Structure prediction of GPCRs using piecewise homologs and application to the human CCR5 chemokine receptor: validation through agonist and antagonist docking

Karthik Arumugam^{a,*}, Serge Crouzy^b, Andy Chevigne^a, Carole Seguin-Devaux^a, Jean-Claude Schmit^{a,c}

^a Laboratoire de Rétrovirologie, Centre de Recherche Public-Santé, 84, Val Fleuri, L-1526 Luxembourg, Luxembourg. karthik.arumugam@crp-sante.lu, andy.chevigne@crp-sante.lu, carole.devaux@crp-sante.lu, jc.schmit@crp-sante.lu

^b Commissariat à l'Energie Atomique/CNRS/Université Joseph Fourier, CEA, iRTSV, LCBM, 38054 Grenoble, France. , serge.crouzy@cea.fr

^c Service National des Maladies Infectieuses, Centre Hospitalier Luxembourg, 4, rue E. Barblé, L-1210 Luxembourg, Luxembourg

*** Corresponding author:** Tel: +352 26 970 911, Fax: +352 26970 221,

E-mail: karthik.arumugam@crp-sante.lu, www.crp-sante.lu

Abstract

This article describes the construction and validation of a three-dimensional model of the human CCR5 receptor using multiple homology modeling. A new methodology is presented where we built each secondary structural model of the protein separately from even distantly related homologs of known structure. The reliability of our approach for GPCRs was assessed through the building of the human CXCR4 receptor of known crystal structure. The models are refined using molecular dynamics simulations and energy minimizations using CHARMM, a classical force field for proteins. Finally, docking models of both the natural agonists and antagonists of the receptors CCR5 and CXCR4 are proposed. This study explores the possible binding process of ligands to the receptor cavity of chemokine receptors at molecular and atomic levels. We proposed few crucial residues in receptors binding to agonist/antagonist for further validation through experimental analysis. In particular, our study provides better understanding of the blockage mechanism of the chemokine receptors CCR5 and CXCR4 and may help the identification of new lead compounds for drug development in HIV infection, inflammatory diseases and cancer metastasis.

Keywords: CXCR4, CCR5, pairwise homologs, antagonist, agonist, Molecular dynamics simulation, Docking

Abbreviations:

MD: Molecular dynamics

RMSD: Root Mean Square Deviations

RMSF: Root Mean Square Fluctuations

ABNR: Adopted Basis Newton-Raphson

MMFP: Miscellaneous Mean Field Potential

EEF1: Effective Energy Function 1

TMH: Transmembrane Helix

CXCR4: C-X-C chemokine receptor type 4

CCR5: CC chemokine receptor 5

Introduction

Chemokine receptors play a crucial role in several physiological and pathological processes, including hematopoiesis, leukocyte homing and trafficking, metastasis, and angiogenesis (Murdoch & Finn 2000). This type of receptor belongs to the superfamily of seven Transmembrane (TM) coupled to a G-protein signaling pathway (Murphy 1994). These receptors have complex membrane topologies consisting of an N-terminal region, three extracellular and intracellular loops, and a cytoplasmic C-terminal tail. The CC chemokine receptor 5 (CCR5) and C-X-C chemokine receptor type 4 (CXCR4) are members of this family. The entry of HIV-1 into the target cell is mediated by CD4 as the primary receptor and the chemokine receptors CCR5 and CXCR4 as the major's necessary receptors (Alkhatib et al. 1996; Trkola et al. 1996; Wu et al. 1996) *in vivo*. Macrophage-tropic strains use CCR5, whereas T cell-tropic strains use CXCR4. Dual tropic strains are those HIV-1 isolates that are capable of using both CCR5 and CXCR4 (Kinter et al. 1998). Chemokine's are the largest superfamily of cytokines that regulate the recruitment of various types of leukocytes sites associated with inflammation and many other responses (Zlotnik & Yoshie 2000; Mackay 2001; Moser & Loetscher 2001). CCR5 functions physiologically as a receptor for the leukocyte chemo attractants CCL5 (RANTES), MIP-1 α , and MIP-1 β (Samson, Labbe, Mollereau, Vassart & Parmentier 1996) and CXCL12 (SDF1 α) functions as a potent chemo attractant for the CXCR4 receptor (Gerard & Rollins 2001). Moreover, a number of small molecules have been identified as CCR5 and CXCR4 antagonists that demonstrated potent antiviral effects in cell culture but only the CCR5 antagonist maraviroc (MVC) (Dorr et al. 2005; Ray 2009) has been approved to date in clinical practice for the treatment of patients

infected with R5 strain viruses. AMD3100 (Hendrix et al. 2000) failed in phase I/II clinical trials for HIV infection due to poor bioavailability and toxicities and are currently evaluated in clinical trials for other diseases such as in high grade Glioma (Hendrix CW, JAIDS 2004). These examples have illustrated the therapeutic potential of CCR5 and CXCR4 agonists and antagonists in the treatment and confirmed that these receptors still constitute attractive targets for drug discovery research.

In order to be suitable for a structure-based drug design process, the structure of the target protein must be known to a high level of confidence. This is typically achieved by the X-ray crystallography protocols. Up to now, attaining a high resolution structure is a particularly complicated achievement when working with membrane proteins, whose crystallization does not take place as simply as for aqueous proteins. The lipid environment in the membrane has an obvious effect on the conformation of the membrane protein (Ivetac & Sansom 2008), and hence compared with the large number of known aqueous protein structures, structural details of very few membrane proteins are available. In the lack of an experimentally established crystal structure, one of the most important approaches to overcome this deficiency is the use of computational methods. The computational methods such as homology modeling, molecular dynamics simulation (MD) and docking are powerful tools to construct a reasonable three dimensional (3D) model of the target and to understand the mechanism of action at the molecular level of their agonists and antagonists.

Understanding the blockage mechanism of chemokine receptors especially CCR5 and CXCR4 would allow the identification of new lead compounds for drug development to tackle HIV infection, inflammatory diseases and cancer metastasis. Previously, several studies have been reported on molecular modeling of CCR5(Liu, Shi, Liu & Sun 2004; Fano, Ritchie & Carrieri 2006; Kondru et al. 2008; Shahlaei et al. 2011) and CXCR4(Huang et al. 2003; Trent et al. 2003; Neves, Simões & Sá e Melo 2010). The authors constructed the CCR5 models associated with MIP-1 β , CCL5 and the CXCR4 models associated with CXCL12 to explore the crucial residues for the interaction between receptors and their natural ligands. In another study,(Kondru et al. 2008) the 3D model of CCR5 was constructed to clarify the binding site and selectivity of the antagonist maraviroc (MVC). Arseniev *et al* also explored the conformational characteristics of CCR5 extracellular domain by structural modeling of this receptor(Efremov et al. 1998). In another report, Liu *et al* built 3D models of CCR5 using a homology modeling approach followed by 1 ns MD simulation(Liu et al. 2004). Recently, great achievement has been attained in the field of G-protein coupled receptor (GPCR) drug design because the structures of CXCR4(Wu et al. 2010) and CXCR1(Park et al. 2012) have been resolved. Berchanski *et al*, built a model of the antagonist AMD3100-CXCR4 complex based on the crystal structure, in order to validate the binding site and its selectivity(Berchanski, Kalinkovich, Ludin, Lapidot & Lapidot 2011). Recently, Changdev *et al* used the crystal structure of CXCR4 to generate a homology model of CCR5 and did docking and membrane molecular dynamics simulation for various CCR5 antagonists(Gadhe, Kothandan & Cho 2012). Nevertheless a common drawback shared by the CXCR4 crystal and CXCR1 solution structures is the spatial

arrangement of their N-terminus, that could not be determined which makes it impossible to understand the ligand and receptor interactions mechanism thoroughly (Szpakowska et al. 2012). Several experiments have proved that the N-terminal region and second extracellular loop (ECL2) of GPCRs play a critical role in ligand binding and signal transduction (LaRosa et al. 1992; Gayle et al. 1993; Hebert et al. 1993; Wu et al. 1996; Blanpain et al. 1999; Colvin, Campanella, Manice & Luster 2006; Prado et al. 2007). The N-terminus and ECL2 of receptors are also thought to play an important role in interactions with the HIV-1 envelope glycoprotein (LaRosa et al. 1992; Gayle et al. 1993; Hebert et al. 1993; Wu et al. 1996; Blanpain et al. 1999; Colvin et al. 2006; Prado et al. 2007). Chemokine's are generally believed to interact with their cognate receptors according to a two-step model (Montecarlo & Charo 1996; Crump et al. 1997). The anchoring of the chemokine to the N-terminus of the receptor is followed by the binding of the flexible chemokine N-terminus to the extracellular loops and the transmembrane segments of the receptors which implies that both the N-terminus and ECL2 of the receptor are important in the binding process. We have recently described the importance and the role of N-terminus in CXCR4 by introducing the sulfo tyrosine at the N-terminus region (Szpakowska et al. 2012). Our guided MD simulation suggest that in absence of sulfate groups that the N-terminal tends to collapse forming a condensed structure, whereas tyrosine sulfation creates a repulsive interactions promoting the adoption of an extended structure largely accessible for chemokines binding (Szpakowska et al. 2012).

The aim of this study was to construct a general computational approach to model 3D structure of GPCRs and also other transmembrane protein families which present a low homology with available structures in the Protein databank (PDB); a case study being the

prediction of CXCR4 and CCR5 model structures along with their N-terminus. We then ran molecular dynamics simulations and molecular docking on both the crystal and model structures of CXCR4 to validate our computational approach. Implicit membrane MD simulations were employed to simulate the binding process of ligand(s) to the receptor cavity of protein(s), to discover the binding conformation, and to explore the protein binding mechanism of ligand at molecular and atomic levels. This kind of calculations has a benefit of iteratively tracking the trajectory of conformational changes. Thereby, we think that our method is valid not only to predict the structure of GPCR proteins but also other transmembrane proteins, which are not homologous to any of the available experimental structures. Some of our computational results on CXCR4 and CCR5 were compared to previous site directed mutagenesis) studies and earlier results for validation.

Materials and Methods

Construction of the initial structure of the receptors involved several steps: sequence alignment, secondary structure prediction, assignment of cartesian coordinates for amino acid residues in structurally conserved regions, assignment of coordinates for amino acid residues in structurally non-conserved regions, and refinement of coordinates for the backbone of the TM helices and correction with respect to hydrogen bonding orientations. The various steps involved in building and validation of receptors model are represented as a flow chart in Fig 1.

Fig. 1 Flow chart showing the scheme for multiple homology model building and the validation of receptors.

Sequence analysis of CXCR4 and CCR5

The sequences of human CCR5 (**C-C chemokine receptor type 5**) and CXCR4 (**C-X-C chemokine receptor type 4**) receptors were retrieved from the Uniprot KB/TrEMBL database (<http://www.uniprot.org/>) (accession number **P61073** and **P51681**). The sequences of CCR5 and CXCR4 which consist of 352 amino acids were subjected to NCBI PSI BLAST search (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the Protein Data Bank (PDB) in order to find a template with suitable identity. Obtaining a correct sequence alignment is the cornerstone of success in all protein modeling procedures. Several hits homologous to CCR5 and CXCR4 which include β 2-adrenergic receptor (PDB ID: 2RH1), turkey β 1-adrenergic receptor (PDB ID:2VT4), A2A adenosine receptor (PDB ID: 3EML), bovine rhodopsin receptor (PDB ID: 1U19), and squid rhodopsin (PDB ID: 2Z73) were obtained. After the search, the alignment between the

template and the target sequences was performed using clustalW 2.0 (Thompson, Higgins & Gibson 1994) with default parameters.

Secondary structure prediction

Prediction of transmembrane regions remains a major challenge in the 3D modeling of GPCRs. To overcome this challenge, we used different transmembrane helix prediction servers like TMHMM(Krogh, Larsson, Von Heijne & Sonnhammer 2001), PSIPred(McGuffin, Bryson & Jones 2000), DAS(Cserző, Eisenhaber, Eisenhaber & Simon 2002), HMMTOP(Tusnady & Simon 2001), SOSUI(Hirokawa, Boon-Chieng & Mitaku 1998), TMPred(HOFMANN 1993), Polyphobius(Käll, Krogh & Sonnhammer 2005) and SPLIT(Juretic, Zoranic & Zucic 2002) to locate the transmembrane helices (where do they begin and end) in CCR5 and CXCR4. Most transmembrane prediction servers use evolutionary information from multiple sequence alignments either directly or indirectly and they often predict different sets of sequences as transmembrane helices. The quality of prediction of the servers was validated by predicting the transmembrane regions of the known structure of CXCR4 and CXCR1. As we already knew the transmembrane regions for this protein from its crystal structure, we compared the predicted and experimental results and identified the best server for our purpose which shows closest results to the original crystal structure.

Model building

In previous studies, bovine rhodopsin and $\beta 2$ adrenergic receptor were used to build models of CXCR4 and CCR5 receptors (Fano et al. 2006; Kondru et al. 2008), but the sequence identity between these receptors and bovine rhodopsin and $\beta 2$ adrenergic receptor is less than 25% which makes the models unreliable (See supplementary Table S1).

Consequently we have used multiple homology modeling. As an initial step, each predicted TM domain, ECL, ICL, N- and C-terminal was aligned individually with the sequences of GPCRs of known crystal structure using clustalW and the similarity percentage or cluster score was calculated. The sequences of GPCRs with known experimental structure (high resolution) known as PDB codes 1U19 (Okada et al. 2004), 2RH1 (Cherezov et al. 2007), 2VT4 (Warne et al. 2008), 2Z73 (Murakami & Kouyama 2008) and 3EML (Jaakola et al. 2008) were used for alignment of the CXCR4 and CCR5 domain sequences. The natural agonists bind to the extracellular region or transmembrane cavity of the receptor, therefore we kept the initial 303 residues for CXCR4 and the initial 341 residues for CCR5 and the remaining C-terminal part was truncated. MODELLER version 9v2 (Sali & Blundell 1994) was used to build the multiple homology models of CXCR4 and CCR5. We also constructed CXCR4 models based on bovin rhodopsin (PDB 1U19) or CXCR1 (PDB 2LNL), only, to evaluate our procedure.

Model refinement with CHARMM

The 3D structures of the receptors, obtained through multiple homology modeling were further refined using the molecular modeling program CHARMM (version 35)(Brooks, Bruccoleri, Olafson, Swaminathan & Karplus 2004). Disulfide bonds were introduced between C101 (ECL1) - C178 (ECL2) and C20 (N-terminus) - C269 (ECL3) in the CCR5 receptor and between C109 (ECL1) - C186 (ECL2) and C28 (N-terminus) - C274 (ECL3) in CXCR4. The CCR5 and CXCR4 models were capped with the NTER and CTER standard termini for proteins. They were then relaxed in vacuum by 2000-step energy minimization using the conjugated gradient method until the root-mean-square (RMS) energy gradient was lower than $0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. Then the receptors were further simulated inside an implicit membrane using the implicit solvent method EEF1(Lazaridis & Karplus 2000) for a neutral (zwitterionic) membrane of 30 Å width and the adapted extended atom CHARMM19_EEF1 force field. Drift of the protein in the X-Y plane was avoided with the introduction of a “MMFP” type restraining potential applied on the X and Y components of the center of mass of the TMH bundle with force constant $100 \text{ kcal/mol/\AA}^2$. The systems were then subject to 50 steps steepest descent followed by ABNR minimization down to an energy gradient of $0.1 \text{ kcal/mol/\AA}$. Finally 20 ns MD simulations were run using Langevin dynamics with a time step of 2 fs. Throughout the MD simulations, harmonic restraints with force constant of $5 \text{ kcal/mol/\AA}^2$ were applied on C α atoms of the transmembrane regions of the receptors. No restraints were applied on loop regions. The same MD protocol was followed starting from the crystal structure of CXCR4 in order to evaluate our results. The stabilized

structural conformation of the receptors CXCR4 and CCR5 was further used in computational docking.

Docking of antagonists on receptors

Molecular docking is an efficient technique to predict the potential ligand binding site(s) on the whole protein target. Here, in order to explore the predictability as well as the characteristics of the binding pocket of the developed model and to make the rational design of novel and more selective antagonists of chemokine receptors possible, molecular docking and longtime MD simulations were carried out on CXCR4 and CCR5 complexes with the antagonists AMD3100 and MVC respectively using the Autodock4.2 and CHARMM programs.

The initial structure of AMD3100 (Drugbank: DB06809) and MVC (Drugbank: DB04835) were obtained from the drug databank (Wishart et al. 2006). The AutoDock 4.2 software package (Morris et al. 2009) was employed to dock AMD3100 and MVC on the stabilized model structures of the receptors CXCR4-pred (for the predicted model of CXCR4) and CCR5, respectively. AMD3100 was also docked on the stabilized crystal structure of CXCR4 (referred to as CXCR4-Xray) for comparison. Based on the size of the drug molecules used here, the number of rotatable bonds varied from four to nine. AutoGrid was used to generate grid maps for each ligand. The dimensions of the grid box were set to 92 x 68 x 38 Å and 54 x 104 x 72 Å in the x, y and z axis for AMD3100 and

MVC systems respectively. The same grid box was used for CXCR4-pred and CXCR4-Xray. For Autodock, Kollman united atom charges were used for the receptor protein atoms and Gasteiger charges for the drug atoms.

The conformer search method used was the Lamarckian genetic algorithm (LGA). LGA is a combination of a genetic algorithm with a local search method to perform energy minimization. We chose 256 runs of LGA (default is ten conformers) for each drug molecule. The number of individuals in each population was set to 150. The maximum number of energy evaluations was set to 400,00000. The maximum number of generations was set to 1,000. Throughout the docking, CXCR4 and CCR5 protein structures were kept rigid and the ligands flexible. The lowest binding free energy protein-ligand complex structures (CXCR4-Pred-AMD3100, CXCR4-Xray-AMD3100, and CCR5-MVC) were selected and used as starting conformations for further implicit membrane MD simulations in the presence of antagonists.

Docking of agonists on receptors

In addition to the previous antagonists, the natural agonists CXCL12 and CCL5 were docked on CXCR4 and CCR5, respectively, using the program HADDOCK (High Ambiguity Driven protein-protein DOCKing) ([http:// haddock.science.uu.nl/services/HADDOCK/haddock.php](http://haddock.science.uu.nl/services/HADDOCK/haddock.php))(Dominguez, Boelens & Bonvin 2003). Initial structures of the natural ligands chosen for these dockings were taken from the NMR structures of CXCL12 PDB ID: 2KEE(Veldkamp et al. 2009) and CCL5 PDB ID: 1RTO(Skelton, Aspiras, Ogez & Schall 1995). CXCL12 and CCL5 are two small

proteins of around 70 amino acids structured into 1 α -helix and 3 β -strands and stabilized by the presence of 3 disulfide bonds.

Both structures were energy minimized using CHARMM and the united atom CHARMM19 force field with ABNR minimization down to a gradient of 0.1 kcal/mol/Å. Minimizations were done in vacuum with strong harmonic restraints to the initial structure (force constant: 5 kcal/mol/Å²). From previous studies, it's known that CXCL12 and CCL5 have high probability to bind the N-terminal and ECL2 extracellular region of the receptors. Therefore, we selected a few random residues in this region as active and the surrounding residues as passive to help the HADDOCK docking procedure. All parameters of the program were kept to their default values and the docking solutions were clustered and ranked according to their HADDOCK scores. The complexes with the best HADDOCK score (CXCR4-Pred-CXCL12, CXCR4-Xray-CXCL12, and CCR5-CCL5) were selected. These protein-ligand complexes were further used in implicit solvent MD simulations aimed at testing their stability.

Molecular dynamics simulations of protein- complexes

The receptor-antagonist docking complexes with lowest binding free energy (CXCR4-Pred-AMD3100, CXCR4-Xray-AMD3100, and CCR5-MVC) and the receptor-agonist docking complexes with best HADDOCK scores (CXCR4-Pred-CXCL12, CXCR4-Xray-CXCL12, and CCR5-CCL5) were used as starting structures for MD simulations. The CHARMM compatible topology and parameters for AMD3100 and MVC were obtained from the SWISSDOCK server (Grosdidier, Zoete & Michielin 2011) (Charges are shown

in Fig S1). Again, after initial energy minimizations of the complexes, 20 ns MD simulations were run in an implicit membrane (EEF1) using Langevin dynamics with a time step of 2 fs. The structures were visualized using VMD (Visual Molecular Dynamics)(Humphrey, Dalke & Schulten 1996).

Results

Secondary structure prediction

Table 1a and Table 1b summarize the secondary structure prediction results from the different servers: the number of transmembrane (TM) domains and their location in the CCR5, CXCR1 and CXCR4 sequences. We compared the results of the servers for CXCR1 and CXCR4 with the secondary structural domains seen in the X-ray structures.

Method	No. of TMH	TM1	TM2	TM3	TM4	TM5	TM6	TM7	% good
Polyphobius	7	42-64	81-104	114-135	155-175	197-222	242-262	284-305	66
HMMTOP	7	44-63	80-99	114-132	155-177	204-223	242-261	288-307	58
TMpred	7	47-63	78-96	111-132	155-173	204-223	240-259	283-305	59
DAS	7	41-63	81-96	111-133	156-173	204-223	240-267	291-304	61
TMHMM	7	43-65	78-96	111-132	155-174	199-221	242-264	284-306	65
SPLIT	7	40-65	80-99	107-138	154-176	195-225	239-270	284-306	77
CXCR4-Xray	7	36-63	71-100	104-140	144-174	192-226	230-267	273-302	100
CXCR1-NMR	7	37-67	74-101	107-139	148-174	200-226 ^a	239-268	275-309	100
CXCR1-SPLIT	7	38-68	75-98	108-136	152-173	203-226	242-265	292-308	81

Table 1a. Prediction of Transmembrane regions of CXCR4 by various methodologies.

Line Xray corresponds to the secondary structural domains seen in the CXCR4 X-ray structure. The two last lines give the position in the sequence of CXCR1 of its seven transmembrane helices from the split server and deduced from the crystal structure. Last column gives the percentage of amino acids with correctly predicted secondary structure:

(CXCR4 X-ray 100 % = 229 aa, CXCR1 X-ray 100 % = 211 aa). ^a Helix 5 of CXCR1 is broken at P206.

The server SPLIT predicted the TMHs in best agreement with the original crystal structure with 77% and 81% of correctly predicted amino acids for CXCR4 and CXCR1, respectively. For CXCR4, 56 amino acids have been incorrectly predicted as coil instead of α -helix (no β -strands) and 11 amino acids have been attributed a structure while they should have been in coils. Based on these results, the server SPLIT was selected to predict the transmembrane segments of CCR5 shown in Table 1b. (For CXCR1, 41 amino acids have been incorrectly predicted as coil).

Method	No. of TMHs	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Polyphobius	6	36-56	68-87	107-129	142-166	198-221	234-257	-
HMMTOP	7	36-56	67-86	103-124	143-167	198-217	236-260	291-312
TMpred	7	39-56	68-87	106-124	146-166	198-218	236-256	285-310
DAS	7	34-57	69-86	103-136	143-164	196-218	235-257	294-297
TMHMM	6	34-56	68-90	105-124	144-166	199-221	234-256	-
SPLIT	7	30-57	64-87	101-131	140-165	195-223	233-263	286-301

Table 1b. Prediction of Transmembrane Helical regions of CCR5 by various programs. Results from SPLIT have been retained based on the results of the previous study on CXCR4.

From the SPLIT server prediction results in Table 1b, 185 amino acids were predicted in α -helices to be compared with 187 for CXCR4 (229 real) and 163 for CXCR1 (211 real) in Table 2a. Due to the obvious similarity between the three receptors, we can expect between 42 and 48 more amino acids in α -helix for CCR5. As a consequence, we could

“manually” extend the boundaries of the TMHs of CCR5, for instance by keeping charged and aromatic residues at the boundaries of the helices. But we chose not to do so assuming that we know nothing about the structure of the receptor to be predicted except the SPLIT results. The resulting predicted secondary structure of CCR5 is shown in the alignment of Fig 2.

Fig 2: Sequence alignment of CCR5, CXCR1 and CXCR4. Secondary structure labeling (A for α) corresponds to the α -helical transmembrane regions of CCR5 predicted by the SPLIT server and used in the following model building.

Model building

The individual domains of CXCR4-pred and CCR5-pred were aligned with low homology GPCR structures found in the PDB: 1U19, 2RH1, 2VT4, 2Z73 and 3EML using clustalW and yielding “clustal” scores are displayed in tables 2a and 2b (clustal scores are simply the number of identities between the two sequences divided by the length of the alignment, and represented as a percentage). Table 2a shows that each separate TM domain of CXCR4 (except TM3) has a higher percentage of similarity to the known GPCR structures than the similarity measured on the whole protein (less than 25%, see Table S1). In the case of CCR5 (Table 2b), most of the TM domains share good homologies with the known GPCR structures, sufficient for conducting multiple homology modeling.

CXCR4 Domains	1U19	2RH1	2VT4	2Z73	3EML
N-terminus	8	8	7	<u>9</u>	42*
TM1	16	<u>33</u>	24	32	24
ICL1	<u>40</u>	15	22	16	20
TM2	<u>31</u>	18	9	18	13
ECL1	27	<u>54</u>	45	45	9
TM3	<u>15</u>	10	10	15	10
ICL2	15	23	<u>33</u>	20	25
TM4	20	<u>25</u>	20	20	5
ECL2	9	14	<u>14</u>	4	4
TM5	<u>33</u>	9	4	14	9
ICL3	8	16	12	<u>16</u>	12
TM6	<u>40</u>	35	30	20	35
ECL3	25	28	0	<u>37</u>	12
TM7	25	15	<u>25</u>	15	20
C-terminus	5	14	20	4	8

Table 2a: Comparison of the sequence identity (clustal score) between GPCRs of known crystal structure and CXCR4-pred for separate secondary structural domains. The highest similarity for each domain is underlined and written in bold characters. TM: transmembrane domains, ECL: extracellular loops, ICL: Intracellular loops. TM domains (transmembrane helices) are defined in Fig. 2.

CCR5 Domains	1U19	2RH1	2VT4	2Z73	3EML
N-terminus	26	6	3	3	<u>28</u>
TM1	22	<u>29</u>	28	20	16
ICL1	20	30	<u>33</u>	20	20
TM2	<u>42</u>	28	23	38	14
ECL1	30	<u>45</u>	25	30	9
TM3	13	9	18	<u>26</u>	13
ICL2	17	23	<u>29</u>	29	23
TM4	<u>23</u>	20	16	12	17
ECL2	13	13	3	<u>16</u>	10
TM5	25	35	30	<u>40</u>	25
ICL3	0	17	11	11	17
TM6	16	29	<u>36</u>	33	25
ECL3	16	14	<u>40</u>	25	25
TM7	27	20	30	25	<u>33</u>
C-terminus	12	3	7	9	5

Table 2b: Comparison of the sequence identity score between GPCRs of known crystal structure and CCR5 for separate secondary structural domains.

Since the receptors have much longer N- and C- termini than those of known GPCRs, it is not surprising to find a very low identity in these regions. Moreover, we were mainly interested in identifying ligand interactions with the receptors which normally bind to the extracellular region or the transmembrane regions of the receptors, far from the C-terminus. Therefore, we have deliberately truncated the badly predicted C-terminal region of the receptors. Other regions of the receptors have sufficient sequence identity with other GPCRs except the extracellular loop ECL2. The importance of the extracellular (EC) regions of the receptors in the binding and recognition of receptor ligands makes it necessary to construct a model that includes those regions: they were consequently refined using the loop refinement methods. Table 2c summarizes the domains from the particular GPCRs which were used to build the entire three dimensional coordinates of the CXCR4 and CCR5 models. These models were built domain after domain, each domain having more than 25 % sequence homology (except ECL2) with the structure used to build it.

Target protein	N-ter	TM1	ICL1	TM2	ECL1	TM3	ICL2	TM4	ECL2	TM5	ICL3	TM6	ECL3	TM7
CXCR4	2Z73	2RH1	1U19	1U19	2RH1	1U19	2VT4	2RH1	2VT4	1U19	2Z73	1U19	1U19	2VT4
CCR5	1U19	2RH1	2VT4	1U19	2RH1	2Z73	2Z73	1U19	2Z73	2RH1	2RH1	2VT4	2VT4	3EML

Table 2c: Consensus table showing the different protein used to build multiple homology models of CXCR4 and CCR5 based on the alignment score.

Once the best template-target alignment was obtained, initial models of the human chemokine receptors CXCR4 and CCR5 (truncated at their C-terminal) were built using multiple homology-modeling methods with the MODELLER program and loop refinement for the intra- and extra-cellular loops (ICL and ECL). Among the 20 models produced, the best CXCR4 and CCR5 receptor models were selected based on the lowest MODELLER objective function. To validate our methodology, a model of CXCR4 was also built in a “traditional way” with MODELLER, from a single alignment with the sequence of bovine rhodopsin and CXCR1.

Model refinement with CHARMM

Time evolution of energy during production MD

An equilibrium in CHARMM potential energy vs time was rapidly obtained, (a plot for the CXCR4 and CCR5 simulations is shown in Fig. S2 of the supplementary material section), with low standard deviations of the potential energy lying between 68 and 70 kcal/mol for all the systems. The simulation started from the CXCR4 crystal structure showed the least fluctuation in potential energy. We estimate that energy was correctly equilibrated for all systems after 4 ns MD.

Root mean square deviation vs time

In order to examine conformational variations of the receptors within the implicit membrane environment, the root-mean-square deviation (RMSD) with respect to the starting structure was calculated. The RMSD for backbone atoms of the TM domains of the proteins CXCR4-pred (model using our new methodology), CXCR4-Xray (from X-ray structure), and CCR5-pred are shown in Figure 3 as a function of the simulation time (20 ns molecular dynamics). Overall, the RMSD of backbone atom rose to about 1.5 Å after 1 ns simulations and then leveled off after 2 ns. This indicates that after an initial increase in the magnitude of residue fluctuation, the receptor protein reached an equilibrium state characterized by the RMSD profile. In addition, the RMSD of CXCR4-pred with respect to the X-ray structure shows how different the model is from the experimental structure: the RMSD starts at 3 Å and gradually decreases to 2.6 Å meaning that our predicted model is reasonably close to the X-ray structure and that this model is stable over time. In addition, we also built CXCR4 models based on the structures of bovine rhodopsin (CXCR4-bV) and CXCR1 (CXCR4-CXCR1) using the direct homology modeling method with MODELLER and refined them with CHARMM. The backbone atom coordinate RMSDs between the final models CXCR4-bV and CXCR4-CXCR1 and the CXCR4 experimental structure were 3.13 Å and 5.13 Å, respectively. These values are to be compared with the 2.6 Å found for CXCR4-pred using our new protocol. Our method thus gives a lower RMSD between the model and the X-ray structure and can reliably be used for the prediction of the structure of CCR5 or other GPCRs.

Fig 3 : Backbone atom RMSD of the TM domains vs time for receptors in 20 ns production MD simulations. (Red: CXCR4-Xray; Green: CXCR4-Pred; Blue: CCR5-pred; all with respect to initial coordinates; Pink: CXCR4-bv with respect to the X-ray structure)

Whole protein backbone atom coordinate RMSD from initial structure are shown for completeness in Fig.S2(supplementary information section). As expected, the RMSDs show a marked increase in value, especially for the predicted CXCR4 structure. This higher deviation is due to the inclusion of the very mobile N-terminal region absent in the X-ray structure and of the ICL and ECL. The same observation holds for CCR5 for which the RMSD increases from 1.2 Å (TM domains) to 2.5 Å (whole protein). In summary, the homology models of the receptors have been optimized in an implicit membrane matrix.

After 20 ns dynamics simulation, the conformation of TMHs and loop domains of the receptor models have reached a newly equilibrated status in the matrix.

Conservation of secondary structure

Fig: 4: Secondary structure superimposition between the X-ray structure of CXCR4: CXCR4-Xray (Red) and our predicted model CXCR4-Pred (Yellow). The all-atom superimposition was made with swiss PDB viewer(Johansson, Zoete, Michielin & Guex 2012).

The conformational change between the refined homology model of CXCR4: CXCR4-Pred and the original CXCR4 crystal structure is shown in Fig. 4. First, the arrangement of α -helices in the TMH domains of CXCR4-Pred became more compact, which may be

an effect of the implicit membrane simulation. Second, the C α atoms in the upper portion of TM3, TM5, TM6 and TM7: four critical helices forming the ligand binding pocket of CXCR4-Pred, shifted little inward. This shift yielded a more compact binding pocket for CXCR4 ligands which may be important for the interactions of the ligand with amino acid residues in the binding site. Finally, the ECL and ICL s of the receptor have been rearranged between the crystal and predicted structure. In summary, modeling and dynamics refinement of the receptor in an implicit membrane system have produced an interesting starting structure for molecular docking.

Radius of Gyration:

The radius of gyration: R_g of a protein over time is a measure of its compactness. If a protein is stably folded, it will likely maintain a relatively steady value of R_g . If a protein unfolds, its R_g will change over time. The radii of gyration for receptors CXCR4 crystal structure, CXCR4 predicted structure and CCR5 predicted structure were calculated (and are presented in Fig. S4 of the supplementary information section). The radius of gyration of CXCR4-Xray was close to 21 Å; CXCR4 predicted and CCR5 were little bit less compact with radii of 22 Å and 23 Å respectively. The increase in R_g value for CXCR4 predicted structure compared to X-ray is due to the presence of the N-terminal residues.

Antagonist docking on receptors

In the docking process, all receptor structures were kept rigid and the ligands flexible. The best protein-ligand complex conformations based on the lowest binding free energy (ΔG value) were selected from the top ranked clusters of conformations. The binding free energies of the best complexes were: -10.43 kcal/mol for CXCR4-Xray-AMD3100, -9.27 kcal/mol for CXCR4-Pred-AMD3100 and -10.21 kcal/mol for CCR5-MVC. To further consider the flexibility of the proteins and ligands, implicit solvent MD simulations were conducted for each ligand-protein complex obtained from the docking analysis. After energy minimization and MD simulation with CHARMM of the complexes output from Autodock, the 3D conformations of CXCR4-Pred-AMD3100 and CCR5-MVC were obtained and displayed in Fig 5a and Fig. 5b, respectively. Interaction energies were calculated between each amino acid of receptors and the ligands aiming at identifying the significant binding-site residues in the models. A residue is proposed as a key residue in the ligand binding complex if it binds the ligand with interaction energy lower than -4 kcal/mol. Key residues in the CXCR4-AMD3100 complex are Asp97 (TM2), His113 (TM3), Asp171 (TM4), Trp195 and Val196 (TM5), Tyr116 and Asp262 (TM6), Ser285 and Glu288 (TM7) which agrees with the previously published results(Wong et al. 2008). In the case of the MVC-CCR5 complex, key residues are Tyr37 (TM1), Leu72, Asp76 and Trp86 (TM2), Phe112 (TM3), Trp248 and Asn252 (TM6), Gln280, Glu283, Gly286 and His289 (TM7). Comparing the receptors with and without the ligands revealed important conformational changes, redefining the shape of the binding pocket, creating extra space for ligand binding and optimizing protein-ligand complementarity. In a previous study, a large binding site in the CXCR4 model was found that can be divided into two buried sub pockets defined by TM2, TM3 and TM7 (pocket P1) and TM4, TM5

and TM6 (pocket P2)(Neves et al. 2010). In our predicted model of the complex between CXCR4 and AMD3100 (CXCR4-Pred/AMD3100), the residues Asp171 in TM4, Asp262 in TM6 and Glu288 in TM7 form hydrogen bonds with the cyclam rings, in good agreement with known mutagenesis data and previous homology modeling studies(Kothandan, Gadhe & Cho 2012). AMD3100 binds to pocket P2 establishing a hydrogen bond and charge-charge interactions between one of the cyclam rings and Asp171. The second positively charged cyclam ring binds between Asp262 and Glu288. Most binding pocket residues known to be important for AMD3100 binding(Neves et al. 2010) are within a 4 Å distance. In the case of CCR5 inhibitor MVC a binding pocket was determined based on the previously published mutagenesis studies(Kothandan et al. 2012). The binding pocket is located in the extracellular region and is partly covered by the extracellular loop 2 (ECL2). In our analysis it is composed of conserved residues Tyr37 (TM1), Asp76 and Trp86 (TM2), Tyr108, Phe109 and Phe112 (TM3), Gln194 and Ile198 (TM5), Trp248, Tyr251 and Asn25 (TM6), Gln280, Glu282, Glu283 and Met287 (TM7). This binding pocket is shown in Figure 5b. A literature review suggests that the crucial acidic residue (Glu282) in TM7 of the binding pocket could establish ionic interactions with the tertiary/quaternary nitrogens of inhibitors(Kothandan et al. 2012). In this study, MVC establishes crucial interactions with amino acids in the binding pocket of CCR5: Glu283 (TM7), Tyr37 (TM1), and Tyr108 (TM3) through hydrogen bonds.

Fig. 5a: Detail of a structural model of the CXCR4-AMD3100 complex. The important residues interacting with AMD3100 are highlighted. AMD3100 is shown in red licorice.

Fig. 5b: Detail of a structural model of the CCR5-MVC complex. The important residues in the interaction are displayed in CPK. MVC is shown as red licorice.

Agonist docking on receptors

After the initial round of energy minimizations and MD simulations of the receptors, CXCL12, the natural chemokine of CXCR4 and CCL5 the chemokine associated to CCR5 were docked into the cavity of the receptors identified previously with the docking program HADDOCK. The best HADDOCK scores were: -89.8 +/- 3.5 for the CXCR4-Xray-CXCL12 complex, 131.8 +/- 2.7 for the CXCR4-Pred-CXCL12 complex and -85.1 +/- 9.9 for the CCR5-CCL5 complex. The big difference in HADDOCK scores between CXCR4-Xray and CXCR4-Pred is probably due to the presence of the N-terminal region in the predicted CXCR4 structure. As previous literature studies indicate, the N-terminal plays a crucial role in the binding of the agonist CXCL12. The selected receptor-agonist complexes were energy minimized with CHARMM and molecular dynamics simulations were carried out for the chemokine-receptor systems to relax and optimize binding interactions between the residues of the chemokines and amino acid residues in the binding cavity of the receptor. Details of the chemokine-receptor complex structures obtained after the MD simulations are shown in Fig. 6a and Fig. 6b and a full model of CCL5 interacting with CCR5 in Fig. 6a. From a previous study, CXCL12 binds in the pocket located near the extracellular side of TM domains in between TMHs 3, 5, 6 and 7²². In particular, Glu2, Asp10, Glu14, Glu15, Asp20, Asp22, Glu26, Glu31, and Glu32 in the N-terminal region, Asp97 in ECL1, Glu179, Asp181, Asp182, Asp187 and Asp193 in ECL2, and Glu269, Glu276 and Glu278 in ECL3 have been shown to interact with CXCL12²². From our MD analysis, Glu2, Asp10, Glu14, Glu15, Asp20, Asp22, Glu26, Glu31, Glu32 in the N-terminal region Glu179, Asp182, Arg188, Phe189 in ECL2,

Trp94, Asp97 in TM2 and Asp193, Glu288 in TM7 are the important residues of the CXCR4 receptor which interact with the CXCL12 chemokine residues Phe13, His17, Val18, Asp20, Tyr21, Lys27, Asn45 and Val49. In the above mentioned list, a residue was proposed as a key residue in the ligand binding complex if it binds the ligand with CHARMM interaction energy lower than -10 kcal/mol. The residues from the ECL3 regions show least interaction with CXCL12.

CCR5 responds to the natural chemokine ligand CCL5 and the available site-directed mutagenesis data on chemokine-CCR5 binding and activation suggest that the core domains of the chemokine binds to the extracellular domain of CCR5 while the N-terminal of the chemokine interacts with the transmembrane helix bundle of CCR5, particularly the aromatic cluster in the middle of TM3 (Fano et al. 2006). The negatively charged residues Asp11, Glu18 and Asp95 are critical in chemokine binding (Fano et al. 2006). From our MD analysis with CHARMM, Ser7, Pro8, Tyr10, Asp11 and Glu18 (N-ter), Gln93, Asp95 and Phe96 (ECL1), Thr167, Arg168, Lys171, Tyr176 and Pro183 (ECL2), Gln186, Phe189 and Lys191 (TM5) were identified as the key interacting residues between CCR5 and CCL5 residues, Phe12, Arg17, Lys45 and Arg47.

The mobility of agonists and antagonists bound to their receptor was evaluated through the calculation of coordinate RMSDs from initial positions. CCL5 was the most tightly bound to its receptor with an average backbone RMSD of 0.45 Å compared to 1.05 Å for CXCL12. The antagonists MVC and AMD3100 were less tightly bound with average RMSDs of 2.3 Å and 2.5 Å, respectively from their initial coordinates.

Fig. 6a: Detail of a structural model of the CXCR4-CXCL12 complex. The interacting residues from CXCR4 are displayed in CPK and the residues from CXCL12 are shown as blue licorice. Labels for CXCL12 residues are highlighted in red.

Fig. 6b: Detail of a structural model of the CCR5-CCL5 (RANTES) complex. The interacting residues from CCR5 are displayed in CPK and the residues from CCL5 (RANTES) are shown as blue licorice. Labels for CCL5 residues are highlighted in red.

Fig. 7: Minimum energy model of the complex between CCR5 and CCL5 (RANTES). The secondary structure elements are displayed.

Root mean square fluctuations

Root Mean Square Fluctuations (RMSF) of atom coordinates corresponding to standard deviations of these coordinates around their average value in time were finally calculated. The C α atom coordinate RMSFs averaged over the last 5 ns of the MD runs for CXCR4-Pred alone (with N-terminal), CXCR4-Pred with CXCL12, CCR5 alone (with N-terminal) and CCR5 complexed with CCL5 are plotted in Fig. 8a and Fig. 8b. Logically, the residues in the α -helical regions or the TM domains are the least mobile, with mean RMSFs around 0.2 Å for both receptors. Large fluctuations observed in the N-terminal region and in the ECL2 between TM4 and TM5 of the CXCR4 protein largely reduced in

the presence of CXCL12 agree with the experimental evidence for a strong interaction of this agonist with the N-terminal region and ECL2 of CXCR4. Conversely; increased fluctuations of the ICL2 residues of CXCR4 complexed with CXCL12 are seen which might trigger the binding of the G-protein. The ECL3 of CXCR4 shows higher mobility in the presence of CXCL12. In the case of CCR5, there are large fluctuations in ECL2 and the N- terminal regions largely reduced in the presence of CCL5. However, the C-terminal region shows high fluctuations with or without bound agonist. Apart from that the ICL's and other ECL's are shows reduced fluctuation when it interacts with CCL5.

Fig 8a: Root mean square fluctuations of the C α atom coordinates of CXCR4 averaged over the 5 last ns of 20 ns MD simulations. Red curve for CXCR4 alone and Green curve for CXCR4 complexed with CXCL12.

Fig 8b: Root mean square fluctuations of the C α atom coordinates of CCR5 averaged over the 5 last ns of 20 ns MD simulations. Red curve for CCR5 alone and Green curve for CCR5 complexed with CCL5.

Discussion

CXCR4 and CCR5 are CXC and CC- chemokine receptors that are important in the trafficking of monocytes/macrophages and in the functions of other cell types relevant to disease pathogenesis (Ness, Kunkel & Hogaboam 2006). X-ray crystal structure of CXCR4 and solution NMR structure of CXCR1 were recently obtained, but N-terminal regions were not resolved due to their high intrinsic flexibility. The low sequence similarity between these GPCRs of known structure and the other members of the family makes it difficult to construct valid all-residue 3D models of CC-chemokine receptors using the classical homology modeling methods. In previous works, bovine rhodopsin and β 2 adrenergic receptor were used to construct the 3D structure of CXCR4 and CCR5

with less than 25% homology(Fano, Ritchie & Carrieri 2006; Kondru et al. 2008), thereby reliability of the model was always doubted. To overcome this problem of low homology and hence possibly unreliable models, we have used combinations of all known GPCR crystal structures available to construct our models based on piecewise homolog method and choose the highest similarity protein sequence for each extracellular or intracellular loops and transmembrane domain. We have finally validated our *in silico* methodologies using the CXCR4 crystal structure.

As an initial step, several online transmembrane prediction tools were employed to predict the secondary structure of the proteins of interest. The quality of the prediction results was evaluated with the known structures of CXCR1(Park et al. 2012) and CXCR4(Wu et al. 2010). The retained secondary structure assignment for CCR5 was taken from the server giving best results on CXCR1 and CXCR4. Obviously, a better assignment would have been obtained by simple “copy” of the secondary structures of CXCR1 and CXCR4. Here, we have chosen to propose a more general methodology of transmembrane protein modeling where closely related structures would not be available. The sequence of each individual secondary structure domain of the receptors of interest was aligned with corresponding sequences of GPCRs of known structure (templates) and the clustalW score was calculated. Finally, selected templates were used to construct multiple homology models of CXCR4 and CCR5 using MODELLER. The model structures having the lowest objective function were retained and improved by energy minimization and molecular dynamics simulations using CHARMM. For CXCR4, our method of predicting the structure of the GPCRs by piecing together secondary structure domains gives lower atom coordinate RMSD from the X-ray structure than standard

homology modeling. Moreover in our approach we have constructed the N-terminal regions of the receptors which play a crucial role in ligand binding.

Classical ligands were then docked to the receptor models using AUTODOCK and HADDOCK. The potent antagonists AMD3100 and MVC and agonists CXCL12 and CCL5, were docked into the binding sites of CXCR4 and CCR5 respectively. Throughout the docking procedure, the target proteins were kept rigid and the ligands were flexible. The best protein-ligand complexes based on the binding free energy and HADDOCK scores were selected from the top ranked clusters of conformations. Molecular dynamics simulations provide intimate details on the protein-ligand interactions at both atomic and molecular levels. To further consider the flexibility of the proteins and the ligands, implicit membrane MD simulations were conducted. The calculation of atom coordinate root mean square fluctuations of the complexes proved that the mobility of the N-terminal region and the loops of the receptors, especially ECL2, decreased drastically in presence of the agonists CXCL12 and CCL5.

CXCR4/AMD3100

Previously, Alexander *et al* studied the binding of AMD3100 to the crystal structure of CXCR4. His113, Tyr116, Thr117, Leu120, Arg188, His203, Tyr255 and Ser285 were proved to contribute to AMD3100 binding. In a site directed mutagenesis study, it was shown that Tyr45, Trp94, Asp97, Tyr116, Asp171, Tyr190, Asp262 and Glu288 are critical residues for AMD3100 binding(Dragic et al. 2000). Among these residues, we have seen that Asp171 on TM4, Asp262 on TM6 and Glu288 on TM7 have a high

probability to form hydrogen bonds with AMD3100. Our docking results highlight two additional potentially important residues: Trp195 and Val196 which may participate in the binding of AMD3100. In another study, a large binding site in the CXCR4 model was found that can be divided into two buried sub pockets defined by TM2, TM3 and TM7 (pocket P1) and TM4, TM5 and TM6 (pocket P2)(Trent et al. 2003) was confirmed by our molecular docking and MD simulation analyses.

CCR5/MVC

From the previously published mutagenesis and computational studies(Dragic et al. 2000; Castonguay et al. 2003; Kothandan et al. 2012) the critical binding residues for the CCR5 inhibitors were reported as Tyr37, Trp86, Tyr108, Phe109, Phe112, Gln194, Ile198, Trp248, Tyr251, Gln280, Glu283 and Met287. Our docking analyses highlight five additional potentially important residues: Leu72, Asp76, Asn252, Gly286 and His289. Our computational analysis confirms that the interactions between MVC and CCR5 are mainly electrostatic and hydrophobic in nature. We suggest that mutations of the above mentioned residues will be effective in changing the affinity of MVC for CCR5.

CXCR4/CXCL12

Numerous studies conducted on whole receptors(LaRosa et al. 1992; Gayle et al. 1993; Hebert et al. 1993; Wu et al. 1996; Blanpain et al. 1999; Colvin et al. 2006; Prado et al. 2007) or receptor-derived synthetic peptides have demonstrated that the N-terminal domain and ECL2 of chemokine receptors play an important role in agonist binding. It

was suggested that the chemokine CXCL12 interacts with Glu2, Asp10, Glu14, Glu15, Asp20, Asp22, Glu26, Glu31, and Glu32 in the N-terminal region of CXCR4 and Asp97 in ECL1, Glu179, Asp181, Asp182, Asp187 and Asp193 in ECL2, and Glu269, Glu276 and Glu278 in ECL3(Huang et al. 2003). Our docking analysis comfort these results with the addition of Arg188 and Phe189 (ECL2), Trp94 (TM2) and Glu288 (TM7) which would be important residues of CXCR4 interacting with the CXCL12 chemokine residues: His17, Val18, Tyr21, Val49 and Asp20. Moreover, we have observed that the N-terminal part of CXCL12 chemokine protrudes inside the transmembrane domain binding pocket of CXCR4 formed by residues Trp94 (TM2), Asp97 (TM2), Asp193 (TM5) and Glu288 (TM7).

These findings support the two-step mechanism previously proposed for the interactions between CXCR4 and CXCL12(Huang et al. 2003; Trent et al. 2003; Fano et al. 2006; Neves, Simoes & Sa e Melo 2010): in a first step (A in Fig. 13): interactions occur between the core domain (site 1) of the chemokine and the N-terminal domain of the receptor; the CXCL12 N-terminus highly dynamic is used to search for the binding cavities buried within the TM helices. In step two (B and C in Fig. S5), interactions appear between the flexible N-terminus (site 2) of the chemokine and the extracellular loops as well as the transmembrane segments of the receptor. This interaction would trigger the conformational changes in the CXCR4 TM region inducing G-protein signalling.

In our previous study(Szpakowska et al. 2012) to validate the importance of the N-terminal region of CXCR4, we sulfo tyrosinated the tyrosine residues of the N-terminus and did long term MD simulations over the entire model with or without sulfate groups at

position 7, 12 and 21 which demonstrated that repulsive interactions caused by the negative charges of the sulfate groups prevent the internal collapse of the N-terminal domain thereby maintaining it in an open conformation accessible for ligand binding. But on other hand, in absence of the sulfate groups the N-terminus tends to form a condensed structure. Comparatively the sulfated N-termini CXCR4 has higher and fastest probability to bind the chemokine's due to its negative cluster formation. Thereby suggesting the role of N-termini of the CXCR4 in chemokine binding will be enhanced by mutating the regions of tyrosine.

CCR5/CCL5

Concerning the chemokine CCL5 binding to the receptor CCR5, we identified several residues as key residues for the interaction. In particular the CCR5 residues Asp11, Glu18, Asp95, Arg168, Lys171 and Lys191 would interact with CCL5 residues Lys45, Arg47, Phe12 and Arg17. Again, our results are in very good agreement with previous site directed mutagenesis binding studies using monoclonal antibodies. To date, little information about the structure of chemokine receptor N-termini is available. The chemokine receptor N-termini display net negative charges and their binding to chemokines is proposed to be typically driven by electrostatic interactions (Szpakowska et al. 2012). In our computational analysis, we found, accordingly, that the N-terminal region of receptors fluctuates a lot without the chemokines CCL5 or CXCL12 and shows significant decrease in fluctuation in the presence of the chemokines. This is also true for

some ECL. The important interaction between the N-terminal region of the chemokines and the transmembrane domain of the chemokine receptors are also reflected in our interaction energy results. However, while the N-terminus ensures the predominant role in the initial chemokine binding, other extracellular parts have also been shown to participate in chemokine binding(Szpakowska et al. 2012). From our results, we expect a two-step mechanism for CCL5binding to CCR5 similar to that described for CXCL12 with CXCR4.

Conclusions

We have constructed multiple homology models of the CCR5 receptor based on the available GPCR crystal structures and further refined the models by molecular modeling and molecular dynamics simulations in an implicit membrane environment. In order to validate our new computational approach, we predicted the 3D structure of CXCR4 in the same way as CCR5 and compared this model with the crystal structure of CXCR4. Overall, our best model based on piecewise multiple homology modeling provides better outcome compared to the traditional homology modeling approach.

CCR5 and CXCR4 antagonists and agonists were then docked to the receptor models and, to understand the flexibility of the receptors and their antagonists/agonists, the docked receptor-antagonist complexes were simulated for 20 ns and new specific ligand-receptor interactions were found. In all cases, our results are in good agreement with previously reported results. Through molecular docking and molecular dynamics simulation of the receptor-ligand complexes, we propose some new important amino

acids for further site directed mutagenesis which may unravel the underlying secret of the binding mechanism of chemokine receptors and which may be useful in the rational design of HIV-1 entry blockers. Our computational approach and analysis should be useful to interpret and understand the structure and mechanism of many membrane proteins whose sequence presents a low homology with existing structures in the PDB.

Acknowledgments

This manuscript is supported by the “Fonds National de la Recherche”, Luxembourg, grant: 10/BM/822389 (3D-SPCAD project). The authors thank Dr. Norbert Garnier (CNRS, Orleans, France) for his helpful advice

References

- Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M. & Berger, E. A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science (New York, NY)* 272: 1955.
- Berchanski, A., Kalinkovich, A., Ludin, A., Lapidot, T. & Lapidot, A. (2011). Insights into the mechanism of enhanced mobilization of hematopoietic progenitor cells and release of CXCL12 by a combination of AMD3100 and aminoglycoside–polyarginine conjugates. *FEBS Journal* 278: 4150-4165.
- Blanpain, C., Doranz, B. J., Vakili, J., Rucker, J., Govaerts, C., Baik, S. S. W., Lorthioir, O., Migeotte, I., Libert, F. & Baleux, F. (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. *Journal of Biological Chemistry* 274: 34719-34727.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., Swaminathan, S. & Karplus, M. (2004). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* 4: 187-217.
- Castonguay, L. A., Weng, Y., Adolfsen, W., Di Salvo, J., Kilburn, R., Caldwell, C. G., Daugherty, B. L., Finke, P. E., Hale, J. J. & Lynch, C. L. (2003). Binding of 2-aryl-4-(piperidin-1-yl) butanamines and 1, 3, 4-trisubstituted pyrrolidines to human CCR5: a molecular modeling-guided mutagenesis study of the binding pocket. *Biochemistry* 42: 1544-1550.
- Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I. & Kobilka, B. K. (2007). High-resolution crystal structure of an engineered human 2-adrenergic G protein coupled receptor. *Science Signalling* 318: 1258.
- Colvin, R. A., Campanella, G. S. V., Manice, L. A. & Luster, A. D. (2006). CXCR3 requires tyrosine sulfation for ligand binding and a second extracellular loop arginine residue for ligand-induced chemotaxis. *Molecular and cellular biology* 26: 5838-5849.
- Crump, M. P., Gong, J. H., Loetscher, P., Rajarathnam, K., Amara, A., Arenzana-Seisdedos, F., Virelizier, J. L., Baggiolini, M., Sykes, B. D. & Clark-Lewis, I. (1997). Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *The EMBO journal* 16: 6996-7007.

- Cserző, M., Eisenhaber, F., Eisenhaber, B. & Simon, I. (2002). On filtering false positive transmembrane protein predictions. *Protein engineering* 15: 745-752.
- Dominguez, C., Boelens, R. & Bonvin, A. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical data. *NMR-based docking of protein-protein complexes* 125: 51.
- Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., Rickett, G., Smith-Burchnell, C., Napier, C., Webster, R., Armour, D., Price, D., Stammen, B., Wood, A. & Perros, M. (2005). Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob Agents Chemother* 49: 4721-4732.
- Dragic, T., Trkola, A., Thompson, D. A. D., Cormier, E. G., Kajumo, F. A., Maxwell, E., Lin, S. W., Ying, W., Smith, S. O. & Sakmar, T. P. (2000). A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *Proceedings of the National Academy of Sciences* 97: 5639-5644.
- Efremov, R. G., Legret, F., Vergoten, G., Capron, A., Bahr, G. M. & Arseniev, A. S. (1998). Molecular modeling of HIV-1 coreceptor CCR5 and exploring of conformational space of its extracellular domain in molecular dynamics simulation. *Journal of Biomolecular Structure and Dynamics* 16: 77-90.
- Fano, A., Ritchie, D. & Carrieri, A. (2006). Modeling the structural basis of human CCR5 chemokine receptor function: from homology model building and molecular dynamics validation to agonist and antagonist docking. *Journal of chemical information and modeling* 46: 1223-1235.
- Fano, A., Ritchie, D. W. & Carrieri, A. (2006). Modeling the structural basis of human CCR5 chemokine receptor function: from homology model building and molecular dynamics validation to agonist and antagonist docking. *J Chem Inf Model* 46: 1223-1235.
- Gadhe, C. G., Kothandan, G. & Cho, S. J. (2012). Computational modeling of human coreceptor CCR5 antagonist as a HIV-1 entry inhibitor: using an integrated homology modeling, docking, and membrane molecular dynamics simulation analysis approach.
- Gayle, R. B., Sleath, P. R., Srinivason, S., Birks, C. W., Weerawarna, K., Cerretti, D., Kozlosky, C., Nelson, N., Bos, T. V. & Beckmann, M. (1993). Importance of the amino terminus of the interleukin-8 receptor in ligand interactions. *Journal of Biological Chemistry* 268: 7283-7289.
- Gerard, C. & Rollins, B. J. (2001). Chemokines and disease. *Nat Immunol* 2: 108-115.
- Grosdidier, A., Zoete, V. & Michielin, O. (2011). Fast docking using the CHARMM force field with EADock DSS. *Journal of computational chemistry* 32: 2149-2159.
- Hebert, C., Chuntharapai, A., Smith, M., Colby, T., Kim, J. & Horuk, R. (1993). Partial functional mapping of the human interleukin-8 type A receptor. Identification of a major ligand binding domain. *Journal of Biological Chemistry* 268: 18549-18553.
- Hendrix, C. W., Flexner, C., MacFarland, R. T., Giandomenico, C., Fuchs, E. J., Redpath, E., Bridger, G. & Henson, G. W. (2000). Pharmacokinetics and safety of AMD-3100, a novel antagonist of the CXCR-4 chemokine receptor, in human volunteers. *Antimicrob Agents Chemother* 44: 1667-1673.

- Hirokawa, T., Boon-Chieng, S. & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14: 378-379.
- HOFMANN, K. (1993). TMbase-A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* 374: 166.
- Huang, X., Shen, J., Cui, M., Shen, L., Luo, X., Ling, K., Pei, G., Jiang, H. & Chen, K. (2003). Molecular Dynamics Simulations on SDF-1 α : Binding with CXCR4 Receptor. *Biophysical journal* 84: 171-184.
- Huang, X., Shen, J., Cui, M., Shen, L., Luo, X., Ling, K., Pei, G., Jiang, H. & Chen, K. (2003). Molecular dynamics simulations on SDF-1 α : binding with CXCR4 receptor. *Biophys J* 84: 171-184.
- Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics* 14: 33-38.
- Ivetac, A. & Sansom, M. S. P. (2008). Molecular dynamics simulations and membrane protein structure quality. *European Biophysics Journal* 37: 403-409.
- Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y. T., Lane, J. R., IJzerman, A. P. & Stevens, R. C. (2008). The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science Signalling* 322: 1211.
- Johansson, M. U., Zoete, V., Michielin, O. & Guex, N. (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics* 13: 173.
- Juretic, D., Zoranic, L. & Zucic, D. (2002). Basic charge clusters and predictions of membrane protein topology. *Journal of chemical information and computer sciences* 42: 620-632.
- Käll, L., Krogh, A. & Sonnhammer, E. L. L. (2005). An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* 21: i251-i257.
- Kinter, A., Catanzaro, A., Monaco, J. A., Ruiz, M., Justement, J., Moir, S., Arthos, J., Oliva, A., Ehler, L. & Mizell, S. (1998). CC-chemokines enhance the replication of T-tropic strains of HIV-1 in CD4+ T cells: role of signal transduction. *Proceedings of the National Academy of Sciences* 95: 11880-11885.
- Kondru, R., Zhang, J., Ji, C., Mirzadegan, T., Rotstein, D., Sankuratri, S. & Dioszegi, M. (2008). Molecular interactions of CCR5 with major classes of small-molecule anti-HIV CCR5 antagonists. *Molecular pharmacology* 73: 789-800.
- Kondru, R., Zhang, J., Ji, C., Mirzadegan, T., Rotstein, D., Sankuratri, S. & Dioszegi, M. (2008). Molecular interactions of CCR5 with major classes of small-molecule anti-HIV CCR5 antagonists. *Mol Pharmacol* 73: 789-800.
- Kothandan, G., Gadhe, C. G. & Cho, S. J. (2012). Structural insights from binding poses of CCR2 and CCR5 with clinically important antagonists: a combined in silico study. *PLoS One* 7: e32864.
- Krogh, A., Larsson, B. È., Von Heijne, G. & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 305: 567-580.
- LaRosa, G. J., Thomas, K., Kaufmann, M., 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. *Journal of Biological Chemistry* 267: 25402-25406.
- Lazaridis, T. & Karplus, M. (2000). Effective energy functions for protein structure prediction. *Current opinion in structural biology* 10: 139-145.
- Liu, S., Shi, X., Liu, C. & Sun, Z. (2004). Characterize dynamic conformational space of human CCR5 extracellular domain by molecular modeling and molecular dynamics simulation. *Journal of Molecular Structure: THEOCHEM* 673: 133-143.
- Mackay, C. R. (2001). Chemokines: immunology's high impact factors. *Nat Immunol* 2: 95-101.
- McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404-405.
- Montecarlo, F. S. & Charo, I. F. (1996). The amino-terminal extracellular domain of the MCP-1 receptor, but not the RANTES/MIP-1 α receptor, confers chemokine selectivity. *Journal of Biological Chemistry* 271: 19084-19092.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* 30: 2785-2791.
- Moser, B. & Loetscher, P. (2001). Lymphocyte traffic control by chemokines. *Nat Immunol* 2: 123-128.
- Murakami, M. & Kouyama, T. (2008). Crystal structure of squid rhodopsin. *Nature* 453: 363-367.
- Murdoch, C. & Finn, A. (2000). Chemokine receptors and their role in inflammation and infectious diseases. *Blood* 95: 3032-3043.
- Murphy, P. M. (1994). The molecular biology of leukocyte chemoattractant receptors. *Annu Rev Immunol* 12: 593-633.
- Ness, T. L., Kunkel, S. L. & Hogaboam, C. M. (2006). CCR5 antagonists: the answer to inflammatory disease?
- Neves, M. A., Simoes, S. & Sa e Melo, M. L. (2010). Ligand-guided optimization of CXCR4 homology models for virtual screening using a multiple chemotype approach. *J Comput Aided Mol Des* 24: 1023-1033.
- Neves, M. A. C., Simões, S. & Sá e Melo, M. L. (2010). Ligand-guided optimization of CXCR4 homology models for virtual screening using a multiple chemotype approach. *J Comput Aided Mol Des* 24: 1023-1033.
- Okada, T., Sugihara, M., Bondar, A. N., Elstner, M., Entel, P. & Buss, V. (2004). The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *Journal of molecular biology* 342: 571-583.
- Park, S. H., Das, B. B., Casagrande, F., Tian, Y., Nothnagel, H. J., Chu, M., Kiefer, H., Maier, K., De Angelis, A. A. & Marassi, F. M. (2012). Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature*.
- Park, S. H., Das, B. B., Casagrande, F., Tian, Y., Nothnagel, H. J., Chu, M., Kiefer, H., Maier, K., De Angelis, A. A., Marassi, F. M. & Opella, S. J. (2012). Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491: 779-783.
- Prado, G. N., Suetomi, K., Shumate, D., Maxwell, C., Ravindran, A., Rajarathnam, K. & Navarro, J. (2007). Chemokine signaling specificity: essential role for the N-terminal domain of chemokine receptors. *Biochemistry* 46: 8961-8968.

- Ray, N. (2009). Maraviroc in the treatment of HIV infection. *Drug Des Devel Ther* 2: 151-161.
- Sali, A. & Blundell, T. (1994). Comparative protein modelling by satisfaction of spatial restraints. *Protein Structure by Distance Analysis* 64: C86.
- Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. (1996). Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry* 35: 3362-3367.
- Shahlaei, M., Madadkar-Sobhani, A., Mahnam, K., Fassihi, A., Saghaie, L. & Mansourian, M. (2011). Homology modeling of human CCR5 and analysis of its binding properties through molecular docking and molecular dynamics simulation. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1808: 802-817.
- Skelton, N. J., Aspiras, F., Ogez, J. & Schall, T. J. (1995). Proton NMR assignments and solution conformation of RANTES, a chemokine of the CC type. *Biochemistry* 34: 5329-5342.
- Szpakowska, M., Fievez, V., Arumugan, K., van Nuland, N., Schmit, J. C. & Chevigne, A. (2012). Function, diversity and therapeutic potential of the N-terminal domain of human chemokine receptors. *Biochem Pharmacol* 84: 1366-1380.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22: 4673-4680.
- Trent, J. O., Wang, Z., Murray, J. L., Shao, W., Tamamura, H., Fujii, N. & Peiper, S. C. (2003). Lipid bilayer simulations of CXCR4 with inverse agonists and weak partial agonists. *Journal of Biological Chemistry* 278: 47136-47144.
- Trkola, A., Dragic, T., Arthos, J., Binley, J. M., Olson, W. C., Allaway, G. P., Cheng-Mayer, C., Robinson, J., Maddon, P. J. & Moore, J. P. (1996). CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* 384: 184-187.
- Tusnady, G. E. & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17: 849-850.
- Veldkamp, C. T., Ziarek, J. J., Su, J., Basnet, H., Lennertz, R., Weiner, J. J., Peterson, F. C., Baker, J. E. & Volkman, B. F. (2009). Monomeric structure of the cardioprotective chemokine SDF-1/CXCL12. *Protein Science* 18: 1359-1369.
- Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G. W., Tate, C. G. & Schertler, G. F. X. (2008). Structure of a β 1-adrenergic G-protein-coupled receptor. *Nature* 454: 486-491.
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. & Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* 34: D668-D672.
- Wong, R. S. Y., Bodart, V., Metz, M., Labrecque, J., Bridger, G. & Fricker, S. P. (2008). Comparison of the potential multiple binding modes of bicyclam, monocyclam, and noncyclam small-molecule CXC chemokine receptor 4 inhibitors. *Molecular pharmacology* 74: 1485-1495.
- Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V. &

- Stevens, R. C. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330: 1066-1071.
- Wu, B., Chien, E. Y. T., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P. & Bi, F. C. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science Signalling* 330: 1066.
- Wu, L., Gerard, N. P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A. A., Desjardin, E. & Newman, W. (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* 384: 179-183.
- Wu, L., Ruffing, N., Shi, X., Newman, W., Soler, D., Mackay, C. R. & Qin, S. (1996). Discrete steps in binding and signaling of interleukin-8 with its receptor. *Journal of Biological Chemistry* 271: 31202-31209.
- Zlotnik, A. & Yoshie, O. (2000). Chemokines: A New Classification Review System and Their Role in Immunity. *Immunity* 12: 121-127.