsin, respecting the correspondence of the sequence alignment. Interestingly, rhodopsin has two successive glycines in positions *2.56* and *2.57* (with a Phe and not a Pro at *2.58*, forming a GGF motif). Most probably as a result of the conjunction of these two flexible residues, its TMH2 is strongly distorted, so that its extracellular part leans towards TMH1.

Figure 3.3.2 (C and D) shows the result of superimposing the representative structures from the TAP *g+* (red) and AAP (yellow) simulations on TMH2 of rhodopsin. Strikingly, the kink induced by Pro$^{2.58}$ in AAP (yellow) orients the extracellular moiety of TMH2 towards TMH3 and away from TMH1 (Fig. 3.3.2C). The presence of Thr in TAP which, as shown above, increases the helix bend angle by about 10°, causes the extracellular side of TMH2 to lean even more towards TMH3 and slightly towards the center of the bundle (Fig. 3.3.2D). The differences in the amino acid sequences of TMH2 in the opsin (GGF) *versus* the chemokine (TxP) families may thus be related to structural differences in this region. In particular, in the chemokine receptors, the extracellular sides of TMH2 and TMH3 would come into close contact. It is noteworthy that chemokine receptors have a cluster of aromatic residues at the extracellular end of TMH2 and TMH3. In other GPCRs, helix-helix interactions mediated by aromatic clusters are believed to play a role in ligand-induced receptor activation (53).

### *Effects of mutations in the conserved TxP motif on CCR5 expression and function*

To investigate the possible role of the TxP motif in CCR5 expression and function, several mutants were generated in the corresponding positions. Mutant P84$^{2.58}$A was built in order to completely eliminate the Pro-kink, while mutants T82$^{2.56}$S, T82$^{2.56}$C, T82$^{2.56}$V and T82$^{2.56}$A were aimed at investigating the kink modulation effects produced by the same residues as those studied in the model peptide MD simulations.

### *Cell surface expression of the CCR5 mutants*

Cell surface expression of the CCR5 mutants was measured by fluorescence-activated cell sorting analysis using a set of monoclonal antibodies, recognizing various epitopes of the receptor, ranging from well defined linear epitopes in the $NH_2$-terminal domain (MC-5) to complex conformational epitopes spanning multiple domains (MC-6). As shown in Fig. 3.3.3, all mutants were properly expressed at the cell surface, as compared with the WT receptor. With the exception of P84$^{2.58}$A, the mutant receptors were recognized as similar levels by all monoclonals, suggesting that the mutations did not alter significantly the folding of the extracellular domain.

The pattern observed for the P84$^{2.58}$A mutant seems to indicate a deeper conformational modification for this mutant, which could in turn cause the alteration of the extracellular domain conformation and eventually affect conformational epitopes. Nevertheless, the antibodies recognizing the amino-terminal part of the receptor did detect the P84$^{2.58}$A mutant at the cell surface at levels similar to those observed for WT CCR5.

### *Chemokine binding properties of WT CCR5 and mutant receptors*

The ability of the different CCR5 mutants to bind the four high affinity CCR5 agonists, the chemokines RANTES, MIP-1α, MIP-1β and MCP-2, was tested by competition binding assays, using $^{125}$I-RANTES as tracer. Fig. 3.3.4 shows the competition curves for the various constructs.

As shown earlier, RANTES appears as the strongest ligand for WT CCR5(4), with an $IC_{50}$ of 0.28 nM (Table 3.3.3). The $IC_{50}$ for MIP-1α (3.7 nM), MIP-1β (1.3 nM) and MCP-2 (2.4 nM) are shifted to slightly higher values as compared to those obtained previously with $^{125}$I-MIP-1β as tracer, but the order of potencies is conserved (Table 3.3.3). Changes in apparent affinities as a function of the tracer used have been observed previously for chemokine and other receptors (9).

P84$^{2.58}$A is able to bind RANTES with WT affinity ($IC_{50}$ = 0.25 nM) while MIP-1β and MCP-2 show significantly decreased binding, with $IC_{50}$ values right-shifted by 3 orders of magnitude (Fig. 3.3.4 and Table 3.3.3). MIP-1α does not compete for the bound tracer, even at the highest concentration tested (1 $\mu$M).

RANTES displays an unaffected affinity for the CCR5 mutants T82$^{2.56}$S, T82$^{2.56}$C, T82$^{2.56}$A and T82$^{2.56}$V, with $IC_{50}$ values ranging from 0.2 to 0.3 *n*M. The three other ligands show affinities that are about 10 times lower than RANTES for all Thr82$^{2.56}$ mutants, as already observed for WT CCR5. MIP-1β, MIP-1α and MCP-2 display roughly WT affinities for all four Thr82$^{2.56}$ mutants, with only mild variations according to the mutant. Their $IC_{50}$ values are all in the range of 1.5-5.7 *n*M, confirming that none of the Thr82$^{2.56}$ mutants have significantly affected binding properties. The largest change is a 3-fold increase in average $IC_{50}$ for MIP-1β binding to T82$^{2.56}$V.

In summary, we find that all the analyzed mutants bind the agonists RANTES, MIP-1α, MIP-1β and MCP-2, with WT-affinities. One exception is the P84$^{2.58}$A mutants. Although it binds RANTES as well as the WT, its affinity for MIP-1β and MCP-2 is reduced, and it does not bind MIP-1α.

### *Functional activation of the mutants by CCR5 agonists*

A third and crucial set of tests was performed in order to investigate the ability of the four CCR5 mutants to be activated by the same four agonists. This was done using a sensitive assay based on the use of apoaequorin as a reporter system for intracellular calcium release. Activation of chemokine receptors, including CCR5, is known to result in calcium signaling. Control stimulation of the cell lines was achieved with a saturating concentration of ATP, which activates endogenously expressed $P_2Y_2$ receptors and generates

a strong luminescent signal. We measured the cell response to ATP in all experiments and normalized the results as percentage of this signal.

Fig. 3.3.5 shows typical activation curves obtained for the six WT and mutant receptors using the four agonists. In agreement with previous observations (4), we find that RANTES is the most potent agonist of WT CCR5, with an $EC_{50}$ of 3.5 *n*M (Table 3.3.3), whereas the $EC_{50}$ of MIP-1β is twice as large on average (7.2 *n*M), and MCP-2 and MIP-1α are somewhat less efficient , with average $EC_{50}$ values of 23.4 and 25.0 *n*M.

However, unlike for the binding assays, in which all mutants except P84$^{2.58}$A, displayed WT-like binding behavior, the functional assays show various degrees of impairment for the different mutants. The CCR5 mutants with most impaired function are P84$^{2.58}$A and T82$^{2.56}$V. P84$^{2.58}$A is activated by RANTES, with an average $EC_{50}$ right-shifted by about one order of magnitude (one log unit in Fig. 3.3.5), and an $E_{max}$ reaching only 20 % of the ATP signal. But strikingly, none of the other three agonists elicits any detectable signal in this functional assay. This is not too surprising for MIP-1*α*, considering its severely decreased affinity for this mutant ($IC_{50}$ > 1000 *n*M). MIP-1*β* and MCP-2 however, did not activate P84$^{2.58}$A, even at 300 *n*M (MCP-2) or 1000 *n*M (MIP-1*β*) concentrations (data not shown), despite full $^{125}$I-RANTES displacement at these concentrations in the binding assay.

The strongly reduced functional response of the T82$^{2.56}$V mutant is somewhat surprising, especially in light of the milder functional impairment observed for T82$^{2.56}$A (Fig. 3.3.5). Indeed a substitution of Thr by a Val is isosteric. It preserves *β*-branched character of the side chain and hence presents a less drastic change than a substitution by Ala, which truncates the side chain beyond the $C_\beta$ atom. T82$^{2.56}$V displays a dramatically reduced response to RANTES, with the activation curve for this ligand shifted to the right by more than one order of magnitude. With MIP-1*α*, high concentrations are required ($EC_{50}$=136 *n*M) to achieve moderate signals, whereas MIP-1*β* is unable to elicit detectable signals from this mutant in most experiments, although in some cases, a very weak response is measured at the highest concentrations. MCP-2 induces no response in T82$^{2.56}$V, even though a normal binding affinity is measured for this chemokine, as shown above.

The T82$^{2.56}$A mutant comes next in the degree of functional impairment. RANTES remains the best agonist for this mutant, followed by MIP-1*α*, ($EC_{50}$ : 52 *n*M), and both ligands display a reduced $E_{max}$ (~30% of maximum response). MIP-1*β* stimulates T82$^{2.56}$A with and $EC_{50}$ of 63 *n*M and MCP-2 with an $EC_{50}$ averaging 147 *n*M, both with a very low efficacy ($E_{max}$ values below 15%, Fig. 3.3.5, Table 3.3.3).

The mutants whose activities are least affected are T82$^{2.56}$S and T82$^{2.56}$C. T82$^{2.56}$S is stimulated by RANTES to a similar degree as WT CCR5, with an $EC_{50}$ of 4.5 *n*M and an $E_{max}$ of 60 %. MIP-1*β*, MIP-1*α* and MCP-2 show similar activation profiles with this mutant, characterized by a strong functional response, with an $EC_{50}$ moderately displaced to the right and an $E_{max}$ somewhat lower than for WT CCR5. RANTES and MIP-1*α* stimulate the T82$^{2.56}$C mutant with WT potencies but reduced $E_{max}$. MIP-1*β* elicits a poor response on this mutant, while MCP-2 is almost inactive.

These results taken together provide evidence on the important role played by Thr82$^{2.56}$ and Pro84$^{2.58}$ in CCR5 activation. The extent to which the activation of these mutants is affected by the series of agonists, measured under similar conditions, leads to the following ranking in terms of impairment of receptor activation: P84$^{2.58}$A>T82$^{2.56}$V>T82$^{2.56}$A>T82$^{2.56}$C>T82$^{2.56}$S.

## 3.3.4. Discussion

In this study, we identified a sequence motif TxP, in TMH2 of chemokine receptors, which is conserved throughout this important sub-family of GPCRs. We made the hypothesis that it plays an important role in receptor function and investigated this putative role by using a combination of theoretical and experimental techniques. Here, the findings from the two types of techniques are brought together and rationalized in light of our current knowledge of the CCR5 receptor structural and functional properties. This leads to the proposal of a mechanism for the implication of the TxP motif in CCR5 activation.

### *The TxP motif is a structural determinant specific of chemokine receptors*

The molecular dynamic simulations, whose aim was to investigate how the side chain at position *i-2* from the Pro residue affects the intrinsic conformational properties of a Pro-containing $\alpha$-helix, lead to the following main conclusions : the presence of Thr, Ser, and Cys side chains at position at *i-2* increases the average Pro-kink angle by around 10º, whereas Val at that position has a negligible effect. The fact that the Val side chain is nonpolar and the observation that the other analyzed side chains, which were all polar, formed hydrogen bonds with the helix backbone during the simulations suggests that the observed effect on the bending angle arises from local deformations in helix geometry induced by these bonds.

Our modeling exercise, using the high resolution three-dimensional structure of rhodopsin, suggests that the presence of the TxP motif in TMH2 would require a rearrangement of the TM helix bundle interactions in the chemokine receptors relative to rhodopsin.

In particular, Pro in position *2.58* would orient the extracellular part of TMH2 towards TMH3 and not close to TMH1 as in rhodopsin; the addition of a Thr at *i-2*, as in the TxP motif, furthermore directs this part of TMH2 toward the center of the bundle (Fig. 3.3.2, *C* and *D*). Our modeling study also suggests that in chemokine receptors, the extracellular region of TMH2 would interact with TMH3 and possibly with TMH7, which is not feasible in the rhodopsin structure.

Our results thus lead us to suggest that the TxP motif in TMH2 is a structural determinant, in chemokine receptors, by virtue of significant local effects on the helix conformation, which propagate though a lever action to the extracellular parts of the TM bundle. It is noteworthy that there is ample experimental evidence that this region of GPCRs, is involved

in their functional properties (16), which suggests in turn that mutants modifying these local comformational effects should also modify receptor function.

### *Role of the TxP motif in ligand-induced activation of CCR5*

Our functional studies of TxP mutants of CCR5 were aimed at verifying this suggestion. P84$^{2.58}$A mutant displayed a significantly affected pattern of binding for some ligands. We hypothesize that the profound structural perturbation in this mutant, caused by the abrogation of the Pro-kink, might change the conformation of some part of the extracellular domain, thus modifying the interactions with ligands. In agreement with this view, this putative conformational change could also be responsible for the observed decrease in recognition of the P84$^{2.58}$A mutant by monoclonal antibodies directed at conformational epitopes. This suggests that this Pro-kink is mandatory for the structural integrity of the protein. Alanine replacement of Pro84$^{2.58}$ abolished the functional response of CCR5 to any of its agonists with the exception of RANTES, which induced minute activity, demonstrating the central role of this Pro-kink in the activation process.

All tested chemokines bound the various Thr82$^{2.56}$ mutants with unchanged affinity, indicating that this residue is not involved in direct interaction with the ligands. In contrast to the binding properties, the chemokine-induced activation of CCR5 was quite sensitive to mutations of Thr82$^{2.56}$, showing a gradation of the effects corresponding to that observed in the simulation.

Activation profiles of T82$^{2.56}$S showed mild differences with those of WT CCR5. However, this mutant had a reduced $E_{max,}$ when stimulated with MIP-1$\alpha$ MIP-1$\beta$ and MCP-2. This behaviour is reminiscent of that observed with partial agonists on WT receptors, and hence suggests that these chemokines cannot fully activate this mutant.

This could occur if the equilibrium between the active and inactive forms of T82$^{2.56}$S is somewhat shifted towards the inactive one. The fact that RANTES activates this mutant normally, could mean that it stabilizes more efficiently the active forms. The results of our MD simulations on the SAP-containing peptide suggest a molecular explanation to this shift in equilibrium. They show that this peptide features a similar $\alpha$-helix kink geometry as in the TAP-containing peptide, when Ser is in the $g$+ conformation, but adopts a very different orientation of the Pro-kink when this side chain is in $g$- (see Fig. 3.3.2B). This latter conformation may stabilize the inactive form of mutant receptors. This effect of Ser might possibly also explain the paucity of Ser residues at position 2.56 in the chemokine receptors (only 4 out of 55 members have it).

The isosteric mutant T82$^{2.56}$V was even more impaired in the functional tests than T82$^{2.56}$A. T82$^{2.56}$V required high chemokine concentrations to trigger very modest activity, although the binding properties were not significantly affected for any of the ligands tested. We cannot provide a simple explanation for the difference of phenotype observed between these two mutants, but considering the observations made in the MD simulations, the hydrogen bonding capacity of the side chain at position

*2.56* seems to be a more important feature for activation than the *β*-branched character of the side chain.

Our results show that, first, the chemokine receptors, the Pro in TMH2 is crucial for proper receptor activation and that the conserved Thr located two positions before the Pro modulates the function of the receptor. The experiments therefore confirm the theoretical hypothesis stating that the TxP motif plays a key structural role in chemokine receptors, the Pro-kink constituting the main element and the Thr acting as a modulator of this Pro-kink. Moreover, this motif is mainly involved in receptor activation, but plays little role in ligand binding. These results also demonstrate that high affinity binding of chemokines by CCR5 is not dependent on the coupling state with G proteins. A previous description of other non-functional CCR5 mutants characterized by unimpaired affinities for their chemokine ligands (10) supports this hypothesis. Along the same line of evidence, it is well established that $NH_2$-terminally truncated chemokines often keep their high affinities for their cognate receptors, while becoming antagonists or weak partial agonists.

### *New insights into chemokine-induced activation*

Surprisingly, although the TxP motif is present in all chemokine receptors, functional alteration observed in the T82$^{2.56}$ mutants is strongly chemokine-dependent. RANTES is the least affected agonist, MCP-2 being strongly sensitive to all mutations, while MIP-1$\alpha$ and MIP-1$\beta$ behave as partial agonists for the various mutants. How can we explain these observations? Interestingly, RANTES differs structurally from the other CCR5 agonists in its amino terminus, with 9 residues before the first conserved cysteine for RANTES, and 10 residues for the three other ligands. Also, the sequence of MCP-2 $NH_2$-terminus is somewhat different from that of the other agonists, whereas those of MIP-1$\alpha$ and MIP-1$\beta$ are quite similar to each other. Hence there seems to be a relation between the sensitivity of receptor activation by the different chemokines to the mutations in TxP and the differences between the $NH_2$-terminal regions of these molecules. It is significant that these regions were shown to be important in receptor activation by mutation and deletion experiments (6;7).

This leads us to formulate the hypothesis that chemokine induced activation involves interactions of the ligand $NH_2$ terminus with the portion of transmembrane domain whose conformation may be modulated by the TxP motif. Thus the differences in behavior among ligands could result from differences in these interactions. This hypothesis could be investigated by testing mutant receptor activation by chimeric chemokines exchanging the $NH_2$ terminus or the core region.

As mentioned above, chemokines truncated in their $NH_2$ terminus often still elicit some functional responses at high concentrations, and hence behave as partial agonists. This parallels the effects of full-size chemokines on our Thr82$^{2.56}$ mutant series, and could result from the binding of the chemokine core region to the extracellular loops of the receptor. It was proposed that this interaction would induce partial signal transduction, characterised by inhibition of cAMP, but would not trigger calcium influx nor chemotaxis (6).

Thus, it does not contradict our proposal that full activation involves interactions with the $NH_2$-terminal region of the chemokine.

Based on chimeras between CCR2b and CCR1, it was recently proposed that activation of chemokine receptors involves the first extracellular loop ECL1 (54). Having shown here that TMH2 in this receptor family extends to position *2.67*, the effects of these chimeras are most probably due to the exchange of variable residues at the extracellular portion of this helix rather than of the swap of ECL1, which is well conserved in these receptors. Interestingly, these changes produced mutant receptors with good ligand affinity, but impaired chemokine-induced functional response. These data parallel our finding and hence agree with our hypothesis on the importance of the extracellular part of TMH2 in receptor activation.

Other recent studies have described low molecular weight chemical compounds acting as antagonists (55-57) with promising applications in the treatment of inflammatory diseases and HIV infection. Interestingly, the binding sites of these antagonists have been located within the transmembrane bundle of the receptors (58;59). Insights provided here on the role of structural determinants in this region may therefore be helpful in further elucidating the action of these compounds and in designing effective drugs. The availability of the high resolution structure of rhodopsin will greatly facilitate this task and allow more detailed analyses of the type described here to be extended to other rhodopsin-like GPCRs.

## References

1. Baggiolini M 1998 Chemokines and leukocyte traffic. Nature 392:565-8

2. Littman DR 1998 Chemokine receptors: keys to AIDS pathogenesis? Cell 93:677-80

3. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG, Doms RW, Vassart G, Parmentier M 1996 Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722-5

4. Blanpain C, Migeotte I, Lee B, Vakili J, Doranz BJ, Govaerts C, Vassart G, Doms RW, Parmentier M 1999 CCR5 binds multiple CC-chemokines: MCP-3 acts as a natural antagonist. Blood 94:1899-905

5. Hemmerich S, Paavola C, Bloom A, Bhakta S, Freedman R, Grunberger D, Krstenansky J, Lee S, McCarley D, Mulkins M, Wong B, Pease J, Mizoue L, Mirzadegan T, Polsky I, Thompson K, Handel TM, Jarnagin K 1999 Identification of residues in the monocyte chemotactic protein-1 that contact the MCP-1 receptor, CCR2. Biochemistry 38:13013-25

6. Jarnagin K, Grunberger D, Mulkins M, Wong B, Hemmerich S, Paavola C, Bloom A, Bhakta S, Diehl F, Freedman R, McCarley D, Polsky I, Ping-

Tsou A, Kosaka A, Handel TM 1999 Identification of surface residues of the monocyte chemotactic protein 1 that affect signaling through the receptor CCR2. Biochemistry 38:16167-77

7. Gong JH, Clark-Lewis I 1995 Antagonists of monocyte chemoattractant protein 1 identified by modification of functionally critical NH2-terminal residues. J Exp Med 181:631-40

8. LaRosa GJ, Thomas KM, Kaufmann ME, Mark R, White M, Taylor L, Gray G, Witt D, Navarro J 1992 Amino terminus of the interleukin-8 receptor is a major determinant of receptor subtype specificity. J Biol Chem 267:25402-6

9. Ahuja SK, Murphy PM 1996 The CXC chemokines growth-regulated oncogene (GRO) alpha, GRObeta, GROgamma, neutrophil-activating peptide-2, and epithelial cell-derived neutrophil-activating peptide-78 are potent agonists for the type B, but not the type A, human interleukin-8 receptor. J Biol Chem 271:20545-50

10. Farzan M, Choe H, Martin KA, Sun Y, Sidelko M, Mackay CR, Gerard NP, Sodroski J, Gerard C 1997 HIV-1 entry and macrophage inflammatory protein-1beta-mediated signaling are independent functions of the chemokine receptor CCR5. J Biol Chem 272:6854-7

11. Blanpain C, Doranz BJ, Vakili J, Rucker J, Govaerts C, Baik SS, Lorthioir O, Migeotte I, Libert F, Baleux F, Vassart G, Doms RW, Parmentier M 1999 Multiple charged and aromatic residues in CCR5 amino-terminal domain are involved in high affinity binding of both chemokines and HIV-1 Env protein. J Biol Chem 274:34719-27

12. Samson M, LaRosa G, Libert F, Paindavoine P, Detheux M, Vassart G, Parmentier M 1997 The second extracellular loop of CCR5 is the major determinant of ligand specificity. J Biol Chem 272:24934-41

13. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M 2000 Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289:739-45

14. Unger VM, Hargrave PA, Baldwin JM, Schertler GF 1997 Arrangement of rhodopsin transmembrane alpha-helices. Nature 389:203-206

15. Strader CD, Fong TM, Tota MR, Underwood D, Dixon RA 1994 Structure and function of G protein-coupled receptors. Annual Review of Biochemistry 63:101-132

16. Ji TH, Grossmann M, Ji I 1998 G protein-coupled receptors. I. Diversity of receptor-ligand interactions. J Biol Chem 273:17299-302

17. Gether U, Kobilka BK 1998 G protein-coupled receptors. II. Mechanism of agonist activation. Journal of Biological Chemistry 273:17979-17982

18. Elling CE, Nielsen SM, Schwartz TW 1995 Conversion of antagonist-binding site to metal-ion site in the tachykinin NK-1 receptor. Nature 374:74-7

19. Sheikh SP, Vilardarga JP, Baranski TJ, Lichtarge O, Iiri T, Meng EC, Nissenson RA, Bourne HR 1999 Similar structures and shared switch mechanisms of the beta2- adrenoceptor and the parathyroid hormone

receptor. Zn(II) bridges between helices III and VI block activation. J Biol Chem 274:17033-41

20. Sheikh SP, Zvyaga TA, Lichtarge O, Sakmar TP, Bourne HR 1996 Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F. Nature 383:347-350

21. Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG 1996 Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. Science 274:768-770

22. Dunham TD, Farrens DL 1999 Conformational changes in rhodopsin. Movement of helix f detected by site-specific chemical labeling and fluorescence spectroscopy. J Biol Chem 274:1683-90

23. Javitch JA, Fu D, Liapakis G, Chen J 1997 Constitutive activation of the beta2 adrenergic receptor alters the orientation of its sixth membrane-spanning segment. J Biol Chem 272:18546-9

24. Rasmussen SG, Jensen AD, Liapakis G, Ghanouni P, Javitch JA, Gether U 1999 Mutation of a highly conserved aspartic acid in the beta2 adrenergic receptor: constitutive activation, structural instability, and conformational rearrangement of transmembrane segment 6. Mol Pharmacol 56:175-84

25. Gether U, Lin S, Ghanouni P, Ballesteros JA, Weinstein H, Kobilka BK 1997 Agonists induce conformational changes in transmembrane domains III and VI of the beta2 adrenoceptor. EMBO Journal 16:6737-6747

26. Barlow DJ, Thornton JM 1988 Helix geometry in proteins. Journal of Molecular Biology 201:601-619

27. Woolfson DN, Williams DH 1990 The influence of proline residues on alpha-helical structure. FEBS 277:185-8

28. Ballesteros JA, Weinstein H 1995 Integrated Methods for Modeling G-Protein Coupled Receptors. Method Neurosci 25:366-428

29. Kolakowski LF Jr, Lu B, Gerard C, Gerard NP 1995 Probing the "message:address" sites for chemoattractant binding to the C5a receptor. Mutagenesis of hydrophilic and proline residues within the transmembrane segments. J Biol Chem 270:18077-82

30. Nakayama TA, Khorana HG 1991 Mapping of the amino acids in membrane-embedded helices that interact with the retinal chromophore in bovine rhodopsin. J Biol Chem 266:4269-75

31. Tonacchera M, Chiovato L, Pinchera A, Agretti P, Fiore E, Cetani F, Rocchi R, Viacava P, Miccoli P, Vitti P 1998 Hyperfunctioning thyroid nodules in toxic multinodular goiter share activating thyrotropin receptor mutations with solitary toxic adenoma. J Clin Endocrinol Metab 83:492-8

32. Javitch JA, Fu D, Chen J 1995 Residues in the fifth membrane-spanning segment of the dopamine D2 receptor exposed in the binding-site crevice. Biochemistry 34:16433-9

33. Wess J, Nanavati S, Vogel Z, Maggio R 1993 Functional role of proline and tryptophan residues highly conserved among G protein-coupled receptors studied by mutational analysis of the m3 muscarinic receptor. EMBO J 12:331-8

34. Hong S, Ryu KS, Oh MS, Ji I, Ji TH 1997 Roles of transmembrane prolines and proline-induced kinks of the lutropin/choriogonadotropin receptor. J 272:4166-71

35. Vichi P, Whelchel A, Posada J 1999 Transmembrane helix 7 of the endothelin B receptor regulates downstream signaling. J Biol Chem 274:10331-8

36. Barak LS, Menard L, Ferguson SS, Colapietro AM, Caron MG 1995 The conserved seven-transmembrane sequence NP(X)2,3Y of the G-protein-coupled receptor superfamily regulates multiple properties of the beta 2-adrenergic receptor. Biochemistry 34:15407-14

37. Blundell T, Barlow D, Borkakoti N, Thornton J 1983 Solvent-induced distortions and the curvature of alpha-helices. Nature 306:281-283

38. Ballesteros JA, Deupi X, Olivella M, Haaksma EE, Pardo L 2000 Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. Biophys J 79:2754-60

39. Ri Y, Ballesteros JA, Abrams CK, Oh S, Verselis VK, Weinstein H, Bargiello TA 1999 The role of a conserved proline residue in mediating conformational changes associated with voltage gating of Cx32 gap junctions. Biophys J 76:2887-98

40. Wintjens RT, Rooman MJ, Wodak SJ 1996 Automatic classification and analysis of alpha alpha-turn motifs in proteins. J Mol Biol 255:235-53

41. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M 1977 The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112:535-42

42. Gray TM, Matthews BW 1984 Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. Journal of Molecular Biology 175:75-81

43. McGregor MJ, Islam SA, Sternberg MJ 1987 Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. J Mol Biol 198:295-310

44. Darden T.A., York D, Pedersen P 1993 Particle Mesh Ewald: An N log(N) method for Ewald sums in large systems. J. Chem.Phys. 98:10089-10092

45. AMBER 5. Case DA, Pearlman DA, Caldwell JW, Cheatham TE, Ross WS, Simmerling TA, Darden TA, Merz KM, Stanton RV, Cheng AL, Vicent JJ, Crowley M, Ferguson DM, Radmer RJ, Seibel GL, Singh UC, Weiner PK, Kollman PA, editors. 1997. University of California San Francisco.

46. Cornell WD, Cieplak P 1995 A Second generation Force Field for the simulation of proteins, nucleic acids and organic molecules. Journal of American Chemistry society 117:5179-5197

47. Kelley L.A., Gardner SP, Sutcliffe MJ 1996 An Automated Approach for clustering an ensemble of NMR-Derived Protein Structures Into Conformationally-Related subfamilies. Protein Eng. 9:1063-1065

48. Rizzuto R, Pinton P, Carrington W, Fay FS, Fogarty KE, Lifshitz LM, Tuft RA, Pozzan T 1998 Close contacts with the endoplasmic reticulum as

determinants of mitochondrial Ca2+ responses. Science 280:1763-6

49. Stables J, Green A, Marshall F, Fraser N, Knight E, Sautel M, Milligan G, Lee M, Rees S 1997 A bioluminescent assay for agonist activity at potentially any G- protein-coupled receptor. Anal Biochem 252:115-26

50. Blanpain C, Lee B, Vakili J, Doranz BJ, Govaerts C, Migeotte I, Sharron M, Dupriez V, Vassart G, Doms RW, Parmentier M 1999 Extracellular cysteines of CCR5 are required for chemokine binding, but dispensable for HIV-1 coreceptor activity. J Biol Chem 274:18902-8

51. Horn F, Weare J, Beukers MW, Horsch S, Bairoch A, Chen W, Edvardsen O, Campagne F, Vriend G 1998 GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 26:275-9

52. Yun RH, Anderson A, Hermans J 1991 Proline in alpha-helix: stability and conformation studied by dynamics simulation. Proteins 10:219-28

53. Javitch JA, Ballesteros JA, Chen J, Chiappa V, Simpson MM 1999 Electrostatic and aromatic microdomains within the binding-site crevice of the D2 receptor: contributions of the second membrane-spanning segment. Biochemistry 38:7961-8

54. Han KH, Green SR, Tangirala RK, Tanaka S, Quehenberger O 1999 Role of the first extracellular loop in the functional activation of CCR2. The first extracellular loop contains distinct domains necessary for both agonist binding and transmembrane signaling. J Biol Chem 274:32055-62

55. Donzella GA, Schols D, Lin SW, Este JA, Nagashima KA, Maddon PJ, Allaway GP, Sakmar TP, Henson G, De Clercq E, Moore JP 1998 AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 co- receptor. Nat Med 4:72-7

56. Baba M, Nishimura O, Kanzaki N, Okamoto M, Sawada H, Iizawa Y, Shiraishi M, Aramaki Y, Okonogi K, Ogawa Y, Meguro K, Fujino M 1999 A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity. Proc Natl Acad Sci USA 96:5698-703

57. Hesselgesser J, Ng HP, Liang M, Zheng W, May K, Bauman JG, Monahan S, Islam I, Wei GP, Ghannam A, Taub DD, Rosser M, Snider RM, Morrissey MM, Perez HD, Horuk R 1998 Identification and characterization of small molecule functional antagonists of the CCR1 chemokine receptor. J Biol Chem 273:15687-92

58. Dragic T, Trkola A, Thompson DA, Cormier EG, Kajumo FA, Maxwell E, Lin SW, Ying W, Smith SO, Sakmar TP, Moore JP 2000 A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. Proc Natl Acad Sci USA 97:5639-44

59. Mirzadegan T, Diehl F, Ebi B, Bhakta S, Polsky I, McCarley D, Mulkins M, Weatherhead GS, Lapierre JM, Dankwardt J, Morgans D Jr, Wilhelm R, Jarnagin K 2000 Identification of the binding site for a novel class of CCR2b chemokine receptor antagonists: binding to a common chemokine receptor motif within the helical bundle. J Biol Chem 275:25562-71

## *3.4. Activation of CCR5 by chemokines involves an aromatic cluster between transmembrane helices 2 and 3.*

Cédric Govaerts[1], Antoine Bondue[1], Jean-Yves Springael[1], Mireia Olivella[2], Xavier Deupi[2], Emmanuel Le Poul[3], Shoshana J. Wodak[4], Marc Parmentier[1], Leonardo Pardo[2] and Cédric Blanpain[1].

[1]IRIBHN, Université Libre de Bruxelles, Campus Erasme, 808 route de Lennik, B-1070 Bruxelles, Belgium.

[2]Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

[3]Euroscreen s.a., B-1070 Bruxelles, Belgium

[4]SCMB, Université Libre de Bruxelles, CP 160/16, Avenue F. Roosevelt, 1050 Bruxelles, Belgium.

## Abstract

CCR5 is a G protein-coupled receptor responding to four natural agonists, the chemokines RANTES (regulated on activation normal T cellexpressed and secreted), macrophage inflmmatory proteins MIP-1$\alpha$, MIP-1$\beta$ and monocyte chemotactic protein MCP-2, and is the main co-receptor for the macrophage-tropic human inmunodeficiency virus strains. We have previously identified a structural motif in the second transmembrane helix of CCR5 which plays a crucial role in the mechanism of receptor activation. We now report the specific role of aromatic residues in helices 2 and 3 of CCR5 in this mechanism. Using site-directed mutagenesis and molecular modeling in a combined approach, we demonstrate that a cluster of aromatic residues at the extracellular border of these two helices are involved in chemokine-induced activation. These aromatic residues are involved in interhelical interactions that are key for the conformation of the helices and govern the functional response to chemokines in a ligand-specific manner. We therefore suggest that helices 2 and 3 contain important structural elements for the activation mechanism of chemokine receptors, and possibly other related receptors as well.

## 3.4.1. Introduction

Chemokine receptors have been dragging more and more attention since the cloning of the first member of the family a decade ago. Not only chemokines and their receptors are now considered as the main organizers of leukocyte trafficking, they have also been associated to an ever increasing number of physiopathological disorders (for review see Refs. 1 and 2). In particular, some chemokine receptors are used by the human immunodeficiency virus (HIV) as coreceptors (1) (in addition to CD4) to infect target cells (reviewed in Ref. 3). Among these, CCR5 has been shown to be essential for HIV pathogenesis, as individual homozygous for the CCR5Δ32 mutation, which results in the synthesis of a non-functional receptor, are highly (although not fully) resistant to HIV infection. Chemokine receptors belong to the rhodopsin-like family (family A) of G protein-coupled receptors (GPCR). CCR5 binds and responds to four natural chemokines, RANTES, MIP-1$\alpha$, MIP-1$\beta$ and MCP-2, with nanomolar affinities (4).

Our current understanding of the activation mechanisms of GPCRs is rapidly evolving, thanks to the availability of the crystal structure of the inactive state of one of its members, rhodopsin, and to the growing amount of biochemical and physicochemical data that can help in identifying the key aspects of this process (for review, see Refs. 5 and 6). It is now well accepted that the transition from inactive to active states requires the reorganization of the transmembrane bundle made of seven imperfect alpha helices. In the rhodopsin-like family, motions of transmembrane helix 3 (TMH3) and TMH6 during the activation process have been identified (7-9). Rigid-body movements have also been proposed for TMH5 and TMH7 (10,11). Most of these conformational changes have been observed in a diverse set of receptors (mainly rhodopsin and the $\beta_2$-adrenergic receptor), and it is believed that they constitute a common conformational path in the activation process, ultimately leading to the release of GDP in the bound G protein and its exchange for GTP.

A striking feature within the rhodopsin-like GPCR family is that, despite a strong sequence conservation of the transmembrane helices, there is a wide structural diversity amongst extracellular ligands, ranging from small neurotransmitters to large glycoproteins. The structural adaptation of a receptor to its cognate ligand is expected to involve, in most cases, sequence specificity in the extracellular domains (for which sequence conservation is much lower across the family), but also in the transmembrane region, which holds the binding pocket of small ligands (for reviews see Refs. 6 and 12). Therefore, each receptor must have evolved specific structural characteristics in order to link the specific recognition of its cognate ligand to what is believed to constitute a common activation process (e.g. motion of TMH6). Thus, for each receptor subfamily, specific structural features needed for the activation by agonists are expected to be found.

In the case of chemokine receptors, we have recently identified a structural motif in TMH2 which is central for chemokine-induced activation (13). This motif, which consists in a proline preceded by a threonine two residues ahead (TxP), is found only in chemokine receptors and a few related peptide receptors. Our study indicated that the extracellular part of TMH2, whose conformation is governed by the TxP motif, is clearly involved in the

activation process, as specific mutations of the motif led to unaffected chemokine binding but strong impairment of receptor activation (13). Moreover, modeling studies performed on this region suggested that, due to the action of the proline, the extracellular part of TMH2 would strongly interact with TMH3. This organization is structurally different from that of bovine rhodopsin, in which a TMH2-TMH1 interaction is found.

As a follow-up of these observations, we have now investigated the possible role of the TMH2-TMH3 interface in the activation process of chemokine receptors. A sequence alignment of chemokine receptors (Fig. 3.4.1) reveals that the extracellular parts of TMH2 and TMH3 contain many aromatic residues. Within the rhodopsin template, these residues are located at relatively short distances, suggesting that they might form an aromatic cluster within the tridimentional structure of the receptors. Aromatic residues have been proposed to be involved in the activation mechanism in various GPCRs (14-22). Among other examples, a role was attributed to such residues in the ligand selectivity and ligand-induced activation of the D2 and D4 dopamine receptors (20,23). The high density of aromatic residues at the top of TMH2 and TMH3 in chemokine receptors suggested that aromatic side chains could mediate interactions between these helices.

In the present study, we have mutated the aromatic residues of CCR5 TMH2 and TMH3 into their CCR2 counterparts, either individually or in combination, and the mutants were tested for cell-surface expression, receptor conformation, ligand binding and functional response. Molecular modeling of the transmembrane region of CCR5 has been performed, providing a structural framework to interpret these data. Integration of the experimental and molecular modeling data indicates that aromatic residues at the TMH2-TMH3 interface are crucial to the mechanism of receptor activation and suggests that this aromatic cluster plays a key role in the conformational changes of CCR5, leading from ligand recognition to receptor activation.

## 3.4.2. Methods

### *Numbering scheme of GPCRs*

In this work, we use a general numbering scheme to identify residues in the transmembrane segments of different receptors (24). Each residue is numbered according to the helix (1 through 7) in which it is located and to the position relative to the most conserved residue in that helix, arbitrarily assigned to 50. For instance Pro$^{2.58}$ is the proline in the transmembrane helix 2 (TMH2), eight residues following the highly conserved aspartic acid Asp$^{2.50}$.

### *Survey of transmembrane helices containing a FWxxY motif in known membrane protein structures*

We surveyed the atomic coordinates of the membrane proteins bacteriorhodopsin (PDB access number 1c3w, 1.55 Å resolution), aa3 (1occ,

2.8 Å) and ba3 (1ehk, 2.4 Å) cytochrome c oxidases, photosyntethic reaction center (1prc, 2.3 Å), potassium channel (1bl8, 3.2 Å), mechanosensitive ion channel (1msl, 3.5 Å), rhodopsin (1f88, 2.8 Å), halorhodopsin (1el2, 1.8 Å), sensory rhodopsin (1h68, 2.1 Å), light harvesting complex (1lgh, 2.4 Å), photosystem I (1jbo, 2.5 Å), AQP1 (1hwo, 3.7 Å) and GlpF (1fx8, 2.2 Å) channels, P-type ATPasa (1eul, 2.6 Å), and fumerate reductase respiratory complex (1qla, 2.2 Å) for α-helical segments featuring the FWxxY motif.

### *Molecular dynamics simulations of transmembrane helices*

The model peptides Ace–$Ala_{11}$–Thr–Ala–Pro–$Ala_{11}$–Nme and Ace–$Ala_7$–Thr–Gly–$Ala_4$–Gly–$Ala_2$–Ser–Gly–$Ala_{15}$–Nme were built in the standard $\alpha$-helical conformation ($\phi$, $\psi$ = –58º, –47º). The amino acid side chains of Ser and Thr were set to the *g+* conformation. Molecular dynamics simulations of these model peptides aim to explore the conformation of TMH2 in CCR5 triggered by the Thr$^{2.56}$xPro$^{2.58}$ motif and TM 3 triggered by the presence of Thr$^{3.29}$, Gly$^{3.30}$, Gly$^{3.35}$, Ser$^{3.38}$, and Gly$^{3.39}$. A similar approach was rcently used to model the conformation of TMH3 in the $5HT_{1A}$ serotonin receptor (25). Ser and Thr residues induce a small bending angle in TM because of the additional hydrogen bond formed between the O$\gamma$ atom of Ser or Thr and the *i–3* or *i–4* peptide carbonyl oxygen (26). Moreover, the additional flexibility provided by the adjacent Pro (due to the absence of the hydrogen bond with the carbonyl oxygen in the preceding turn of the helix) or Gly (due to the lack of the side chain) reinforces this effect. The obtained structures were placed in a rectangular box containing methane molecules to mimic the hydrophobic environment of the TM bundle. The peptide-methane systems were subjected to 500 iterations of energy minimization and then heated to 300 K in 15 ps. This was followed by an equilibration period (15 to 500ps) and a production run (500 to 1500ps) at constant volume using the particle mesh Ewald method to evaluate electrostatic interactions. Structures were colected for analysis every 10 ps during the production run (100 structures per simulation). In order to obtain a rough idea on the possible consequences that the presence of these residues in TMH2 and TM 3 might have on the structure of the receptor, we perform a molecular modeling exercise using the three-dimensional structure of rhodopsin as the template. The backbone of one helical turn preceding the highly conserved Asp$^{2.50}$ in TMH2 and E/DR$^{3.50}$Y motif in TMH3 superimposed the helix bundle of rhodopsin with the computed structures.

### *Molecular dynamics simulations of the CCR5 receptor and mutant receptors*

The three-dimensional model of transmembrane helices 1 and 4-7 of CCR5 was constructed by computer-aided model building techniques from the transmembrane domain of bovine rhodopsin, as determined by Palczewski et al (27). The following conserved residues were employed in the alignment of rhodopsin and human CCR5 transmembrane sequences: Asn55 (55 being the residue number in the 1F88 PDB file of Rhodopsin) and Asn48$^{1.50}$ (48 is the residue number in the CCR5 sequence, *1.50* in the standarized numbering); Trp161 and Trp153$^{4.50}$; Pro215 and Pro206$^{5.50}$; Pro267 and

Pro250$^{6.50}$; and Pro303 and Pro294$^{7.50}$. Representative structures of transmembrane helices 2 and 3, selected by automatically clustering the geometries obtained during the molecular dynamics trajectories into conformationally-related subfamilies with the program NMRCLUST (28), were included into the model (see *Molecular dynamics simulations of transmembrane helices* above). SCWRL-2.1 was employed to add the side chains of the non-conserved residues based on a backbone-dependent rotamer library (29). All ionizable residues in the helices were considered uncharged with the exception of Asp$^{2.40}$, Asp$^{2.50}$, Asp$^{3.49}$, Arg$^{3.50}$, Lys$^{5.50}$, Arg$^{6.30}$, Arg$^{6.32}$ and Glu$^{7.39}$. To relieve residual strain resulting from suboptimal positioning of the side chains and the TMH2-TMH3 interface at the extracellular part, this resulting initial structure was placed in a rectangular box containing methane molecules, energy minimized (500 steps), heated (from 0 to 300 K in 15 ps) and equilibrated (from 15 to 100 ps). During these processes the $C_\alpha$ atoms were kept fixed at their positions in the rhodopsin crystal structure, with the exception of the residues forming the TMH2-TMH3 interface (from *2.58* to *3.29*). The optimized TMH2-TMH3 interface accomplishes: *i)* the distance between the top (C-terminus) of TMH2 and the top (N-terminus) of TMH3 is in the 10-11Å range to allow the first extracellular loop of 4 residues to be shaped; *ii)* Phe85$^{2.59}$ interacts with Leu104$^{3.28}$ and Tyr89$^{2.63}$ interacts with Thr99$^{3.23}$, in a similar manner to the cytochrome C oxidase structure of helices III and VII in subunit III (see "Results"); and *iii)* there are not steric clashes between helices. The interactions of the side chain of Leu104$^{3.28}$ with the side chains of Phe85$^{2.59}$ and Trp86$^{2.60}$ were further characterized by *ab-initio* quantum mechanical calculations at the MP2/6-31G* level of theory, which is capable of describing the proposed C-H$\cdots\pi$ interactions (30).

Molecular models of the helix bundles for the mutant receptors containing the F85$^{2.59}$L, L104$^{3.28}$F, F85$^{2.59}$L-L104$^{3.28}$F, F109$^{3.33}$H, F112$^{3.36}$Y, and F109$^{3.33}$H-F112$^{3.36}$Y substitutions were constructed from the previously obtained structure of CCR5, by changing the atoms implicated in the aminoacid substitutions by interactive computer graphics. Subsequently, wild type and mutant receptors were placed in a rectangular box (~71 Å x 60 Å x 50 Å in size) containing methane molecules (~2850 molecules in addition to the transmembrane domain) to mimic the hydrophobic environment of the transmembrane helices. The density of 0.4-0.5 g cm$^{-3}$ of the methane box is approximately half of the density observed in the hydrophobic core of the membrane. This is due to the different equilibrium distance between carbons in the methane box and in the polycarbon chain of the lipid. However, it has been shown that this procedure reproduce several structural characteristics of membrane embedded proteins (31). The receptor-methane systems were subjected to 500 iterations of energy minimization and then heated to 300 K in 15 ps. This was followed by an equilibration period (15 to 100ps) and a production run (100 to 250ps) at constant volume using the particle mesh Ewald method to evaluate electrostatic interactions (32). Structures were colected for analysis every 10 ps during the production run (15 structures per simulation). The molecular dynamics simulations were run with the Sander module of AMBER 5 (33), the all-atom force field (34), SHAKE bond constraints in all bonds, a 2 fs integration time step, and constant temperature of 300 K coupled to a heat bath.

### *CCR5 mutants*

Plasmids encoding the CCR5 mutants studied here were constructed by site-directed mutagenesis using the Quickchange method (Stratagene). Following sequencing of the constructs, the mutated coding sequences were sub-cloned into the bi-cistronic expression vector pEFIN3 as previously described for generation of stable cell lines (35). All constructs were verified by sequencing prior to transfection.

### *Expression of mutant receptors in CHO-K1 cells*

CHO-K1 cells were cultured in HAM's F12 medium supplemented with 10% fetal calf serum (Invitrogen), 100 units/ml penicillin and 100 μg/ml streptomycin (Invitrogen). Constructs encoding wild-type or mutant CCR5 in the pEFIN3 bicistronic vector were transfected using FuGENE 6 (Roche Molecular Biochemicals) in a CHO-K1 cell line expressing an apoaequorin variant targeted to mitochondria (36). Selection of transfected cells was made for 14 days with 400 μg/ml G418 (Invitrogen) and 250 μg/ml zeocin (Invitrogen, for maintenance of the apoaequorin encoding plasmid), and the population of mixed cell clones expressing wild-type or mutant receptors was used for binding and functional studies. Cell surface expression of the receptor variants was measured by flow cytometry using monoclonal antibodies recognizing different CCR5 epitopes: 2D7 (phycoerythrin-conjugated, PharMingen), MC-1, MC-4, MC-5 and MC-6 (kindly provided by Mathias Mack, Munich, Germany) were detected by anti-mouse IgG phycoerythrin-coupled coupled secondary antibody (Sigma).

### *Binding assays*

CHO-K1 cells expressing wild-type or mutant CCR5 were collected from plates with $Ca^{2+}$ and $Mg^{2+}$-free phosphate-buffered saline supplemented with 5 mM EDTA, gently pelleted for 2 min at 1000 × g and resuspended in binding buffer (50 mM Hepes pH 7.4, 1 mM $CaCl_2$, 5 mM $MgCl_2$, 0.5% bovine serum albumin). Competition binding assays were performed in Minisorb tubes (Nunc), with 40,000 cells in a final volume of 0.1 ml. The mixture contained 0.05 nM [$^{125}$I]-RANTES (2000 Ci/mmol, Amersham Biosciences) or 0.1 nM [$^{125}$I]-MIP-1$\beta$ as tracer, and variable concentrations of competitors (R&D Systems). Total binding was measured in the absence of competitor, and non-specific binding was measured with a 100 fold excess of unlabelled ligand. Samples were incubated for 90 minutes at 27° C, then bound tracer was separated by filtration through GF/B filters pre-soaked in 0.5% polyethylenimine (Sigma) for [$^{125}$I]-RANTES or in 0.1% bovine serum albumine (Sigma) for [$^{125}$I]-MIP-1$\beta$. Filters were counted in a $\beta$-scintillation counter. Binding parameters were determined with the Prism software (GraphPad Software) using non-linear regression applied to a one site competition model.

### *Aequoporin assay*

Functional response to chemokines was analysed by measuring the luminescence of aequorin as described (37,38). Cells were collected from plates with $Ca^{2+}$ and $Mg^{2+}$-free DMEM supplemented with 5 mM EDTA. They were then pelleted for 2 min at 1000 × g, resuspended in DMEM at a density of 5 x $10^6$ cells/ml and incubated for 2 h in the dark in the presence of 5 $\mu$M coelenterazine H (Molecular Probes). Cells were diluted 5 fold before use. Agonists in a volume of 50 µl DMEM were added to 50 $\mu$l of cell suspension (50,000 cells) and luminescence was measured for 30 s in a Berthold Luminometer.

### *GTP$\gamma$S binding assay*

The measurement of chemokine-stimulated GTPγS binding to membranes of cells expressing wt-CCR5 or the Y108mutant were performed as described (39,40). Briefly, membranes (10 or 20 $\mu$g) from wt-CCR5 or Y108A cells were incubated for 15 min at room temperature in GTP$\gamma$S binding buffer (20 mM Hepes, pH 7.4, 100 mM NaCl, 3 mM $MgCl_2$ ,3$\mu$M GDP, 10 $\mu$g/ml saponin) containing different concentrations of chemokines, in 96-well microplates (Basic FlashPlates, PerkinElmer Life Sciences). [$^{35}$S] GTP$\gamma$S (0.1 nM, Amersham Biosciences) was added, microplates were shaken for 1 min and further incubated at 30 °C for 30 min. The incubation was stopped by centrifugation of the microplate for 10 min, at 800 $\mu g$ and 4 °C, and aspiration of the supernatant. Microplates were counted in a TopCount (Packard, Downers, IL) for 1 min/well. Functional parameters were determined with the PRISM software (GraphPad Software) using non-linear regression applied to a sigmoidal dose-response model.

### *MAP Kinase Assay*

Cells, serum-starved for 24 h, were collected and resuspended in serum-free DMEM. After 3 min of stimulation with various concentrations of RANTES and MCP-2, cells were collected by centrifugation and heated to 100 °C for 5 min in lysis buffer (100 mM Tris-HCl, pH 6.8, 4 mM EDTA, 4% SDS, 20% glycerol, and 0.02% $\beta$-mercaptoethanol). For Western blot analysis, solubilized proteins corresponding to ~5 × $10^5$ cells were loaded onto 10% SDS-polyacryl-amide gels in a Tricine buffer system (41). After transfer to nitrocellulose membranes, proteins were probed with mouse anti-phospho-p42/p44 (1:1000) (Cell Signal) or rabbit anti-total p38 (1:2000) antibodies (Santa Cruz).

## 3.4.5. Results

To investigate the possible role of aromatic residues at the TMH2-TMH3 interface in defining the structure and function of the CCR5 receptor, we have compared the aminoacid sequence of CCR2 and CCR5 (Fig. 3.4.1B). These receptors are strongly related, sharing ~85% sequence identity within

their TM helices. However, their extracellular domains are much more divergent, which certainly contributes to their strong selectivity toward their respective ligands (35). The TMH2-TMH3 aromatic cluster was found to be quite divergent between these two receptors. This might suggest that these positions are not important for the structure and/or function of the receptors and therefore highly tolerant to variability. Alternatively, these positions could be functionally important although specific to each receptor, and the various substitutions would be expected in this case to be correlated. To study the functional consequences of the differences at aromatic positions observed between CCR5 and CCR2 (Fig. 3.4.1B), we engineered CCR5 mutants in which aromatic residues were substituted by the corresponding amino acids in CCR2. The F85$^{2.59}$L and Y89$^{2.63}$S mutants affected TMH2, L104$^{3.28}$F, F109$^{3.33}$H and F112$^{3.36}$Y involved TMH3. We also combined these point mutations either within TMH2 (F85$^{2.59}$L-Y89$^{2.63}$S double mutant), within TMH3 (F109$^{3.33}$H-F112$^{3.36}$Y, L104$^{3.28}$F-F109$^{3.33}$H-F112$^{3.36}$Y) or across both helices (F85$^{2.59}$L-L104$^{3.28}$F, Y89$^{2.63}$S-L104$^{3.28}$F, F85$^{2.59}$L-Y89$^{2.63}$S-L104$^{3.28}$F).

### *Cell surface expression of the mutant receptors*

We determined cell surface expression of the CCR5 mutants by fluorescence-activated cell sorting analysis, using five well characterized monoclonal antibodies (40, 42). The epitopes recognized by these monoclonal antibodies have been mapped to the N-terminal domain of the receptor (MC-5 and 3A9), the second extracellular loop (2D7), or a combination of extracellular domains (523 and MC6).

Fig. 3.4.2 illustrates the average surface expression of the different mutants following normalization to wild-type CCR5 expression level. The TMH2 mutant F85$^{2.59}$L exhibited a moderate but significant reduced expression, reaching ~50% of the wt signal. Although the Y89$^{2.63}$S mutant was well expressed, the combination of both substitutions (F85$^{2.59}$L-Y89$^{2.63}$S) led to a similar reduced expression as for F85$^{2.59}$L. Interestingly, while the single mutation L104$^{3.28}$F (in TMH3) did not affect cell surface expression, its combination with F85$^{2.59}$L restored to normal the low expression observed for F85$^{2.59}$L alone (mutant F85$^{2.59}$L-L104$^{3.28}$F). Strikingly, the triple mutant F85$^{2.59}$L-Y89$^{2.63}$S-L104$^{3.28}$F was not found at the cell surface, indicating that, although single and double mutants are expressed, the combination of all three substitutions is not tolerated. This mutant will therefore not be considered further in the following experimental settings. Mutants Y89$^{2.63}$S-L104$^{3.28}$F, F109$^{3.33}$H, F112$^{3.36}$Y, F109$^{3.33}$H-F112$^{3.36}$Y and L104$^{3.28}$F-F109$^{3.33}$H-F112$^{3.36}$Y are all expressed above 50% of the wt signal with a rather uniform pattern of recognition by the different mABs suggesting that these mutations do not alter the conformation of CCR5 extracellular domains.

### *Binding and functional properties of the mutant receptors*

The functional consequences of the mutations and their combinations were characterized in terms of binding and intracellular responses, using the four natural agonists of CCR5: RANTES, MIP-1$\alpha$, MIP-1$\beta$ and MCP-2. The

binding affinities for the various ligands were determined by competition binding assays using $^{125}$I-RANTES or $^{125}$I-MIP-1$\beta$ as labeled tracer, while functional responses were monitored by using the aequorin-based assay as described previously (13). Representative binding curves are shown in Fig. 3.4.3, functional concentration-action curves are shown in Fig. 3.4.4, and the data are summarized in Table 3.4.1.

We first focused on residues F85$^{2.59}$, Y89$^{2.63}$ and L104$^{3.28}$, located at the extracellular ends of TMH2 and TMH3. The F85$^{2.59}$L mutant was clearly impaired both at the level of binding, and in its functional response to chemokines. Although binding of RANTES was almost unaffected, the apparent affinities for MIP-1$\beta$ and MCP-2 were significantly lower than those observed for wt-CCR5 (see Table 3.4.1 and Fig. 3.4.3), and MIP-1$\alpha$ binding was undetectable. The functional responses of F85$^{2.59}$L were mild (Fig. 3.4.4) and grossly consistent with the binding data. The potency of RANTES was moderately affected on this mutant, but its efficacy was reduced by 4 fold; MIP-1$\beta$ and MCP-2 displayed a strong impairment of both their potencies and efficacies, while MIP-1$\alpha$ was almost inactive. The phenotype of the L104$^{3.28}$F mutant was grossly similar in terms of functional responses (with slightly better efficacies) despite binding parameters much closer to the wild-type levels. Interestingly, the addition of the L104$^{3.28}$F substitution to the F85$^{2.59}$L mutant (F85$^{2.59}$L-L104$^{3.28}$F), allowed to restore partly the function of this mutant. Indeed, on the double mutant, MIP-1$\alpha$ was characterized by a binding affinity close to wt, and the functional response was significantly improved. The functional properties of MCP-2 appeared somewhat more affected by the double mutation, while RANTES and MIP-1$\beta$ showed similar behaviours on the double mutant and on both single mutants.

Mutating Y89$^{2.63}$S affected mildly the activity of RANTES and MIP-1$\alpha$ (conserved potency, 2-fold reduction of Emax) decreased moderately the $E_{max}$ and potency of MIP-1$\beta$, but affected strongly the functional parameters of MCP-2, despite the normal affinity of this mutant in binding assays. Combining F85$^{2.59}$L and Y89$^{2.63}$S substitutions led to a severe impairment of functional responses, particularly for MIP-1$\beta$ and MCP-2 (no response), and to a lower extent for MIP-1$\alpha$, whereas RANTES was moderately affected. These effects appeared more than simply additive as compared to the single mutants. The Y89$^{2.63}$S-L104$^{3.28}$F mutant was well expressed at the cell-surface, but no specific binding could be detected by competition binding assays, and this receptor was barely functional. Only RANTES could elicit a small signal at high concentrations, with a strongly reduced $E_{max}$ but a decent potency (Table 3.4.1).

### *Modelling the TMH2-TMH3 interface*

The experimental data presented above demonstrates the importance of the TMH2-TMH3 aromatic cluster in the activation of CCR5. The fact that adding the L104$^{3.28}$F substitution to the F85$^{2.59}$L background partly restores the impaired function of this mutant (cell-surface expression and binding) might suggest a direct interaction between the two residues involved. Moreover, Y89$^{2.63}$ might contribute to this interaction, as mutating this residue, although well tolerated by itself, is highly disruptive in the context

of F85$^{2.59}$L, Y89$^{2.63}$S or F85$^{2.59}$L-Y89$^{2.63}$S. Fig. 3.4.5A shows the location of the side chains of these residues in a CCR5 model using strictly the rhodopsin crystal as a template (27). The orientation of these side chains towards the lipidic environment is in apparent contradiction with the hypothesis that these residues are important for the structure and the activation mechanism of CCR5.

We have however suggested previously that TMH2 of CCR5 would adopt, in its outer half, a conformation which is different from that of bovine rhodopsin (13). In CCR5 and other chemokine receptors, the extracellular side of TMH2 is predicted to be in close contact with TMH3 (and not with TMH1 as in rhodopsin) due to the structural action of a conserved Thr$^{2.56}$xPro$^{2.58}$ motif (13). We now propose that this region of TMH2 would be part of a structural and functional motif involving the neighbouring part of TMH3 (aromatic cluster and surrounding residues).

To provide a structural framework allowing to understand better the experimental data presented above, we adopted a modeling procedure (detailed in the "Experimental Methods" section) based on the following scheme.

### *Independent exploration of the conformation of TMH2 and TMH3 by molecular dynamics simulations*

Fig. 3.4.5 (B and C) shows the computed structures of TMH2 (green) and TMH3 (yellow). The bending of TMH2 towards TMH3, in its outer half, is tolerated in the context of the CCR5 helical bundle as the result of the relocation of TMH3 towards TMH5. It is important to note that these energetically available structures of TMH2 and TMH3 were obtained separately. Thus, the conformational spaces explored by these helices are the consequence of the amino acid sequence of TMH2 and TMH3 and not of steric hindrance between helices.

Fig. 3.4.5 (D and E) shows the result of superimposing the representative structure of the MD simulation and the rhodopsin template. The specific residues in TMH2 and TMH3 of CCR5 generate structural differences in the extracellular part of the receptor, without modifying its more compact cytoplasmic surface.

### *Membrane protein database search*

Because stable structural motifs are likely to recur in proteins of known structure, we surveyed the database of the structure of membrane proteins (see "Experimental Procedures") for $\alpha$-helices segments containing the aromatic cluster of TMH2: the FWxxY motif. This motif is also found in subunit III of the bovine cytochrome C oxidase (PDB identification code 1occ), where transmembrane helix III of the enzyme interacts with the neighboring helix VII. Inspection of this motif in the cytochrome C oxidase structure reveals that the Phe98 side chain in helix III interacts with Leu252 in helix VII, and that the Tyr102 side chain in helix III hydrogen bonds Ser255 in helix VII. We propose a similar pattern to describe the interactions between TMH2 and TMH3 of CCR5: Phe85$^{2.59}$ would interact

with Leu104$^{3.28}$, and Tyr89$^{2.63}$ with Thr99$^{3.23}$. It is important to note that Phe85$^{2.59}$, Tyr89$^{2.63}$, Thr99$^{3.23}$, and Leu104$^{3.28}$ are found specifically in CCR5, but not in CCR2, and contribute therefore to a CCR5-specific motif important for the receptor structure.

Fig. 3.4.6A shows a detailed view of the TMH2-TMH3 interface in the model resulting from MD simulation. In this model, Leu104$^{3.28}$ is located in an aromatic pocket formed by the side chains of Phe85$^{2.59}$ and Trp86$^{2.60}$, the electron–poor C–H hydrogens of Leu104$^{3.28}$ interact with the π electron–rich clouds of the aromatic rings. This type of C-H$\cdot\cdot\pi$ interaction plays a significant role in stabilizing local three-dimensional structures of proteins (43). Moreover, Phe85$^{2.59}$ aromatic ring is located between Leu103$^{3.27}$ and Leu104$^{3.28}$ side chains. Thus, there is a significant interaction between the aromatic residues (Phe85$^{2.59}$ and Trp86$^{2.60}$) in TMH2 and the hydrophobic residues (Leu103$^{3.27}$ and Leu104$^{3.28}$) in TMH3. In addition, the TMH2-TMH3 interface is stabilized by a hydrogen bond between Tyr89$^{2.63}$ and Thr99$^{3.23}$. To evaluate the magnitude of the TMH2-TMH3 interaction that might be attributed to the Phe85$^{2.59}\cdots$Leu104$^{3.28}$ and Trp86$^{2.60}\cdots$Leu104$^{3.28}$ interactions we performed *ab-initio* quantum mechanical calculations on minimal recognition models consisting of the functional groups of the intervening side chains (see “Experimental Procedures”). The energies of interaction of Phe85$^{2.59}$ and Trp86$^{2.60}$ with the multiple C–H hydrogens of Leu104$^{3.28}$ are –2.4 and –2.6 kcal/mol, respectively.

### *Structural and functional role of aromatic residues in TMH3*

Residues F109$^{3.33}$ and F112$^{3.36}$ are located in TMH3 within the outer third of the membrane, and are predicted to face toward the center of the helix bundle as inferred from the molecular model of CCR5 (see above) or the rhodopsin template itself. The F109$^{3.33}$H mutation had little effect on the receptor function. Both the binding properties and the functional response of the mutant were closely similar to those of wt-CCR5 for all four ligands (Table 3.4.1, Fig. 3.4.3 and 3.4.4). In contrast, the conservative F112$^{3.36}$Y substitution influenced strongly the activation of the receptor by its agonists. The potencies of RANTES, MIP-1$\alpha$ and MIP-1$\beta$ were relatively preserved, while that of MCP-2 was decreased by ~1 log. The efficacies of all ligands were however severely affected, with $E_{max}$ values ranging from 10 to 25% of the ATP response (Table 3.4.1). Remarkably, combining these two substitutions in the F109$^{3.33}$H-F112$^{3.36}$Y mutant significantly restored the functional response of the receptor, with $E_{max}$ values ranging from 54 to 90% of the maximal cell response and improved of potencies for all four ligands. The L104$^{3.28}$F-F109$^{3.33}$H-F112$^{3.36}$Y change, which combine the three mutations of TMH3, produced a receptor phenotypically close to that of the single L104$^{3.28}$F. While binding affinities were wt-like, activation was affected differentially for the different agonists : RANTES was almost not affected, MCP-2 severely impaired, with MIP-1$\alpha$ and MIP-1$\beta$ showing intermediate behaviours. This suggest an addition of effects of the L104$^{3.28}$F single mutant (significantly affected) and the F109$^{3.33}$H-F112$^{3.36}$Y double mutant (mild effect) . This is not surprising considering that these motifs are located in distant part of the structure, as, in our model, L104$^{3.28}$

interacts with TMH2, while F109$^{3.33}$ and F112$^{3.36}$ face TMH5 and TMH6 respectively (see “Discussion”).

### *Mutant Y108$^{3.32}$A*

Considering the selective effect on receptor function observed after mutation of transmembrane residues in TMH2 and TMH3, it appeared also interesting to test the putative role of residue Y108$^{3.32}$. This locus is known as an important binding site in a wide variety of GPCRs (reviewed in Refs. 6, 12 and 24). Site-directed mutagenesis has been used to demonstrate its central role in different receptors for neurotransmitters and peptides. The Y108$^{3.32}$A mutant was well recognized at the cell surface by antibodies targeting the N-terminus of the receptor (mean fluorescence above 60% of signal for wt-CCR5) while fluorescence was significantly decreased for antibodies recognizing either the second loop, or a combination of extracellular domains, possibly underlying a conformational modification of the extracellular loops of the receptor. This mutant exhibited a rather wild-type behavior following stimulation by RANTES, MIP-1$\beta$ or MIP-1$\alpha$, with slightly reduced efficacies, but was completely unreactive to MCP-2. We could however measure high affinity binding of this ligand by competition binding assay (Fig. 3.4.3 and Table 3.4.1). As for several other mutants described above, this mutant became unresponsive to MCP-2 withour affecting the affinity for this ligand, whereas other CCR5 agonists (e.g. RANTES) were still able to stimulate it. It is known that GPCRs, including CCR5 (40), can adopt multiple active states that trigger different intracellular cascades and differentially stabilized by various agonsists. In this line, one could hypothesize that MCP-2 would not be able to induce $Ca^{2++}$ increase through this mutant receptor, while being able to trigger other intracellular cascades. As shown in Fig. 3.4.7A, the four natural agonists could trigger GTP$\gamma$S binding through the wt receptor. In this assay, RANTES was the most potent agonist, whereas MCP-2 appeared more potent than MIP-1$\alpha$ and MIP-1β. As observed with the aequorin assay, MCP-2 was unable to stimulate the Y108$^{3.32}$A mutant in the GTP$\gamma$S assay. It has been shown that chemokine-induced activation of CCR5 can lead to activation of p42/p44 (44, 45). As shown in Fig. 3.4.7B, stimulation of CCR5 by RANTES or MCP-2 led to the dose-dependant phosphorylation of p42/p44 MAP kinases. In this assay, RANTES stimulated the Y108$^{3.32}$A mutant, although with a reduced efficiency as compared with wt-CCR5, and MCP-2 was almost inactive, except for a weak response observed at 100 nM.

The specific alteration of the mutant response to MCP-2 is reminiscent of the behavior observed in our previous study, in which mutations of the TxP motif also affected preferentially the biological response to MCP-2, without significantly affecting the binding of the chemokine. In the wt model, the Tyr108$^{3.32}$ side chain is positioned in the face-to-edge orientation (T-shaped) with the indole ring of Trp86$^{2.60}$ (see Fig. 3.4.6A). This type of $\pi$-$\sigma$ aromatic-aromatic interaction has been described as stabilizing a protein structure (46). It is expected that the conformational changes induced in TMH2 by mutations affecting the TxP motif would relocate the side chain of Trp86$^{2.60}$, located four residues apart, and in consequence affect its interaction with Tyr108$^{3.32}$.

## 3.4.4. Discussion

The three-dimensional model of CCR5 based on the rhodopsin crystal structure (including the specific modeling of the TMH2-TMH3 interface as described above; Fig. 3.4.5) provides a coherent framework to interpret the mutagenesis data detailed in the "Results" section. We propose that the residues varying between CCR5 and CCR2 at the level of the TMH2-TMH3 aromatic cluster are forming a specific interaction motif. In the wt receptor, Phe85$^{2.59}$ would interact with Leu104$^{3.28}$, while Tyr89$^{2.63}$ would hydrogen bond with Thr99$^{3.23}$, and we suggest that the substitutions observed in the CCR2 sequence are indeed correlated. It is important to note that this modeled TMH2-TMH3 interface provides a hydrophobic environment for Thr82$^{2.56}$ of the TxP motif in TMH2, essentially through the side chain of Leu104$^{3.28}$ (Fig. 3.4.6A). Thus, Thr82$^{2.56}$ orients the $C_\gamma$ atom towards this hydrophobic environment and the $O_\gamma$ atom towards the polar peptide bond of the backbone. The additional hydrogen bond between the $O_\gamma$ atom of Thr82$^{2.56}$ and the carbonyl group increases the magnitude of the Pro-kink and this TxP motif is a structural determinant involved in chemokine-induced activation (13).

In order to analyze in this structural framework the consequences of the amino acid substitutions explored experimentally, several of the mutations were introduced in our model, and we specifically studied the structural and dynamical properties of the extracellular part of the TMH2-TMH3 interface using MD simulations (see "Experimental Procedures"). This modeling procedure allows to propose a description of the TM interface for the mutants and suggests the nature of the structural changes that might lead to alterations of the receptor function (Fig. 3.4.6).

### *Structural interpretation of mutations located at the extracellular end of TMH2 and TMH3*

The F85$^{2.59}$L mutation weakens the interaction between the side chain of this TMH2 residue and L103$^{3.27}$ and L104$^{3.28}$ in TMH3, as a Leu-Leu interaction is not of the same magnitude as a Leu-Phe interaction. As a consequence, in the simulation, Leu85$^{2.59}$ side chain reorients away from TMH3 (Fig. 3.4.6B).

Replacing L104$^{3.28}$ by Phe dramatically modifies the TMH2-TMH3 interface. The aromatic side chain of Phe104$^{3.28}$ would optimally interact with the other aromatic side chains in a face to edge configuration (46). Thus, in the molecular dynamic simulations, the side chain of Phe104$^{3.28}$ tends to achieve this interaction with the aromatic side chains of both Phe85$^{2.59}$ and Trp86$^{2.60}$ (Fig. 3.4.6C). However, the side chain of Phe104$^{3.28}$ in this conformation is bulkier than that of Leu, resulting in a significant reorientation of Phe85$^{2.59}$ side chain towards the periphery of the TM bundle, and in a weakening of the TMH2-TMH3 interface. It is important to note that the L104$^{3.28}$F mutation increases the polarity of the Thr82$^{2.56}$ side chain environment. Thus, the presence of the electron-rich clouds of Phe104$^{3.28}$ aromatic ring might facilitate hydrogen bonding to the $O_\gamma$ atom of Thr82$^{2.56}$, in a manner similar to that proposed for hydrogen bonding between benzene and water (47). This would contrast with the wild type

receptor, in which a hydrogen bond is formed with the polar peptide bond of the backbone. As already discussed, the additional hydrogen bond of the Thr residue to the peptide bond has a significant influence on the conformation of the helix (13, 26).

In the double F85$^{2.59}$L-L104$^{3.28}$F mutant, the aromatic side chain of Phe104$^{3.28}$ is now located between Leu85$^{2.59}$ and Leu103$^{3.27}$ (Fig. 3.4.5D). Thus, the electron–rich clouds of the aromatic ring interact with the electron–poor C–H hydrogens of both Leu85$^{2.59}$ and Leu104$^{3.28}$. Notably, this mode of interaction of Phe104$^{3.28}$ places Leu85$^{2.59}$ in the proximity of Thr82$^{2.56}$, mimicking the hydrophobic environment of Thr82$^{2.56}$ in the wild type receptor (see above). Moreover, the aromatic side chain of Tyr89$^{2.63}$ is interacting in a face to edge configuration with Phe104$^{3.28}$. Thus, it appears from this simulation that the F85$^{2.59}$L-L104$^{3.28}$F mutant would partly restores the packing of the TMH2-TMH3 interface.

This could explain the improvement in cell-surface expression and binding affinity for MIP-1$\alpha$ and MIP-1β, between the F85$^{2.59}$L and F85$^{2.59}$L-L104$^{3.28}$F mutants (see Table 3.4.1). As the conformation of the short ECL1 (four residues) is likely influenced by the packing of the TMH2-TMH3 interface, we would suggest that point mutations (like F85$^{2.59}$L) could modify the conformation of the EC domain and affect the binding of ligands.

The triple mutant F85$^{2.59}$L-L104$^{3.28}$F-Y89$^{2.63}$S is intriguing. Intuitively, one would expect that, having replaced all differing aromatic positions into the CCR2 corresponding residues, the resulting mutant would show a better functional phenotype than the single or double mutant (being closer to the functional CCR2 receptor). However, it appeared that this triple mutant is not expressed at the cell surface (see Fig. 3.4.2), suggesting severe misfolding of the protein. Our modeling suggests that the addition of the Y89$^{2.63}$S substitution to the double F85$^{2.59}$L-L104$^{3.28}$F mutant would strongly modify the TMH2-TMH3 interface, because the shorter and non-aromatic side chain of Ser89$^{2.63}$ can not fulfill the interaction with both Thr99$^{3.23}$ and Phe104$^{3.28}$ (Fig. 3.4.6D). We hypothesize that this major structural difference between this mutant interface and the WT structure could be a factor responsible for the weak expression, either through a perturbation of the folding process, or through destabilization of the folded protein, leading to rapid internalization and/or degradation. In the single Y89$^{2.63}$S mutant, this modification of the packing of TMH2 and TMH3 would not happen as the Phe85$^{2.59}$- Leu104$^{3.28}$ interaction maintains a proper distance between the two helices. Moreover, Thr99$^{3.23}$ would tend to interact with the side chain of Gln93$^{2.67}$ (Fig. 3.4.6C), changing moderately the interface. Notably, the Y89$^{2.63}$S mutant shows a normal level of expression and is moderately affected in the functional tests, suggesting that the Y89$^{2.63}$S-Thr99$^{3.23}$ interaction proposed here is necessary for full functional efficiency, but not for the structural stability of the receptor. The double mutants F85$^{2.59}$L-Y89$^{2.63}$S and Y89$^{2.63}$S-L104$^{3.28}$F show a level of expression reduced by ~50%, a functional response strongly affected for F85$^{2.59}$L-Y89$^{2.63}$S and almost completely abolished for Y89$^{2.63}$S-L104$^{3.28}$F, underlining the increase in structural perturbation due to the additional mutations.

Combining the experimental results with our modeling approach leads to the following picture: the function of the CCR5 receptor is strongly dependant on the interface between the extracellular ends of TMH2 and

TMH3, which is determined by a series of interhelical interactions. The structural and dynamical properties of this interface are maintained by, at least, two polar interactions, Tyr89$^{2.63}$- Thr99$^{3.23}$ and Phe85$^{2.59}$- Leu104$^{3.28}$. These interactions are in equilibrium with each other, providing balanced structural constraints. Perturbating this interface by mutating one of these residues would modify this equilibrium and affect the function of the receptor.

This model implies that, in the case of the CCR2 receptor, the TMH2-TMH3 interface is organized differently, and that the substitutions tested here are counterbalanced by other changes in the sequence. In particular, one can spot the Thr99$^{3.23}$Ala change between the two receptors (see Fig. 3.4.1), but other changes (like Ala92$^{2.66}$Asn or Gln93$^{2.67}$Glu) could also be important. Interestingly, a chimeric construct involving ICL1, TMH2, ECL1 and TMH3 of CCR2 in a CCR5 background appeared to be well expressed and functional (35), suggesting that interactions between TMH2 and TMH3 are sufficient to obtain structural stability and functionality (as the other helices were strictly from CCR5).

### *The extracellular ends of TMH2 and TMH3 are involved in the activation mechanism*

By mutating the Thr$^{2.56}$xP$^{2.58}$ motif, we previously have shown that the structural integrity of TMH2 is crucial for CCR5 function (13). According to the nature of the substitution (various residues in place of Thr82$^{2.56}$, or Ala instead of Pro84$^{2.58}$), we could modulate the extent of the structural perturbation, which was translated into a functional defect. However, the different agonists of CCR5 were affected differently by these mutations, suggesting that this part of the receptor is involved in the activation process in a ligand-specific manner.

We proposed that this TxP motif governs the structural and dynamical properties of the extracellular end of TMH2. By looking at the specific role of Phe85$^{2.59}$ and Tyr89$^{2.63}$, we now probe the elements involved in the ligand-induced activation in this particular part of the TM bundle. Interestingly, we find the same trend as that observed after mutating the TxP motif : the ligand sensitivity to the mutations is variable, RANTES being the least affected agonist, while MCP-2 is the most sensitive.

### *Structural modeling of mutants involving the middle of TMH3*

Phe109$^{3.33}$ and Phe112$^{3.36}$, variable between CCR5 and CCR2, are not facing TMH2 in the CCR5 model, but are oriented towards TMH5 and TMH6. We therefore undertook full bundle MD simulation in order to model (see “Experimental Procedures”) the possible effects of the F109$^{3.33}$H, F112$^{3.36}$Y and F109$^{3.33}$H-F112$^{3.36}$Y mutations, and the putative interaction between these residues and TMH5 and/or TMH6.

While the F109$^{3.33}$H change does not modify the pattern of helix-helix interactions as compared to the wt receptor (not shown), changing Phe112$^{3.36}$ for Tyr modifies the interactions between TMH3 and TMH6 (Fig. 3.4.6, panels E and F). In our simulation, the hydroxyl group added by the

mutation forms a hydrogen bond with Asn252$^{6.52}$ in TMH6. In the WT receptor, this Asn hydrogen bonds back to the backbone carbonyl of Trp248$^{6.48}$, and also interacts through its $H_{\delta 1}$ atom with the aromatic ring of Phe112$^{3.36}$ (Fig. 3.4.6E). Pointing inside the bundle, residue Asn252$^{6.52}$ is highly conserved among CC-chemokine receptors, and position *6.52* is known to be functionally important in several rhodopsin-like GPCRs (6). In particular, this position is involved in ligand binding, ligand selectivity, or receptor activation in a significant number of peptide (48-52) and neurotransmitter (16, 17, 53) receptors, suggesting a possible functional role of the corresponding residue in chemokine receptors as well. In addition, the hydrogen bond proposed here to form with the backbone could also be important. The conserved Pro, present in TMH6 of most rhodopsin-like receptors (Pro$^{6.50}$) is known to be crucial for their function, most probably through the structural action of the proline. It has been proposed that, in the context of an $\alpha$-helix, polar residues hydrogen bonding the backbone in the vincinity of a Pro could significantly modulate the structure of the proline-induced deformation, and therefore be functionally relevant (13, 54). Interestingly, in the F112$^{3.36}$Y mutant, the polar interaction between Y112$^{3.36}$ and Asn252$^{6.52}$ changes the conformation of the asparagine away from the backbone, hence preventing it from hydrogen bonding with the backbone carbonyl of Trp248$^{6.48}$.

Therefore, the polar interaction between Tyr112$^{3.36}$ and Asn252$^{6.52}$ could strongly modify the functional properties of Asn252$^{6.52}$ by interfering with its putative interaction with the ligand and/or its structural role. It is important to note that, in contrast to the mutations located in the TMH2-TMH3 interface, the F112$^{3.36}$Y mutation affects all four agonists in a similar manner. The activation profile is qualitatively similar to that of the WT receptor (the potency order is preserved) but the potency and efficacy of agonists are reduced ~3 fold.

As detailed in the “Results” section, the double F109$^{3.33}$H-F112$^{3.36}$Y mutant exhibits a behaviour similar to that of the WT receptor, hence showing that adding the F109$^{3.33}$H mutation restores the function of lost in the F112$^{3.36}$Y mutant. While such functional recovery may suggest a direct interaction between the two residues, this is very unlikely in the context of an $\alpha$-helix, in which a His and a Tyr cannot interact together when separated by 3 positions. In our model, the F109$^{3.33}$H substitution modifies *indirectly* the effect of the F112$^{3.36}$Y mutation (Fig. 3.4.6G). While Tyr112$^{3.36}$ now interacts with the backbone carbonyl of Gly202$^{5.46}$, His109$^{3.33}$ interacts with Asn252$^{6.52}$, leading to a reorientation of the side chain which mimics the WT situation, including a recovery of the hydrogen bond between Asn252$^{6.52}$ and the carbonyl of Trp248$^{6.48}$. The effect of the single F112$^{3.36}$Y mutation is therefore counterbalanced by the addition of the histidine, which allows the reorientation of Asn252$^{6.52}$, leading to an almost normal behaviour of this putative functional motif.

### 3.4.5. Conclusion

The wealth of structural, biophysical and biochemical data accumulated during the last years provides valuable insights into the activation mechanism of GPCRs. It is well established that the activation process requires helix motions; in particular, movements of TMH3, TMH6 and TMH7 have been evidenced (5,6). The importance of the TMH2-TMH3 interface in the activation of CCR5 support the concept that other parts of the TM bundle are also involved in this mechanism. As suggested previously (13), we hypothesize that the active role of the TMH2-TMH3 interface in the activation process is somehow specific to chemokine receptors, as suggested by the conservation of the TxP motif in TMH2, and of the aromatic cluster in the extracellular ends of TMH2 and TMH3. While our work suggests that TMH2 (and possibly TMH3) would actually move upon chemokine-induced activation, biophysical approaches would be necessary to demonstrate the motion of these helices during activation.

## References

1. Baggiolini, M. (1998) Nature 392, 565-568

2. Gerard, C. and Rollins, B. J. (2001) Nat.Immunol. 2, 108-115

3. Berger, E. A., Murphy, P. M., and Farber, J. M. (1999) Annu.Rev.Immunol. 17, 657-700

4. Blanpain, C., Migeotte, I., Lee, B., Vakili, J., Doranz, B. J., Govaerts, C., Vassart, G., Doms, R. W., and Parmentier, M. (1999) Blood 94, 1899-1905

5. Meng, E. C. and Bourne, H. R. (2001) Trends Pharmacol.Sci. 22, 587-593

6. Gether, U. (2000) Endocr.Rev. 21, 90-113

7. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L., and Khorana, H. G. (1996) Science 274, 768-770

8. Dunham, T. D. and Farrens, D. L. (1999) J.Biol.Chem. 274, 1683-1690

9. Gether, U., Lin, S., Ghanouni, P., Ballesteros, J. A., Weinstein, H., and Kobilka, B. K. (1997) EMBO J. 16, 6737-6747

10. Marjamaki, A., Frang, H., Pihlavisto, M., Hoffren, A. M., Salminen, T., Johnson, M. S., Kallio, J., Javitch, J. A., and Scheinin, M. (1999) J.Biol.Chem. 274, 21867-21872

11. Abdulaev, N. G. and Ridge, K. D. (1998) Proc.Natl.Acad.Sci.U.S.A 95, 12854-12859

12. Ji, T. H., Grossmann, M., and Ji, I. (1998) J.Biol.Chem. 273, 17299-17302

13. Govaerts, C., Blanpain, C., Deupi, X., Ballet, S., Ballesteros, J. A., Wodak, S. J., Vassart, G., Pardo, L., and Parmentier, M. (2001) J.Biol.Chem. 276, 13217-13225

14. Wess, J., Gdula, D., and Brann, M. R. (1991) EMBO J. 10, 3729-3734

15. Fong, T. M., Cascieri, M. A., Yu, H., Bansal, A., Swain, C., and Strader, C. D. (1993) Nature 362, 350-353

16. Kim, J., Wess, J., van Rhee, A. M., Schoneberg, T., and Jacobson, K. A. (1995) J.Biol.Chem. 270, 13987-13997

17. Cho, W., Taylor, L. P., Mansour, A., and Akil, H. (1995) J.Neurochem. 65, 2105-2115

18. Lee, S. Y., Zhu, S. Z., and el Fakahany, E. E. (1996) Recept.Signal.Transduct. 6, 43-52

19. Roth, B. L., Shoham, M., Choudhary, M. S., and Khan, N. (1997) Mol.Pharmacol. 52, 259-266

20. Javitch, J. A., Ballesteros, J. A., Weinstein, H., and Chen, J. (1998) Biochemistry 37, 998-1006

21. Colson, A. O., Perlman, J. H., Jinsi-Parimoo, A., Nussenzveig, D. R., Osman, R., and Gershengorn, M. C. (1998) Mol.Pharmacol. 54, 968-978

22. Rhee, M. H., Nevo, I., Bayewitch, M. L., Zagoory, O., and Vogel, Z. (2000) J.Neurochem. 75, 2485-2491

23. Simpson, M. M., Ballesteros, J. A., Chiappa, V., Chen, J., Suehiro, M., Hartman, D. S., Godel, T., Snyder, L. A., Sakmar, T. P., and Javitch, J. A. (1999) Mol.Pharmacol. 56, 1116-1126

24. Ballesteros, J. A. and Weinstein, H. (1995) Methods in Neurosci. 25, 366-428

25. López-Rodríguez, M. L., Vicente, B., Deupí, X., Barrondo, S., Olivella, M., Morcillo, M. J., Benhamu, B., Ballesteros, J. A., Sallés, J., and Pardo, L. (2002) Mol. Pharmacol., 62, 14-21

26. Ballesteros, J., Deupi, X., Olivella, M., Haaksma E., and Pardo, L. (2000) Biophys.J. 79, 2754-2760

27. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le, T., I, Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) Science 289, 739-745

28. Kelley, L. A., Gardner, S. P., and Sutcliffe, M. J. (1996) Protein Eng. 9,1063-1065

29. Bower, M. J., Cohen, F. E., and Dunbrack, R. L. (1997) J.Mol.Biol. 267, 1268-1282

30. Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., and Tanabe, K. (2000) J.Am.Chem.Soc 122, 3746-3753

31. Olivella, M., Deupi, X., Govaerts, C., and Pardo, L. (2002) Biophys.J. 82, 3207-3213

32. Darden, T., York, D., and Pedersen, L. (1993) J.Chem.Phys. 98, 10089-10092

33. Case, D. A, Pearlman, D. A, Caldwell, J. W, Cheatham, T. E, Ross, W. S, Simmerling, C. L, Darden, T. A, Merz, K. M, Stanton, R. V, Cheng, A.

L, Vincent, J., Crowley, M, Ferguson, D. M, Radmer, R. J, Seibel, G. L, Singh, U. C, Weiner, P. K, and Kollman, P. A. AMBER5, University of California, San Francisco. 1997.

34. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) J.Am.Chem.Soc. 117, 5179-5197

35. Samson, M., LaRosa, G., Libert, F., Paindavoine, P., Detheux, M., Vassart, G., and Parmentier, M. (1997) J.Biol.Chem. 272, 24934-24941

36. Rizzuto, R., Pinton, P., Carrington, W., Fay, F. S., Fogarty, K. E., Lifshitz, L. M., Tuft, R. A., and Pozzan, T. (1998) Science 280, 1763-1766

37. Stables, J., Green, A., Marshall, F., Fraser, N., Knight, E., Sautel, M., Milligan, G., Lee, M., and Rees, S. (1997) Anal.Biochem. 252, 115-126

38. Blanpain, C., Lee, B., Vakili, J., Doranz, B. J., Govaerts, C., Migeotte, I., Sharron, M., Dupriez, V., Vassart, G., Doms, R. W., and Parmentier, M. (1999) J.Biol.Chem. 274, 18902-18908

39. Blanpain, C., Wittamer, V., Vanderwinden, J. M., Boom, A., Renneboog, B., Lee, B., Le Poul, E., El Asmar, L., Govaerts, C., Vassart, G., Doms, R. W., and Parmentier, M. (2001) J. Biol. Chem. 276, 23795–23804

40. Blanpain, C., Vanderwinden, J. M., Cihak, J., Wittamer, V., Le Poul, E., Issafras, H., Stangassinger, M., Vassart, G., Marullo, S., Schlndorff, D., Parmentier, M., and Mack, M. (2002) Mol. Biol. Cell 13, 723–737

41. Schagger, H., and von Jagow, G. (1987) Anal. Biochem. 166, 368–379

42. Lee, B., Sharron, M., Blanpain, C., Doranz, B. J., Vakili, J., Setoh, P., Berg, E., Liu, G., Guy, H. R., Durell, S. R., Parmentier, M., Chang, C. N., Price, K., Tsang, M., and Doms, R. W. (1999) J.Biol.Chem. 274, 9617-9626

43. Steiner, T. and Koellner, G. (2001) J.Mol.Biol. 305, 535-557

44. Dairaghi, D. J., Franz-Bacon, K., Callas, E., Cupp, J., Schall, T. J., Tamraz, S. A., Boehme, S. A., Taylor, N., and Bacon, K. B. (1998) Blood 91, 2905–2913

45. Misse, D., Esteve, P. O., Renneboog, B., Vidal, M., Cerutti, M., St Pierre, Y., Yssel, H., Parmentier, M., and Veas, F. (2001) Blood 98, 541–547

46. Burley, S. K. and Petsko, G. A. (1985) Science 229, 23-28

47. Suzuki, S., Green, P. G., Bumgarner, R. E., Dasgupta, S., Goddard III, W. A., and Blake, G. A. (1992) Science 257, 942-945

48. Zoffmann, S., Gether, U., and Schwartz, T. W. (1993) FEBS Lett. 336, 506-510

49. Fong, T. M., Yu, H., Cascieri, M. A., Underwood, D., Swain, C. J., and Strader, C. D. (1994) J.Biol.Chem. 269, 2728-2732

50. Kask, K., Berthold, M., Kahl, U., Nordvall, G., and Bartfai, T. (1996) EMBO J. 15, 236-244

51. Turner, C. A., Cooper, S., and Pulakat, L. (1999) Biochem.Biophys.Res.Commun. 257, 704-707

52. Huang, X. P., Nagy, P. I., Williams, F. E., Peseckis, S. M., and Messer, W. S., Jr. (1999) Br.J.Pharmacol. 126, 735-745

53. Granas, C., Nordvall, G., and Larhammar, D. (1998) J.Recept.Signal.Transduct.Res. 18, 225-241

54. Ri, Y., Ballesteros, J. A., Abrams, C. K., Oh, S., Verselis, V. K., Weinstein, H., and Bargiello, T. A. (1999) Biophys.J. 76, 2887-2898

# Chapter 4
# Discussion

The influence of Ser and Thr residues on the geometry of TMHs has been studied employing theoretical tools (Chapter 3.1). These studies have been extended to Pro-containing TMHs in Chapter 3.2. Experimental support to these theoretically derived hypothesis (Chapter 3.3) is provided through a collaboration with Prof. G. Vassart and Prof. M. Parmentier from the Institut de Recherche Interdisciplinaire en Biologie Humaine et Nucléaire (IRIBHN), Université Libre de Bruxelles. Finally, using site-directed mutagenesis and molecular modeling in a combined approach, a three-dimensional model of the G protein-coupled receptor CCR5 was constructed (Chapter 3.4).

## *Serine and threonine residues in their g- sidechain conformation induce a bend in $\alpha$-helices*

In addition to Pro (1) other amino acids may be also implicated in modulating the conformation of TMHs. Sequence analysis of transmembrane regions reveals that Ser and Thr are the most frequently occurring polar residues in TMHs (2), and that these residues tend to be conserved in certain positions of the TM bundle of the Class A family of GPCRs (3;4). Moreover, analysis of structural databases show that there is an increase of the population of the *g-* conformation of both Ser and Thr in membrane proteins relative to globular proteins (5). Moreover, Ser and Thr residues have the highest probability to be found in different rotamer conformations (6). These results suggest a possible structural or functional role for Ser/Thr (5;7;8).

We propose that these particular features of Ser/Thr are related with their putative role in the regulation of structure and function of TMHs in membrane proteins. In this thesis, analysis of structural databases has been used to show how the presence of Ser and Thr residues adopting the *g-* conformation correlates with a significant bending of the $\alpha$-helix at this position. Therefore, local alterations of the rotamer configurations of these residues may result in significant conformational changes across TMHs, thus participating in the molecular mechanisms underlying transmembrane signalling.

There is indirect experimental support for this hypothesis: site-directed mutagenesis experiments of $Ser^{5.43}$ and $Ser^{5.46}$ in the $\beta_2$-adrenergic receptor lead to the conclusion that these two residues are involved in the control of the equilibrium between active and inactive forms of the receptor (8). Our theoretical results allow an explanation of this putative role at the molecular level. These residues will probably be in its *g-* or *g+* conformations, in absence of ligand, thus satisfying their hydrogen bond potential with the helical backbone. But in order to bind the ligand, they need to undergo a conformational transition to the *t* rotamer, allowing the $OH\gamma$ group to interact with the ligand. This simple change is able to lead to a modification in the overall structure of the $\alpha$-helix, through a modulation of the helical bend. This mechanism could also be present, for instance, in the amine receptors subfamily of GPCRs, where $Ser^{5.43}$ and $Ser^{5.46}$ in TMH5 also have been shown to bind the ligand (3). We propose that this mechanism could be related to the first steps of receptor activation.

Finally, it is worth to note that there are also experimental results in non-GPCR membrane proteins that support a structural role for Ser and Thr. For instance, mutation $Ser^{254}$ by a Thr in the M1 region of the guinea pig $5HT_{3A}$ receptor, an ion channel, leads to a dramatic change in desensitization process of the receptor (9).

## *Serine and threonine residues modulate the geometry of Pro-kinked $\alpha$-helices*

There is broad experimental evidence that Pro residues disrupt the geometry of $\alpha$-helices, due to the steric clash between the pyrrolidine ring of Pro and the backbone carbonyl in the previous turn of the helix. As a consequence the intra-helical hydrogen bond network that stabilizes the $\alpha$-helix is also disrupted (10-14). Because Ser and Thr residues also modify this hydrogen bond network, we hypothesize a synergic effect between nearby Ser/Thr and Pro residues on the structure of $\alpha$-helices.

The analysis of the molecular dynamics simulations (Chapter 3.2) has revealed how Ser and Thr residues modulate the conformation of Pro-containing α-helices. The changes in the local structure include: *i)* an increase of the helical bend angle in (S/T)AP and (S/T)P motifs; *ii)* a decrease of the helical bend angle in (S/T)AAP, PA(S/T), and PAA(S/T) motifs; and *iii)* a change in the direction of the helix in SP and SAP motifs in the *g-* rotamer when the OγH group of Ser hydrogen bonds the carbonyl oxygen at position i-3.

It has also been shown how some of these Ser/Thr and Pro combinations, like TxP and SxP, are over-represented on a database of transmembrane domains. In agreement, sequence analysis on a database of Class A GPCRs shows that the most common combination is the (S/T)xP motif, found in 16% of all Pro-kink in GPCRs (9% for TxP, 7% for SxP). Moreover Ser and Thr are not equally found with specific sites clearly favoring one residue over the other. For example, $P^{2.58}$ predominantly associates with a Thr

(forming a $TxP^{2.58}$ motifs) while $P^{4.59}$ display four times more $SxP^{4.59}$ motifs than $TxP^{4.59}$. This indicates that Thr and Ser are not equivalent in such motif and already suggests (as their H-bonding propensities are identical) that the methyl group differentiating their sidechains may have a defined structural or functional role in the context of the motif. Therefore, we hypothesize that alterations in the structure of Pro-kinked TMHs induced by nearby Ser/Thr residues may be related to functional purposes.

## *The TxP motif in TMH2 is a structural and functional determinant of the CCR5 receptor*

In order to reinforce the hypothesis of the role of nearby Ser/Thr and Pro residues on the structure of $\alpha$-helices, a combined theoretical and experimental study was carried out in collaboration with the IRIBHN (Chapter 3.2). Through this cooperation, our group has provided the theoretical framework to interpret the experimental results at a molecular level.

The system used for the study was the chemokine CCR5 receptor, which has been shown to contain the $TxP^{2.58}$ motif, highly conserved in the peptidergic receptor family and almost fully conserved in the chemokine receptors family. Our theoretical study shows how the combined effect of Thr and Pro leans the extracellular side of TMH 2 towards TMH 3, illustrating how local modifications can be propagated through the whole helical structure and results in a modification of the overall structure of the receptor. The experimental results suggest that the hydrogen bonding potential of the Thr residue is crucial for receptor structure and function. It is interesting to observe how the functional response of the different CCR5 mutants correlates with the structural modifications observed in the molecular dynamics simulations.

Finally, it has to be stressed that these results may be extrapolated to other systems featuring TxP motifs, as other chemokine receptors, other peptidergic receptors or, even, different transmembrane systems, as ionic channels.

## *Molecular modelling of the CCR5 receptor*

Our previous theoretical study leaded to the hypothesis of an interaction between TMH2 and TMH3 in the CCR5 receptor. Therefore, we applied a set of bioinformatic tools, as multiple sequence alignment, homology modelling, molecular dynamics simulations, and search in structural databases, to the building of a three-dimensional model of the CCR5 chemokine receptor. The proposed inter-helical interactions between TMH2 and TMH3, which are not feasible according to the rhodopsin three-dimensional template, received

experimental support through side-directed mutagenesis and functional assays in collaboration with the IRIBHN.

We propose a three-dimensional model of CCR5 where Phe$^{2.59}$ interacts with Leu$^{3.28}$, while Tyr$^{2.63}$ interacts with Thr$^{3.23}$. In this model, Thr$^{2.56}$ of the TxP motif is located in a local hydrophobic environment, forcing the side chain of Thr$^{2.58}$ to interact with the backbone carbonyl of the previous turn, which has been previously shown to induce a distortion in the $\alpha$-helix. Combination of the theoretical and experimental results indicate that the function of the CCR5 receptor is strongly dependent on the interface between the extracellular ends of TMH2 and TMH3, which is determined by a series of polar and hydrophobic interactions.

These results illustrate the importance of the transmembrane bundle structure and dynamics in the activation mechanism of GPCRs. While movement of other helices have been shown to be implicated, we suggest that movement of TMH2 and, possibly, TMH3, would be specific of chemokine-induced activation.

It is likely that CCR5 and rhodopsin differ in other regions of the TM bundle. However, the important structural difference in the extracellular half of TMH2 in both receptors, where the GG motif in rhodopsin is replaced by a TxP motif in the chemokine family, appears as the most striking difference. We propose that this divergence in the external region of the TM bundle is due to the evolutive adaptation of the CCR5 receptor to its ligands.

This study also illustrates an important issue in the field of homology modelling of GPCRs: the sequence specificities of a particular receptor have to be explicitly taken into account when building a three-dimensional theoretical models. For instance, bovine rhodopsin, the structural template used in homology modelling of GPCRs, lacks the TxP sequence motif present in the TMH2 of CCR5, and its crystal structure shows how its TMH2 is bent towards TMH1. The strict use of the rhodopsin template would not allow the interpretation of the experimental results obtained for the CCR5 receptor.

## *Implication of Ser/Thr and Pro motifs in the mechanism of activation of GPCRs*

As it has been pointed out in Chapter 1.3, GPCRs display a high diversity in the structure of their ligands. Therefore, ligand-receptor interactions will be strongly family-dependent (15;16), which suggests that the first steps of GPCRs activation, those involving ligand-receptor recognition and the initial conformational rearrangements of the extracellular domains and the region of the TM bundle next to the extracellular part, will be divergent in the different classes of GPCRs and specific for each subfamily. On the other hand, GPCRs interact with only a few G protein subtypes, so receptor-G protein interactions are proposed to be more conserved through the whole family (17). Moreover, higher sequence conservation in the cytoplasmatic

part of the transmembrane region of Class A GPCRs is detected (18). These facts seem to indicate that the process of GPCR activation is common in its last steps, those involving interaction with the G protein and rearrangement of the cytoplasmatic domains and the region of the TM bundle next to the cytoplasm.

Both conserved or family-specific reorganizations of the transmembrane bundle involve movements of the TMHs. This thesis is framed in the study at a molecular level of these structural rearrangements. Our results clearly illustrate how the combined effect of Ser/Thr and Pro influence the geometry of TMHs, which, in turn, can be involved in the modulation of the function of membrane proteins.

## References

1. Cordes FS, Bright JN, Sansom MS 2002 Proline-induced distortions of transmembrane helices. J Mol Biol 323:951-60

2. Senes A, Gerstein M, Engelman DM 2000 Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. J Mol Biol 296:921-36

3. Strader CD, Candelore MR, Hill WS, Sigal IS, Dixon RA 1989 Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor. Journal of Biological Chemistry 264:13572-13578

4. Chanda PK, Minchin MC, Davis AR, Greenberg L, Reilly Y, McGregor WH, Bhat R, Lubeck MD, Mizutani S, Hung PP 1993 Identification of residues important for ligand binding to the human 5-hydroxytryptamine1A serotonin receptor. Molecular Pharmacology 43:516-520

5. Ballesteros JA, Deupi X, Olivella M, Haaksma EE, Pardo L 2000 Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. Biophys J 79:2754-60

6. Chakrabarti P, Pal D 1998 Main-chain conformational features at different conformations of the side-chains in proteins. Protein Eng 11:631-47

7. Liapakis G, Ballesteros JA, Papachristou S, Chan WC, Chen X, Javitch JA 2000 The forgotten serine. A critical role for Ser-2035.42 in ligans binding to and activation of the beta 2-adrenergic receptor. J Biol Chem 275:37779-88

8. Ambrosio C, Molinari P, Cotecchia S, Costa T 2000 Catechol-binding serines of beta(2)-adrenergic receptors control the equilibrium between active and inactive receptor states. Mol Pharmacol. 57:198-210

9. Lobitz N, Gisselmann G, Hatt H, Wetzel CH 2001 A single amino-acid in the TM1 domain is an important determinant of the desensitization kinetics of recombinant human and guinea pig alpha- homomeric 5-hydroxytryptamine type 3 receptors. Mol Pharmacol 59:844-51

10. Wess J 1993 Molecular basis of muscarinic acetylcholine receptor function. Trends in Pharmacological Sciences 14:308-313

11. Hong S, Ryu KS, Oh MS, Ji I, Ji TH 1997 Roles of transmembrane prolines and proline-induced kinks of the lutropin/choriogonadotropin receptor. J 272:4166-71

12. Ri Y, Ballesteros JA, Abrams CK, Oh S, Verselis VK, Weinstein H, Bargiello TA 1999 The role of a conserved proline residue in mediating conformational changes associated with voltage gating of Cx32 gap junctions. Biophys J 76:2887-98

13. Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA, Javitch JA 2002 Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. J Biol Chem 277:40989-96

14. Stitham J, Martin KA, Hwa J 2002 The critical role of transmembrane prolines in human prostacyclin receptor activation. Mol Pharmacol 61:1202-10

15. Ji TH, Grossmann M, Ji I 1998 G Protein-coupled Receptors. I. Diversity of Receptor-Ligand Interactions. J. Biol. Chem. 273:17299-17302

16. Gether U 2000 Uncovering molecular mechanisms involved in activation of G protein- coupled receptors. Endocr Rev 21:90-113

17. Horn F, van der Wenden EM, Oliveira L, IJzerman AP, Vriend G 2000 Receptors coupling to G proteins: is there a signal behind the sequence? Proteins 41:448-59

18. Mirzadegan T, Benko G, Filipek S, Palczewski K 2003 Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. Biochemistry 42:2759-67

# Conclusions

- This thesis aims to show how through the use of bioinformatics tools, primary sequences of proteins and interactions at an atomic level can be translated to three-dimensional protein structures.

- The three-dimensional structure of bovine rhodopsin is a very useful template for modelling of GPCRs, as it can be used as starting point in homology modelling techniques. However, we want to show how the sequence specificities of each family have to be explicitly taken into account in order to build models, which require experimental support to be reliable.

- The sequence specificities consist in sequence patterns conserved within certain families, which are translated into structural divergences. Among these sequence patterns, we hypothesize that Ser and Thr, alone or combined with nearby Pro residues, can modulate the geometry of TMHs, due to its capability to interfere with the hydrogen bond network that stabilize $\alpha$-helices. When Pro and Ser/Thr appear in combined motifs, there is a synergic effect that reinforces the structural consequences of these motifs in the structure of $\alpha$-helices. This hypothesis has received indirect experimental support, through the design and analysis of mutagenesis studies on the CCR5 receptor.

- The influence of Ser and Thr and Ser/Thr-Pro motifs in the structure of TMHs may be related to processes of activation in the Class A of GPCRs, and, possibly, other membrane proteins as well. In GPCRs, these motifs may have evolved in order to adapt a conserved mechanism of activation of the G protein to the cognate ligands of each receptor family.

**Figure 1.1.** Ramachandran plot showing the φ and ψ angles of all residues of the crystal structure of bovine rhodopsin (green triangles for glycines and black squares for the rest of aminoacid types). Red and yellow regions of the map correspond to the allowed regions (α-helices and β-sheets) of the polypeptide conformational space. Diagram created with the Sting Millennium Suite (RCSB/PDB-San Diego).

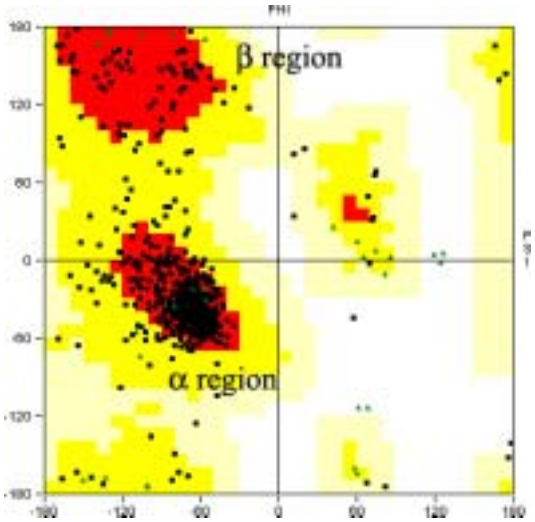



**Figure 1.2.** Three-dimensional structure of an α-helix. Only the peptidic backbone is depicted, where grey spheres correspond to carbon atoms, blue to nitrogen, red to oxygen, and small black spheres to hydrogen. The intramolecular network of hydrogen bonds that stabilizes this secondary structure is shown as yellow dotted lines. Each NH group at *i* position establishes a hydrogen bond interaction with the CO group at the *i-4* position in the previous turn of the helix. Figure created using MolScript v2.1.1 (103) and Raster3D v2.5 (104).

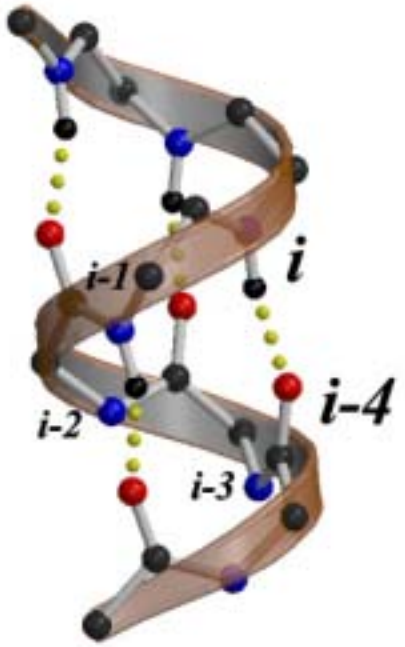



**Figure 1.3.** Three-dimensional structure of a Pro-kinked *α*-helix. The bulky pyrrolidine ring at position *i* faces the backbone carbonyl at position *i-4* in the previous turn of the helix. In order to avoid the steric clash between these two groups, the helix bends and opens up at this region, resulting in a strong local deformation, the *Pro-kink*. The intra-helical hydrogen bond network is also affected. For instance, the hydrogen bond between residues *i+1* and *i-3* (yellow dotted line) is disrupted. Figure created using MolScript v2.1.1 (103) and Raster3D v2.5 (104).

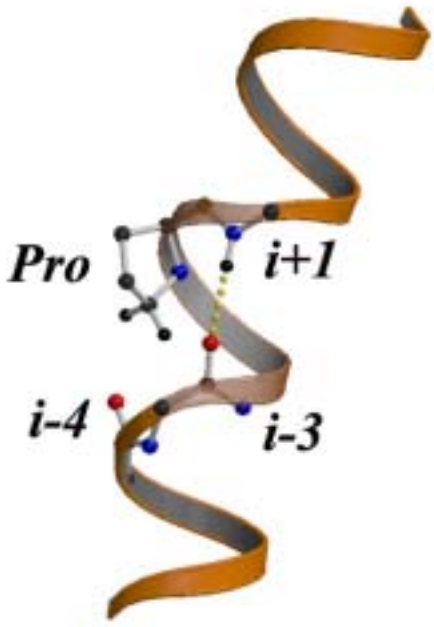


**Figure 1.6.** Preferred conformations of Ser side-chain. In the *gauche*– (*g*-) conformation an intrahelical hydrogen bond between the $H_\gamma$ of Ser (at position *i*) and the backbone carbonyls of the previous turn of the helix (at positions *i-3* or *i-4*) can be established. When Ser is in its *gauche*+ (*g*+) conformation, this hydrogen bond is only possible with the carbonyl at *i-4*. Finally, if Ser is in its *trans* (*t*) conformation, this interaction is not possible. Thr and Cys residues can, in principle, display these same sidechain conformations. However, some particular cases are forbidden (see text). Figure created using MolScript v2.1.1 (103) and Raster3D v2.5 (104).

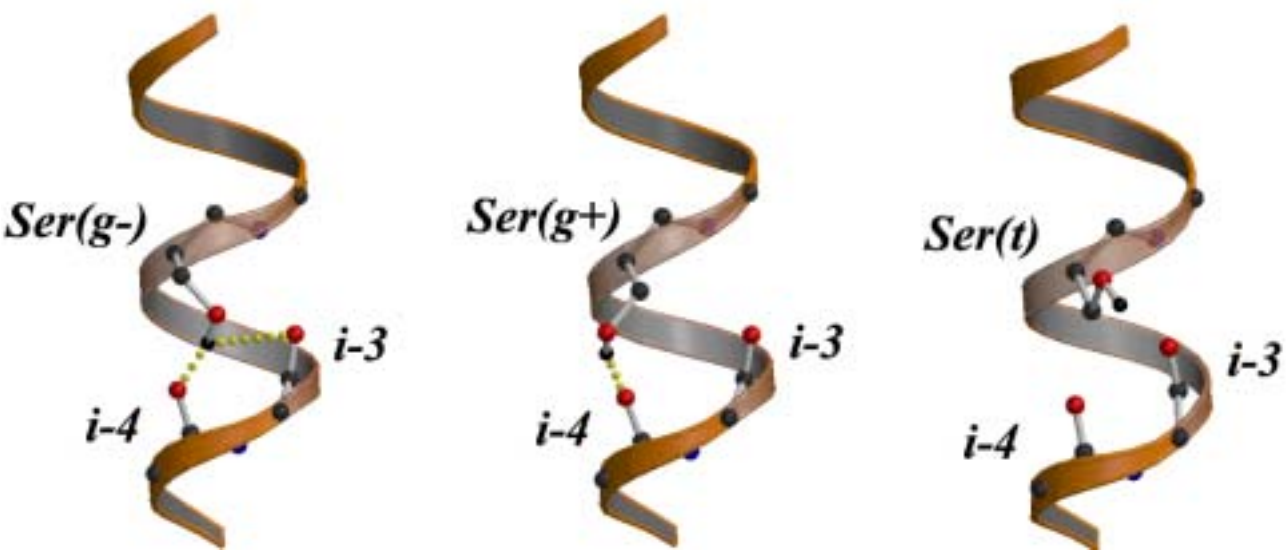


**Figure 1.9.** Three-dimensional structure of bovine rhodopsin, the structural prototype of Class A GPCRs. The top of the figure corresponds to the extracellular region. The seven transmembrane helices are arranged in a counterclockwise fashion as viewed from the extracellular side. The colour code is the following: TMH I red, TMH II yellow, TMH III orange, TMH IV green, TMH V blue, TMH VI dark blue and TMH VII purple. Helix VIII is shown in magenta and the non-helical regions (N-terminus, cytoplasmic and extracellular loops and C-terminus) are shown in grey. Figure created using MolScript v2.1.1 (103) and Raster3D v2.5 (104).

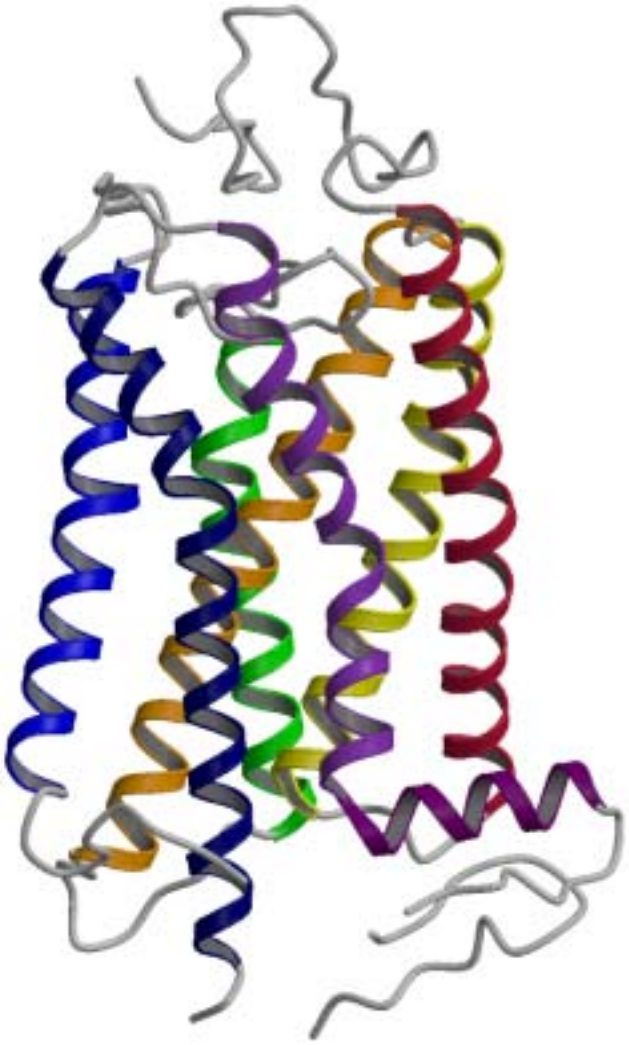


**Figure 1.10.** This figure illustrates the distinct receptor-ligand interactions for the different ligand types of GPCRs. Figure extracted from reference (34).

- *Panel A:* general structure of GPCRs.
- *Panel B:* ligand binding within the transmembrane bundle.
- *Panel C:* ligand binding in the transmembrane bundle, exoloops, and N-terminal segment.
- *Panel D:* ligands that bind to and cleave the N-terminal segment. The resulting shorter N-terminal segment interacts with exoloops to generate a signal.
- *Panel E*: ligand binding to the N-terminal segment, which interacts with exoloops to generate a signal.
- *Panel F*: ligand binding to the N-terminal segment, which interacts with the membrane-associated domain, thus generating a signal.

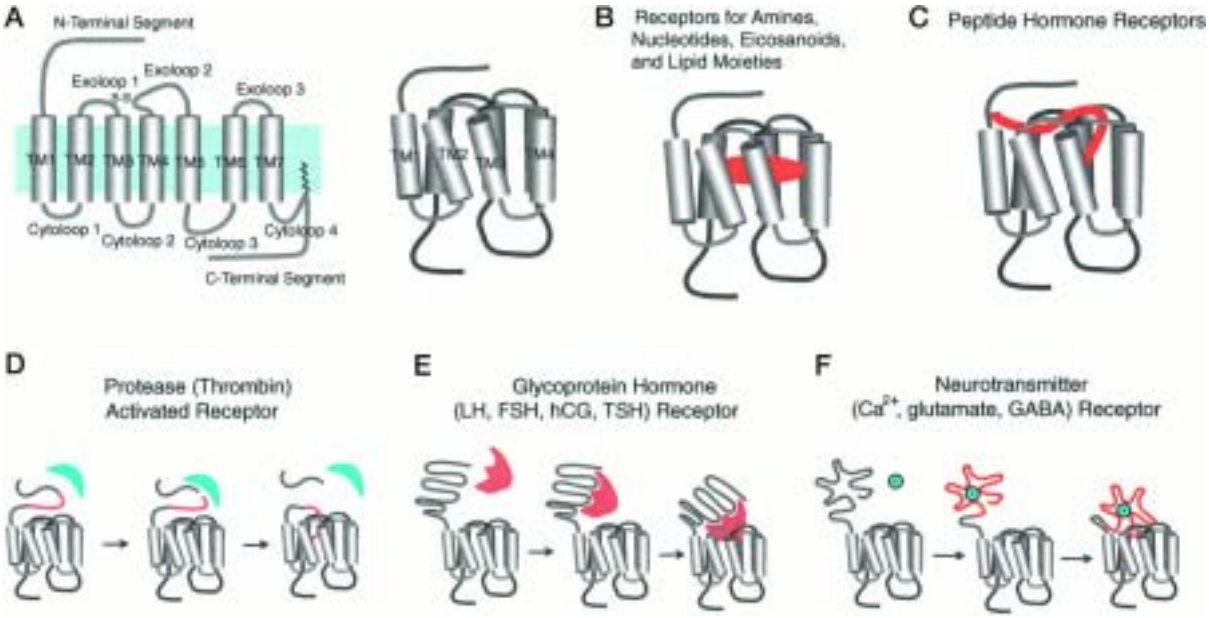


**Figure 1.11.** a. Theoretical model of the transducin-rhodopsin dimer complex. Helices of rhodopsin are colored from helix I in blue, helix II in light blue, helix III in green, helix IV in light green, helix V in yellow, and helix VI in dark yellow to helix VII and cytoplasmic helix 8 in red. Transducin is represented in yellow space-filling background for the α subunit, in red for the β subunit and in green for the γ subunit. b. Organization and topography of the cytoplasmic surface of rhodopsin. This image, obtained using atomic-force microscopy, shows the organization of rhodopsin in rows of dimers, as well as individual dimers (inside dashed ellipse), presumably broken away from one of the rows, and occasional rhodopsin monomers (arrowheads) in the native disc membrane. Figures extracted from reference (100).

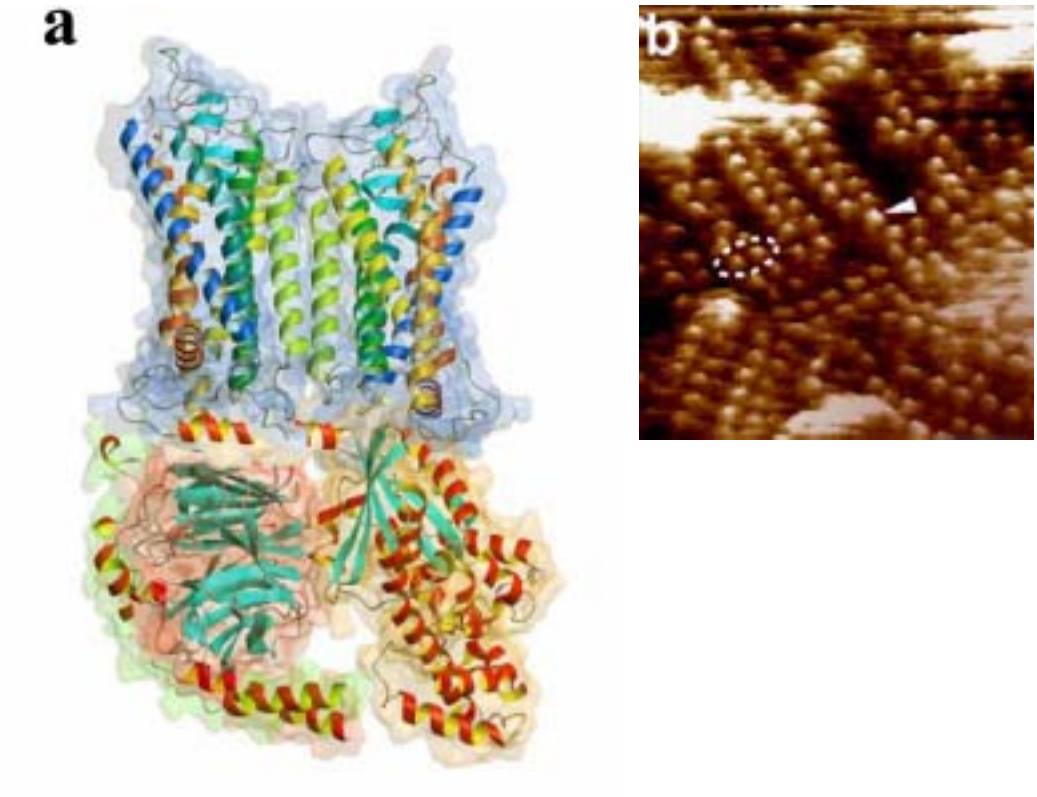


**Figure 3.1.2.** Comparison of helix bending between a polyAla α–helix (red) and (a) a polyAla α–helix with a single Ser or Thr residue (the $C_\alpha$–$C_\beta$ bond is shown) in the gauche– (g–) conformation (blue); (b) helix 32, which contains $Thr^{277}$ and $Thr^{279}$ (the $C_\alpha$–$C_\beta$ bonds are shown) in g–, from the photosyntethic reaction center; (c) helix 1, which contains $Thr^{33}$ also in g–, from the potassium channel; and (d) a polyAla α–helix with a Ser residue (the $C_\alpha$–$C_\beta$ bond is shown) in trans (t) conformation (green) or g– (blue) conformation. (c) A detailed view of the amino acids from i (the residue in g–) to i–4 are shown as ball and stick. Figures were created using MOLSCRIPT (24).

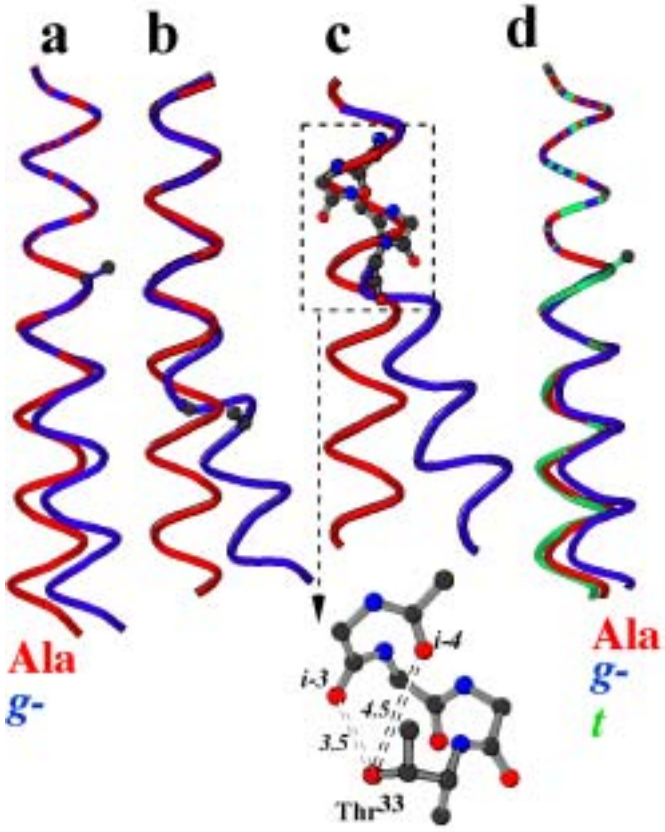


**Figure 3.2.1.** Detailed view of the hydrogen bond pattern, in the Pro-kink turn, of model peptides containing Pro and the (S/T)P, (S/T)AP, and (S/T)AAP patterns, with Ser and Thr side chains built in both *g*+ and *g*- conformations. Only polar hydrogens are depicted to offer a better view. The OγH group of Ser/Thr is interacting with the carbonyl oxygen at position j-4, relative to Ser/Thr, except when noted. Figures were created using MolScript v2.1.1 (29) and Raster3D v2.5 (30).

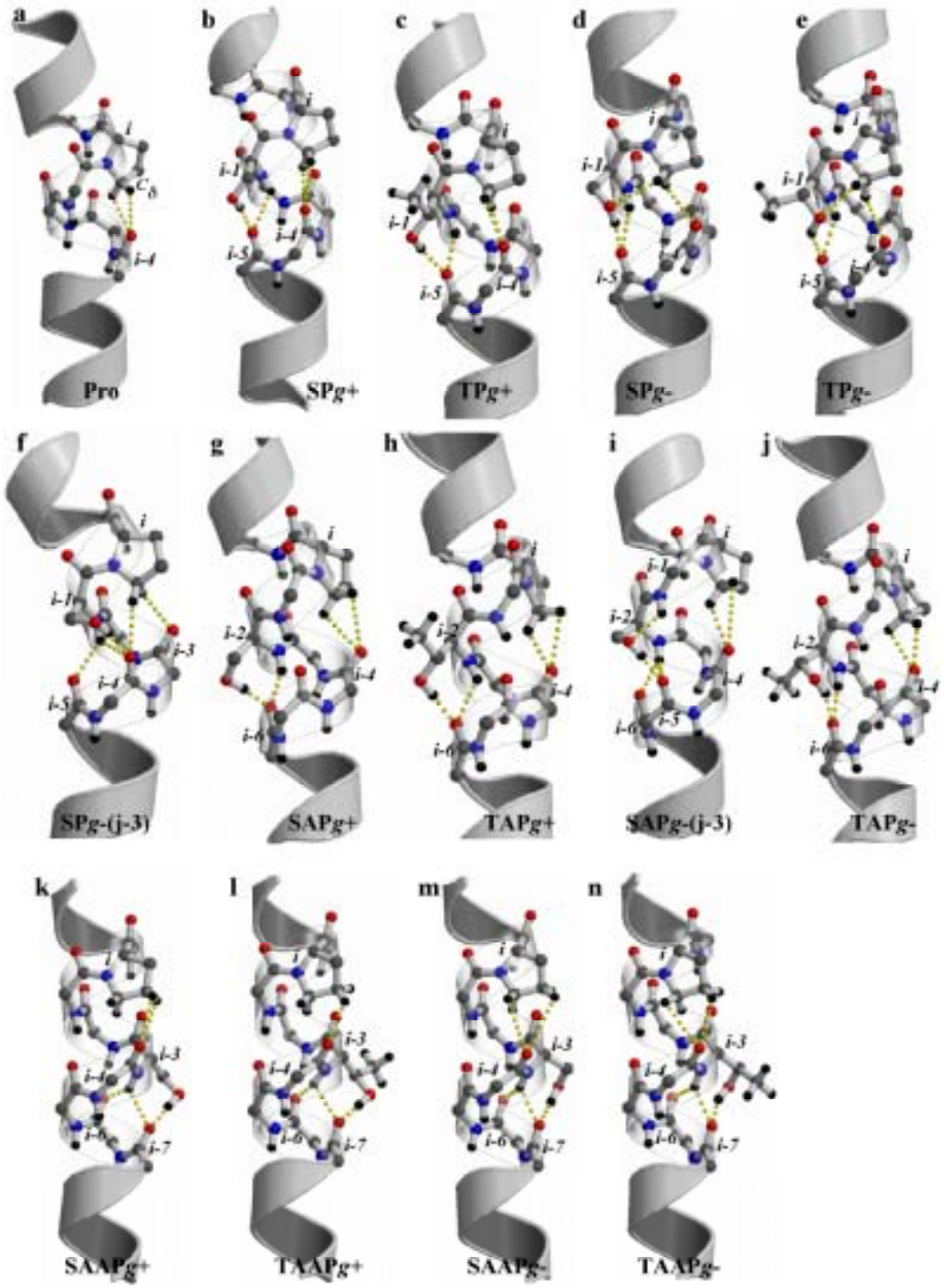


**Figure 3.2.2.** Detailed view of the hydrogen bond pattern, in the Pro-kink turn, of model peptides containing the P(S/T), PA(S/T), and PAA(S/T) patterns, with Ser and Thr side chains built in both *g*+ and *g*- conformations. Only polar hydrogens are depicted to offer a better view. The OγH group of Ser/Thr is interacting with the carbonyl oxygen at position j-4, relative to Ser/Thr, except when noted. Figures were created using MolScript v2.1.1 (29) and Raster3D v2.5 (30).

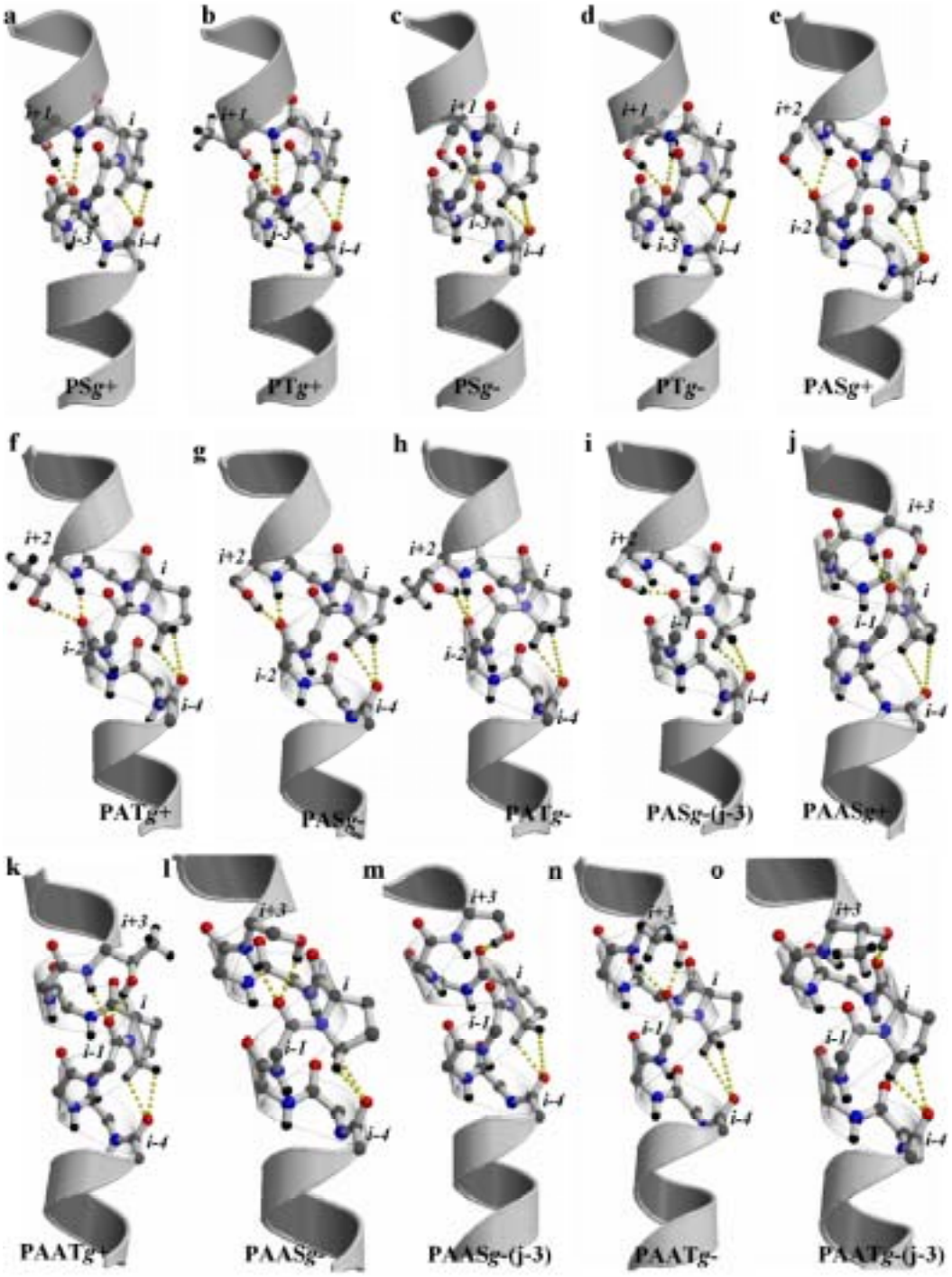


**Figure 3.2.3.** Comparison of the evolution of the unit twist angles (º) along the α-helix, from turns (i-11, i-8) to (i+4, i+7), between a Pro-containing polyAla α-helix (control, solid line in black) and (a) (S/T)P, (b) (S/T)AP, (c) (S/T)AAP, (d) P(S/T), (e) PA(S/T), and (f) PAA(S/T) motifs. The color and style codes are: Ser (red), Thr (blue), $g^+$ rotamer (solid line), $g^-$ rotamer interacting with the carbonyl oxygen at position j-4 (dot line), and $g^-$ rotamer interacting with the carbonyl oxygen at position j-3 (dash line).

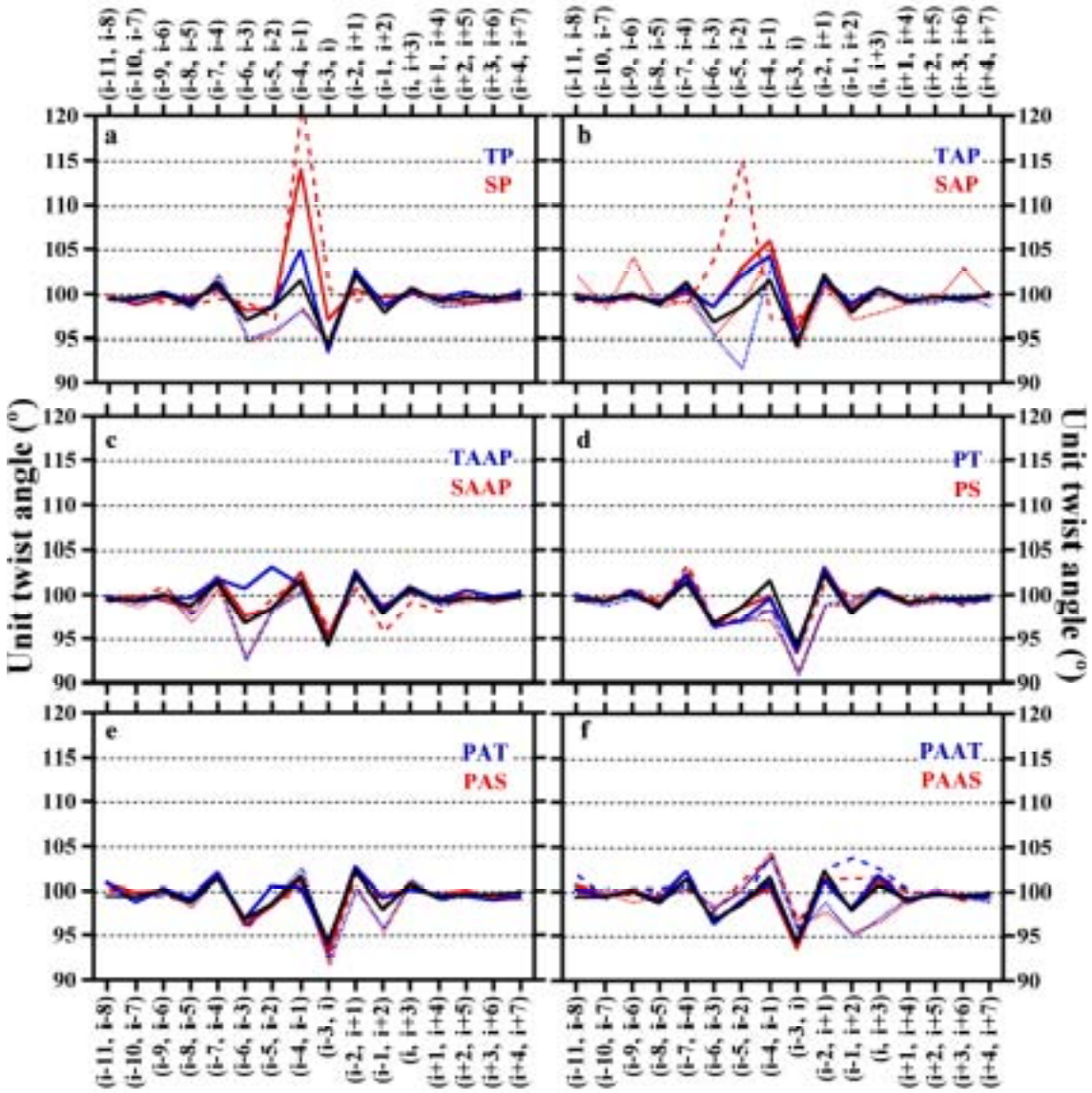


**Figure 3.2.4.** Comparison of helix bending between a Pro-containing polyAla α-helix (control, yellow) and (a) SP (red) and SAP (dark green) in the *g*+ rotamer, and TAP (light green) and PAAS (purple) in the *g*- rotamer transmembrane α-helices (the OγH group of S/T is interacting with the carbonyl oxygen at position j-4, relative to S/T, in all the cases); and (b, c) SP (orange) and SAP (light green) in the *g*- rotamer interacting with the carbonyl oxygen at position j-3 relative to Ser (different views rotated by 90° along the helical axis). The side chains of Pro and Ser/Thr are shown. Figures were created using MolScript v2.1.1 (29) and Raster3D v2.5 (30).

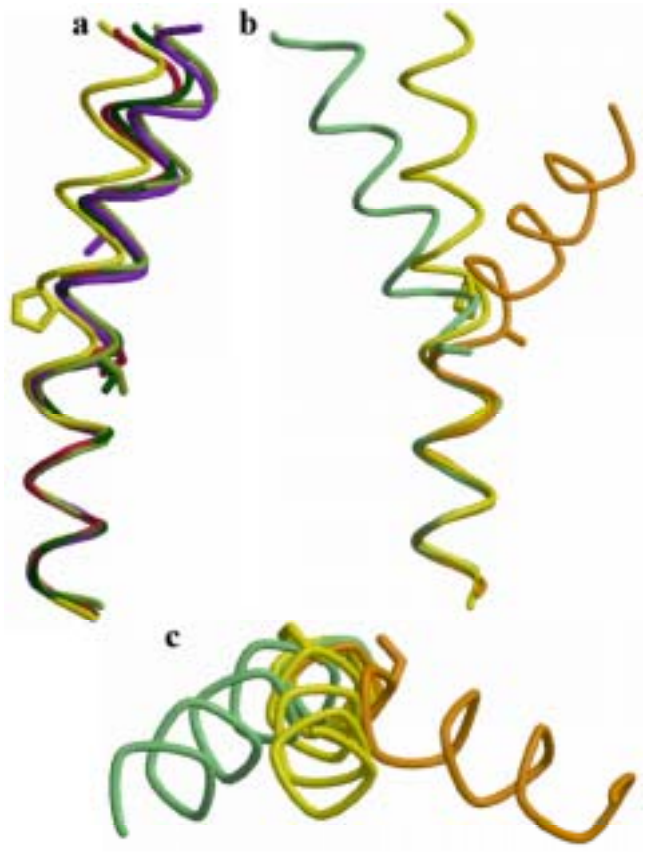


**Figure 3.2.6.** Molecular mechanisms of decreasing the bending of Pro-kinked transmembrane helices. (a) (S/T)AAP motifs in both *g*+ and *g*- rotamers moderate the steric clash between the pyrrolidine ring and the O=$C_{i-4}$ carbonyl oxygen by modifying the conformation of the $N_{i-3}$-H amide group, throughout the additional hydrogen bond formed between $O_{\gamma}H$ of $S/T_{i-3}$ and the O=$C_{i-7}$ carbonyl oxygen. (b) PA(S/T), in both *g*+ and *g*-, and PAA(S/T) in *g*+ rotamers obstruct the bending of the helix by adding a constrain in the form of a hydrogen bond (in green) in the other side of the helix where the steric $C_{\delta}H\bullet\bullet\bullet O=C_{i-4}$ clash (in red) occur. Figures were created using MolScript v2.1.1 (29) and Raster3D v2.5 (30).

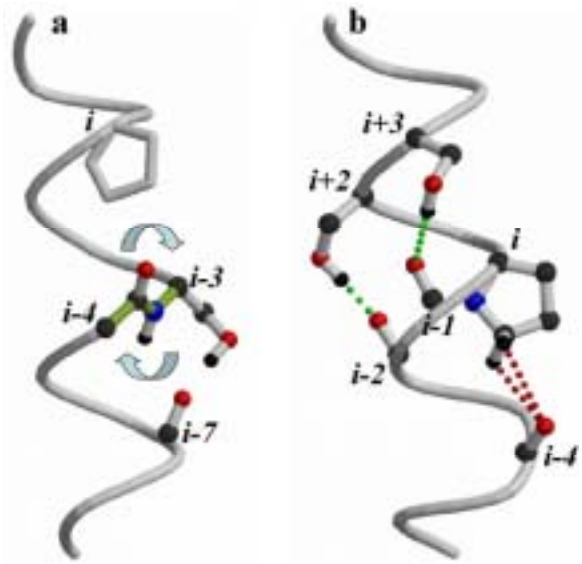



**Figure 3.2.7.** Comparison of helix bending between (a) SAP in the *g*+ rotamer (dark green), *t* rotamer (yellow), and *g*- rotamer interacting with the carbonyl oxygen at position j-3 relative to Ser (light green); and (b) SP in the *g*+ rotamer (red) and *g*- rotamer interacting with the carbonyl oxygen at position j-4 (dark orange) and j-3 (light orange). The conformation of the *t* rotamer was taken as a regular Pro-kinked α-helix. The side chains of Pro and Ser/Thr are shown. Figures were created using MolScript v2.1.1 (29) and Raster3D v2.5 (30).

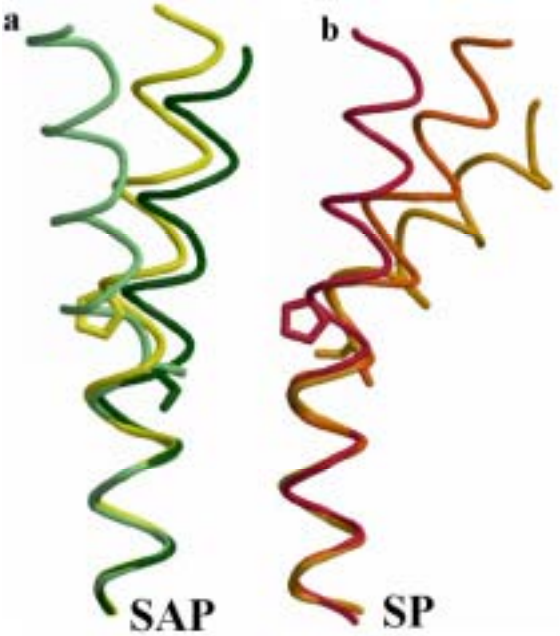


**Figure 3.2.8.** Comparison of helix bending between the computationally determined conformations of Pro (control, yellow) and TP in the *g*+ rotamer (red) polyAla α-helices; and the experimentally determined conformation of helix C of bacteriorhodopsin that contains $TP^{91}$ in the *g*+ rotamer (orange), as observed in the PDB structure 1AP9. The side chains of Pro and Ser/Thr are shown. Figures were created using MolScript v2.1.1 (29) and Raster3D v2.5 (30).

**Figure 3.3.2.** Effect of AAP and TAP g1 motif on the conformation of α-helices. A, representative structures for AAP (yellow), TAP g+ (red), and an ideal α-helix (white). Backbones are represented as ribbons, and side chains of Pro and Thr are shown as solid sticks (with polar hydrogen in white) as well as the hydrogen bonding carbonyl, 4 residues before the Thr. Hydrogen bonds are indicated by dotted white lines. B, representative structures for AAP, TAP g+, SAP g- (orange), and an ideal α-helix. This view is rotated by 90° along the helical axis relative to A. In the g- rotamer, the Ser hydrogen bonds with carbonyl situated 3 residues upstream, which induces clearly a different orientation of the Pro-kink. C, the representative structure of the AAP (yellow ribbon) and TAP g+ (red ribbon) motifs are positioned in the rhodopsin template, respecting the homology between CCR5 and rhodopsin. The two representative helices were superimposed on the cytoplasmic end of TM2 in the rhodopsin structure (Protein Data Bank number 1F88), using backbone atoms up to position 2.54. Rhodopsin (turquoise) helices are shown as cylinders, except for the extracellular part of TM2, which is shown as a ribbon. This panel is viewed from the side of the protein. D, same representation as in C but viewed from the extracellular side. The influence of the Thr on the PK is clearly visible, as it bends the helix inside the bundle. Figures were created using using the InsightII software (Accelrys, San Diego).

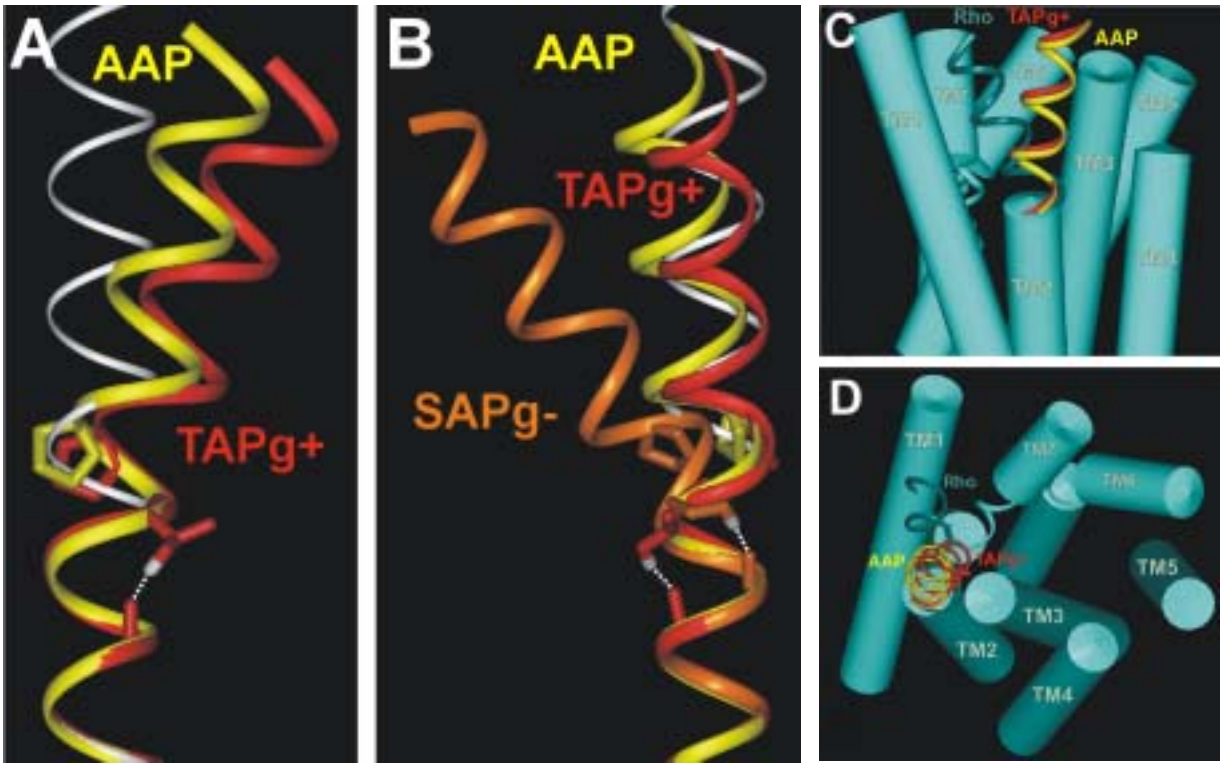


**Figure 3.4.5.** Modelling procedure of the TM2-TM3 interface. A, side-chain positions of residues 85, 89, and 104 in a model of the transmembrane domain of CCR5 based rigorously on the rhodopsin template. The $C_\alpha$ trace is shown as ribbon, in red for TM2 and TM3, in blue for the other helices. In such a model, these side chains face the membrane, which is hardly compatible with the functional role suggested by the mutagenesis experiments. B, superimposition of 100 structures from independent Molecular Dynamics simulations of TM2 (in green) and TM3 (in yellow) onto the rhodopsin template (see "Experimental Procedures" for details). Only the $C_\alpha$ trace is shown for each structure, represented as ribbon. Viewed from the extracellular side. C, same as the previous panel but viewed from the side. It appears clearly that the conformational spaces are almost not overlapping. In particular, the conformational space of TM3 moves away from that of rhodopsin, allowing room for the kinked TM2 of CCR5. D and E, extracellular and side views of the representative structures of TM2 (in green) and TM3 (in yellow) selected from the simulations and used to construct the interface. Figures were created using using the InsightII software (Accelrys, San Diego).

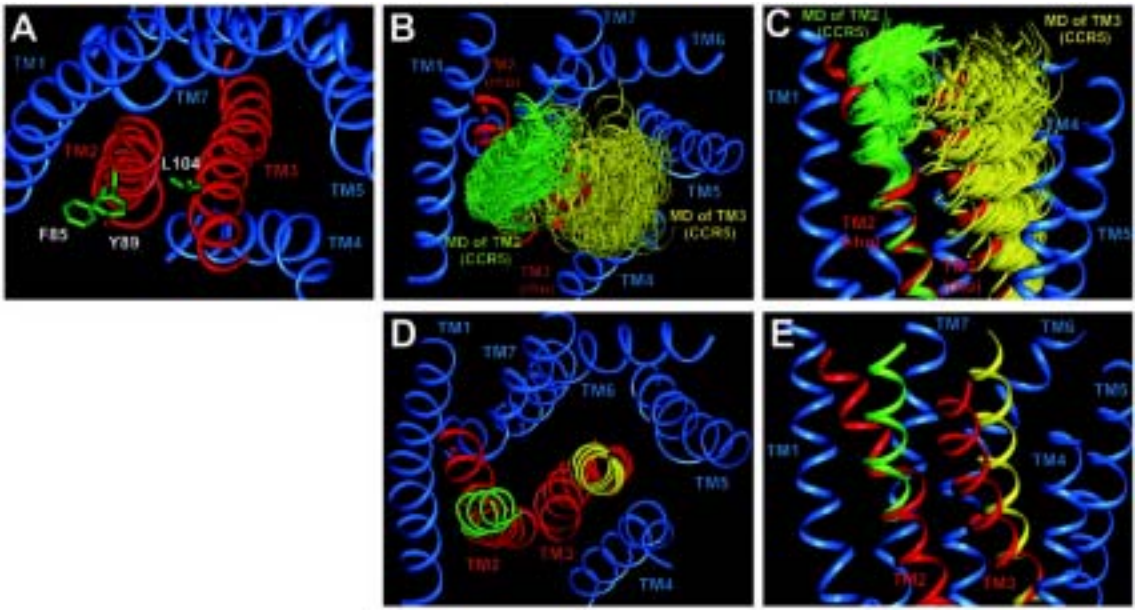


**Figure 3.4.6.** Molecular models of TM2-TM3 interface and the TM3-TM6 interaction for wt and mutant receptors. A, model of the TM2-TM3 interface obtained for the wt receptor. Only interacting side chains are shown, highlighting the proposed Phe-85-Leu-104 and Tyr-89-Thr-99 interactions (see text for details). Panels B and C show the models for F85L and L104F mutants, respectively, illustrating the loss of interaction at the 85–104 locus for these receptors. D, model of the double F85L/L104F mutant showing the proposed recovered interaction. E, wt model showing the location of Phe-109 and Phe-112 and featuring the conformation of Asn-252 interacting with the backbone in TM6. F, model of the F112Y mutant, showing how this mutant could modify the conformation of Asn-252 in TM6. G, model of the double F109H/F112Y, illustrating how the F109H additional mutation would recover the structural effect of the F112Y mutant. Figures were created using using the InsightII software (Accelrys, San Diego).

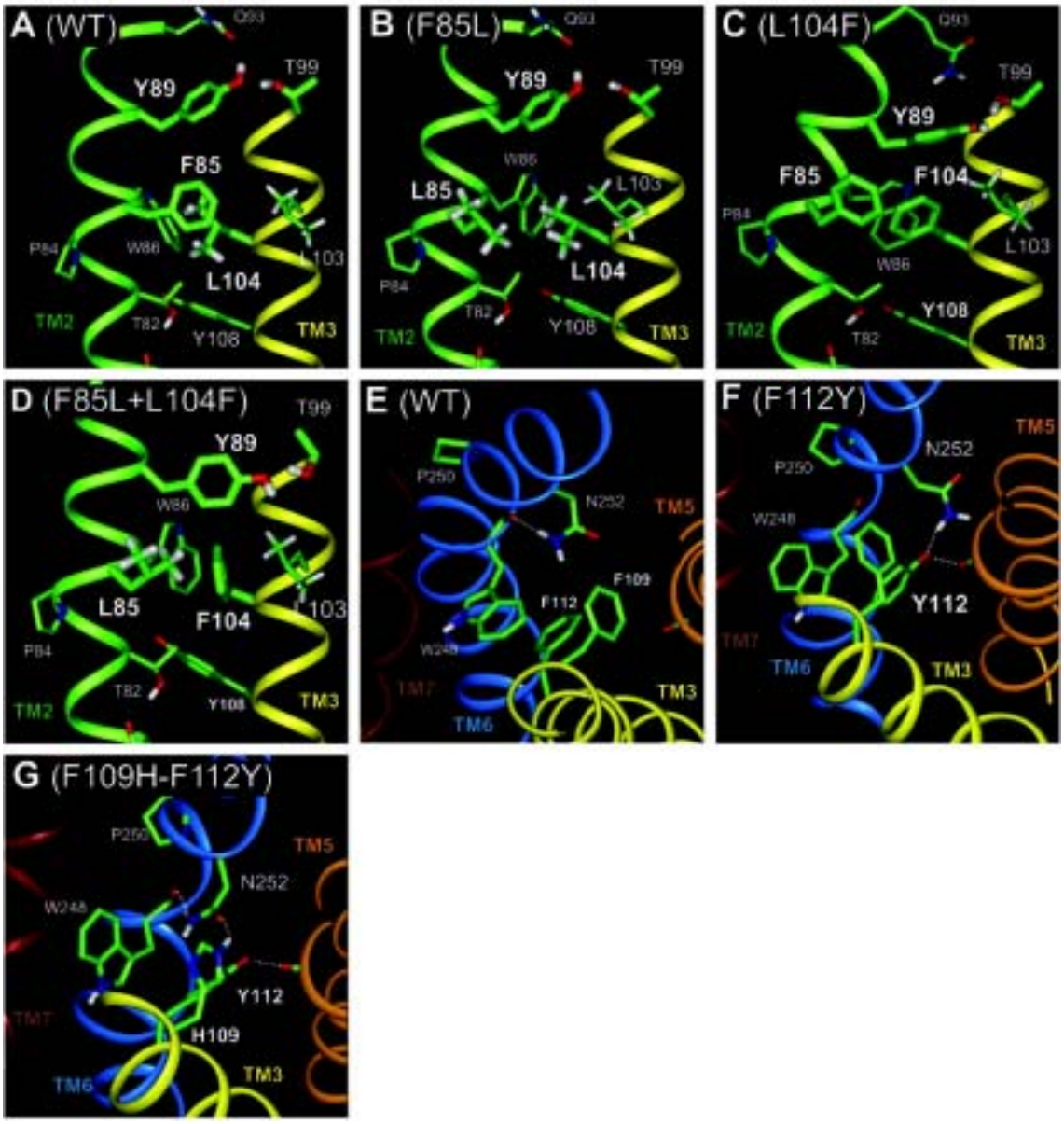


**Figure 1.4.** Mean values of the φ (dotted line) and ψ (solid line) torsional angles along a Pro-containing α-helix as observed during molecular dynamics simulations. The *x* axis shows the position of the residues in the helix, where *i* corresponds to the position of the Pro residue. The region corresponding to the Pro-kink turn (residues from *i-4* to *i*) is squared.

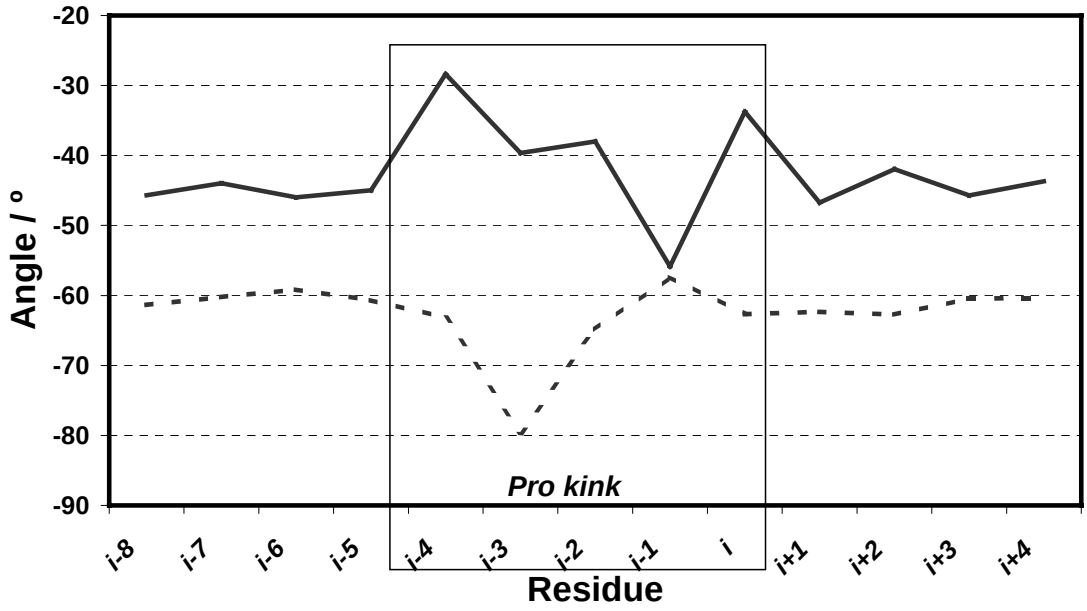



**Figure 1.5.** Mean values of the unit twist angle along a Pro-containing α-helix as observed during molecular dynamics simulations. The *x* axis shows the position of each helical turn (composed by four residues), where (*i-3*, *i*) corresponds to the Pro-kink turn (circled). The region corresponding to those turns with at least one residue in the Pro-kink region (turns from (*i-5*, *i-2*) to (*i*, *i+3*)) is squared. The decrease of the unit twist angle at the Pro-kink turn corresponds to an opening of the helix at this region.

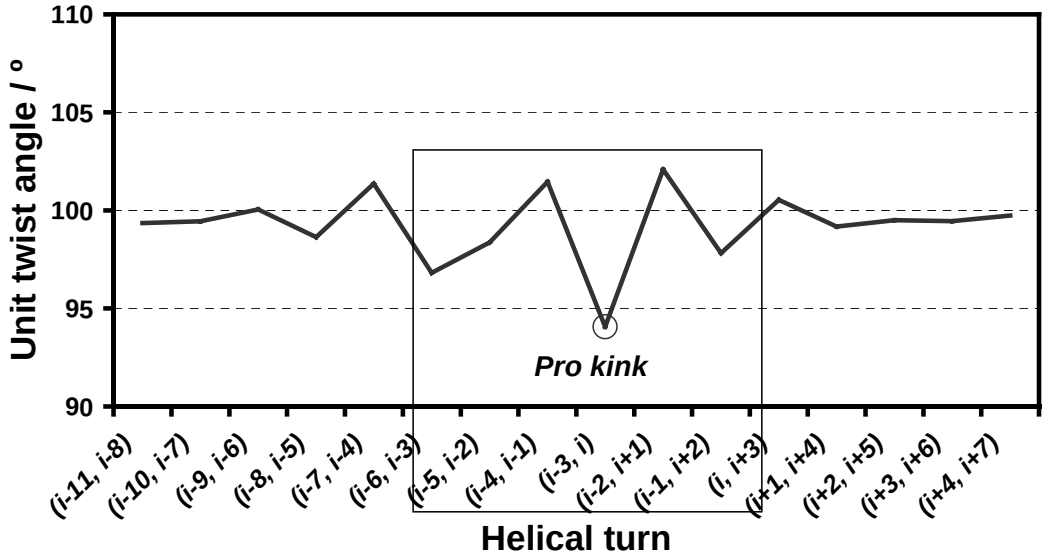
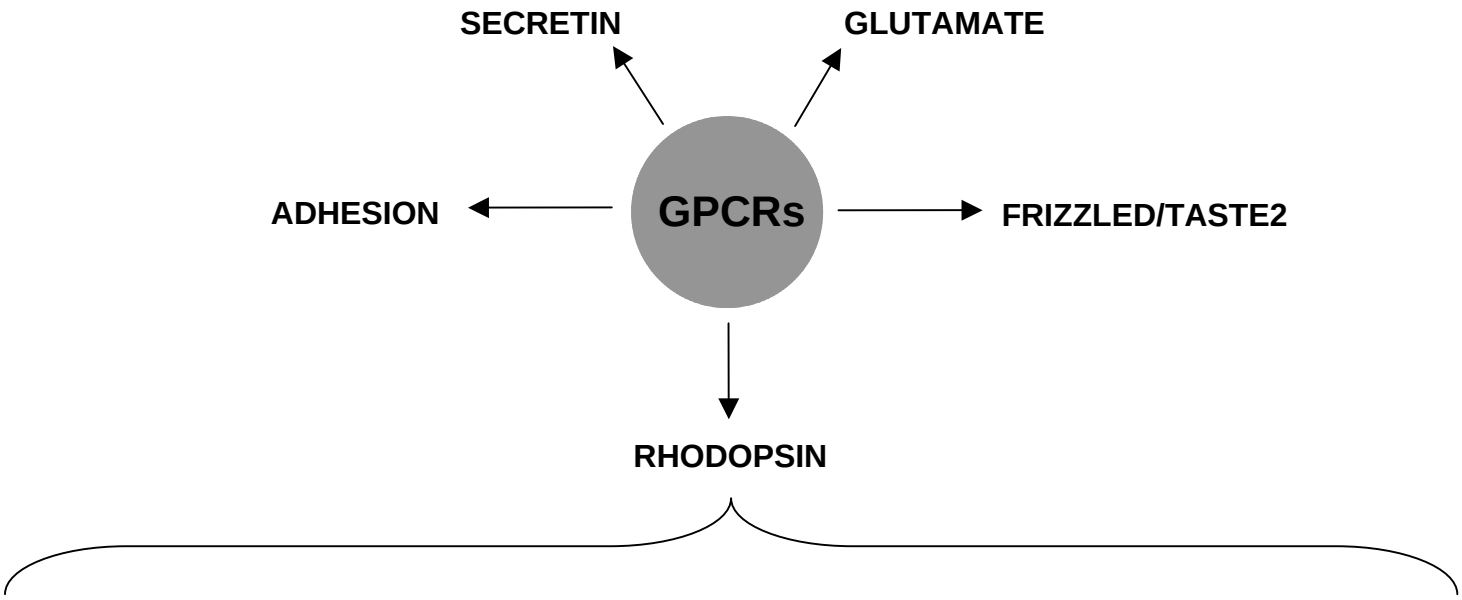


**Figure 1.7.** Classification of GPCRs obtained through a phylogenetic analysis of ~800 sequences of human receptors (105). There are five main groups, the larger being the rhodopsin family, which is subdivided in four main groups and thirteen subclasses.




**α**

**prostaglandin cluster**
prostaglandin

**amine cluster**
serotonin
dopamine
muscarinic
histamine
adrenergic
trace amine

**opsin cluster**
rod visual pigment
cone visual pigment
per-opsin
encephalopsin
melanopisn
retinal GPCR

**melatonin cluster**
melatonin

**MECA cluster**
melanocortin
endothelial differenciation GPCR
cannabionid
adenosin

**β**

**(peptides)**
hypocretin
neuropeptide FF
tachykinin
cholecystokinin
neuropeptide Y
endothelin-related
gastrin-releasing peptide
neuromedin B
uterinbombesin
neurotensin
growth hormone secretagogues
neuromedin
thyrotropin releasing hormone
ghrelin
arginine vasopresin
gonadotropin-releasing hormone

**γ**

**SOG cluster (peptides)**
galanin
RF-amide binding
somatostatin
opioid

**MCH cluster**
melanin-concentrating hormone

**chemokine cluster**
chemokine
angiotensin/bradykinin-related

**δ**

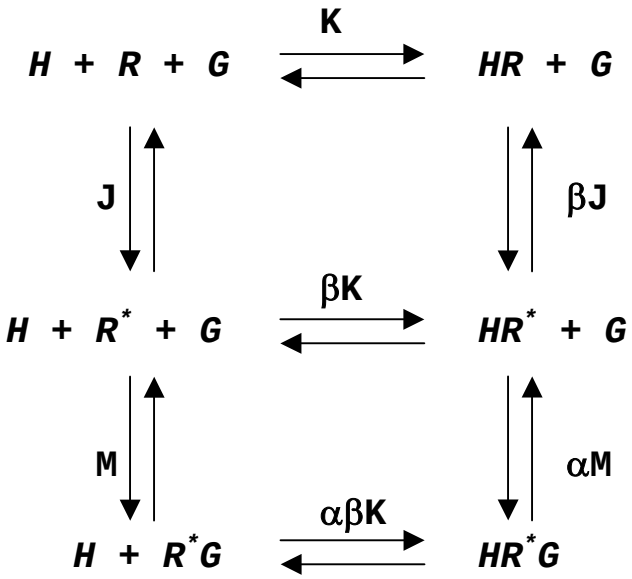
**MAS related cluster**
MAS1 oncogene
MAS-related

**glycoprotein**
glycoprotein hormone
leucine-rich-repeat containing GPCRs

**purin cluster**
formyl peptide
nucleotide

**olfactory cluster**

**Figure 1.8.** Diagram of the extended ternary complex model of GPCRs activation. This process is described through the equilibrium among the different chemical species: receptor (R), activated receptor ($R^*$), G protein (G) and ligand (H). Each reaction of equilibrium is characterized by an equilibrium constant.




These constants are:

$J = [R^*] / [R]$

$K = [HR] / [H][R]$

$M = [R^*G] / [R^*][G]$

$\beta = [HR^*][R] / [HR][R^*]$

$\alpha = [HR^*G][R^*] / [HR^*][R^*G]$

**Figure 2.1.** Highly conserved sequence motifs in the TMHs of Class A GPCRs. The highest conserved positions are highlighted in boldface.

| | |
|---|---|
| TMH1 | G**N** |
| TMH2 | NLAxA**D** |
| TMH3 | D**R**Y |
| TMH4 | **W** |
| TMH5 | **P** |
| TMH6 | **P** |
| TMH7 | N**P**xxY |

**Figure 2.2.** Energy terms in molecular mechanics calculations. The total energy is computed as the sum of all the contributions over all atoms in the system.

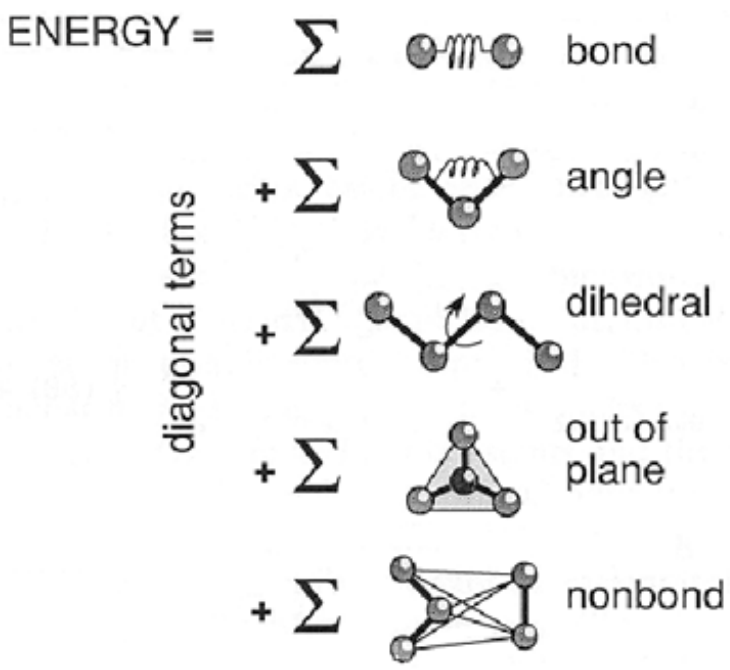


**Table 3.1.1.** Means ($\bar{x}$), standard deviations (s), and the difference in degrees (Δ) relative to Ala (in bold) of the backbone $\phi_i$ and $\psi_i$ dihedral angles and bend angle of α–helices containing Ala and Ser, Thr, and Cys residues in the gauche– (g–), trans (t), and gauche+ (g+) rotamer conformations. The results are presented for membrane and soluble proteins.

| | ***Membrane proteins*** | | | | | | | ***Soluble proteins*** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi_i$ | | | $\psi_i$ | | | | $\phi_i$ | | | $\psi_i$ | | |
| | n | $\bar{x}$ | s | Δ | $\bar{x}$ | s | Δ | n | $\bar{x}$ | s | Δ | $\bar{x}$ | s | Δ |
| **Ala** | **48** | **-60.9** | **6.5** | | **-44.4** | **8.6** | | **730** | **-63.1** | **5.6** | | **-41.2** | **5.6** | |
| *g–* | *22* | *-65.2* | *8,6* | *-4.3*$^{*}$ | *-34.3* | *8,5* | *10.1*$^{***}$ | *76* | *-67.9* | *9.2* | *-4.8*$^{***}$ | *-34.1* | *9.4* | *7.0*$^{***}$ |
| Thr | 9 | -66.3 | 7.8 | -5.4 | -35.7 | 8.9 | 8.7$^{***}$ | 34 | -69.6 | 9.2 | -6.5$^{***}$ | -35.0 | 10.5 | 6.2$^{***}$ |
| Ser | 13 | -64.5 | 9.3 | -3.6 | -33.3 | 8.4 | 11.1$^{***}$ | 42 | -66.5 | 9.0 | -3.4$^{**}$ | -33.5 | 8.4 | 7.7$^{***}$ |
| Cys | 0 | —— | — | — | —— | — | — | 0 | —— | — | — | —— | — | — |
| *g+* | *52* | *-61.8* | *6.6* | *-0.9* | *-45.3* | *9.3* | *-0.9* | *386* | *-63.4* | *5.6* | *-0.3* | *-43.2* | *6.3* | *-2.0*$^{***}$ |
| Thr | 32 | -61.8 | 7.6 | -0.9 | -46.3 | 9.3 | -1.9 | 211 | -63.4 | 5.3 | -0.3 | -44.2 | 5.2 | -3.0$^{***}$ |
| Ser | 16 | -62.5 | 4.9 | -1.6 | -43.9 | 9.9 | 0.5 | 129 | -63.3 | 6.1 | -0.2 | -42.2 | 7.6 | -1.0 |
| Cys | 4 | -58.9 | 1.5 | 2.0 | -42.9 | 6.4 | 1.5 | 46 | -63.7 | 5.7 | -0.6 | -41.6 | 6.8 | -0.4 |
| *t* | *5* | *-56.9* | *9.6* | *4.0* | *-44.4* | *8.0* | *0.0* | *96* | *-62.6* | *5.7* | *0.4* | *-44.4* | *5.3* | *-3.2*$^{***}$ |
| Thr | 0 | —— | — | — | —— | — | — | 2 | -68.8 | 8.8 | -5.7 | -40.1 | 6.1 | 1.1 |
| Ser | 5 | -56.9 | 9.6 | 4.0 | -44.4 | 8.0 | 0.0 | 74 | -62.7 | 5.3 | 0.4 | -44.0 | 4.9 | -2.8$^{**}$ |
| Cys | 0 | —— | — | — | —— | — | — | 20 | -62.0 | 6.7 | 1.1 | -46.1 | 6.2 | -5.0$^{**}$ |

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

**Figure 3.1.1.** Analysis of the torsion ϕ (panels a and b) and ψ (panels c and d) angles of α–helices, in membrane (panels a and c) and soluble (panels b and d) proteins, containing Ala (control in black) and Ser, Thr, and Cys residues in the gauche–, trans, and gauche+ rotamer conformations. Histograms depict the mean values and the lines extending from the bar represent the standard deviation.

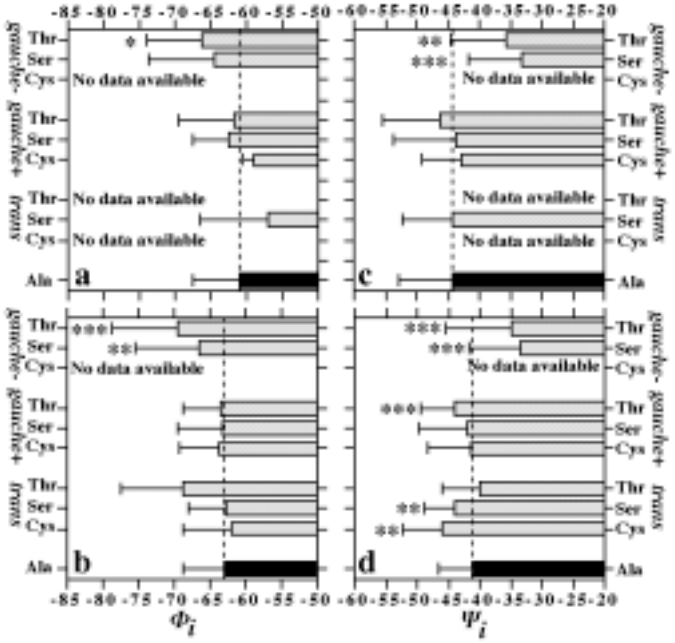


**Figure 3.1.3.** Analysis of the bend angle of α–helices containing Ala (control in black) and Ser, Thr, and Cys residues in the *gauche–*, *trans*, and *gauche+* rotamer conformations in soluble proteins. Histograms depict the mean values and the lines extending from the bar represent the standard deviation.

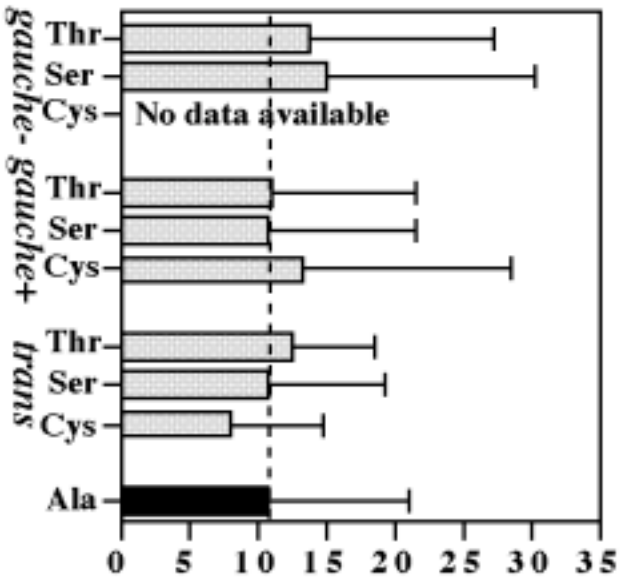



**Table 3.1.2.** Means ($\bar{x}$), standard deviations (s), and the difference in degrees (Δ) relative to Ala (in bold) of the bend angle of α-helices in soluble proteins containing Ala and Ser, Thr, and Cys residues in the *gauche–* (*g–*), *trans* (*t*), and *gauche+* (*g+*) conformations.

| | | n | $\bar{x}$ | s | Δ |
|---|---|---|---|---|---|
| | **Ala** | **730** | **10.7** | **10.2** | |
| *g–* | | *76* | *14.5* | *14.4* | *3.8** |
| | Thr | 34 | 13.9 | 13.4 | 3.2 |
| | Ser | 42 | 15.0 | 15.3 | 4.3 |
| | Cys | 0 | — | — | — |
| *g+* | | *386* | *11.2* | *11.3* | *0.5* |
| | Thr | 211 | 10.9 | 10.7 | 0.2 |
| | Ser | 129 | 10.8 | 10.7 | 0.1 |
| | Cys | 46 | 13.2 | 15.4 | 2.5 |
| *t* | | *96* | *10.3* | *8.2* | *-0.4* |
| | Thr | 2 | 12.5 | 6.1 | 1.8 |
| | Ser | 74 | 10.8 | 8.5 | 0.1 |
| | Cys | 20 | 8.1 | 6.7 | -2.6 |

$^{*}p < 0.05$

**Table 3.2.1.** Observed and expected, calculated with the TMSTAT formalism, number of occurences of the (S/T)xxP, (S/T)xP, (S/T)P, P(S/T), Px(S/T), and Pxx(S/T) motifs in a non-homologous database of transmembrane helices; and the odds ratio and its significance.

| Pair | Observed | Expected | Odds Ratio | Significance |
|---|---|---|---|---|
| **SxxP** | 292 | 294.9 | 0.99 | 0.88 |
| **TxxP** | 267 | 268.7 | 0.99 | 0.95 |
| **SxP** | 353 | 314.6 | 1.12 | 0.02 |
| **TxP** | 316 | 286.6 | 1.10 | 0.07 |
| **SP** | 291 | 334.2 | 0.87 | 0.01 |
| **TP** | 325 | 304.6 | 1.07 | 0.21 |
| **PS** | 280 | 334.2 | 0.84 | 0.001 |
| **PT** | 308 | 304.6 | 1.01 | 0.83 |
| **PxS** | 297 | 314.6 | 0.94 | 0.31 |
| **PxT** | 283 | 286.6 | 0.99 | 0.85 |
| **PxxS** | 249 | 294.9 | 0.84 | 0.004 |
| **PxxT** | 250 | 268.7 | 0.93 | 0.23 |

**Table 3.2.2.** Percentage of occurences of the (S/T)xxP, (S/T)xP, (S/T)P, P(S/T), Px(S/T), and Pxx(S/T) motifs in the rhodopsin-like family of G protein-coupled receptors.

| position[a] | number of occurences of Pro | SxxP | TxxP | SxP | TxP | SP | TP | PS | PT | PxS | PxT | PxxS | PxxT | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.36 | 378 | 2 | 4 | 0 | 1 | 5 | 0 | 6 | 12 | 2 | 1 | 3 | 0 | 35 |
| 1.48 | 203 | 0 | 13 | 9 | 0 | 0 | 12 | 2 | 1 | 1 | 0 | 3 | 2 | 44 |
| 2.58 | 546 | 6 | 1 | 10 | 71 | 0 | 1 | 0 | 1 | 1 | 0 | 5 | 2 | 100 |
| 2.59 | 875 | 6 | 16 | 3 | 15 | 5 | 6 | 2 | 3 | 16 | 10 | 1 | 3 | 86 |
| 2.60 | 103 | 5 | 0 | 3 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 8 | 28 | 47 |
| 4.59 | 1139 | 2 | 1 | 42 | 11 | 6 | 4 | 3 | 3 | 0 | 2 | 2 | 0 | 76 |
| 4.60 | 634 | 34 | 12 | 3 | 2 | 1 | 3 | 1 | 1 | 1 | 2 | 5 | 0 | 66 |
| 5.50 | 1588 | 0 | 1 | 3 | 3 | 0 | 3 | 2 | 4 | 11 | 9 | 1 | 5 | 40 |
| 6.50 | 1603 | 8 | 3 | 1 | 0 | 3 | 15 | 1 | 0 | 4 | 4 | 7 | 4 | 51 |
| 6.59 | 97 | 0 | 6 | 1 | 4 | 6 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 21 |
| 7.38 | 235 | 3 | 8 | 9 | 53 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 3 | 94 |
| 7.45 | 101 | 13 | 4 | 0 | 13 | 0 | 0 | 3 | 0 | 2 | 4 | 0 | 0 | 39 |
| 7.46 | 325 | 1 | 10 | 0 | 0 | 2 | 40 | 4 | 4 | 0 | 0 | 1 | 0 | 62 |
| 7.50 | 1888 | 5 | 4 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| | 9715 | 6 | 5 | 7 | 9 | 2 | 6 | 2 | 2 | 4 | 3 | 2 | 2 | 52 |

[a]Positions in the generalized numbering scheme of Ballesteros & Weinstein (17).

**Table 3.2.3.** Percentage (%) of the intra-helical hydrogen bond between the $O_\gamma H$ side chain of Ser/Thr and the $O=C_{j-4}$ and $O=C_{j-3}$ carbonyl, and bend angle (mean±standard deviation) as observed during the molecular dynamics simulations of polyAla α-helices containing the (T/S)AAP, (T/S)AP, (T/S)P, P(T/S), PA(T/S), and PAA(T/S) motifs, with Ser and Thr side chains built in both *gauche+* and *gauche-* conformations.

| | ***gauche +*** | | | | ***gauche -*** | | | |
|---|---|---|---|---|---|---|---|---|
| | $O=C_{j-4}$ | | $O=C_{j-3}$ | | $O=C_{j-4}$ | | $O=C_{j-3}$ | |
| | % | bend | % | bend | % | bend | % | bend |
| **SAAP** | 100 | 15.8±6.4** | 0 | - | 94 | 14.2±5.4*** | 6 | - |
| **TAAP** | 100 | 15.1±7.3*** | 0 | - | 89 | 13.5±4.7*** | 11 | - |
| **SAP** | 100 | 30.3±11.6*** | 0 | - | 2 | - | 98 | 34.8±9.6*** |
| **TAP** | 98 | 27.2±7.6*** | 2 | - | 88 | 26.5±7.4*** | 12 | - |
| **SP** | 99 | 30.5±8.2*** | 1 | - | 67 | 19.6±4.8 | 33 | 49.8±10.1*** |
| **TP** | 99 | 26.9±6.9*** | 1 | - | 92 | 20.9±6.1 | 8 | - |
| **P** [a] | **19.7±7.3** | | | | | | | |
| **PS** | 100 | 18.9±6.4 | 0 | - | 83 | 17.5±5.8 | 17 | - |
| **PT** | 100 | 18.9±6.8 | 0 | - | 98 | 19.7±6.5 | 2 | - |
| **PAS** | 100 | 16.0±6.5** | 0 | - | 65 | 14.2±5.6*** | 35 | 14.6±6.6** |
| **PAT** | 100 | 14.1±6.8*** | 0 | - | 93 | 16.4±5.1* | 7 | - |
| **PAAS** | 100 | 14.0±6.2*** | 0 | - | 35 | 29.6±8.6*** | 65 | 30.0±10.9*** |
| **PAAT** | 100 | 15.6±6.6** | 0 | - | 74 | 24.7±8.8** | 26 | 32.4±10.0*** |

[a]A Pro-containing polyAla α–helix is used as a control. One–way analysis of variance plus a posteriori two–sided Dunnett's T tests was employed to contrast if the bend angle of the model peptides differs from the control simulation.

(*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$)

**Figure 3.2.5.** Evolution of the distance (bottom panels) between the $H_\gamma$ atom of Ser and the oxygen of the $O=C_{j-4}$ (solid line) or $O=C_{j-3}$ (broken line) carbonyl, relative to Ser, and the values of $\chi_1$ (top panels) of Ser obtained during the production run of the molecular dynamics simulations of SP (left panels) and SAP (right panels) in the *g*- rotamer.

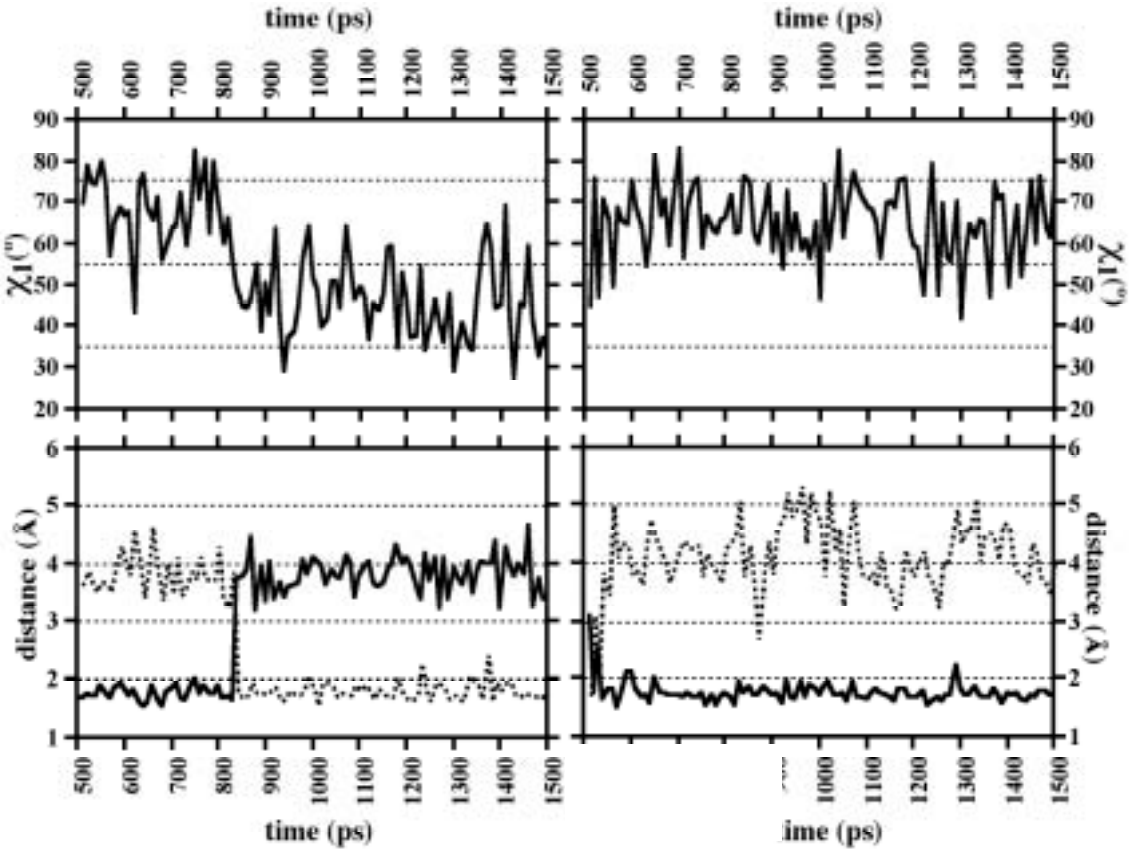


**Table 3.3.1.** Bending angles for TXP motifs found in the Protein Data Bank (PDB) Two segments are taken from the structure of cytochrome c oxidase (1AR1), which is a membrane protein. The other proteins are hemoglobin (1VHB), a phosphoenolpyruvate carboxylase (1FIY), a transposase inhibitor (1B7E), an endonuclease (1RVE), and the clathrin heavy chain (1BPO).

| PDB no. | Motif | Helical segment | Position of the Pro | Bend angle (degrees) |
|---|---|---|---|---|
| 1VHB | TVP | A7→A20 | A15 | 28 |
| 1FIY | TDP | 823→856 | 838 | 44 |
| 1B7E | TLP | 349→379 | 373 | 49 |
| 1RVE | SRP | B37→B59 | B50 | 26 |
| 1AR1 | SLP | A219→A248 | A236 | 35 |
| 1AR1 | SLP | B85→B103 | B98 | 29 |
| 1BPO | CRP | 428→442 | 438 | 41 |

**Table 3.3.2.** Average and standard deviation of the bending angle obtained during the MD simulations. Each value is obtained over a sample of 100 structures equally distant during the MD run (see Methods). One-way analysis of variance plus a posteriori one-sided Dunnett's T tests were used to determine wether the bending angles of the different motifs are significantly larger than that of the control AAP. Asterisks denotes statistically significant deviation from the AAP control ($p < 0.001$).

| | Bend angle (degrees) |
|---|---|
| AAP | 19.9 (7.3) |
| TAP g+ | 28.4 (7.1)* |
| TAP g- | 26.6 (7.8)* |
| SAP g+ | 30.1 (11.9)* |
| SAP g- | 35.0 (10.0)* |
| CAP g+ | 31.3 (10.8)* |
| VAP | 21.0 (5.5) |

**Figure 3.3.1.** Alignment of TM2 sequences from chemokine receptors and bovine rhodopsin. For the sake of clarity, only human and murine sequences are shown. The generalized numbering scheme (see "Experimental Procedures") is used to label the alignment. The TXP (or SXP) motif is indicated in boldface characters, and its conserved residues are highlighted. The sequence of TM2 of bovine rhodopsin is also aligned, showing the high homology between CCR5 and bovine rhodopsin in the cytosolic part of TM2, up to the TXP motif. Note that we make the assumption that, in chemokine receptors, TM2 extends to position 2.67 (included). This is based on the observation that TM2 extends to this residue in the three-dimensional structure of rhodopsin (13) and on the suggestion that position 2.67 terminates TM2 in the dopamine D2 receptor, on the basis of the substituted cysteine accessibility method (53).

```
                2.38        2.50    2.58     2.67
                  |           |       |        |
Rho  (bovin)    PLNYILLNLAVADLFMVFGGFTTTLYTSLH

CCR5 (human)    MTDIYLLNLAISDLFFLLTVPFWAHYAAAQ
CCR5 (mouse)    VTDIYLLNLAISDLLFLLTLPFWAHYAANE
CCR5 (rat)      MTDIYLFNLAISDLLFLLTLPFWAHYAANE
CCR2 (mouse)    MTDIYLLNLAISDLLFLLTLPFWAHYAANE
CCR2 (human)    LTDIYLLNLAISDLLFLITLPLWAHSAANE
CCR1 (human)    MTSIYLLNLAISDLLFLFTLPFWIDYKLKD
CCR1 (mouse)    MTSIYLFNLAVSDLVFLFTLPFWIDYKLKD
CCRV (mouse)    MTSIYLFNLAISDLVFLSTLPFWVDYIMKG
CCR3 (human)    MTNIYLLNLAISDLLFLVTLPFWIHYVRGH
CCR3 (mouse)    MTNIYLFNLAISDLLFLFTVPFWIHYVLWN
CCR3 (rat)      MTNIYLLNLAISDLLFLFTVPFWIHYVLWN
CCR4 (human)    MTDVYLLNLAISDLLFVFSLPFWGYYAADQ
CCR4 (mouse)    MTDVYLLNLAISDLLFVLSLPFWGYYAADQ
CCR8 (human)    ITDVYLLNLALSDLLFVFSFPFQTYYLLDQ
CCR8 (mouse)    ITDIYLLNLAASDLLFVLSIPFQTHNLLDQ
CCR6 (human)    MTDVYLLNMAIADILFVLTLPFWAVSHATG
CCR6 (mouse)    MTDVYLLNMAITDILFVLTLPFWAVTHATN
CCR7 (human)    MTDTYLLNLAVADILFLLTLPFWAYSAAKS
CCR7 (mouse)    MTDTYLLNLAVADILFLLILPFWAYSEAKS
CCR9 (human)    MTDMFLLNLAIADLLFLVTLPFWAIAAADQ
CCR9 (mouse)    MTDMFLLNLAIADLLFLATLPFWAIAAAGQ
CCRA (human)    PTSAHLLQLALADLLLALTLPFAAAGALQG
CXCR1 (human)   VTDVYLLNLALADLLFALTLPIWAASKVNG
CXCR1 (rat)     VTDVYVLNLAIADLLFSLTLPFLAVSKWKG
CXCR2 (mouse)   VTDVYLLNLAIADLFFALTLPVWAASKVNG
CXCR2_(rat)     VTDVYLLNLAIADLFFALTLPVWAASKVNG
CXCR2 (human)   VTDVYLLNLALADLLFALTLPIWAASKVNG
CXCR3 (human)   STDTFLLHLAVADTLLVLTLPLWAVDAAVQ
CXCR5 (mouse)   STETFLFHLAVADLLLVFILPFAVAEGSVG
CXCR5 (rat)     STETFLFHLAVADLLLVFILPFAVAEGSVG
CXCR5 (human)   STETFLFHLAVADLLLVFILPFAVAEGSVG
CXCR4 (bovin)   MTDKYRLHLSVADLLFVLTLPFWAVDAVAN
CXCR4 (human)   MTDKYRLHLSVADLLFVITLPFWAVDAVAN
CXCR4 (mouse)   MTDKYRLHLSVADLLFVITLPFWAVDAMAD
CXCR4 (rat)     MTDKYRLHLSVADLLFVITLPFWAVDAMAD
CX3CR1 (human) VTDIYLLNLALSDLLFVATLPFWTHYLINE
CX3CR1 (mouse) ITDIYLLNLALSDLLFVATLPFWTHYLISH
CX3CR1 (rat)    ITDIYLLNLALSDLLFVATLPFWTHYLISH
XCR1 (human)    LTNIFILNLCLSDLVFACLLPVWISPYHWG
```

**Figure 3.3.3**. Level of expression of the receptors. Cell surface expression of WT CCR5 and the different mutants was measured by fluorescence-activated cell sorting using five different monoclonal antibodies. The data are representative of three different experiments. The 2D7 antibody recognizes a conformational epitope centered on ECL2. We have recently identified the epitopes of the other antibodies tested here (C. Blanpain, M. Mack, J.-M. Vanderwinden, V. Wittamer, E. Le Poul, G. Vassart, and M. Parmentier, manuscript in preparation); MC-1 and MC-6 recognize multidomain conformational epitopes, while MC-4 targets a conformational NH2-terminal epitope and MC-5 a linear epitope also located in the amino-terminal domain of CCR5. Values represent mean cell fluorescence normalized by the value obtained for CCR5 (100%) separately for each antibody.

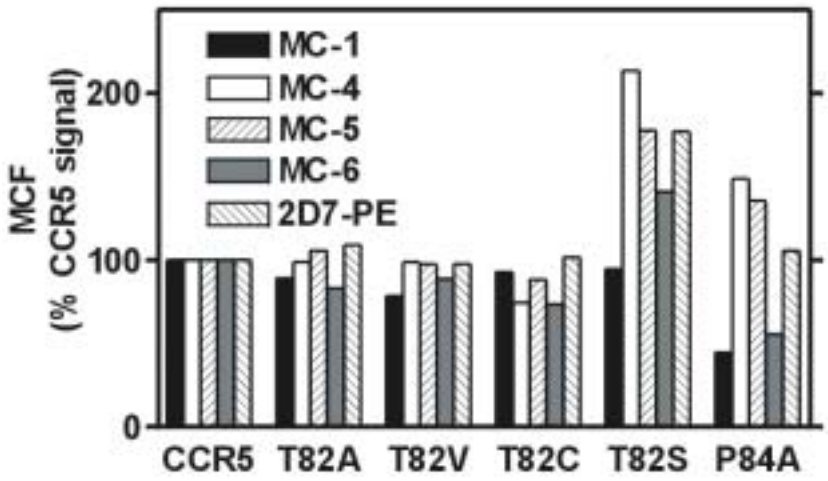



**Figure 3.3.4.** Binding properties of the WT and mutant receptors. Competition binding assays were performed on CHO-K1 cell lines expressing WT CCR5 and the different mutants using 125I-RANTES as tracer. The data are representative of at least two experiments. Results were analyzed by the GraphPad Prism software, using a single-site model, and the data were normalized for nonspecific (0%) and specific binding in the absence of competitor (100%). All points were run in triplicate (error bars represent S.E.). Unlabeled ligands are as follows: RANTES (•), MIP-1β (□), MIP-1α (▲), and MCP2 (♦).

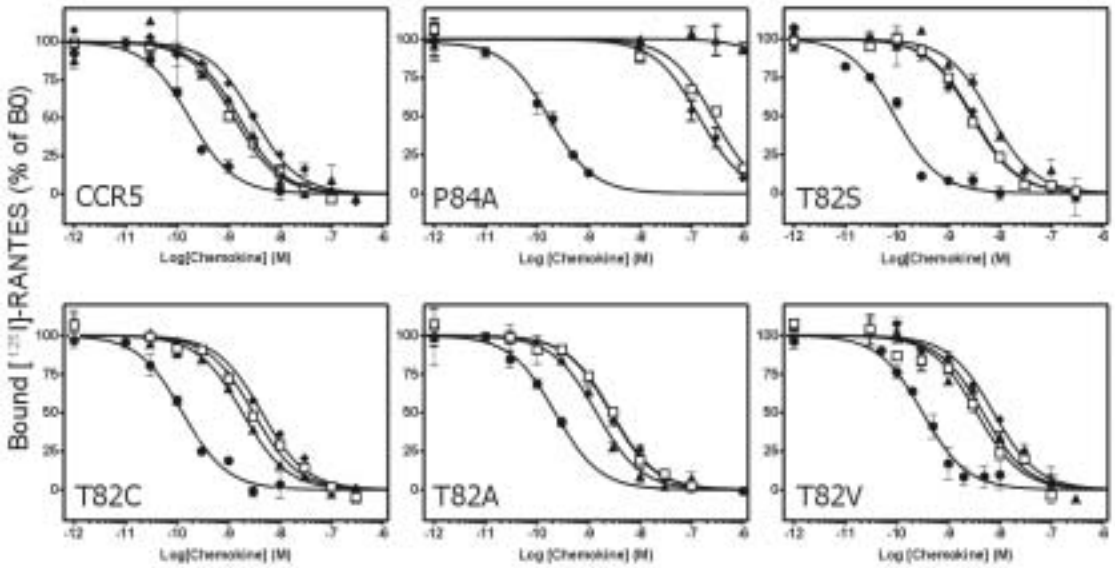


**Table 3.3.3.** Binding and functional properties of WT CCR5 and mutant receptors. The pIC50 (-log M) values were obtained from competition binding experiments, using 125I-RANTES as a tracer (as displayed in Fig. 4). Values represent the mean 6 S.E. of at least two independent determinations. The pEC50 (-log M) values were obtained from functional dose-response curves using the aequorin assay (as displayed in Fig. 5). Values represent the mean 6 S.E. of at least three independent determinations.

| | RANTES | | MIP-1β | | MIP-1α | | MCP-2 | |
|---|---|---|---|---|---|---|---|---|
| | $pIC_{50}$ | $pEC_{50}$ | $pIC_{50}$ | $pEC_{50}$ | $pIC_{50}$ | $pEC_{50}$ | $pIC_{50}$ | $pEC_{50}$ |
| CCR5 | 9.61 ± 0.14 | 8.48 ± 0.07 | 8.88 ± 0.01 | 8.15 ± 0.03 | 8.52 ± 0.23 | 7.62 ± 0.09 | 8.65 ± 0.11 | 7.70 ± 0.19 |
| P84A | 9.73 ± 0.03 | 7.82 ± 0.02 | 6.61 ± 0.03 | <6 | <6 | <6 | 6.92 ± 0.02 | <6 |
| T82S | 9.79 ± 0.13 | 8.36 ± 0.06 | 8.71 ± 0.16 | 7.54 ± 0.09 | 8.31 ± 0.14 | 7.43 ± 0.11 | 8.54 ± 0.01 | 7.47 ± 0.12 |
| T82C | 9.61 ± 0.20 | 8.14 ± 0.06 | 8.49 ± 0.03 | 7.38 ± 0.14 | 8.84 ± 0.11 | 7.83 ± 0.09 | 8.37 ± 0.05 | <6 |
| T82A | 9.56 ± 0.09 | 8.22 ± 0.06 | 8.52 ± 0.03 | 7.22 ± 0.04 | 8.81 ± 0.04 | 7.36 ± 0.18 | 8.36 ± 0.22 | 6.84 ± 0.06 |
| T82V | 9.55 ± 0.10 | 7.76 ± 0.08 | 8.44 ± 0.01 | 6.36 ± 0.06 | 8.64 ± 0.31 | 6.87 ± 0.03 | 8.25 ± 0.06 | <6 |

**Figure 3.3.5.** Activation of the different receptors by the four CCR5 agonists. Shown is the functional response to RANTES (•), MIP-1β (□), MIP-1α (▲), and MCP2 (♦) of CHO-K1 cells expressing WT CCR5 or the various mutants using the aequorin assay. All points were run in triplicate (error bars represent S.E.). The displayed curves represent a typical experiment out of at least three performed independently. Results were analyzed by nonlinear regression using the GraphPad Prism software. Data were normalized to maximal cell line stimulation by a saturating concentration of ATP. Note that the vertical scales of the graphs have been adapted to the maximal responses obtained for each line.

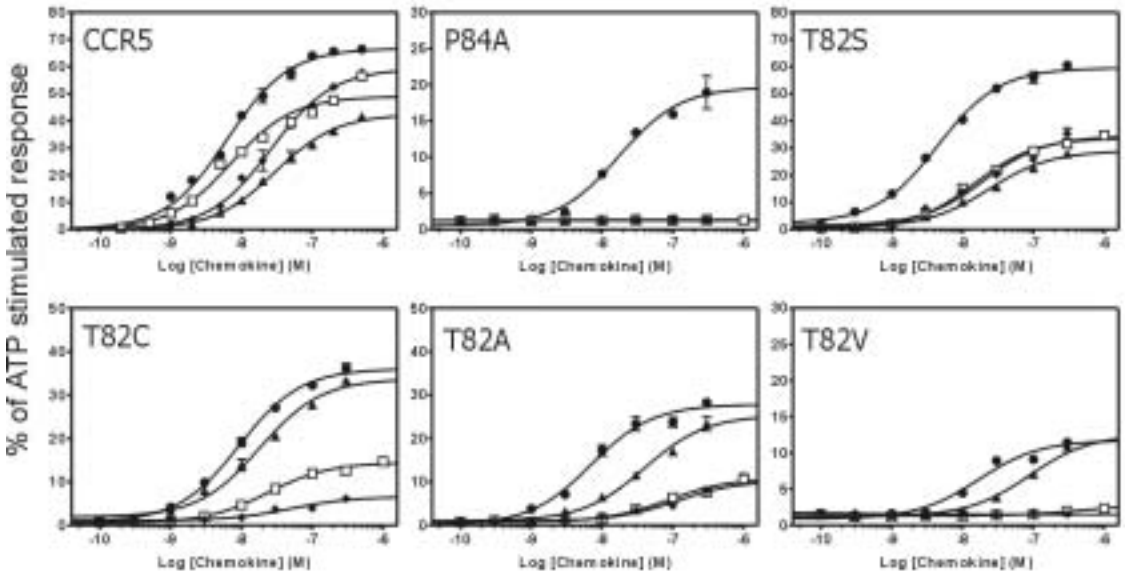


**Figure 3.4.1.** Alignment of TM2 and TM3 sequences from chemokine receptors. A, alignment comprising TM2 and TM3 for the different chemokine receptors. The generalized numbering scheme (see "Experimental Procedures") is used to label the alignment. Aromatic residues are highlighted in bold characters, whereas the TXP motif is boxed. The limits of the helices were defined as those observed for bovine rhodopsin (27). B, subset of the alignment focusing on the sequence differences between CCR5 and CCR2. Aromatic residues differing between the two receptors are in bold and marked by arrows. The CCR5 numbering for these positions is given above the sequence.

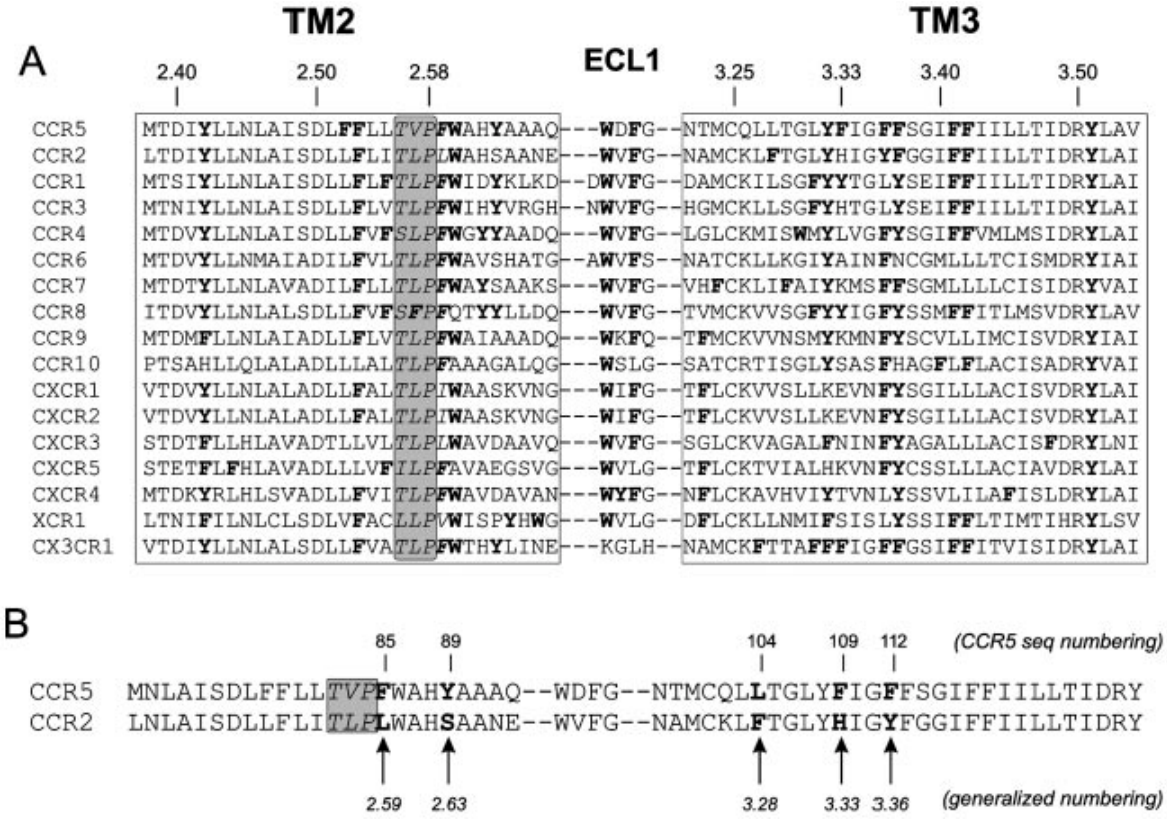


**Figure 3.4.2.** Expression efficiency of the receptor mutants. Cellsurface expression of wt-CCR5 and the different mutants measured by fluorescence-activated cell sorting using five different monoclonal antibodies. The data are representative of three independent experiments. The 2D7 antibody recognizes a conformational epitope centered on ECL2, MC-6 and 523 recognize multidomain conformational epitopes, whereas MC-5 and 3A9 target epitopes located in the N-terminal domain of CCR5. Values represent the mean (error bars: S.E.) of the mean cell fluorescence obtained in three different experiments, normalized by the value obtained for CCR5 (100%) separately for each antibody.

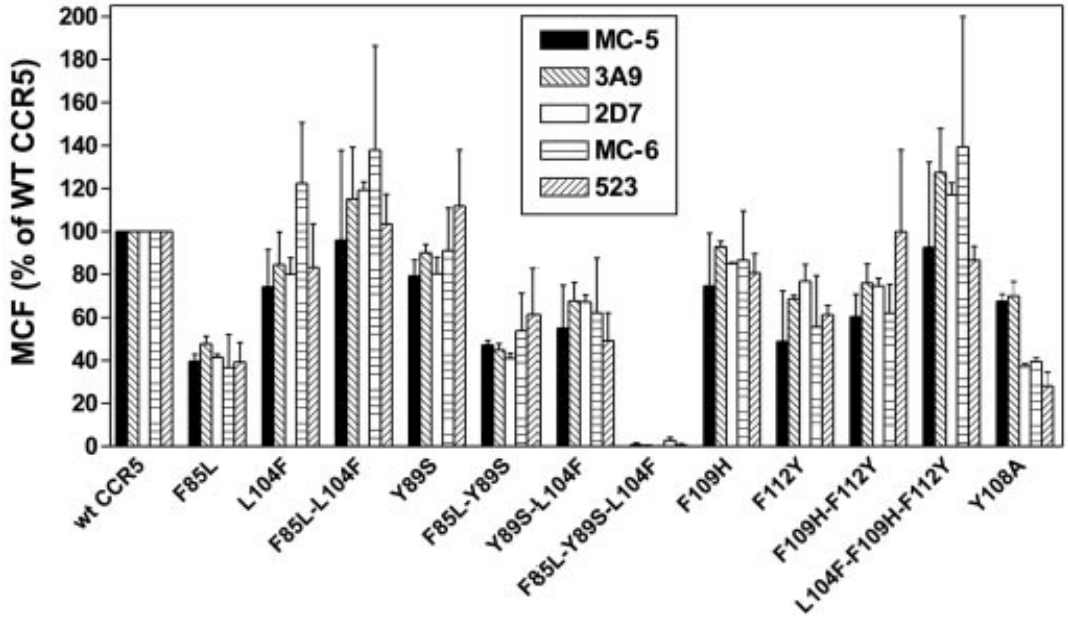


**Figure 3.4.3.** Binding properties of the wt and mutant receptors. Competition binding assays on CHO-K1 cell lines expressing wt-CCR5 and the different mutants using $^{125}$I-RANTES or $^{125}$I-MIP-1α as tracer. The data are representative of at least two experiments. Results were analyzed by the GraphPad Prism software, using a single-site model, and the data were normalized for nonspecific (0%) and specific binding in the absence of competitor (100%). All points were run in triplicate (error bars: S.E.). Unlabeled ligands are RANTES (■), MIP-1β (▲), MIP-1α (∇ ), and MCP2 (◊).

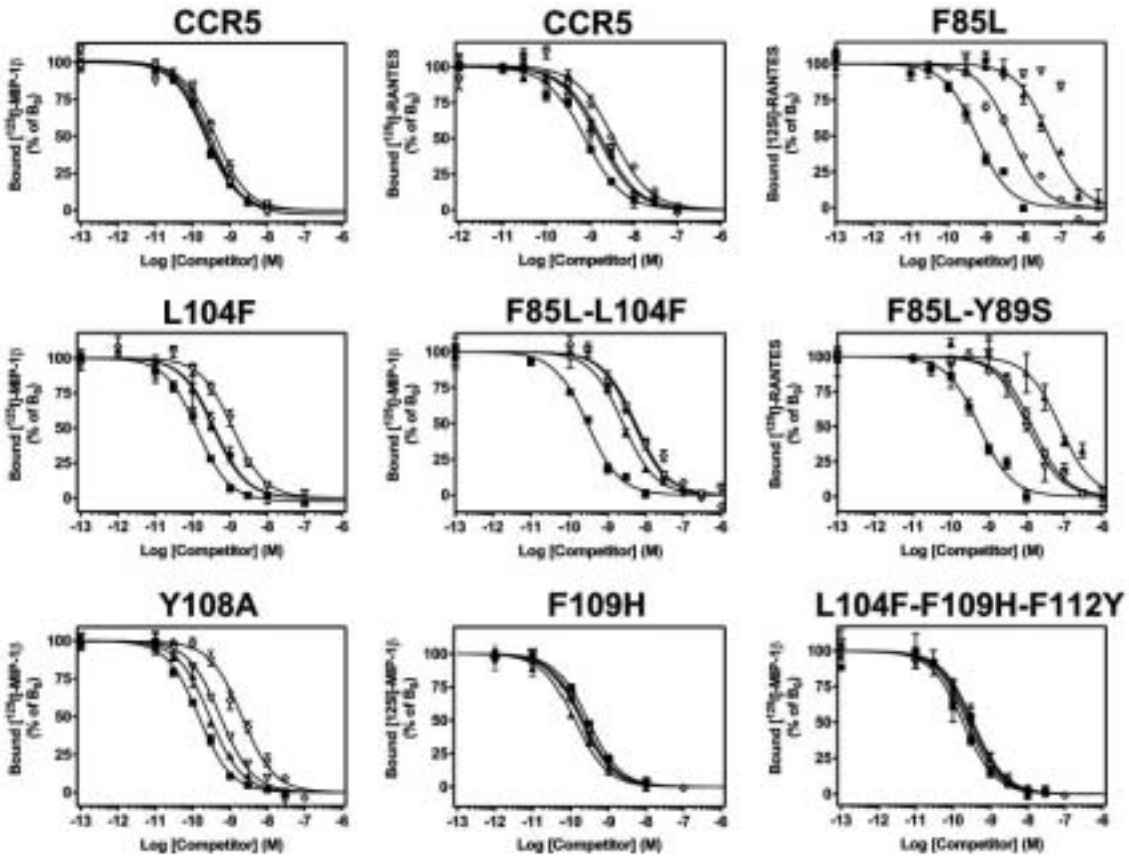


**Figure 3.4.4.** Activation of the different receptors by the four CCR5 agonists. Functional response to RANTES (■), MIP-1β (▲), MIP-1α (∇ ), and MCP2 (◊) of CHO-K1 cells expressing wt-CCR5 or the various mutants, using the aequorin assay. All points were run in triplicate (error bars: S.E.). The displayed curves represent a typical experiment out of at least three performed independently. Results were analyzed by nonlinear regression using the GraphPad Prism software. Data were normalized to maximal cell line stimulation by a saturating concentration of ATP. Note that the vertical scales of the graphs have been adapted to the maximal responses obtained for each line.

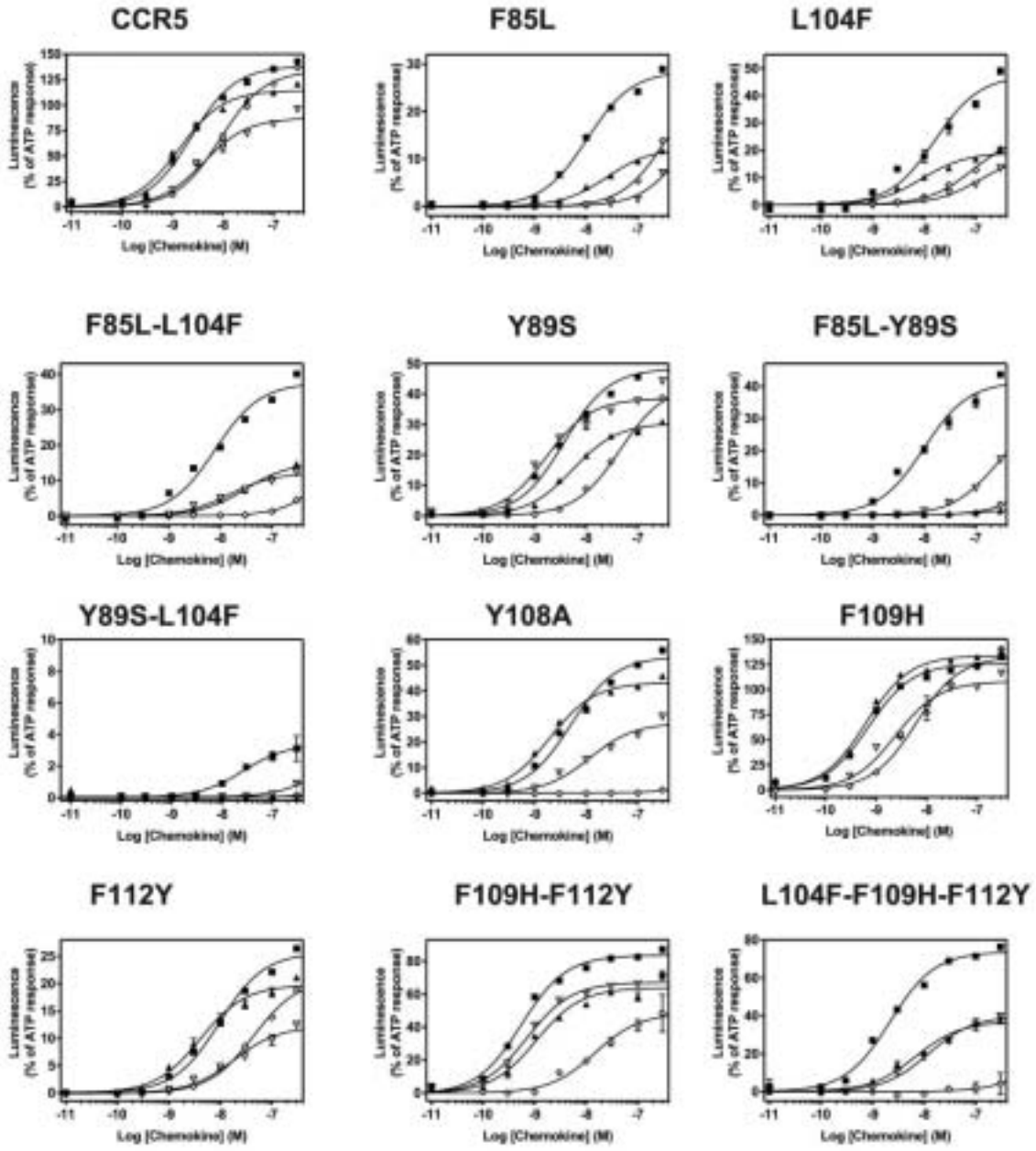


**Table 3.4.1.** Binding and functional properties of wt-CCR5 and mutant receptors. The $pIC_{50}$ (-log M) values were obtained from competition binding experiments, using $^{125}$I-RANTES or $^{125}$I-MIP-1β as a tracer. Values represent the mean ± S.E. of at least two independent determinations. The $pEC_{50}$ (-log M) values were obtained from functional dose-response curves using the aequorin assay. Values represent the mean ± S.E. of at least three independent determinations.
—, not done. NM, not measurable; no specific binding could be detected.

| | Binding | | Function | | Binding | | Function | | Binding | | Function | | Binding | | Function | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MIP-1β $pIC_{50}$ | RANTES $pIC_{50}$ | $pEC_{50}$ | $E_{max}$ | MIP-1β $pIC_{50}$ | RANTES $pIC_{50}$ | $pEC_{50}$ | $E_{max}$ | MIP-1β $pIC_{50}$ | RANTES $pIC_{50}$ | $pEC_{50}$ | $E_{max}$ | MIP-1β $pIC_{50}$ | RANTES $pIC_{50}$ | $pEC_{50}$ | $E_{max}$ |
| CCR5 | 9.64 ± 0.08 | 9.13 ± 0.00 | 8.49 ± 0.12 | 95 ± 9 | 9.59 ± 0.06 | 8.75 ± 0.11 | 8.74 ± 0.08 | 73 ± 10 | 9.42 ± 0.05 | 8.77 ± 0.11 | 8.50 ± 0.09 | 63 ± 6 | 9.59 ± 0.06 | 8.68 ± 0.21 | 8.21 ± 0.12 | 96 ± 10 |
| F85L | — | 9.18 ± 0.06 | 7.93 ± 0.04 | 26 ± 3 | 9.09[a] | 7.68 ± 0.36 | 7.72 ± 0.14 | 10 ± 1 | — | <6.5 | <6.5 | | — | 8.10 ± 0.11 | <6.5 | |
| L104F | 9.78 ± 0.17 | — | 8.08 ± 0.12 | 46 ± 5 | 9.50 ± 0.06 | — | 8.22 ± 0.07 | 27 ± 5 | 8.99 ± 0.06 | — | 7.25 ± 0.11 | 18 ± 2 | 9.35 ± 0.06 | — | 7.26 ± 0.10 | 37 ± 6 |
| F85L/L104F | — | 9.36 ± 0.18 | 8.02 ± 0.06 | 31 ± 5 | 9.80[a] | 8.40 ± 0.19 | 7.91 ± 0.09 | 7 ± 4 | — | 8.37 ± 0.06 | 7.35 ± 0.17 | 43 ± 28 | — | 8.01 ± 0.25 | <6.5 | |
| Y89S | 9.31[a] | 9.29[a] | 8.18 ± 0.13 | 40 ± 4 | 9.63 ± 0.25 | 8.62[a] | 8.11 ± 0.16 | 19 ± 6 | 9.56[a] | — | 8.53 ± 0.13 | 33 ± 3 | 9.88[a] | — | 7.01 ± 0.23 | 32 ± 5 |
| F85L/Y89S | NM | 9.28 ± 0.30 | 8.02 ± 0.10 | 41 ± 8 | NM | 7.16 ± 0.04 | <6.5 | | NM | 7.86 ± 0.10 | <6.5 | | NM | 7.93 ± 0.09 | <6.5 | |
| Y89S/L104F | NM | NM | 7.65 ± 0.04 | 7 ± 3 | NM | NM | <6.5 | | NM | NM | <6.5 | | NM | NM | <6.5 | |
| Y108A | 9.92 ± 0.09 | — | 8.30 ± 0.06 | 53 ± 2 | 9.78 ± 0.08 | — | 8.45 ± 0.24 | 40 ± 4 | 9.39 ± 0.13 | — | 7.93 ± 0.07 | 30 ± 2 | 8.73 ± 0.16 | — | <6.5 | |
| F109H | 9.55[a] | — | 8.99 ± 0.10 | 95 ± 16 | 9.89 ± 0.04 | — | 9.01 ± 0.11 | 84 ± 25 | 9.67[a] | — | 8.83 ± 0.12 | 78 ± 15 | 9.74[a] | — | 8.03 ± 0.12 | 83 ± 26 |
| F112Y | — | — | 8.26 ± 0.10 | 32 ± 5 | 9.76 ± 0.05 | — | 8.39 ± 0.05 | 18 ± 3 | — | — | 8.15 ± 0.23 | 20 ± 5 | — | — | 7.36 ± 0.09 | 25 ± 3 |
| F109H/F112Y | — | — | 9.00 ± 0.14 | 80 ± 6 | 9.91 ± 0.08 | — | 8.94 ± 0.07 | 60 ± 2 | — | — | 8.97 ± 0.09 | 62 ± 4 | — | — | 7.67 ± 0.11 | 54 ± 4 |
| L104F/F109H/F112Y | 9.82 ± 0.07 | — | 8.62 ± 0.13 | 67 ± 7 | 9.69 ± 0.13 | — | 8.35 ± 0.10 | 38 ± 3 | 9.75 ± 0.13 | — | 7.88 ± 0.02 | 38 ± 3 | 9.75 ± 0.28 | — | <6.5 | |

[a] For practical reasons, some binding experiments yielding wild-type-like results both in the preliminary tests and dose-response curves were not further investigated; these values are given for illustration purposes.

**Figure 3.4.7.** GTP$_\gamma$S binding and MAP kinase activation assays of wt-CCR5 and the Y108A mutant. A, functional response to RANTES (■), MIP-1β (▲), MIP-1α (∇ ), and MCP2 (◊) of CHO-K1 cells expressing wt-CCR5 and the Y108A mutant in the GTP□S assay (see "Experimental Procedures"). Data are presented as raw counts/min, and all points were run in triplicate (error bars: S.E.). The displayed curves represent a typical experiment out of at least two performed independently. B, immunoblot detection of activated p42/p44 MAP kinase revealed with anti-phosphop42/ p44. CHO-K1 cells expressing either the wt-CCR5 of the Y108A mutant were stimulated with RANTES or MCP-2 at three different concentrations (1, 10, and 100 nM). Detection of p38 by Western blotting was used to control that an equal amount of material was loaded in each lane. A typical experiment of three performed independently is shown.