Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard <u>Terms & Conditions</u> and the <u>Ethical guidelines</u> still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/molecularbiosystems

Graphical Abstract for

Insights into the binding modes of CC chemokine receptor 4 (CCR4) inhibitors: A combined approach involving homology modelling, docking, and molecular dynamics simulation studies

Changdev G. Gadhe¹ and Mi-hyun Kim^{1,2,*}

¹Department of Pharmacy, College of Pharmacy, Gachon University, 155 Gaetbeol-ro, Yeonsu-gu, Incheon, Republic of Korea

²Gachon Institute of Pharmaceutical Science, Gachon University, Yeonsu-gu, Incheon, Republic of Korea



Insights into the binding modes of CC chemokine receptor 4 (CCR4) inhibitors: A combined approach involving homology modelling, docking, and molecular dynamics simulation studies

Changdev G. Gadhe^{1,2} and Mi-hyun Kim^{1,2,*}

¹Department of Pharmacy, College of Pharmacy, Gachon University, 155 Gaetbeol-ro, Yeonsu-gu, Incheon, Republic of Korea

²Gachon Institute of Pharmaceutical Science, Gachon University, Yeonsu-gu, Incheon, Republic of Korea

*Corresponding author

Tel.: +82 32 899 6439; Fax: +82 32 899 6039

Email: kmh0515@gachon.ac.kr

Abstract

CC chemokine receptor 4 (CCR4), a G protein-coupled receptor (GPCR), plays a vital role in the progression of asthma, T-cell lymphoma, inflammation, and Alzheimer's disease. To date, the structure of CCR4 has not been determined. Therefore, the nature of the interactions between inhibitor-CCR4 is not well known. In this study, we used the CCR5 as a template to model the structure of CCR4. Docking studies were performed for four naphthalenesulphonamide derivatives and crucial ligand-protein interactions were analysed. Molecular dynamics (MD) simulations of these complexes (100 ns each) were carried out to gain insights into the interactions between ligands-CCR4. MD simulations revealed that the residues identified by the docking were displaced and new residues were inserted near the ligands. Results of a principal component analysis (PCA) suggested that CCR4 unfolds at the extracellular site surrounding the ligands. Our simulations identified crucial residues involved in CCR4 antagonism, which were supported by previous mutational studies. Additionally, we identified Ser3.29, Leu3.33, Ser5.39, Phe6.47, Ile7.35, Thr7.38, Thr7.40, and Ala7.42 as residues that play crucial roles in CCR4 antagonism. Mutational studies will help elucidate the significance of these residues in CCR4 antagonism. An understanding of ligand-CCR4 CCR4 interactions might aid in the design of novel inhibitors.

Keywords: CC chemokine receptor 4, homology modelling, docking, molecular dynamics simulation, binding modes of inhibitors

Introduction

Chemokines are small-molecular-weight (8–10 kDa) proteins that control the cellular trafficking of various leucocytes. They play important roles in angiogenesis, autoimmune responses, cancer progression, embryonic development, host defence mechanisms, inflammation, and surveillance.¹⁻³ Chemokines have been classified into four main subfamilies based on the position of their cysteine residues: C, CC, CXC and CX3C.⁴ Chemokines exert their effects via chemokine receptors present on the cell surface. These receptors belong to the G protein-coupled receptor (GPCR) superfamily. GPCRs exert their biological effects through coupled heterotrimeric G proteins (Gαβγ).⁵ All members of the GPCR family share common structural features. For example, all GPCRs have seven transmembrane (TM) helices that are connected by three extracellular and three intracellular loops of variable lengths. Among the chemokines, only CCR2, CCR5, CXCR1 and CXCR4 have been characterised in previous studies, while the others have remained uncharacterised. The aim of this study was to predict the structure of CCR4 and analyse the interactions between CCR4 and its inhibitors. CCR4 is the sole chemokine receptor for the chemokine ligands, CCL22 and CCL17, and is predominantly expressed by Th2 cells.⁶ CCR4 probably plays a role in regulating the intrathymic movement of thymocytes during T-lymphocyte education and differentiation. CCR4 may also be involved in the maintenance of immune homeostasis in the lungs and in innate immunity.^{7,8}

CCR4 has been implicated in the progression of asthma, inflammation, T-cell lymphoma, and Alzheimer's disease.⁹⁻¹¹ Therefore, CCR4 is an attractive drug target, and various approaches such as inhibition of CCR4 using antibodies and small molecular inhibitors have been studied to prevent disease progression. To this end, the humanised monoclonal antibody, Mogamulizumab (KW-0761), which has anti-inflammatory and antineoplastic activities against CCR4, has been developed. Although numerous antagonists have been developed and characterised in clinical trials by leading pharmaceuticals, none has emerged as an approved drug. Therefore, the search for CCR4 inhibitors to combat the progression of the above-mentioned diseases continues. Progress in the field of CCR4 antagonists has been reviewed in previous reports.^{12, 13}

CCR4 is an important drug target because of its involvement in numerous diseases. However, the unavailability of X-ray- or nuclear magnetic resonance (NMR)-derived structures has hindered the development of drugs against CCR4. The exact mechanism of action of the inhibitors of CCR4 is not clearly understood. In the absence of a three-dimensional (3D) structure, homology modelling is a powerful approach that may be used to generate a reliable model based on a suitable template. Although, the derived homology model is not identical to the template structure, in the absence a crystal structure, it may be used to identify the binding modes of agonists or antagonists. In addition, 3D-protein modelling, along with other computational techniques such as docking, molecular dynamics simulation, mutational analyses, and calculations of the free energy of binding have been useful in the characterization of numerous drug targets.¹⁴⁻¹⁶ Such computational modelling approaches have been adopted to study the structure of CCR4.¹⁷⁻¹⁹ However, to the best of our knowledge, molecular dynamics (MD) simulations of the ligand-CCR4 complex in its natural environment (i.e. lipid environment) over long time scales have not been performed to date. Therefore, the underlying mechanism of the interaction between CCR4 and its ligands in a lipid environment is still

unclear. The aim of this study was to (1) develop a 3D model of human CCR4 based on a recent template structure, (2) identify the active site and dock inhibitors in this site, and (3) perform MD simulations of ligand-CCR4 complexes to identify the residues involved in the interactions.

Experimental

Data sets

Data on CCR4 inhibitors were collected from a previous study.²⁰ The chemical structures of four randomly selected compounds and their biological activities are reported in **Table 1**. The chemical structures of the inhibitors were drawn using the Marvin Sketch plugin of the ChemAxon suite (<u>www.chemaxon.com</u>), and the coordinates were saved in the mol2 format. These molecules were then imported into the Jaguar module²¹ of the Schrödinger package and optimised using the OPLS2005 force field with conjugate gradient minimization until the convergence threshold reached 0.05 kcal/mol or the maximum number of steps exceeded 5,000. More accurate partial atomic charges were calculated at the 6-31G** DFT/B3LYP level using the Jaguar module, and this information was incorporated into the mol2 coordinate file for further analysis.

Three-dimensional (3D) protein modelling

The primary sequence of human CCR4 was retrieved from the UniProt KB database (accession code: P51679). The National Center for Biotechnology Information-Basic Local Alignment Search Tool (NCBI-BLAST)²² was employed to identify a suitable human homolog for the template-based modelling of CCR4. CCR5 was found to have high sequence identity with CCR4. Therefore, the X-ray crystal structure of human CCR5²³ (Protein Database [PDB] code: 4MBS; resolution, 2.71 Å) was identified as the most suitable template.

Because the accurate prediction of TM helices is difficult, we employed nine different servers to predict the TM helices of CCR4—RHYTHM,²⁴ DGPRED,²⁵ MEMSAT3,²⁶ TMPRED,²⁷ TMHMM,²⁸ OCTOPUS,²⁹ DAS-TMfilter,³⁰ TMMOD,³¹ POLYPHOBIUS,³² and DSSP.³³ The results obtained from these servers are listed in **Table 2**. Each residue predicted to be part of a helix by these servers was accorded a score. A residue predicted to be part of a TM helix by all the servers was given a score of 100%. A residue was considered part of a TM helix only if its score was >50% i.e. predicted by \geq 5 servers. We verified our criterion by assigning secondary structures to the template using the DSSP server.

To obtain a 3D model of CCR4, the sequences of CCR4 and the template were aligned using the ClustalW2 program.³⁴ The alignment was checked for correctness by comparing it with the alignment reported in a previous study.³⁵ The residues were numbered according to the method described by Ballesteros and Weinstein.³⁶ 3D-modeling of CCR4 was performed using the Modeller9v6 program.³⁷ A total of 100 models were derived. These models were evaluated on the basis of their discrete optimised protein energy (DOPE) scores and root mean square

deviations (RMSDs) with respect to the C α trace of the template. A final model was selected and validated for phipsi angle distribution, G-factor,³⁸ and non-bonded interaction patterns,³⁹

Molecular docking of inhibitors into the CCR4 model

Molecular docking studies for the four randomly selected inhibitors were carried out using the AutoDock 4^{40} program. Glu7.39, which occurs in TM7, is strongly conserved among chemokines and is known to form electrostatic interactions with antagonists.⁴¹⁻⁴⁴ Therefore, the centre of the binding site was assigned on the coordinates of co-crystal ligand, and the grid also consisted of Glu7.39. The AutoGrid program was used to assign grid dimensions of $50 \times 50 \times 50$ Å along the XYZ directions with a grid spacing of 0.375 Å (roughly quarter the length of the C-C bond). The Lamarckian genetic approach (LGA) was applied with the default docking parameters and one hundred docking solutions were generated and clustered according to RMSD tolerance (1.5 Å).

Molecular dynamics (MD) simulation in lipid bilayers

We carried out MD simulations using the GROMACS 4.6 package,⁴⁵ (installed on 64 bits Linux system containing 16 CPU and CentOS 6.4) applying the GROMOS96 53a6 force field.⁴⁶ Ligand topologies and parameters were developed using the PRODRG server.⁴⁷ All the ligand-CCR4 complexes were immersed in dipalmitoylphosphatidylcholine (DPPC) bilayers while keeping their long axis (z-axis) parallel to the DPPC bilayers. After ensuring that the alpha helices of CCR4 were parallel with DPPC, overlapping lipids were removed using the InflateGRO program. Lipid parameters and topology files were obtained from Tieleman's website (http://moose.bio.ucalgary.ca).⁴⁸ Next, the DPPC head groups were hydrated by adding sufficient solvent molecules on either side of the lipids. In order to make the system electro-neutral, appropriate numbers of solvent molecules were replaced with counter ions (Na⁺ and Cl⁻). A simple point charge (SPC) solvent model was used to represent the solvent (water) molecules. Periodic boundary conditions (PBC) were applied to all the simulated systems. The dimension of the box containing the simulated systems (ligand, CCR4, lipids, solvent, Na⁺, and Cl⁻) were 9.45 × 8.46 × 9.95 nm.

Next, we set-up the actual simulation by minimising whole systems to ensure the removal of steric clashes introduced during the process. The maximum number of steps for minimization was set to 50,000 using a steepest descent algorithm. After system relaxation, a constant number, volume, and temperature (NVT) simulation was performed for 100 ps at a temperature of 323 K and a coupling constant of 0.5 ps. After temperature stability was attained, a constant number, pressure, and temperature (NPT) simulation was performed for 1,000 ps. During this phase of the simulation, the temperature was set to 323 K and the pressure to 1 bar, with coupling constants of 0.5 and 5.0 ps, respectively. During the NVT and NPT ensemble simulations, position restraints were applied to the ligands and CCR4. Long range electrostatic interactions were calculated using the particle mesh Ewald method, and the cut-off for short-range van der Waals was set to 1.2 nm.⁴⁹ The LINCS algorithm was used to constraint all covalent bonds,⁵⁰ and the Berendsen coupling scheme was employed to equilibrate the ensembles.⁵¹ Final production MD simulations were performed for 100 ns each.

The MD simulations were analysed by extracting information from the trajectory files obtained during the simulations. The RMSD, root mean square fluctuation (RMSF), radius of gyration (R_g), solvent accessible surface area (SASA), and hydrogen bonds (H-bonds) were analysed. Molecular images were produced using the open source PyMOL visualization package (http://sourceforge.net/projects/pymol/?source=directory).

Essential dynamics (ED) analysis

ED or principle component analysis (PCA) is a powerful tool for the advanced analysis of MD simulation trajectories. The concerted motion of a protein may be extracted from the numerous MD simulation frames using ED. ED analysis may be performed by constructing and diagonalising a mass-weighted covariance matrix. In this study, ED analysis was performed using the GROMCAS software.⁵² The covariance matrix was constructed by removing the rotational and translation motions from the trajectory, and by analysing atomic fluctuations. ED analysis provides a set of eigenvectors characterised by eigenvalues, which illustrate the direction and amplitude of fluctuation. The first eigenvector represents maximum motion. The first few eigenvectors (5–10%) explain >90% of the fluctuations, and these eigenvectors represent relevant biological motions. In this study, ED analysis was performed on backbone atoms using the trajectory generated by the MD simulation.

Results

Helices prediction and homology model analysis

Nine different secondary structure prediction servers were used to predict the topology of CCR4 (the TM helices, extracellular [ECL] and intracellular [ICL] regions) and the lengths of the secondary structure elements accurately. The nine servers predicted different lengths for the TM region. This may be attributed to the fact that the methods adopted by the servers are based on different assumptions and algorithms. The secondary structures of CCR4, as predicted by the nine servers, are depicted in **Fig. 1**.

Fig. 2 depicts the sequence alignment between the target (CCR4) and the template (CCR5). This alignment was used by the 3D-modelling program, Modeller9v6, to develop 100 models of CCR4. Modeller9v6 correctly transformed the coordinates of the template into the target structure. In the resultant models, every TM helix was composed of at least 20 residues (TM1, 34–68; TM2, 73–100; TM3, 107–140; TM4, 151–175; TM5, 196–231; TM6, 237–272; and TM7, 276–307). This length is compatible with the hydrophobic thickness of the lipid slab. Moreover, the length of the TM helices in the models was in accordance with the results obtained from the secondary structure prediction methods (**Fig. 1**). Disulphide bonds were introduced between Cys29-Cys276 and Cys110^{3.25}-Cys187 in the models. The disulphide bonds have significant roles in maintaining structural integrity at outer domain of the receptor. This structural organization is necessary for natural ligand recognition, their binding and further signalling process.

After confirming the correctness of the TM predictions, we selected one model for further analysis. The model with the lowest molecular probability density function (MolPDF) score, DOPE score, and Ca RMSD of the target

and template structures was selected as the final model. The selected model was validated using the PROCHECK plot, z-score, knowledge based potential energy, and ERRAT score. This model had a C α RMSD of 0.14 Å with respect to the template structure. The Ramachandran plot was used to assess the quality of the model by comparing the distribution of the phi-psi angles of the model with that obtained for a database of high-resolution structures. The objective of this evaluation was to examine the distribution pattern of the phi-psi torsion angles for the protein structure (normal and/or unusual). The Ramachandran plot for the generated model is shown in **Fig. 3**. A majority of the residues (99.3%) occupied the most favoured and additionally allowed regions, while a few residues (0.7%) were present in the generously allowed region. Surprisingly, no residue was present in the disallowed region, indicating that the model was of sufficiently good quality. The non-bonded interaction pattern of the model was assessed using the ERRAT plot, which indicated an overall model quality of 90.87% (**Supplementary Fig. S1**). The template structure had uncharacterised N- and C-terminal regions; the corresponding regions were deleted from the model. The results of model evaluation before and after the MD simulation are listed in **Table 3**.

The ProSA⁵³ web server was used for further model validation. This server calculates the z-score and knowledge-based energy of the model. The z-score is a measure of the quality of the model and provides an estimate of the differences between the random conformational energy distributions and total energies of the two structures (model and template). A z-score value outside the range of the native protein structure indicates a flawed structure. The z-score of the CCR4 model (-3.76) was similar to that of the template structure (-3.05). The total energies of the two structures were calculated from their distance-based pair potentials, and the results were plotted as a function of the amino acid residues (**Supplementary Fig. S2**). All the above validation parameters indicated that our model was of acceptable quality.

Docking analyses of the CCR4 inhibitors

Molecular docking studies for the four randomly selected CCR4 inhibitors were carried out in a predefined binding site using the AutoDock program. The binding pocket for these inhibitors was defined as the extracellular part of the TM (1–3 and 5–7) regions, including a part of ECL2 (same as co-crystal ligand). One hundred poses per ligand were generated and clustered with an RMSD tolerance limit of 1.5 Å. We selected binding poses for the four inhibitors from the most populated cluster, using our knowledge of crucial interacting residues and similar binding area as of co-crystal ligand. The docking summary for the four inhibitors is listed in **Table 4.** The selected docked poses of the inhibitors overlapped the space occupied by the ligand, Maraviroc, in the crystal structure of the CCR5-Maraviroc complex (**Supplementary Fig. S3**). The modes of binding of the inhibitors in the binding cavity of CCR4 are shown in **Figs. 4** and **5**, and the two-dimensional (2D)-representations of the interactions between the inhibitors and CCR4 are shown in **Supplementary Fig. S4** and **55**.

The docked poses of compounds **1** and **2** (which are enantiomers) are shown in **Fig. 4**. The only difference in the binding modes of these two enantiomers was the orientations of the pyrrole and N-benzyl moieties. Both molecules shared the same space in the active site, i.e., in the vicinity of Tyr1.39, Trp2.60, Ser3.29, Tyr3.32, Leu3.33, Phe3.36, Tyr3.37, Gly4.60, Ser5.39, Ile5.43, Trp6.48, Tyr6.51, Glu7.39, Thr7.40, and Phe7.43. However,

compound **2** interacted with a few additional residues, including Leu2.57 and two residues from ECL2 (Ser176 and Thr189). Moreover, the piperidine nitrogen of both inhibitors was H-bonded with the crucial Glu7.39 residue. In addition, the naphthalene moiety of both inhibitors was positioned in the major binding pocket and was surrounded by hydrophobic residues. The phenyl ring was positioned in the minor binding pocket, and it interacted with the surrounding residues via lipophilic interactions. These results suggest that hydrophobic interactions play a major role in ligand binding, whereas the H-bond interactions between the protonated 'N' and the acidic residue (Glu7.39) anchor the ligand in a specific conformation.

The binding pose of compound **3** is shown in **Fig. 5A**. Compound **3**, which is a small ligand, was positioned in the minor binding pocket. It interacted via H-bonds with the conserved Glu7.39 residue. The minor binding pocket for compound **3** was formed by residues from the N-terminus (Lys35) and the TM helices (Leu1.35, Tyr1.39, Trp2.60, Tyr2.63, Ala2.64, Tyr3.32, Asn7.36, Glu7.39, Thr7.40, and Phe7.43). The naphthalene ring interacted with Tyr1.39, Trp2.60, and Tyr2.63 via edge-to-face π stacking interactions and with Leu1.35 and Ala2.64 via hydrophobic interactions.

Compound 4, which is the most potent of the four inhibitors, was docked in the TM region (Fig. 5B). The binding of Compound 4 in the two sub-pockets (major and minor) was mainly stabilised by hydrophobic interactions. The quaternary nitrogen of the piperidine ring formed a H-bond with the acidic Glu7.39 residue. The minor binding pocket comprised residues from TM1, TM2, TM3, and TM7, whereas the major binding pocket was lined by TM3, TM4, TM5, TM6, and TM7.^{54, 55} The aromatic Tyr3.32 and acidic Glu7.39 residues were located at the interface of the major and minor binding pockets, and they protruded inside the pocket. The ligand adopted an L-shaped conformation inside the binding pocket. However, the N-substituted ethyl phenyl moiety was docked into the minor binding pocket (composed of residues Tyr1.39, Leu2.57, Trp2.60, Glu7.39, Thr7.40, and Phe7.43) and was stabilised by strong hydrophobic interactions. On the other side, the naphthalene sulphonamide moiety of compound **4** was located in a hydrophobic pocket formed by residues Tyr3.32, Leu3.33, Phe3.36, Tyr3.37, Gly4.60, Ser5.39, Ile5.43, Trp6.48, and Tyr6.51. The base of this pocket was highly hydrophobic and accommodated the planar naphthalene ring. Residues from ECL2 (Thr189 and Tyr191) interacted with the pyrrolidine moiety of compound **4**. Edge-to-face π stacking interactions were observed between the naphthalene ring and residues Tyr3.32, Phe3.36, Tyr3.37, Trp6.48, and Tyr6.51. π - π stacking interactions were observed between the ethyl phenyl moiety and residues Tyr1.39 and Phe7.43. The ethyl phenyl moiety was sandwiched between these two residues.

MD simulation analysis

Backbone and ligand RMSD analyses

The stability of the MD simulation was measured in terms of deviations and fluctuations from the initial ligand-CCR4 structure. The time evolution of the RMSDs of CCR4 (backbone atoms only) and ligands (all atoms) was monitored as a function of time. The RMSDs of CCR4 and the ligands in the four simulated systems are shown in **Figs. 6A** and **6B**, respectively. The CCR4 backbone showed variable RMSD in different simulated systems. System4 (compound 4-CCR4) took approximately 20,000 ps to equilibrate. After equilibration, the system showed an average RMSD of 0.4 nm until 60,000 ps. The RMSD then increased to 0.5 nm, and finally stabilised at an average value of 0.475 nm for the remainder of the simulation period. System-2 (compound 2-CCR4) achieved equilibrium rapidly until ~10,000 ps, and then more gradually until ~38,000 ps. After equilibration, a plateau was observed for the rest of the simulation time, with an average RMSD of 0.42 nm. Among all the simulated systems, CCR4 displayed the lowest RMSD in system-1 (compound 1-CCR4). In this system, the maximum deviation attained was 0.25 nm, although the average RMSD was maintained at 0.15 nm during the entire simulation period. System-3 (compound 3-CCR4) took ~30,000 ps to equilibrate, and subsequently maintained a plateau with an average RMSD of 0.365 nm for the remainder of the simulation period.

The RMSDs of the four ligands were monitored throughout the simulations and were plotted as time evolution of RMSDs (**Fig. 6B**). The graph revealed that ligand **3** from system-3 showed a drastic change in the RMSD pattern compared with the other ligands. Compound **3** showed an RMSD of 8.4 nm; this may be attributed to its small size with respect to the large binding pocket. Because of the ligand's small size, it was not stable during the simulation and moved within the pocket, resulting in a continuous rise in the RMSD. Therefore, it may be concluded that in order to elicit an inhibitory effect on CCR4, the ligand should bind tightly within the cavity in a stable conformation. Accordingly, compound **3** had the lowest inhibition constant ($K_i \ge 33 \mu M$) among the four simulated ligands.

The RMSD plot of compound **4** (Fig. 6B), the most active compound, revealed that the ligand was bound to CCR4 stably. The RMSD reached a maximum value (0.258 nm) at approximately 50,000 ps. The average deviation was 0.175 nm for first 50,000 ps, and 0.16 nm for the remainder of the simulation period. The RMSD plot suggested that compound **4** was bound tightly in the active site via hydrophilic and hydrophobic interactions with the surrounding residues. This result indicated the structural stability of CCR4 and the importance of long MD simulations.

The RMSD plot for compound **2** is depicted in **Fig. 6B**. The plot revealed that the ligand deviated from its initial conformation. It must be noted here that the RMSD of the ligand fluctuated with time. In the initial stage, it showed an RMSD of ~0.1 nm until ~30,000 ps, but later, it stabilised with an RMSD of ~0.075 nm at ~67,000 ps. However, after this period the RMSD increased suddenly until ~70,000 ps, and attained a maximum value of ~0.26 nm. After 70,000 ps, the RMSD decreased gradually until ~86,000 ps, and then rose gradually during the remainder of the simulation period.

The RMSD plot of the enantiomeric compound **1** (Fig. 6B) revealed that the compound deviated from its initial structure. However, it had an average structure at the end of the simulation. The RMSD value plateaued after \sim 70,000 ps. This observation indicated that, although the simulation for compound 07 had a high standard error, the compound was bound stably to CCR4 at the end of the simulation time.

The temperature and pressure of the simulated systems were monitored as a function of time and were 323K and 1 bar, respectively (**Supplementary Fig. S6**).

Vibrations observed under conditions of equilibrium are not random in nature, but depend on the flexibility of the local structure. Therefore, RMSF may be used to assess the stability and accuracy of a simulation at equilibrium. The RMSFs of the backbone atoms (N, $C\alpha$, and C) of CCR4 in all the simulated systems were calculated from the MD trajectories (Fig. 7A). The last 20,000 ps (when it was assumed that the systems had attained equilibrium) were used to calculate the RMSF values. The results revealed that most of the residues fluctuated by less than 0.15 nm. The RMSF value of only a few residues, which were not part of the helices, exceeded 0.2 nm, indicating that most of the residues were stable. The residues in the intracellular and extracellular loop regions as well as those in the Nand C-terminal regions fluctuated to a greater extent. The highest RMSF values were observed for the residues of ECL2. This may be attributed to the fact that this loop is involved in binding the agonist/antagonist. Previous studies on CCR5 have also reported that ECL2 has the highest fluctuation intensity among the loop regions.¹⁴ The Nterminal (1-39), C-terminal (309-360), intracellular (ICL1, 68-77; ICL2, 134-150; and ICL3, 227-242) and extracellular (ECL1, 99-111; ECL2, 176-206; and ECL3, 268-284) loops of CCR4 are of variable lengths. All the loop regions of CCR4 fluctuated to a greater extent (Fig. 7A.). ECL2 showed maximum fluctuations in system-2 (-0.49 nm). Among the simulated systems, the overall fluctuation of CCR4 was lowest in system 3. This may be attributed to the small size of the ligand, which enabled it to move freely within the binding pocket. As a result, ligand binding did not induce any changes in the overall structure of CCR4. The loop regions of system-3 also had the lowest fluctuation values. Therefore, it may be concluded that a tightly interacting ligand probably changes the overall topology of CCR4, resulting in increased RMSF values. These results are consistent with previous reports of ligand-induced conformational changes in GPCRs.⁵⁶

The R_g for the backbone atoms of CCR4 was calculated and plotted against simulation time (**Fig. 7B**). The R_g value is an estimation of the overall compactness of the polypeptide chain in a simulated system. R_g is defined as the root mean square distance between all the atoms in the polypeptide and the centroid. Calculation of R_g provides an indication of the structural changes that occur during simulations. CCR4 in system-1 (green) had a higher R_g value than those in the other systems (**Fig. 7B**). Initially, the structure of CCR4 was quite compact. However, as the simulation progressed, the structure unfolded, resulting in a high R_g . In contrast, system-3 (blue) showed fewer changes in the structure of CCR4 and maintained an overall R_g of ~2.06 nm. These results were consistent with the low backbone RMSDs observed for CCR4 in system-3 and indicated that binding of ligand **3** did not change the overall topology and structure of CCR4. This observation coincides with the overall low RMSF of CCR4 backbone atoms in system-3 (**Fig. 7A**). However, in the case of compound **4** (system-4), the R_g of CCR4 was low until 40,000 ps, and increased gradually thereafter. It suggests unfolding of CCR4 structure. In the case of the enantiomeric compound **2** (system-2), initially the structure was unfolded. However, after attaining equilibrium, it maintained its compact structure with an average R_g of 2.05 nm.

Potential energy and solvent accessible surface area (SASA) analyses

The potential energies and SASAs of the simulated systems were calculated, and they are shown in Fig. 8. The stability of the simulated systems was evaluated based on their potential energy plots (where the potential energy was plotted as a function of time; **Fig. 8A**). All the systems were stable during the simulation, and had comparable potential energy values. In the case of system-4, the potential energy fluctuated between -9.24e+05 kJ/mol and -9.35e+05 kJ/mol. After equilibration (at ~10,000 ps), the potential energy decreased gradually to an average value of -9.31e+05 kJ/mol. Similarly, the potential energy of system-1 fluctuated between -9.24e+05 kJ/mol and -9.34e+05 kJ/mol, with an average of -9.29e+05 kJ/mol; the potential energy of system-2 fluctuated between -9.24e+05 kJ/mol and -9.35e+05 kJ/mol and -9.35e+05 kJ/mol and -9.35e+05 kJ/mol, with an average of -9.29e+05 kJ/mol; and the potential energy of system-3 fluctuated between -9.24e+05 kJ/mol and -9.34e+05 kJ/mol and -9.34e+05 kJ/mol. These energy profiles indicated that all the simulated systems were stable after the equilibration period.

The SASAs of the simulated systems were monitored to analyse the overall change in the shape of CCR4 during the simulation (Fig. 8B). The SASA plot revealed that the SASA of system-4 (115–135 nm²) increased gradually until the equilibration period and reached a plateau thereafter, with an average SASA of 125 nm². This indicates that after equilibration, the ligand was bound tightly in the binding pocket, rendering the pocket inaccessible to bulk solvent. The SASA of system-3 fluctuated between 114 and 131 nm², with an average value of 122 nm². The SASA of system-3 decreased initially until the equilibration period, and thereafter showed large fluctuations. This indicated that after equilibration, solvent molecules could access the binding pocket. However, the ligand displaced the solvent from the pocket and reduced the overall SASA of the system. This also indicated that ligand 3 moved freely within the binding pocket without interacting with the surrounding residues. The SASA of system-1 reached a plateau after the equilibration period. However, after ~50,000 ps, the SASA decreased before increasing to attain a plateau at a higher level. This observation indicated that the ligand was able to move within the binding pocket. The SASA plot for system-2 revealed that the SASA changed from 117 to 133 nm² during the course of the MD simulation. The SASA rose gradually until 50,000 ps, and then decreased gradually until ~90,000 ps. This indicated that the protein unfolded during the simulation, exposing buried hydrophobic amino acid residues to the solvent. This result also indicated that the ligand reoriented itself in the pocket. Subsequently, CCR4 folded into its final state and the size of the binding pocket reduced in the remaining time.

ED analyses

ED analysis was performed for the backbone atoms of CCR4 using the trajectory and 2D-projections of principal component (PC)1 and PC2 (**Fig. 9**). The top three eigenvectors, which represent maximum variance in simulated systems, are depicted in **Supplementary Fig. S7**. PC1 represented the motion of the extracellular part of CCR4 (in the inhibitor binding area; **Fig. S7**). This suggested an opening of the binding pocket at this region and unfolding of CCR4. In systems-1 and 4 (**Fig. S7**), PC1 showed movement in same direction (arrow); however, the overall shape of the binding pocket at this region was different in the two systems, as is evident from their R_g values (**Fig. 7B**) and SASA (**Fig. 8B**) plots. In system-1, after equilibration, the SASA rose gradually until 50,000 ps, before decreasing and then increasing again until the end of the simulation. This suggested the presence of two

energy basins in the free energy landscape (**Fig. 9A**). The free energy landscape of system-4 (**Fig. 9D**) had only one dominant basin corresponding to the conformation of CCR4. This was consistent with the fact that the SASA plot of system-4 attained a plateau after equilibration. In the case of system-3, two free energy basins were favoured (**Fig. 9C**), indicating that two conformations were dominant in the simulation. These results were consistent with the SASA plot obtained for system-3 (**Fig. 8B**). In addition, both energy basins were located in the folded region, indicating that the ligand did not induce more changes in the conformation of CCR4. This may be attributed to the small size of the ligand and is evident through the comparative movement of PC1 for other systems (**Fig. S7**). For system-2, two dominant energy basins were observed (**Fig. 9B**), indicating two preferred conformational states. This observation was also supported by the SASA plot. These results indicated that the ligand flipped inside the pocket in the middle of the simulation and attained a plateau for the remainder of the simulation period.

Predicted binding mode of the inhibitors after MD simulation

After the MD simulations, the binding modes of the simulated inhibitors were analysed to identify the residues involved in ligand-CCR4 interaction. The binding modes were calculated from the average structure obtained from the simulation. The last 20,000 ps of simulation time for each system were selected after analysing the RMSDs of both the backbone and ligand atoms. The average structure was crude, and to remove any strain introduced in structure during the course of the simulation, energy minimization was performed for each complex structure. The binding modes of the simulated ligands and the interacting residues (within a 4-Å radius) are shown in **Figs. 10** and **11**.

The binding mode of compound 1 is depicted in Fig. 10A. It was observed that after the MD simulation, the ligand adopted a more open conformation. This conformation was different from that observed in the docking studies. In addition, the interacting residues were different in the MD simulation. Residues from TM1, TM4, and TM5, which were identified as interacting residues in the docking studies, disappeared in the MD simulation. The Nbenzyl moiety, which was located in the minor binding pocket (TM1, TM2 and TM3), formed hydrophobic interactions with residues Phe2.53, Trp2.60, Tyr2.63, Cys3.25, Ile3.28, and Ser3.29, in addition to residues from ECL2 (Cys187). The 2-naphthyl moiety was bound in a deep crevice and formed face-to-edge π -stacking interactions with Tyr3.32, Phe3.36, Phe6.47, Trp6.48, and Tyr6.51, and hydrophobic interactions with Thr7.38 and Ala7.42. The ligand adopted a conformation in which the quaternary nitrogen moved away from the crucial Glu7.39 residue. This might explain the lower biological activity ($K_i = 20.2 \mu M$) of this compound compared with its enantiomer (compound 2; $K_i = 4.6 \mu$ M). Moreover, the N-benzylpiperidine moiety moved away from TM7 and TM1, and closer to TM2 and TM3 during the course of the MD simulation. As a result, the ligand did not interact with the crucial Glu7.39 residue. These results suggest that in order to have a higher inhibitory effect, the ligand must closely interact (via electrostatic or salt-bridge interactions) with Glu7.39. The movement of the N-benzylpiperidine can be viewed in **Supplementary movie 1**. Visual molecular dynamics (VMD) simulation package was used to create the animation movies.⁵⁷ The H-bond interactions between compound 1 and CCR4 were plotted against time evolution (Supplementary Fig. S8A and 2D-schematic plot in Fig. S9A).

The refined binding mode of compound **2** is shown in **Fig. 10B**. During the MD simulation, the ligand moved inside the TM region and assumed a compact L-shaped conformation. While the docking studies suggested that the ligand interacted with all the TM helices, the results of the MD simulation indicated that only TM2, 3, and 5–7 interacted with the ligand. The ligand interacted with residues crucial for CCR4 inhibition. Because of its compact L-shaped conformation, the ligand fit easily in the major binding pocket. The quaternary 'N' of the ligand formed a salt bridge with Glu7.39, indicating the importance of this interaction for CCR4 antagonism. Careful scrutiny of the binding mode revealed that the naphthyl ring was displaced from the deep pocket (composed of TM6 and TM7) into the shallow pocket (composed of TM3, TM5, and TM6), and the *N*-benzylpiperidine was situated in the deep pocket formed by residues Phe6.47, Trp6.48, Asn6.52, Ile7.35, Thr7.38, and Glu7.39. Tyr3.32, Trp6.48, and Glu7.39 pointed towards the centre of the pocket. These residues were probably inaccessible to the agonist, and therefore, hindered the activation of CCR4. In spite of the conformational changes observed in the docking and MD simulation studies, the ligand maintained crucial interactions with the residues comprising the binding pocket (Trp2.60, Tyr2.63, Tyr3.32, Ile5.43, Trp6.48, and Glu7.39). The H-bond and 2D-schematic interaction plots are shown in **Supplementary Figs. S8B** and **S9B**, respectively. The process of the ligand reorientation inside CCR4 is shown in **Supplementary movie 2**.

The binding mode of compound **3** is shown in **Fig. 11A**. The ligand moved out of its original position in the binding pocket to interact with Thr177, Tyr179, Lys188, Thr189, Lys190, and Thr191 of ECL2. From previous studies, it is known that this compound has the lowest binding affinity ($K_i > 33 \mu$ M). This may be attributed to its small size and to the fact that it binds loosely in the binding pocket of CCR4. These results suggest that in the design of a suitable inhibitor for CCR4, it is essential that the ligand fits snugly in the binding pocket of the protein and forms stable interactions with the surrounding residues. The H-bond and 2D-interaction plots for compound **3**-CCR4 are shown in **Supplementary Figs. S8C** and **S10A**, respectively. The movement of the ligand from inside-to-outside of the binding pocket was mapped and shown in the **Supplementary movie 3**.

In the case of compound **4** (**Fig. 11B**), the ligand moved from its initial position and into the CCR4 cavity. This emphasised the importance of performing MD simulations after rigid ligand docking. Some of the residues identified in docking study disappeared in the MD simulation, and new residues were introduced in the vicinity of the ligand. Previous studies on other chemokines (CCR2 and CCR5) have revealed that residues from TM1 (Tyr1.39), TM2 (Trp2.60, Tyr2.63), TM3 (Tyr3.33, Phe3.36), TM5 (Ile5.43), TM6 (Trp6.48, Tyr6.51), and TM7 (Glu7.39) are crucial in ligand interactions. In this study, we identified other interacting residues such as Leu1.35, Phe2.53, Ala2.64, Tyr3.37, Phe6.44, Phe6.47, Asn6.52, Thr7.38, Thr7.40, Ala7.42, and His7.45. Mutational studies on these newly identified residues might help elucidate their roles in CCR4 antagonism. Compound **4** mostly interacted with CCR4 through hydrophobic and hydrogen bond interactions. The sulfonyl oxygen of **4** formed two hydrogen bonds with Trp6.48 and Asn6.52. The naphthyl ring was located in a hydrophobic pocket formed by residues Phe2.57, Phe3.36, Phe6.44, Phe6.47, Trp6.48, Tyr6.51, and Ala7.42. The naphthyl ring displaced Trp6.48 from its position and stabilised in the deep crevice by forming face-to-edge π -stacking interactions with the surrounding residues. Compound **4** was anchored to TM7 (Glu7.39) via salt-bridges. The *N*-ethyl phenyl moiety

assumed a relaxed conformation and extended into TM1, 2 and 7. This moiety formed hydrophobic contacts with Leu1.35, Tyr1.39, Trp2.60, Tyr2.63, Ala2.64, and Ala7.42, and face-to-edge π -stacking interactions with Tyr1.39, Trp2.60, and Tyr2.63. The H-bonds and 2D-interaction plots between compound 33-CCR4 are shown in **Supplementary Figs. S8D** and **S10B**, respectively. Compound **4** movement inside CCR4 was mapped and displayed in **Supplementary movie 4**.

In order to support our finding of crucial residues, we calculated short range Lennard-Jones and Coulombic interaction energies between ligands and surrounding 4 Å residues using gromacs utility, and listed in **Table 5**. Simulation trajectories from last 20,000 ps were used to calculate the energies. Obtained results revealed that the Glu7.39 is of high importance in CCR4 antagonism, which is reflected by the Coulombic interaction energy for compound **2** (-42.80 kJ/mol) and **4** (-47.43 kJ/mol). Our finding has been supported by the previous reports on CCR4 and other related chemokines (CCR2 and CCR5), which highlighted Glu7.39 makes an electrostatic interaction with the positive centre of ligands. In our simulations, positive nitrogen of the compounds **2** and **4** was in the vicinity of Glu7.39, which is evident through **Figs. 10B** and **11B**, respectively. In addition, Trp6.48 (-17.99 kJ/mol) and Asn6.52 (-11.90 kJ/mol) shows higher electrostatics interaction energies, suggesting H-bond with **4**, and evident through **Fig. 11B**. However, compound **2** showed an additional electrostatic interactions with Tyr3.32 (-12.94 kJ/mol), Ser5.39 (-18.03 kJ/mol) and Asn6.52 (-18.47 kJ/mol), hints toward formation of H-bond and supported by **Fig.10B**. On the other side, compound **1** does not have good electrostatic interactions with active site residues, but had close van der Waals interactions (**Table 5**).

In the case of compound **3**, highest van der Waals interaction energy (**Table 5**) arises from the ECL2 residues (Thr177, Tyr179, Lys190 and Tyr191). As the small size of ligand allowed it freely move inside binding pocket and finally stabilized at the ECL2. Van der Waals interactions between compound **4** and Trp2.60, Tyr2.63, Tyr3.32, Phe3.36, Phe6.47, Trp6.48, Tyr6.51, Glu7.39 and Thr7.40 indicates strong attractive forces between them. Summation of the total short range Lennard-Jones and Coulombic interaction energies between ligand-CCR4 proposed a trend in energy (**Supplementary Fig. S11**). However, an effect of the solvent molecules was not considered during the calculation, therefore, obtained results may not be exactly accurate.

Discussion

The CCR4 receptor has been identified as an important drug target in various diseases. To date, many antagonists against CCR4 have been developed. However, none has emerged as a drug molecule. Recently, CCR4 was modelled based on the structure of bovine rhodopsin. Subsequently, docking, MD simulation, and virtual screening were performed to identify important binding site residues and lead molecules.¹⁷⁻¹⁹

In the present study, we used the recently reported structure of CCR5 as a template to model the structure of CCR4. CCR5 has a higher sequence identity with CCR4 than bovine rhodopsin and other GPCRs. Numerous models were generated by homology modelling, and the final model was selected based on the C α RMSD, MolPDF, and DOPE scores, and further evaluated based on the Ramachandran plot, ERRAT plot, z-score, and knowledge-

based energy score. Then, four inhibitors were randomly selected and docked into the model to obtain an initial impression of the binding conformation. The binding conformation was selected on the basis of the most populated cluster, knowledge of vital residues, and docking score. To overcome the limitations of the docking program, the protein was assumed to be a rigid moiety, while the ligands were considered flexible. However, the obtained docking results were not conclusive because the ligand-protein interaction is dynamic in nature. Therefore, to analyse the ligand-protein complex system in its natural environment and gain further insights into the mechanism of interaction, the system was immersed in a box containing DPPC lipid bilayers, and an all-atom MD simulation was performed for 100 ns for each complex. A long simulation time was chosen because membrane proteins require a long time to equilibrate. To avoid bias while selecting the binding mode, the average binding modes were calculated for the last 20,000 ps of the simulation.

PCA was performed for the backbone atoms and RMSD matrices were generated. The first two PCs were extracted and reported in terms of the free energy landscape. The PCs suggested that the CCR4s in systems 1, 2, and 4 underwent movement and unfolding. However, system-3 was less affected by ligand-induced unfolding because of the small size of the ligand.

The calculated binding modes for the ligands revealed that the ligands had moved from their original docked positions into the binding pocket during the MD simulation. These results emphasise the importance of MD simulations in the identification of better binding modes of inhibitors. Our simulations identified the following residues as crucial for CCR4 antagonism: Tyr1.39, Trp2.60, Tyr2.63, Tyr3.32, Phe3.36, Ile5.43, Trp6.48, Tyr6.51, Asn6.52, and Glu7.39. These results are consistent with the ligand binding sites identified in previous studies.^{41, 44, 58-60} Each of the four randomly selected ligands interacted differently with CCR4. Compound **4** interacted with TM1-3, 6, and 7; compound **2** with TM2, 3, 5–7, and ECL2; compound **1** with TM2, 3, 6, 7, and ECL2; and compound **3** (least active) with ECL2 only. Inhibitor **3** moved out of the binding pocket during the MD simulation, indicating that an effective inhibitor must have an optimum size for binding and inhibition. Compounds **2** and **4** assumed a compact L-shaped conformation in the binding pocket while maintaining crucial contacts. The enantiomer of compound **2**, compound **1**, assumed a more open conformation and moved away from the crucial Glu7.39, and this might be the reason for its lower binding affinity compared with compound **2**. Compound **3** did not interact with any of the crucial residues listed above and had the lowest affinity. In addition to the residues reported in previous studies, our simulation identified Ser3.29, Leu3.33, Ser5.39, Phe6.47, Ile7.35, Thr7.38, Thr7.40, and Ala7.42, as important residues in CCR4 antagonism. Prospective mutational studies on these residues will help reveal their significance.

Our simulation studies suggests that the electrostatic interaction between Glu7.39 and the piperidine nitrogen of the ligand is crucial for CCR4 antagonism, which is consistent with previous reports.⁴¹ This observation was supported by the Lennard-Jones and Coulombic interaction energies (**Table 5**). All the simulated inhibitors mostly interacted with hydrophobic residues. Therefore, it may be concluded that, for efficient CCR4 antagonism, the ligand must have both hydrophobic and hydrophilic moieties.

The binding modes of compounds **1**, **2**, and **4** revealed that they could occupy both the binding pockets (minor and major), indicating some degree of allosteric inhibition. This might hinder the binding of CCR4 agonists and CCR4 activation. Our results are consistent with the observations reported by Wu *et al.*⁶¹ that small-molecule and cyclic-peptide antagonists bind to both the minor and major binding pockets, and prevent the binding of the agonist. To date, various approaches have been adopted to study GPCRs, such as (1) initial relaxation of the model by MD simulation followed by docking-guided MD simulation,^{62, 63} and (2) docking-guided MD simulation alone.^{14, 64, 65} Both approaches use different protocols and may produce variable results. In the current study, we used the second approach and obtained plausible binding modes for all the simulated inhibitors, with the exception of compound **3**.

Conclusions

In this study, we performed 3D-modelling, docking of inhibitors, and MD simulations of CCR4-inhibitor complexes to identify the residues involved in binding and interaction and to analyse the time-dependent changes in the conformation of the antagonists. Although numerous GPCR structures have been solved to date, we selected the recently reported structure of CCR5 as our template because it shares high sequence similarity with CCR4. However, the possibility of obtaining slightly different binding modes for the inhibitors with other template GPCRs cannot be ruled out. In this study, our aim was to identify important binding site residues and predict the binding modes of the inhibitors. We performed short simulation (400 ns) because of limited resources. Nevertheless, we obtained stable binding modes of the inhibitors, except for compound 3. In MD simulations, compound 3 moved out from its original position to extracellular site and stabilised by ECL2. Small size of the compound 3 and inadequate contacts with the active site drove it to move away from active site, which explains its lowest activity among simulated inhibitors. Our simulation results indicated that the ligands fit snugly in the binding pocket and were stabilised through numerous interactions. However, the interacting residues identified in the docking studies disappeared and new residues were introduced in the vicinity of the ligands during the MD simulation. The PCA results indicated that CCR4 unfolds at the ligand-binding region. Apart from previously reported residues, we identified Ser3.29, Leu3.33, Ser5.39, Phe6.47, Ile7.35, Thr7.38, Thr7.40, and Ala7.42, as residues that might play crucial roles in ligand binding and supported by the calculated Lennard-Jones and Coulombic interaction energies. Mutational studies will help to exploit the significance of these residues in CCR4 antagonism. The results of this study might provide important insights into structure activity relationship analysis and the mechanism of action of CCR4 antagonists. An understanding of the underlying mechanism will aid drug design and development.

Acknowledgments

This study was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF), which is funded by the Ministry of Education, Science, and Technology (No: 2012R1A6A3A04038302).

Conflicts of Interest

The authors have no conflicts of interest to declare.

Bibliographic references and Notes

- 1. C. Gerard and B. J. Rollins, *Nat. Immunol.*, 2001, **2**, 108-115.
- 2. N. Godessart and S. L. Kunkel, Curr. Opin. Immunol., 2001, 13, 670-675.
- 3. A. Zlotnik and O. Yoshie, *Immunity*, 2000, **12**, 121-127.
- 4. F. H. Epstein and A. D. Luster, New Eng. J. Med., 1998, 338, 436-445.
- F. Bachelerie, A. Ben-Baruch, A. M. Burkhardt, C. Combadiere, J. M. Farber, G. J. Graham, R. Horuk, A. H. Sparre-Ulrich, M. Locati and A. D. Luster, *Pharmacol. Rev.*, 2014, 66, 1-79.
- R. Bonecchi, G. Bianchi, P. P. Bordignon, D. D'Ambrosio, R. Lang, A. Borsatti, S. Sozzani, P. Allavena, P. A. Gray and A. Mantovani, *J. Exp. Med.*, 1998, 187, 129-134.
- F. Annunziato, L. Cosmi, G. Galli, C. Beltrame, P. Romagnani, R. Manetti, S. Romagnani and E. Maggi, J. Leu. Biol., 1999, 65, 691-699.
- Y. Chvatchko, A. J. Hoogewerf, A. Meyer, S. Alouani, P. Juillard, R. Buser, F. Conquet, A. E. Proudfoot, T. N. Wells and C. A. Power, *J. Exp. Med.*, 2000, **191**, 1755-1764.
- 9. J. E. Pease and R. Horuk, *Expert Opin. Ther. Pat.*, 2009, **19**, 39-58.
- O. Yoshie, R. Fujisawa, T. Nakayama, H. Harasawa, H. Tago, D. Izawa, K. Hieshima, Y. Tatsumi, K. Matsushima and H. Hasegawa, *Blood*, 2002, 99, 1505-1511.
- Q. Yuan, S. K. Bromley, T. K. Means, K. J. Jones, F. Hayashi, A. K. Bhan and A. D. Luster, *J. Exp. Med.*, 2007, **204**, 1327-1334.
- 12. J. E. Pease and R. Horuk, *Expert Opin. Ther. Pat.*, 2009, **19**, 199-221.
- 13. J. E. Pease and R. Horuk, *Expert Opin. Drug Discov.*, 2014, 9, 467-483.
- 14. C. G. Gadhe, G. Kothandan and S. J. Cho, J. Biomol. Struct. Dyn., 2013, 31, 1251-1276.
- 15. Y.-T. Wang, R. Schilling, R. H. Fink and W. B. Fischer, *Biophy. Chem.*, 2014, **192**, 33-40.
- 16. G. Tiwari and D. Mohanty, *PloS One*, 2013, **8**, e71340.
- M. N. Davies, J. Bayry, E. Z. Tchilian, J. Vani, M. S. Shaila, E. K. Forbes, S. J. Draper, P. C. Beverley, D. F. Tough and D. R. Flower, *PLoS One*, 2009, 4, e8084.
- K. Yokoyama, N. Ishikawa, S. Igarashi, N. Kawano, N. Masuda, K. Hattori, T. Miyazaki, S.-i. Ogino, M. Orita and Y. Matsumoto, *Bioorg. Med. Chem.*, 2008, 16, 7968-7974.
- J. Bayry, E. Z. Tchilian, M. N. Davies, E. K. Forbes, S. J. Draper, S. V. Kaveri, A. V. Hill, M. D. Kazatchkine, P. C. Beverley and D. R. Flower, *Proc. Natl. Acad. Sci.*, 2008, **105**, 10221-10226.
- D. F. Burdi, S. Chi, K. Mattia, C. Harrington, Z. Shi, S. Chen, S. Jacutin-Porte, R. Bennett, K. Carson and W. Yin, *Bioorg. Med. Chem. Lett.*, 2007, 17, 3141-3145.
- A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang and R. A. Friesner, *Int. J. Quant. Chem.*, 2013, 113, 2110-2142.

- 22. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389-3402.
- Q. Tan, Y. Zhu, J. Li, Z. Chen, G. W. Han, I. Kufareva, T. Li, L. Ma, G. Fenalti and J. Li, *Science*, 2013, 341, 1387-1390.
- 24. M. Punta, L. R. Forrest, H. Bigelow, A. Kernytsky, J. Liu and B. Rost, *Methods*, 2007, 41, 460-474.
- T. Hessa, N. M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S. H. White and G. von Heijne, *Nature*, 2007, 450, 1026-1030.
- D. W. Buchan, S. Ward, A. E. Lobley, T. Nugent, K. Bryson and D. T. Jones, *Nucleic Acids Res.*, 2010, 38, W563-W568.
- 27. K. HOFMANN, Biol. Chem. Hoppe-Seyler, 1993, 374, 166.
- 28. A. Krogh, B. Larsson, G. Von Heijne and E. L. Sonnhammer, J. Mol. Biol., 2001, 305, 567-580.
- 29. H. Viklund and A. Elofsson, *Bioinformatics*, 2008, 24, 1662-1668.
- 30. M. Cserzö, F. Eisenhaber, B. Eisenhaber and I. Simon, *Protein Eng.*, 2002, 15, 745-752.
- 31. R. Y. Kahsay, G. Gao and L. Liao, *Bioinformatics*, 2005, 21, 1853-1858.
- 32. L. Käll, A. Krogh and E. L. Sonnhammer, *Bioinformatics*, 2005, 21, i251-i257.
- 33. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577-2637.
- 34. J. D. Thompson, D. G. Higgins and T. J. Gibson, Nucleic Acids Res., 1994, 22, 4673-4680.
- 35. J. M. Baldwin, G. F. Schertler and V. M. Unger, J. Mol. Biol., 1997, 272, 144-164.
- 36. J. A. Ballesteros and H. Weinstein, *Methods Neurosci.*, 1995, 25, 366-428.
- 37. A. Šali and T. L. Blundell, J. Mol. Biol., 1993, 234, 779-815.
- R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, J. App. Crystallogr., 1993, 26, 283-291.
- 39. C. Colovos and T. O. Yeates, *Protein Sci.*, 1993, **2**, 1511-1519.
- 40. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639-1662.
- T. A. Berkhout, F. E. Blaney, A. M. Bridges, D. G. Cooper, I. T. Forbes, A. D. Gribble, P. H. Groot, A. Hardy, R. J. Ife and R. Kaur, *J. Med. Chem.*, 2003, 46, 4070-4086.
- 42. M. M. Rosenkilde and T. W. Schwartz, Curr. Topics Med. Chem., 2006, 6, 1319-1333.
- 43. E. L. Wise, C. Duchesnes, P. C. da Fonseca, R. A. Allen, T. J. Williams and J. E. Pease, *J. Biol. Chem.*, 2007, **282**, 27935-27943.
- T. Dragic, A. Trkola, D. A. Thompson, E. G. Cormier, F. A. Kajumo, E. Maxwell, S. W. Lin, W. Ying, S. O. Smith and T. P. Sakmar, *Proc. Natl. Acad. Sci.*, 2000, **97**, 5639-5644.
- 45. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson and D. van der Spoel, *Bioinformatics*, 2013, btt055.
- 46. C. Oostenbrink, A. Villa, A. E. Mark and W. F. Van Gunsteren, *J. Comput. Chemistry*, 2004, **25**, 1656-1676.

- 47. A. W. Schuttelkopf and D. M. Van Aalten, *Acta Crystallogr. Section D: Biol. Crystallogr.*, 2004, **60**, 1355-1363.
- 48. D. P. Tieleman, J. L. MacCallum, W. L. Ash, C. Kandt, Z. Xu and L. Monticelli, J. Phys: Condensed Matter, 2006, 18, S1221.
- 49. T. Darden, D. York and L. Pedersen, J. Chem. Phys., 1993, 98, 10089-10092.
- 50. B. Hess, J. Chem. Theory Comput., 2008, 4, 116-122.
- H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, J. Chem. Phys., 1984, 81, 3684-3690.
- 52. A. Amadei, A. Linssen and H. J. Berendsen, *Proteins*, 1993, 17, 412-425.
- 53. M. Wiederstein and M. J. Sippl, *Nucleic Acids Res.*, 2007, **35**, W407-W410.
- 54. J. Pease and R. Horuk, J. Med. Chem., 2012, 55, 9363-9392.
- 55. J. S. Surgand, J. Rodrigo, E. Kellenberger and D. Rognan, *Proteins*, 2006, 62, 509-538.
- 56. C. Hoffmann, A. Zürn, M. Bünemann and M. Lohse, Br. J. Pharmacol., 2008, 153, S358-S366.
- 57. W. Humphrey, A. Dalke and K. Schulten, *Journal of molecular graphics*, 1996, 14, 33-38.
- 58. M. G. Paterlini, *Biophys. J.*, 2002, **83**, 3012-3031.
- 59. G. Kothandan, C. G. Gadhe and S. J. Cho, *PloS One*, 2012, 7, e32864.
- 60. M. Nishikawa, K. Takashima, T. Nishi, R. A. Furuta, N. Kanzaki, Y. Yamamoto and J.-i. Fujisawa, *Antimicrob. Agents Chemother.*, 2005, **49**, 4708-4715.
- B. Wu, E. Y. Chien, C. D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells and F. C. Bi, *Science*, 2010, **330**, 1066-1071.
- 62. M. Shahlaei, A. Madadkar-Sobhani, K. Mahnam, A. Fassihi, L. Saghaie and M. Mansourian, *Biochim. Biophys. Acta (BBA)-Biomembranes*, 2011, **1808**, 802-817.
- M. Shahlaei, A. Madadkar-Sobhani, A. Fassihi and L. Saghaie, J. Chem. Inf. Model., 2011, 51, 2717-2730.64. T. Córdova-Sintjago, N. Villa, L. Fang and R. G. Booth, Mol. Phys., 2014, 112, 398-407.
- 65. B. Li, W. Li, P. Du, K. Q. Yu and W. Fu, J. Phys. Chem. B, 2012, 116, 8121-8130.

Fig. 1. CCR4 TM helices were predicted by the (1) RHYTHM, (2) DGPRED, (3) MEMSAT3, (4) TMPRED, (5) TMHMM, (6) OCTOPUS, (7) DAS-TMFILTER, (8) TMMOD and (9) POLYPHOBIUS secondary structure prediction methods. H-score represent the how many methods predicted amino acid to be a part of helix.

Fig. 2. Sequence alignment between template (CCR5) and target (CCR4).

Fig. 3. Ramachandran's plot for the selected CCR4 model. There was no residue in the disallowed region indicates quality of model in terms of phi-psi angle.

Fig. 4. Docked pose analyses of (A) compound **1** and (B) compound **2** inside CCR4 cavity. TM helices are numbered at the top, active site residues were shown by golden stick and compound **1** by violet stick. CCR4 represented by white helices with transparent surface. Residues were numbered according to Ballesteros-Weinstein method. Docked pose of compound **2** was shown by magenta stick and active site residues by yellow stick. H-bond between protonated N and acidic Glu7.39 was represented by the yellow dash line.

Fig. 5. Docked pose analyses of (A) compound **3** and (B) compound **4** inside CCR4 cavity. TM helices are numbered at the top, active site residues were shown by green stick and compound **3** by golden stick. CCR4 represented by white helices with transparent surface. Docked pose of compound **4** was shown by cyan stick and active site residues by the magenta stick. H-bond between protonated N and acidic Glu7.39 was represented by the yellow dash line.

Fig. 6. CCR4 (A) backbone atoms RMSD and (B) ligands all atom RMSD was computed as a function of time for system-1 (green), system-2 (red), system-3 (blue), and system-4 (black).

Fig. 7. (A) RMSF of CCR4 backbone atoms from last 20000 ps trajectory. (B) Rg of CCR4 backbone atoms calculated as a function of time for system-1 (green), system-2 (red), system-3 (blue), and system-4 (black).

Fig. 8. (A) Potential energy plot as time evolution during MD simulation. (B) Solvent accessible surface area (SASA) during simulations calculated from trajectory files for system-1 (green), system-2 (red), system-3 (blue), and system-4 (black).

Fig. 9. Essential dynamics analysis was performed and first two PCs were projected for (A) system-1, (B) system-2, (C) system-3 and (D) system-4. Graph shown in a Gibbs free energy landscape, red represent energetically favorable and blue represent unfavorable.

Fig. 10. Average binding modes of the (A) **1**-CCR4 and (B) **2**-CCR4 after MD simulation. CCR4 represented by the helices and ligands by stick models. Interacting residues with polar hydrogen in 4 Å surroundings were shown by sticks. TM helices were numbered at the top of each helix and H-bond was represented by the yellow dash line.

Fig. 11. Average binding modes of the (A) **3**-CCR4 and (B) **4**-CCR4 after MD simulation. CCR4 represented by the helices and ligands by stick models. Interacting residues with polar hydrogen in 4 Å surroundings were shown by sticks. TM helices were numbered at the top of each helix and H-bonds were represented by the yellow dash lines.



Fig. 1. CCR4 TM helices were predicted by the (1) RHYTHM, (2) DGPRED, (3) MEMSAT3, (4) TMPRED, (5) TMHMM, (6) OCTOPUS, (7) DAS-TMFILTER, (8) TMMOD and (9) POLYPHOBIUS secondary structure prediction methods. H-score represent the how many methods predicted amino acid to be a part of helix. 177x254mm (96 x 96 DPI)

CCR5 CCR4	MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLV 51 MNPTDIADTTIDESIYSNYYLYESIPKPCTKEGIKAFGELFLPPLYSLVFVFGLLGNSVV 60 * :**. *:** *:**************
CCR5 CCR4	ILILINCKRLKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIG 111 VLVLFKYKRLRSMTDVYLLNLAISDLLFVFSLPFWGYYAADQWVFGLGLCKMISWMYLVG 120 :*:*:: ***:****:****:*************
CCR5 CCR4	FFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFASLPGIIFTRSOK 171 FYSGIFFVMLMSIDRYLAIVHAVFSLRARTLTYGVITSLATWSVAVFASLPGFLFSTCYT 180 *:*****::*::*::******:*****:*:****:*:***:*:
CCR5 CCR4	EGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVICYSGILKTLLRCRNEKKRH 231 ERNHTYCKTKYSLNS-TTWKVLSSLEINILGLVIPLGIMLFCYSMIIRTLQHCKNEKK-N 238 * * * **** ******************
CCR5 CCR4	RAVRLIFTIMIVYFLFWAPYNIVLLINTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHCC 291 KAVKMIFAVVVLFLGFWTPYNIVLFLETLVELEVLQDCTFERYLDYAIQATETLAFVHCC 298 :**::**::::::::::::::::::::::::::::::
CCR5 CCR4	INPIIYAFVGEKFRNYLLVFFQKHIAKR-FCKCCSIFQQEAPERASSVYTRSTGEQEISV 350 INPIIYFFLGEKFRKYILQIFKTCRGLFVLQYCGLLQIYSADTPSSSYTQSTMDHDLHD 358 :***** *:*****:*:*:*:*::*:*:*:*:*:*:*:*
CCR5 CCR4	GL 352 AL 360 .*

Fig. 2. Sequence alignment between template (CCR5) and target (CCR4). 172x97mm (96 x 96 DPI)



Fig. 3. Ramachandran's plot for the selected CCR4 model. There was no residue in the disallowed region indicates quality of model in terms of phi-psi angle. 192x187mm (96 x 96 DPI)



Fig. 4. Docked pose analyses of (A) compound 1 and (B) compound 2 inside CCR4 cavity. TM helices are numbered at the top, active site residues were shown by golden stick and compound 1 by violet stick. CCR4 represented by white helices with transparent surface. Residues were numbered according to Ballesteros-Weinstein method. Docked pose of compound 2 was shown by magenta stick and active site residues by yellow stick. H-bond between protonated N and acidic Glu7.39 was represented by the yellow dash line. 280x103mm (96 x 96 DPI)



Fig. 6. CCR4 (A) backbone atoms RMSD and (B) ligands all atom RMSD was computed as a function of time for system-1 (green), system-2 (red), system-3 (blue), and system-4 (black). 142x202mm (96 x 96 DPI)



Fig. 5. Docked pose analyses of (A) compound 3 and (B) compound 4 inside CCR4 cavity. TM helices are numbered at the top, active site residues were shown by green stick and compound 3 by golden stick. CCR4 represented by white helices with transparent surface. Docked pose of compound 4 was shown by cyan stick and active site residues by the magenta stick. H-bond between protonated N and acidic Glu7.39 was represented by the yellow dash line. 280x101mm (96 x 96 DPI)



Fig. 7. (A) RMSF of CCR4 backbone atoms from last 20000 ps trajectory. (B) Rg of CCR4 backbone atoms calculated as a function of time for system-1 (green), system-2 (red), system-3 (blue), and system-4 (black). 271x116mm (96 x 96 DPI)



Fig. 8. (A) Potential energy plot as time evolution during MD simulation. (B) Solvent accessible surface area (SASA) during simulations calculated from trajectory files for system-1 (green), system-2 (red), system-3 (blue), and system-4 (black). 277x131mm (96 x 96 DPI)



Fig. 9. Essential dynamics analysis was performed and first two principal components (PCs were projected for (A) system-1, (B) system-2, (C) system-3 and (D) system-4. Graph shown in a Gibbs free energy landscape, red represent energetically favorable and blue represent unfavorable. 174x197mm (96 x 96 DPI)



Fig. 10. Average binding modes of the (A) 1-CCR4 and (B) 2-CCR4 after MD simulation. CCR4 represented by the helices and ligands by stick models. Interacting residues with polar hydrogen in 4 Å surroundings were shown by sticks. TM helices were numbered at the top of each helix and H-bond was represented by the yellow dash line. 275x129mm (96 x 96 DPI)



Fig. 11. Average binding modes of the (A) 3-CCR4 and (B) 4-CCR4 after MD simulation. CCR4 represented by the helices and ligands by stick models. Interacting residues with polar hydrogen in 4 Å surroundings were shown by sticks. TM helices were numbered at the top of each helix and H-bonds were represented by the yellow dash lines. 273x127mm (96 x 96 DPI)

Compound ID	Chemical structure	Activity K_{i} , (μ M)
1		20.2
2	HZ Z O O O O	4.6
3		>33
4		0.10

Table 1. Structures and biological activities of the CCR4 inhibitors used in this study.

Table 2. CCR4 TM regions predicted by the nine secondary structure prediction methods. UniProtKB length of CCR4 helices was shown along with the template (4MBS_A) secondary structure length predicted by DSSP method. Secondary structure prediction servers predicted at least 20 residues in every TM.

	RHYTH M	dgpred	MEMS	TMPRE	TMHM M	OCTOPU S	DAS- TMfilte	TMmod	Poly-	UniProt KB	4MBS
	101		AIS	D	101	5	r		phoblus	KD	(D551)
TM1	49-66	45-67	41-65	46-64	42-64	42-62	41-65	45-65	45-65	40-67	26-53
TM2	79-96	77-99	75-99	80-101	76-98	77-97	78-95	78-98	77-98	78-98	64-91
TM3	113-133	115-137	109-133	113-132	113-132	107-137	110-144	113-133	112-133	112-133	98-131
TM4	152-176	159-180	154-178	160-179	153-175	154-174	155-173	151-175	151-174	151-175	142-165
TM5	207-227	208-229	207-226	209-227	201-223	193-213	205-224	207-227	206-226	207-226	187-223
TM6	244-265	240-258	241-264	241-264	244-266	243-263	243-267	245-270	242-264	243-267	229-264
TM7	292-309	288-307	279-302	287-308	286-308	285-305	297-304	290-310	285-307	285-308	269-300
TM8	326-343	318-340									

Table 3. Structural e	evaluation of template a	and CCR4 before an	nd after MD simu	lation by various
parameters.				

		Ramachandi	ran's plot (%) allov	Errat (%)	G-factor	Z-score	
	Core	Additional	Generously	Disallowed			
Template (4MBS)	93.6	6.0	0.4	0.0	93.84	0.15	-3.05
CCR4 before MD	94.8	4.5	0.7	0.0	90.88	0.12	-3.76
CCR4After MD (1)	80.7	19	0.4	0.0	93.25	-0.02	-3.03
CCR4After MD (2)	79.6	19	0.7	0.7	91.91	-0.01	-3.27
CCR4After MD (3)	87.7	10.4	1.1	0.7	86.07	-0.01	-3.78
CCR4After MD (4)	87.7	11.2	0.7	0.4	92.70	0.02	-4.12

Parameters	Compound 1	Compound 2	Compound 3	Compound 4
Binding Energy ∆G (Kcal/mol)	-11.04	-11.57	-9.37	-12.68
Docking Energy (Kcal/mol)	-14.64	-15.69	-12.87	-17.5
Intermol_ene	-12.83	-13.36	-10.56	-14.77
Total_int	-1.81	-2.33	-2.31	-2.73
Torsional_ene	1.79	1.79	1.19	2.09
H-bond	Glu7.39	Glu7.39	Glu7.39	Glu7.39

Intermol_ene- Intermolecular energy Vdw_hb_desolv_ene- van der Waals, h-bond, desolvation energy

Ele_ene-electrostatic energy

Total_int- total internal energy

Torsional_ene- torsional energy

-	Compound 1		Comp	Compound 2		Compound 3		Compound 4	
Residues	LJ-SR	Coul-SR	LJ-SR	Coul-SR	LJ-SR	Coul-SR	LJ-SR	Coul-SR	
Leu1.35	-	-	-	-	-	-	-3.485	0.002	
Tyr1.39	-	-	-	-	-	-	-8.559	-1.819	
Phe2.53	-3.078	0.301	-	-	-	-	-5.776	1.211	
Trp2.60	-16.93	-4.422	-7.556	-0.308	-	-	-19.64	-2.881	
Tyr2.63	-11.85	-0.884	-2.485	-0.094	-	-	-11.42	1.061	
Ala2.64	-	-	-	-	-	-	-4.141	0.099	
Cys3.25	-7.241	-2.647	-	-	-	-	-	-	
Ile3.28	-5.508	0.433	-	-	-	-	-	-	
Ser3.29	-7.805	0.585	-	-	-	-	-	-	
Tyr3.32	-14.92	-5.788	-11.74	-12.94	-	-	-12.61	-0.723	
Leu3.33	-9.548	-0.115	-	-	-	-	-2.641	-0.336	
Phe3.36	-14.07	-0.506	-	-	-	-	-21.33	-4.972	
Tyr3.37	-	-	-	-	-	-	-2.794	0.303	
Ser176	-2.56	-0.635	-	-	-	-	-	-	
Thr177	-	-	-	-	-10.02	-7.016	-	-	
Tyr179	-	-	-	-	-17.77	-0.024	-	-	
Cys187	-9.13	1.839	-	-	-	-	-	-	
Lys188	-	-	-	-	-7.766	2.798	-	-	
Thr189	-	-	-2.539	-3.993	-7.860	-0.88	-	-	
Lys190	-	-	-	-	-23.73	-16.42	-	-	
Tyr191	-	-	-	-	-12.16	-3.56	-	-	
Ser5.39	-	-	-3.665	-18.03	-	-	-	-	
Ile5.43	-	-	-13.41	-0.160	-	-	-	-	
Asn5.44	-	-	-5.585	-1.062	-	-	-	-	
Phe6.44	-	-	-	-	-	-	-3.717	0.096	
Phe6.47	-8.01	-0.907	-5.581	-1.228	-	-	-12.21	1.016	
Trp6.48	-18.73	-1.069	-18.73	-2.591	-	-	-19.92	-17.99	
Tyr6.51	-16.68	-1.997	-	-	-	-	-19.19	-0.51	
Asn6.52	-	-	-8.267	-18.47	-	-	-6.785	-11.90	
Ile7.35	-	-	-10.44	-1.551	-	-	-	-	
Thr7.38	-7.705	-0.610	-6.042	-0.138	-	-	-5.935	0.036	
Glu7.39	-14.56	5.415	-8.167	-42.80			-17.74	-47.43	
Thr7.40	-	-	-	-	-	-	-10.20	-1.013	
Ala7.42	-6.148	-0.088	-	-	-	-	-7.423	0.015	
His7.45	-	_	-	-	-	-	-2.923	0.15	

Table 5. Short range Lennard-Jones and Coulombic interaction energies between ligands and 4 Å residues. Last 20000 ps simulation time was used to calculate the energies.

LJ-SR: short range Lennard-Jones interaction energy, Coul-SR: short range Coulombic interaction energy, unit of energies are in kJ/mol. Based on the interaction energies important residues were identified and marked in the **bold** face.