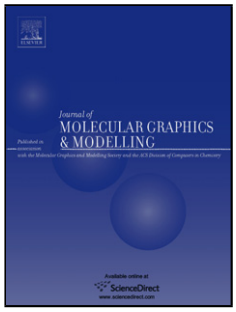


Title: Studying the Binding Interactions of Allosteric Agonists and Antagonists of the CXCR4 Receptor

Author: Jesús M. Planesas Violeta I. Pérez-Nueno José I. Borrell Jordi Teixidó

PII: S1093-3263(15)00082-0
DOI: http://dx.doi.org/doi:10.1016/j.jmgm.2015.05.004
Reference: JMG 6544

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 22-2-2015
Revised date: 6-5-2015
Accepted date: 7-5-2015

Please cite this article as: J.M. Planesas, V.I. Pérez-Nueno, J.I. Borrell, J. Teixidó, Studying the Binding Interactions of Allosteric Agonists and Antagonists of the CXCR4 Receptor, *Journal of Molecular Graphics and Modelling* (2015), http://dx.doi.org/10.1016/j.jmgm.2015.05.004

Studying the Binding Interactions of Allosteric Agonists and Antagonists of the CXCR4 Receptor

Jesús M. Planesas [1], *Violeta I. Pérez-Nueno* [1,2*], *José I. Borrell* [1] *and Jordi Teixidó* [1]

1 Grup d'Enginyeria Molecular, Institut Químic de Sarriá (IQS), Universitat Ramon Llull, Barcelona, Spain

Tel: +34-93-267.20.00. Fax: +34-93-205.62.66. E-mail: violeta.perez@iqs.url.edu, jordi.teixido@iqs.edu

2 Harmonic Pharma, Espace Transfert, 615 rue du Jardin Botanique, 54600 Villers lès Nancy, France

Tel: +33-354 958 604 Fax: +33-383 593 046 E-mail: pereznueno@harmonicpharma.com

HIGHLIGHTS

- Analysis of agonist and antagonist allosteric binding sites in CXCR4.
- Agonist pepducin ATI-2341 binds in the intracellular domain of CXCR4.
- Antagonists AMD11070 and GSK812397 bind in a subsite in the extracellular pocket.

ABSTRACT

Several examples of allosteric modulators of GPCRs have been reported recently in the literature, but understanding their molecular mechanism presents a new challenge for medicinal chemistry. For the specific case of the cellular receptor CXCR4, it is known that pepducins (lipidated fragments of intracellular GPCR loops) such as ATI-2341 modulate CXCR4 activity agonistically via an allosteric mechanism. Moreover, there are also examples of small organic molecules such as AMD11070 and GSK812397 which may also act as allosteric antagonists. However, incomplete knowledge of the ligand-binding sites has hampered a detailed molecular understanding of how these inhibitors work. Here, we attempt to answer this question by analysing the binding interactions between the CXCR4 receptor and the above-mentioned allosteric modulators. We propose two different allosteric binding sites, one located in the intracellular loops 1, 2 and 3 (ICL1, ICL2 and ICL3) which binds the pepducin agonist ATI-2341, and the other at a subsite of the main extracellular orthosteric binding pocket between extracellular loops 1 and 2 and the N-terminus, which binds the antagonists AMD11070 and GSK812397. Allosteric interactions between the CXCR4 and ATI-2341 were predicted by combining different modelling approaches. First, a rotational blind docking search was applied and the best poses were subsequently refined using flexible docking methods and molecular dynamic simulations. For the AMD11070 and GSK812397 antagonists, the entire CXCR4 protein surface was explored by blind docking in order to define the binding region. A second docking analysis by subsites was then performed to refine the allosteric interactions. Finally, we identified the binding residues that appear to be essential for CXCR4 allosteric modulators.

INTRODUCTION

Cellular allosteric regulation is a typical process associated to conformational and functional transitions in protein structures. Allosteric behaviors are found in monomeric proteins, macromolecular complexes, DNA or RNA. Allostery (or allosteric regulation) describes the regulation of the protein function, structure and/or

flexibility induced by the binding of a ligand called allosteric modulator or allosteric effector. Allosteric modulators bind to a site on the structure (allosteric site) located in a different topographical region than the catalytic/active binding site (orthosteric site), but modifying the activity of the catalytic/active binding site[1]. Alteration of the orthosteric ligand affinity is an essential evidence of the presence of allosteric modulators. The result of this alteration is an increment or decrease of the protein-ligand complexation rate[2].

Allosteric regulation is generally due to the non-covalent binding of ions, lipids, proteins, DNA, RNA, or small pharmacological compounds[3,4] to the allosteric sites as activators/agonists, inhibitors/antagonists, allosteric modulators, or other effector types. The receptors when undergo allosteric conformational changes can propagate them to subsequent biochemical processes involving other receptors. Simultaneously, a cascade of allosteric changes generates perturbations or allosteric mutations. These allosteric changes can both cause disease and contribute to development of new therapeutics[5]. Nevertheless, the biocatalytical process of allostery represents an opportunity to regulate abnormal cellular behaviors through the action of ligands playing the role of drug modulators. Chemical allosteric modulators boast several advantages over orthosteric ligands as potential therapeutic agents due to their dormancy in the absence of endogenous-orthosteric activity. Allosteric modulators show greater selectivity as a result of higher sequence divergence in the allosteric site. Also, they show limited positive or negative cooperation imposing a 'ceiling' on the magnitude of their allosteric effect[6]. In recent years, remarkable progress has been made in the optimization, discovery and clinical development of allosteric drugs in kinases, GPCRs and ion channels by the pharmaceutical industry[6].

Regarding the phenomenon of allostery in GPCR proteins[7], there are more than 100 allosteric modulators reported[6], while currently most of the pharmacological compounds whose target is a GPCR act as orthosteric ligands. In these receptors, the known allosteric sites are structurally and evolutionary less conserved than orthosteric sites. Precisely, *Anderson et al.*[8] studied the superposition of 60 different crystal structures verifying this high evolutionary conservation in terms of structure and sequence. In general, allosteric sites in GPCRs can be located in the extracellular side of the receptor, in the transmembrane region, as well as in the intracellular side of the protein. It is observed that depending on the GPCR subfamily, allosteric sites are placed in concrete regions[9]. Thus, chemokine receptors usually have two allosteric binding sites, one located in the transmembrane region and the other in the intracellular region inside the cytoplasm[10]. Attending to these observations a wide range of opportunities to modulate proteins like GPCRs opens up in the forthcoming years.

The mechanism of action of allosteric GPCR modulators has been classified by *Ballesteros et al.*[11] as in-target and off-target, attending to if modulators bind directly to the GPCR. Contrary, if modulators bind to a second

protein which binds to the GPCR forming a heterodimer complex.

Focusing on CXCR4 allostery, *Sachpatzidis et al.*[12] identified two allosteric and agonist peptides (RSVM y ASLW). Both peptides, with 17 residues, do not bind at the same binding site as the small organic antagonist AMD3100 and the antagonist peptide T140 do. Therefore, both peptides exhibit allosteric behavior by binding to CXCR4 in an alternative binding site to the orthosteric site.

Later, *Mosi et al.*[13] proposed the antagonist AMD11070 (Figure 1) as potential inhibitor of the interaction CXCR4-SDF-1, the natural ligand of CXCR4 receptor. AMD11070 would act as a selective noncompetitive inhibitor of isoforms SDF-1α and SDF-1β (or CXCL12a and CXCL12b, respectively) with different modes of action and different IC_{50} values. Hence, AMD11070 would use a mixed mechanism of inhibition without producing allosteric effects to other chemokine receptors like CXCR3, CCR1, CCR2b, CCR4, CCR5 or CCR7[14]. These characteristics, such as the selectivity of AMD11070, are typical from allosteric modulators.

Similarly, *Boggs et al.*[15] synthetized the small organic compound GSK812397 (Figure 1), a bioavailable and oral noncompetitive antagonist of CXCR4, which inhibits HIV entry. This pharmacological compound has a high selectivity and may be able to bind CXCR4 in a distinct and separate site than HIV. Thereby when GSK812397 binds to the receptor may induce conformational changes in CXCR4, inhibiting thermodynamically the entrance of the virus into the cell[16].

Other kind of pharmacological compounds commonly used as allosteric modulators are peptides or their peptidomimetic derivatives. *Oishi et al.*[17] reported a compilation of different peptides and peptidomimetic ligands of CXCR4. Among them, pepducins[18] are a representative family of intracellular regulators of GPCR signaling[19,20]. Pepducins are peptides with 10 to 20 residues, whose protein sequences are derived from the intracellular loops of the targeted GPCRs. Additionally, pepducins have in one of their terminal sequences a lipophilic moiety, like palmitate, myristate or lithocholic substituents, linked by peptide bonds[21]. Those substituents provide pepducins the capacity to penetrate through the lipid bilayer of the cytoplasmic membrane and anchor to it to establish interactions with the intracellular face of the target receptor[22]. Pepducins show the characteristic of allosteric modulators of GPCRs[21], they bind to the receptor stabilizing the conformations of GPCRs between active or inactive states, resulting in positive or negative modulation, respectively. Once stabilized, pepducins may block the ability to undergo further conformational changes, inhibiting the receptor to achieve the active state in the case of negative modulators, or contrarily, enhancing the active state for positive modulators. These blocked conformations could disable the process of receptor dimerization or oligomerization[23]. *Janz et al.*[24] reported three pepducins as allosteric agonists of CXCR4, because they are

competitive among themselves but they do not compete with the antagonist peptide T-140 and neither against the endogenous ligand SDF-1. The non-competitive behavior of pepducins with SDF-1 ligand was corroborated using ATI-2766 (Figure 1), an analogous peptidomimetic derivate of ATI-2341[25]. ATI-2766 has 15 protein residues, one residue less than ATI-2341, and contains a different lipophilic tail, as well as a modified L-Photo-Leucine residue which exhibits photoaffinity in photo-chemical cross-linking experiments[26,27]. The lipophilic tail of ATI-2341 is a palmitic acid substituent while ATI-2766 has two lipophilic groups, 2-undecyloxyethane and the labeling reagent 5-(6)-carboxy-tetramethyl-rhodamine (TAMRA), which is commonly used in fluorescence assays to monitor biomolecular interactions[28].

Figure 1

In this work, we analyze and propose possible allosteric interactions between the CXCR4 receptor and its allosteric modulators. We deal with two different kinds of allosteric modulators attending to the nature of the compounds identified. On the one hand, pepducin modulators whose allosteric behavior has been experimentally verified[24] and establish peptide-protein interactions with CXCR4. On the other hand, the two small organic molecules, AMD11070 and GSK812397, whose allosteric mechanism is still unclear and requires further experimental investigation[13,16].

METHODS

We first used blind docking in order to establish the region with highest probability of involving allosteric interactions between pepducins and the CXCR4 receptor. We then analyzed pepducins binding modes by flexible protein-protein docking. Since many protein-protein or peptide-protein docking protocols work on the basis of residue-residue interactions, and heteroatoms belonging to non-peptidic substituents (HETATM)[29] are not considered part of the biopolymer chain, and neither their associated interactions, we considered convenient to compare the results obtained by flexible docking with results obtained by molecular dynamic simulations (MD).

The representative pepducin chosen to study CXCR4-pepducin allosteric interactions was ATI-2341, as its lipophilic tail is smaller than the ATI-2766, and the palmitate substituent of ATI-2341 has less HETATMs than the ATI-2766 lipophilic substituents ((2-undecyloxyethane) and TAMRA).

Preparation of pepducin ATI-2341[24] consisted of protonation, addition of Gasteiger partial charges[30] and minimization of the structure using CHARMM27[31] force field. Starting from the minimized and energetically stable structure of ATI-2341, a conformational search was applied using the MD protocol *Standard Dynamic*

Cascade implemented in Discovery Studio[32]. The conformational search was performed in an implicit solvent model (GBSW) heating from 50 K to 300 K and using the force field CHARMM22[33] in a time scale of 1 ns.

Regarding the CXCR4 receptor, we used 3OE6 structure, since in our previous work[34] this structure showed the best virtual screening (VS) performance in comparison with the rest of structures published in the Protein Data Bank[35]. PDB 3OE6 at 3.2 Å resolution presents the intracellular loop ICL3 uncompleted and connected to the protein of fusion T4-lysozyme (T4L). T4L is a protein used to facilitate crystallization of CXCR4[36] and other GPCRs. The use of the highly crystallizable T4-lysozyme (T4L) is a common strategy in structure determination of protein membranes like GPCRs[37,38]. For example, T4-lysozyme has been used in the structural determination of β_2 adrenergic receptor (β_2AR)[38,39] and adenosine A_{2A} receptor[40] in addition to CXCR4. This T4-lysozyme protein is fused to the intracellular loop ICL3 of CXCR4 suppressing 33 protein residues from its C-terminal domain[41]. Hence, T4L segments were removed from CXCR4 structure and ICL3 was completed and energetically optimized using Modeller[42].

Peptidomimetic-protein blind Docking using Zdock.

The interactions between CXCR4 protein structure and the pepducin ATI-2341 peptidomimetic structure were first approached by blind docking using ZDock[43]. ZDock is a rigid body docking method based on the Fast Fourier Transform[44] that scans the surface of the protein considered as receptor. The protein considered as ligand is moved rotationally and translationally searching for the best tridimensional shape complementarity between receptor and ligand, and the corresponding electrostatic interactions. ZDock is a recommendable protein-protein docking protocol[45,46,47], mainly when there is no previous information about the residues involved in the interactions between both proteins, neither about the specific region of potential binding sites. In this way, the convenience of using a global systematic search method, like ZDock, in combination with other accurate docking methods has been shown to be a successful strategy to establish protein-protein interactions[48].

The scoring function ZDock Score[45] can be expressed in terms of shape complementarity plus electrostatics and desolvation energy. For the complex CXCR4-ATI-2341, we only consider the shape complementary term, mainly because ZDock was used only as a first approach to search the most favorable binding region. Detailed energy terms were scored later in the flexible docking and MD steps in order to establish accurate interactions.

ZDock was set with a rotation angle step of 6^{o} and without restrictions for the rotational sampling of the pepducin. For each docking run 2000 clustered poses were returned. Poses were grouped and assigned to the same cluster attending to a RMSD cutoff of 5 Å between poses.

Peptide-protein flexible Dockings.

Flexible dockings[49] were performed to refine the pepducin-protein interactions found by blind docking. The best pose clustered in the best binding region returned by the rotational blind docking was used as starting conformation. The use of hybrid analysis of rescoring-refinement has been shown to improve protein-protein docking solutions[50]. For example, *Pierce et al.*[51] showed the improvement of protein docking solutions when a combination of protocols like ZDock and RossetaDock were used.

As pepducins ATI-2341 and ATI-2766 have 105 and 97 rotatable bonds and protein residues compose their active backbone, the use of protein-protein or peptide-protein docking protocols were considered as the most suitable to search for their key interactions.

Accordingly, flexible docking protocols Rosetta FlexPepDock Web Server[52] and HADDOCK Web Server (High Ambiguity Driven protein-protein DOCKing[53]) were used to study and compare the allosteric interactions retrieved among themselves.

FlexPepDock

Rosetta FlexPepDock is a high resolution protocol that refines protein-peptide complexes through a process which enables full flexibility of the peptide and side-chain flexibility of the protein receptor[54]. FlexPepDock web server is mainly orientated to peptide-protein dockings and classifies the complexes predicted by its Rosetta energy score. The search can be performed in low resolution or in high-resolution phase and, consequently, the energy score calculation varies depending on it. The high resolution phase includes all atoms and is controlled by Van der Waals interaction energy, Gaussian solvation, electrostatic energy and orientation-dependent hydrogen bonding potential[50].

HADDOCK

HADDOCK[55] is a high ambiguity driven protein-protein docking protocol used, essentially, when some information is available regarding the interaction between both proteins[48]. This allows guiding the protein conformational search and, hence, the web server protocol[53] requires for each protein to specify a minimum number of residues involved in these interactions. HADDOCK uses a knowledge based method for driving the flexible docking which combines conformational selection and induced fit mechanisms[56].

Peptidomimetic-protein MD.

MD simulations have been used in multiple cases[57,58,59,60] to study protein-ligand interactions as alternative approximations to flexible docking protocols, mainly when reduced sets are tested[61]. In proteins[62], the possibility to design specific force fields to simulate protein motions is also considered as an advantage with respect to flexible docking methods.

Protein-protein flexible docking protocols typically do not incorporate interactions involving organic substituents whose nature is non-peptidic. As pepducins are peptidomimetic compounds whose lipophilic tail is a non-peptide substituent, it was considered relevant to study the CXCR4-ATI-2341 interactions using MD simulations. Comparison with flexible docking results was performed to check if the lipophilic palmitate interacts or clashes with CXCR4 or if modifies the interactions found by the flexible docking protocols.

MD simulations were performed using the *Standard Dynamic Cascade* protocol implemented in Discovery Studio[32] in order to calculate motions and trajectories of the pepducin-CXCR4 complex. The *Standard Dynamic Cascade* includes two minimization steps, a heating followed by an equilibration MD step, and finally the production MD step. This standard protocol usually can be used, among others, to test and verify the stability of protein-ligand binding modes obtained by docking[63], to generate stable conformations of peptides[64], or also to predict flexibility in structurally modified enzymes at different temperatures[65].

Regarding CXCR4, as its physiology corresponds to a protein membrane, all MD simulations were performed adding an implicit membrane GBSW[66] and setting an implicit solvent model GBSW[67] to estimate the solvation energy using the force field CHARMM22[33]. Due to the influence of the membrane thickness in MD results, we also performed simulations at 25.4 Å, 28.5 Å and 30.0 Å in order to find the energetically most stable complex[68].

Small ligand-protein Blind Docking.

For the small molecules AMD11070 and GSK812397, whose allosteric mechanism is still unclear[13,16], an analogous approach was proposed to study their interactions. The entire CXCR4 protein surface was explored by blind docking in order to define the binding region, and a second docking analysis by subsites was then performed to refine the allosteric interactions.

Initially, both structures were prepared using *LigPrep*[69] from Schrödinger. Force field OPLS_2005 was applied to generate possible ionization states at pH 7 +/-2 with *ionizer*, and also to generate tautomers and stereoisomers. All the resulting structures were then blind docked to 3OE6 CXCR4 crystal structure using AutoDock Vina[70]. In the blind docking the grid box volume (27000 cubic Angstroms) covered the whole protein structure. The best

poses were superposed, analyzed and clustered. Clusters were re-grouped in regions attending to the region of CXCR4 where they interact using Discovery Studio.

LigandFit Docking by subsites

Once established the main region with more docked poses, we analyzed the AMD11070 and GSK812397 binding modes by a subsequent docking analysis by subsites. In a docking by subsites, the binding region to explore is divided sequentially into one, two, three or more subsites depending on the partition level defined, and then each structure is docked into all subsites. Hence, if the partition level is equal to three, every structure is docked 6 times and only the docked pose with the best scoring is retrieved with the associated subsite. This method becomes essential for proteins with big pockets, such as CXCR4.

The use of LigandFit[71] docking functions in our previous work[34] showed very good results when screening a test set of CXCR4 actives and decoys by DockScore scoring function. Moreover, regarding the evaluation of pose prediction, self-docking the cognate ligands showed also LigandFit as one of the best performing functions, principally for structures 3ODU and 3OE6. Hence, LigandFit docking function was used to divide the main binding site applying 3 partition levels, which created 6 binding subsites, according to our previous work[34].

RESULTS

Peptidomimetic-protein blind Docking using Zdock.

Blind dockings of ATI-2341 to 3OE6 structure generated 2000 poses with their corresponding ZDock Score, grouped in different clusters attending to the RMSD cutoff defined (5 Å). The best 50 docked poses with the highest ZDock values were retained. The clusters associated to these poses were classified attending to the number of docked poses they included and their corresponding ZDock Score. Finally, adjacent clusters sharing the same topographic protein sectors were grouped as regions of high probability of interactions.

Results showed two different regions where the best clusters were distributed (see Figure 2). The first region was located in the N-terminal domain, just in the entrance of the main pocket of CXCR4, and delimited by the three extracellular loops (called *Extracellular Region*). The second region was found in the intracellular domain of the receptor, which includes the intracellular loops ICL1, ICL2 and ICL3 (called *Intracellular Region*).

Figure 2

As it can be observed in Figure 2, the majority of the best 50 poses was found in the *Intracellular Region* (23 poses) and secondly in the *Extracellular Region* (18 poses). If we only consider the best 10 scored poses, the

Intracellular Region retains 6 poses and the *Extracellular Region* places only 2 docked poses with the best ZDock score.

Peptide-protein flexible Dockings.

According to the blind docking results, we can consider the existence of two different binding sites where pepducin ATI-2341 could bind, one in the extracellular side of the protein and the other in the intracellular domain of CXCR4. Nevertheless, it is relevant to note that all pepducins have lipophilic tails to cross through the plasmatic membrane, hence facilitating their binding in the intracellular face of proteins. Moreover, the conformations retrieved located in ICL1 agree with the fact that pepducins ATI-2341 and ATI-2766 are derived from ICL1 of GPCRs[21]. Attending to these considerations and the fact that the highest number of best ZDock poses was retrieved in the *Intracellular Region*, we explored further the potential binding site located in the intracellular domain of CXCR4. Therefore, the best pose retrieved in the *Intracellular Region* was chosen as starting conformation to further study CXCR4-pepducin interactions.

FlexPepDock

FlexPepDock web server requires one structure containing simultaneously the ligand and the receptor. As input structures, we used the best intracellular ZDock poses of ATI-2341 for 3OE6. The protocol was set to generate 200 structures of high resolution for each FlexPepDock run.

Once completed the simulation, FlexPepDock returned the top 10 models found with their associated scoring values expressed as total Rosetta energy. The energy score obtained was 247.23 and the *rmsBB* was 2.49 (Figure 3).

Distances between CXCR4 residues and ATI-2341 residues are shown in Supporting Information (Figure 1) in a phylogram chart.

It is worth noting that the palmitate substituent in the lipophilic tail was absent in the pepducin ATI-2341 conformations retrieved by FlexPepDock.

HADDOCK

To perform the flexible docking using HADDOCK, the receptor and the ligand must be supplied independently, as well as the starting protein-ligand interaction residues. For the pepducin ATI-2341, all 16 residues were

specified as active residues, and for CXCR4 a distance threshold of 5 Å was defined to select the residues that interact with the conformation obtained by ZDock.

The flexible docking performed in 3OE6 returned 10 clusters and 4 best poses in each cluster. The best cluster (cluster 4 in Figure 3 B) had a HADDOCK score of -61.6 +/-13.1. Distances between these residues and ATI-2341 residues are shown in Supporting Information (Figure 1) in a phylogram chart.

In the same way than using FlexPepDock, the palmitate substituent in the non-peptidic tail of the pepducin ATI-2341 conformations was not returned using HADDOCK flexible docking.

Figure 3

Peptidomimetic-protein MD.

The existence of the lipophilic substituent palmitate in pepducins, which is not taken into consideration by flexible peptide-protein protocols, can introduce reasonable doubts from the standpoint of overall interactions and clashes. The possibility to undergo distinct conformational changes may induce lack of information in the allosteric interaction map. For this reason, MD simulations were performed and compared with flexible peptide-protein docking results in terms of interactions and conformational trajectories generated by MD for the whole pepducin structure including organic substituents.

The MD steps (minimization, heating, equilibration, dynamics and production) were set as follows. An initial minimization step was performed using the steepest descent algorithm, defining 1000 minimization cycles and a RMS gradient of 0.1 kcal/mol·Å to finish the routine. A second minimization step, was performed defining 2000 minimization cycles and a RMS gradient of 0.0001 kcal/mol·Å. The temperature was initially set at 50 K and gradually increased to 300 K during the heating time by increments of 0.001 ps and 2000 iteration steps fixed in a similar way to the equilibrium step at 300 K. The production step was performed at constant temperature of 300 K (canonical ensemble) in time steps of 0.001 ps and 500.000 steps.

Furthermore, MD simulations were performed adding an implicit membrane GBSW to CXCR4 and using the Implicit Solvent Model GBSW. The influence of membrane thickness was tested at 25.4 Å, 28.5 Å and 30.0 Å. Using a thickness of 28.5 Å, the relative position and orientation of the CXCR4 membrane, as well as the results obtained by MD were considered the most suitable.

The MD simulation of the pepducin ZDock poses returned 1000 complex conformations. The energies of the most stable models were equal to 2724.4 kcal/mol at 304.7 K for the complex ATI-2341 - 3OE6 (Figure 3). Distances between CXCR4 residues and ATI-2341 residues are shown in Supporting Information (Figure 1).

It is worth noting that the MD performed kept the complete structure of ATI-2341 including its palmitate

substituent.

Small ligand-protein Blind Docking

LigPrep was used to prepare AMD11070 and GSK812397 small molecules for blind docking. Each molecule generated globally 16 structures among stereoisomers and tautomeric structures. These resulting structures were blind-docked using AutoDock Vina into the CXCR4 monomeric structure 3EO6. A large search space involving the whole protein was defined to explore completely the CXCR4 structure. The blind docking protocol returned 9 poses for every docked structure, being the best-ranked pose the one with the lowest free energy of binding. All the best-ranked poses were arranged in clusters and clusters were gathered into three main regions (Extracellular Region, Transmembrane Region and Intracellular Region). Figure 4 shows the docking regions containing the majority of AMD11070 and GSK812397 best blind-docked poses into CXCR4. The majority of poses were distributed in the Extracellular Region and inside the main pocket of CXCR4. Accordingly, this region was selected to analyze and refine possible allosteric interactions between CXCR4 and the small molecules.

Figure 4

LigandFit Docking by subsites

The binding site was defined using a total volume of 1520 cubic Angstroms that covered the total extracellular pocket. This volume was divided into 3 partition levels (1+2+3 = 6 subsites), where to dock meticulously all AMD11070 and GSK812397 molecules. Figure 5 shows the defined binding subsites. Partition 1 is the original unpartitioned site. Partition 2 divides the original site into two parts, thereby producing two subsites, and partition 3 divides the original site into three parts, thereby producing three subsites. Partitioning a site is performed using a cluster analysis algorithm that involves the use of random numbers. Every molecule was docked into each subsite using LigandFit docking function, hence enabling an exhaustive site search. The best binding pose, its DockScore scoring and the subsite where it was found energetically favorable docked, were kept for each docking. As it can be seen in Figure 5, docking results obtained for CXCR4 structures showed a distribution of binding poses mainly in binding subsites 1-2 and 1-3. Subsite 1-2 covers the volume defined by subsites 1-3 and 3-3 when the binding partition level passes from 2 to 3 partitions, respectively. The following distribution of docked poses was found: 9 binding poses were located in subsite 1-2 (28.1%), 22 poses in subsite 1-3 (68.8%), and only one pose in subsite 2-3 (3.1%), while subsites 2-2 and 3-3 did not contain any pose. The

best-scored pose was found in subsite 1-2.

Figure 5

DISCUSSION

Allostery in CXCR4 has been studied using pepducin agonists and small molecules (AMD11070 and GSK812397), which may act as allosteric antagonists. Regarding the pepducins, ATI-2341 was chosen instead of ATI-2766 due to its relative small lipophilic moiety in comparison to ATI-2766.

Blind dockings were performed using ZDock as the first step to approach how the allosteric pepducin ATI-2341 interacts with CXCR4. Rotational and translational blind docking is highly dependent on the ligand conformation that determines clashes and interactions with the protein receptor. For this reason, a starting ATI-2341 conformation for the blind docking was obtained using MD simulation.

Results generated by Zdock showed mainly two different regions where the highest number of the best poses was located. The extracellular region included poses and clusters that correspond with the orthosteric region of CXCR4 where the endogenous binding site is located, whereas the intracellular region covered intracellular loops of CXCR4, basically the first intracellular loop (ICL1).

Pepducins are protein sequences derived from one of the three intracellular loops[19,21] of different GPCRs, in particular pepducins ATI-2341 and ATI-2766 are protein sequences derived from ICL1. This evidence and the fact that the highest number of best ZDock poses was retrieved in the intracellular domain, makes the *Intracellular Region* the area to explore further pepducin-CXCR4 interactions by flexible protein-protein docking. The role that the lipophilic tail plays in the binding of ATI-2341 to CXCR4 was explored by MD simulations.

Figure 6 shows the interaction map found using the different computational approaches. It can be seen that the resulting pepducin conformations were moved in all protocols from the initial position in the interface region defined mainly by ICL2 and ICL3, to a displaced position near to ICL1 and establishing interactions with the ICL2. Probably the binding of pepducin modulator to this interface pocket could modify the conformational changes that CXCR4 undergoes when it is activated by the endogenous ligand, altering the corresponding signaling response.

Figure 6

Analyzing the typology of the interactions established between pepducin ATI-2341 and CXCR4 (Figures 6 and 7) it can be observed that residues Arg70 and Ser71 from ICL1, Arg134 from ICL2, and Thr240 from ICL3 establish polar interactions in all methods used. In addition, residues Leu69 and Met72 from ICL1, and Ala237

and Leu238 from ICL3 establish lipophilic interactions also in all computational protocols. Attending to the hydrogen bond interactions, Arg134 and Ala237 formed hydrogen bonding in all the different methods used, and residues Thr73 from ICL1, Asp133 and Arg148 from ICL2, and Lys236 as well as Thr240 from ICL3 established hydrogen bonding in accordance with two of the three methods used.

Figure 7

Figure 8 shows the superposition of ATI-2341 conformations obtained by flexible dockings and by MD. It can be observed how the three refining methods present different protein folding for ATI-2341, principally due to the high flexibility of the peptidomimetic compound. Nevertheless, it can be noticed that the three conformers are located in the same CXCR4 binding site and some of the pepducin residues for FlexPepDock and MD show the same interactions with ICL1 and ICL2.

Figure 8

Regarding the small molecules, the blind docking performed by AutoDock Vina on CXCR4, for the overall 32 AMD11070 and GSK812397 isomeric structures, showed mainly 3 different clusters. The main cluster containing the greatest number of docked poses was located in the extracellular side of CXCR4 (*Extracellular Region*). This region corresponds to the main pocket of CXCR4 where the endogenous ligand SDF-1 binds. The other two clusters were placed on the external surface of the transmembrane domain, between helices 6 and 7 (*Transmembrane Region*) similarly to the allosteric binding site recently discovered[72] in purinergic receptor $P2Y_1$, and the second one, between ICL1 and ICL3, in an equivalent region to the one found for pepducin ZDock dockings (*Intracellular Region*).

According to the distribution of blind docked poses, the wide *Extracellular Region* was considered the region with higher probability to establish interactions between the small ligands AMD11070, GSK812397 and CXCR4 receptor, and therefore the region for docking refinement by subsites.

Dockings performed by subsites located all poses in the homologue subsites 1-2 and 1-3, unless one GSK812397 structure which was found in subsite 2-3 (Figure 5). These observations show the existence of minimum 2 different binding sites in CXCR4, corresponding approximately to subsites 1-2 and 2-3, which agrees with the experimental observation that co-crystalized ligands IT1t and CVX15 bind to CXCR4 in different domains of the extracellular binding pocket usually known as minor and major binding sites respectively. In fact, if we superpose the co-crystalized ligand IT1t into our binding subsites, this ligand fills the binding subsite 3-3, suggesting the existence of a third possible binding subsite overlapped[73] to the subsites 1-2 and 1-3 where we found AMD11070 and GSK812397 poses (Supplementary Fig. 2). This hypothesis agrees with the assumptions

proposed by *Srivastava et al.*[74] and *Garcia-Perez et al.*[75] *Srivastava et al.*, which reported recently the crystal structure of hGPCR receptor bound to an allosteric ligand. The authors provide insights into the plausible binding of multiple ligands to the receptor hGPR40 and propose three putative binding sites. One of them should be located near the orthosteric site. In addition, *Garcia-Perez et al.* in their work docked Maraviroc inhibitor into a homology model of CCR5. They obtained multiple poses that roughly delimit three different sites, yet overlapping, in the TM pocket of the protein, in a similar manner that we obtained docked poses by subsites using 3 partitions.

However, it is important to highlight that there were no appreciable differences between the subsite volumes occupied by all the docked poses of AMD11070 and GSK812397 (Figure 5). The visual inspection of the docked poses returned by each docking was relatively similar, without observing bias or distinct subsite distribution induced by the dissimilar structure of the two chemical candidates. In fact, this observation may be indicative about similar patterns in their interaction with CXCR4 and similarities in their allosteric mechanism.

We considered that the best AMD11070 and GSK812397 poses found (Figure 5) should represent better CXCR4-ligand interactions. Indeed, if we analyze these best docked structures we find how they fill the central region of subsite 1-2 (Figure 9). When protein-ligand interaction diagrams are represented, we observe that residues Asp187, His203, Tyr255, His281 and Glu288 interact directly with GSK812397 and these residues delimit specifically the defined binding site 1-2 (Supplementary Fig. 2). In this regard, *Wong et al.*[76] investigated by site-directed mutagenesis key CXCR4 residues that established interactions with AMD11070 and proposed three different binding modes to explain all interactions between AMD11070 and Tyr45, Trp94, Asp97, Asp171, Asp262 and Glu288. Later, *Mosi et al.*[13] confirmed again those residues as important ones for AMD11070 binding, emphasizing residues Tyr45, Trp94 and Asp97 among them. All these observations agree with our best model proposed for AMD11070 and GSK812397.

Figure 9

Interestingly, *Tian et al*[77] and *Zhou et al.*[78] concluded in their mutagenesis analyses that several residues of the CXCR4 transmembrane, such as Tyr45, His79, Asp97, Pro163, Trp252, Tyr255, Asp262, Glu288, His294 and Asn298, as well as from the Extracellular loop 2, such as Asp182, Arg183, Tyr184, Asp187, Phe189, Tyr190 and Pro191 interact specifically with HIV-1 without affecting signaling, nor the binding of SDF-1α. Consequently, it is presupposed their allosteric behavior. The authors noticed specially, how alterations of extracellular loop 2 by mutations had little effect on SDF1- α binding but significantly diminished the activity of HIV-1. Again, this observation agrees with the interactions obtained for GSK812397. Among these listed

residues, GSK812397 interacts directly with Asp97, Asp187, Tyr255, Asp262 and Glu288. It is interesting to note how residue Asp262 delimits subsite 1-2 from subsite 2-2 and consequently can establish interactions with ligands that bind in both subsites (Figure 9). The central position of this residue in the extracellular pocket explains why so many CXCR4 ligands interact with Asp262. However, Asp97 establishes medium-large range interaction with AMD11070. This residue is located at the entrance of subsite 3-3, where cognate ligand IT1t binds, outside from subsite 1-2 (see Figure 9). Therefore, it is not considered a key residue for the allosteric interactions.

Recently, Thiele *et al.*[79] studied the binding mode of peptide FC131 with CXCR4 by mutagenesis techniques. The authors confirmed that binding modes of FC131 and CVX15 were similar and located at the same major binding site but FC131 was positioned deeper into the hydrophobic subpocket. The study indicates how mutations in residues Asp187, His203, Tyr255 and His281, among others, did not impair the antagonist potency of FC131. In addition, the author's support that Glu288 affects signaling properties but not the binding of CXCL12, and they propose that residue Glu288 forms indirect interaction with FC131 via water mediated H-bond network. These observations are consistent with the proposal that residues Asp187, His203, Tyr255, His281 and Glu288 may form a different subsite than the major binding subsite, located between the major and minor binding subsites.

Based on these observations and in agreement with the protein-ligand interaction analysis performed, we propose Glu288 as the main residue that establish interactions of allosteric nature with the small CXCR4 antagonists AMD11070 and GSK812397. Also, residue Asp187 and in second term residues His281, Tyr255 and His203, provide short-range hydrophobic interactions and π interactions with both antagonists. This set of residues defines specifically subsites 1-2 and 1-3 (Figure 9). It is important to note that Asp187 and Glu288, whose interactions are found in both antagonists, also present charge interactions and hydrogen bonding with GSK812397. Charge interactions are considered key and essential in the majority of processes where CXCR4 acts, and therefore residues Asp187 and Glu288 should be considered strategic in CXCR4 allosteric mechanisms.

Overall, we proposed that CXCR4 has minimum two allosteric binding sites in distinct domains attending to the allosteric functionality. Pepducins bind to an intracellular region while small molecules to an extracellular region. Pepducins seem to act by blocking the β-hairpin conformational changes in the active state, which would cause α-helices not to be able to make translational movements between them. This seems to be different from

what happens with small molecules, which can partially block the active site, allowing the natural CXCL12 ligand to enter but not the HIV. Our binding mode hypothesis for small molecules agrees with *Surgand et al.*[80] and *Scholten et al.*[81] works, in which they also propose that, in general, GPCRs such as CXCR4, CXCR2 or CCR4 have two binding pockets in the extracellular side of these receptors. One minor binding pocket formed by residues from helices 1, 2, 3 and 7, corresponding to our binding subsite 3-3 occupied by IT1t and one major binding pocket formed by residues from helices 3, 4, 5 and 6, corresponding to our binding subsites 1-3 plus 2-3 occupied by peptide CVX15. The overlapping region between subsites 1-3 and 3-3 corresponds to the allosteric subsite occupied by allosteric candidates AMD11070 and GSK812397 (see Supplementary Fig. 2). This agrees with, *Scholten et al.* work who state that in the case of CXCR4, both pockets are interconnected and ligands bind in both sites, some of them exclusively in the major or in the minor binding pocket, and ligands with partial overlapping binding sites also show non-competitive allosteric behavior.

Hence, the small molecules AMD11070 and GSK812397 could act both as antagonists and allosteric modulators depending on the position occupied when binding the CXCR4 extracellular pocket. The differences in the molecular mechanism of these small molecules could be due to their possible movement from a low affinity site to another interconnected site of higher affinity. For example, between the minor binding pocket (corresponding to our subsite 3-3), which would produce an agonistic effect, and the overlapping region between subsites 1-3 and 3-3, which would produce an allosteric effect. Similarly, *Lagane, B. et al.*[82] hypothesized for the CCR5 allosteric modulation by Maraviroc (MVC) that "MVC could theoretically pass from the superficial site 1 to the deepest site 3 by a simple translation without necessity of conformational changes for the ligand adaptation to the two sites". Once the small molecules are bound with high affinity to CXCR4, they may prevent efficient binding of orthosteric ligands to the receptor probably by perturbing the transmembrane–extracellular loops interactions, which is the essential characteristic of allosteric modulation.

CONCLUSION

In the present study, we have analyzed the putative key interactions between chemically different allosteric modulators (pepducins and small molecules) and the CXCR4 receptor by combining several methods. Pepducin ATI-2341, a peptidomimetic compound, was selected as the most representative allosteric agonist of CXCR4. AMD11070 and GSK812397, two small organic antagonists of CXCR4 and candidates to be allosteric modulators, were also selected to study their interactions. Given that these are different compounds attending to their chemical nature, structure and mode of action, we proposed different blind docking approximations to

locate the main regions where high affinity binding was more favorable. Next, a refining docking process was carried out to establish the putative key allosteric interactions. For pepducin ATI-2341 two flexible dockings and one MD protocol were performed in order to compare modeling approaches and obtain a consensus result for the key pepducin-CXCR4 binding interactions. The three methods used confirmed similar interactions between pepducin ATI-2341 and CXCR4. For the small organic ligands, AMD11070 and GSK812397, docking refinement was performed by subsites. The corresponding protein-ligand interactions were analyzed and compared finding similar interactions and sharing very similar binding positions in the big extracellular pocket of CXCR4.

In conclusion, we propose that CXCR4 has minimum 2 allosteric binding sites in distinct domains attending to the allosteric functionality, such as antagonists to inhibit inflammation and cancer or agonists to enhance angiogenesis in tissue regeneration. One allosteric binding site would be placed in the intracellular region near to ICL1, interacting with ICL2 and in lower mode with ICL3 and the extreme C-terminal. We state that this would be the natural binging site of pepducin ATI-2341. The allosteric agonist ATI-2341 then would act blocking β-hairpin and impeding the translational movement of the α-helices of CXCR4 in its active state to evolve to the inactive state. The second allosteric site would be accommodated in the extracellular pocket, in a subsite of the main orthosteric binding pocket, hence being different from the orthosteric minor binding subsite where IT1t binds, but adjacent or overlapped to it. This proposed second binding site would interact with residues that do not affect the binding of the endogenous ligand but in contrast inhibit HIV-CXCR4 binding. Residues Asp187 and Glu 288 that do not affect signaling but inhibit HIV seem to interact in all our proposed models.

ACKNOWLEDGEMENTS

The authors thank Accelrys, Chemical Computing Group for providing Academic Licenses for Discovery Studio and MOE respectively. Glide (Schrodinger) was used through the Servei de Disseny de Farmacs (SDF) in the Centre de Supercomputacio de Catalunya (CESCA). This work is supported by the Programa Nacional de Biomedicina (Ministerio de Educación y Ciencia, SAF2010-21617-C02-02).

REFERENCES

[1] K.A. Reynolds, R.N. McLaughlin, R. Ranganathan, Hot spots for allosteric regulation on protein surfaces, Cell. 147 (2011) 1564–1575.

[2] L.T. May, K. Leach, P.M. Sexton, A. Christopoulos, Allosteric modulation of G protein–coupled receptors, Annu. Rev. Pharmacol. Toxicol. 47 (2007) 1–51.

[3] Q. Cui, M. Karplus, Allostery and cooperativity revisited, Protein Sci. 17 (2008) 1295–1307.

[4] P. Csermely, R. Palotai, R. Nussinov, Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, Trends Biochem. Sci. 35 (2010) 539–546.

[5] R. Nussinov, C.-J. Tsai, Allostery in disease and in drug discovery, Cell. 153 (2013) 293–305.

[6] P. Conn, A. Christopoulos, C. Lindsley, Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders, Nat. Rev. Drug Discov. 8 (2009) 41–54.

[7] M. Spedding, J.-P. Pin, R. Hubbard, J. Kiss, P. Keov, P.M. Sexton, et al., Allosteric modulation of G protein-coupled receptors: A pharmacological perspective, Neuropharmacology. 60 (2011) 24–35.

[8] C.-I.A. Wang, R.J. Lewis, Emerging opportunities for allosteric modulation of G-protein coupled receptors, Biochem. Pharmacol. 85 (2012) 153–162.

[9] N. London, B. Raveh, O. Schueler-Furman, Druggable protein-protein interactions - from hot spots to hot segments, Curr. Opin. Chem. Biol. 17 (2013) 952–959.

[10] N. Yanamala, J. Klein-Seetharaman, Allosteric modulation of G protein coupled receptors by cytoplasmic, transmembrane and extracellular ligands, Pharmaceuticals. 3 (2010) 3325–3342.

[11] J. Ballesteros, J. Ransom, In-target versus off-target allosteric modulators of GPCRs, Drug Discov. Today Ther. Strateg. 3 (2007) 445–450.

[12] A. Sachpatzidis, B.K. Benton, J.P. Manfredi, H. Wang, A. Hamilton, H.G. Dohlman, et al., Identification of allosteric peptide agonists of CXCR4, J. Biol. Chem. 278 (2003) 896–907.

[13] R.M. Mosi, V. Anastassova, J. Cox, M.C. Darkes, S.R. Idzan, J. Labrecque, et al., The molecular pharmacology of AMD11070: an orally bioavailable CXCR4 HIV entry inhibitor, Biochem. Pharmacol. 83 (2011) 472–479.

[14] S. Fricker, R. Mosi, V. Anastassova, J. Labrecque, R. Wong, R. Skerlj, et al., The orally bioavailable allosteric CXCR4 HIV-1 entry inhibitor AMD11070, Retrovirology. 9 (2012) P7.

[15] S. Boggs, V.I. Elitzin, K. Gudmundsson, M.T. Martin, M.J. Sharp, Kilogram-scale synthesis of the CXCR4 antagonist GSK812397, Org. Process Res. Dev. 13 (2009) 781–785.

[16] S. Jenkinson, M. Thomson, D. McCoy, M. Edelstein, S. Danehower, W. Lawrence, et al., Blockade of X4-Tropic HIV-1 cellular entry by GSK812397, a potent noncompetitive cxcr4 receptor antagonist, Antimicrob. Agents Chemother. 54 (2010) 817–824.

[17] S. Oishi, N. Fujii, Peptide and peptidomimetic ligands for CXC chemokine receptor 4 (CXCR4), Org. Biomol. Chem. 10 (2012) 5720–5731.

[18] P. Dimond, K. Carlson, M. Bouvier, C. Gerard, L. Xu, L. Covic, et al., G protein–coupled receptor modulation with pepducins: moving closer to the clinic, Ann. N. Y. Acad. Sci. 1226 (2011) 34–49.

[19] J. Lomas-Neira, A. Ayala, Pepducins: an effective means to inhibit GPCR signaling by neutrophils, Trends Immunol. 26 (2005) 619–621.

[20] J. Miller, A. Agarwal, L. Devi, Insider access: pepducin symposium explores a new approach to GPCR modulation, Ann. N. Y. Acad. Sci. 1180 (2009) E1–E12.

[21] K.E. Carlson, T.J. McMurry, S.W. Hunt III, Pepducins: lipopeptide allosteric modulators of GPCR signaling, Drug Discov. Today Technol. 9 (2012) e33–e39.

[22] L. Covic, A.L. Gresser, J. Talavera, S. Swift, A. Kuliopulos, Activation and inhibition of G protein-coupled receptors by cell-penetrating membrane-tethered peptides, Proc. Natl. Acad. Sci. 99 (2002) 643–648.

[23] K. O'Callaghan, A. Kuliopulos, L. Covic, Turning receptors on and off with intracellular pepducins: new insights into G-protein-coupled receptor drug development, J. Biol. Chem. 287 (2012) 12787–12796.

[24] J.M. Janz, Y. Ren, R.J. Looby, M.A. Kazmi, P. Sachdev, A. Grunbeck, et al., Direct interaction between an allosteric agonist pepducin and the chemokine receptor CXCR4, J. Am. Chem. Soc. 40 (2011) 15878–15881.

[25] B. Tchernychev, Y. Ren, P. Sachdev, J.M. Janz, L. Haggis, A. O'Shea, et al., Discovery of a CXCR4 agonist pepducin that mobilizes bone marrow hematopoietic cells, Proc. Natl. Acad. Sci. 107 (2010) 22255–22259.

[26] M. Suchanek, A. Radzikowska, C. Thiele, Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells, Nat. Methods. 2 (2005) 261–268.

[27] M. Trakselis, S. Alley, F. Ishmael, Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics, Bioconjug. Chem. 16 (2005) 741–750.

[28] P. Kask, K. Palo, Fluorescence-intensity distribution analysis and its application in biomolecular detection technology, Proc. Natl. Acad. Sci. U. S. A. 96 (1999).

[29] F. Bernstein, T. Koetzle, G. Williams, Protein Data Bank, Atomic coordinate and Bibliographic entry format description, J. Mol. Biol. 112 (1977) 535–542.

[30] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity--a rapid access to atomic charges, Tetrahedron. 36 (1980) 3219–3228.

[31] A.D. MacKerell, M. Feig, C.L. Brooks, Extending the treatment of backbone energetics in protein force fields: Limitations of gas□phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, J. Comput. Chem. 25 (2004) 1400–1415.

[32] Discovery Studio, Accelrys, Inc, San Diego, 2009.

[33] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B. 102 (1998) 3586–616.

[34] J.M. Planesas, V.I. Pérez-Nueno, J.I. Borrell, J. Teixidó, Impact of the CXCR4 structure on docking-based virtual screening of HIV entry inhibitors, J. Mol. Graph. Model. (2012) 123–136.

[35] H.M. Berman, T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, et al., The protein data bank, Acta Crystallogr. Sect. D Biol. Crystallogr. 58 (2002) 899–907.

[36] B. Wu, E.Y.T. Chien, C.D. Mol, G. Fenalti, W. Liu, V. Katritch, et al., Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists, Science (80-.). 330 (2010) 1066–1071.

[37] R.M. Bill, P.J.F. Henderson, S. Iwata, E.R.S. Kunji, H. Michel, R. Neutze, et al., Overcoming barriers to membrane protein structure determination, Nat. Biotechnol. 29 (2011) 335–340.

[38] Y. Zou, W.I. Weis, B.K. Kobilka, N-Terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor, PLoS One. 7 (2012) e46039.

[39] S.G.F. Rasmussen, H.J. Choi, D.M. Rosenbaum, T.S. Kobilka, F.S. Thian, P.C. Edwards, et al., Crystal structure of the human &bgr; 2 adrenergic G-protein-coupled receptor, Nature. 450 (2007) 383–387.

[40] V.-P. Jaakola, M.T. Griffith, M.A. Hanson, V. Cherezov, E.Y.T. Chien, J.R. Lane, et al., The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist, Science. 322 (2008) 1211–1217.

[41] S.H. Park, B.B. Das, F. Casagrande, Y. Tian, H.J. Nothnagel, M. Chu, et al., Structure of the chemokine receptor CXCR1 in phospholipid bilayers, Nature. 491 (2012) 779–783.

[42] A. Šali, T. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.

[43] T. Cheng, X. Li, Y. Li, Z. Liu, R. Wang, Comparative Assessment of Scoring Functions on a Diverse Test Set, J. Chem. Inf. Model. 49 (2009) 1079–1093.

[44] R. Chen, Z. Weng, A novel shape complementarity scoring function for protein-protein docking, Proteins Struct. Funct. Bioinforma. 51 (2003) 397–408.

[45] R. Chen, L. Li, Z. Weng, ZDOCK: An initial-stage protein-docking algorithm, Proteins Struct. Funct. Bioinforma. 52 (2003) 80–87.

[46] H. Hwang, T. Vreven, B.G. Pierce, J.-H. Hung, Z. Weng, Performance of ZDOCK and ZRANK in CAPRI rounds 13–19, Proteins Struct. Funct. Bioinforma. 78 (2010) 3104–3110.

[47] T. Vreven, B.G. Pierce, H. Hwang, Z. Weng, Performance of ZDOCK in CAPRI rounds 20-26, Proteins. 81 (2013) 2175–2182.

[48] S. Vajda, D. Kozakov, Convergence and combination of methods in protein–protein docking, Curr. Opin. Struct. Biol. 19 (2009) 164–170.

[49] A.M.J.J. Bonvin, Flexible protein–protein docking, Curr. Opin. Struct. Biol. 16 (2006) 194–200.

[50] S. Lyskov, J.J. Gray, The RosettaDock server for local protein–protein docking, Nucleic Acids Res. 36 (2008) W233–W238.

[51] B. Pierce, Z. Weng, A combination of rescoring and refinement significantly improves protein docking performance, Proteins Struct. Funct. Bioinforma. 72 (2008) 270–279.

[52] N. London, B. Raveh, E. Cohen, G. Fathi, O. Schueler-Furman, Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions, Nucleic Acids Res. 39 (2011) W249–W253.

[53] S.J. De Vries, M. van Dijk, A.M.J.J. Bonvin, The HADDOCK web server for data-driven biomolecular docking, Nat. Protoc. 5 (2010) 883–897.

[54] B. Raveh, N. London, O. Schueler□Furman, Sub-angstrom modeling of complexes between flexible peptides and globular proteins, Proteins Struct. Funct. Bioinforma. 78 (2010) 2029–2040.

[55] C. Dominguez, R. Boelens, A.M.J.J. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, J. Am. Chem. Soc. 125 (2003) 1731–1737.

[56] M. Trellet, A.S.J. Melquiond, A.M.J.J. Bonvin, A Unified Conformational Selection and Induced Fit Approach to Protein-Peptide Docking, PLoS One. 8 (2013) e58769.

[57] N. Nakajima, J. Higo, A. Kidera, H. Nakamura, Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation, Chem. Phys. Lett. 278 (1997) 297–301.

[58] F.L. Gervasio, A. Laio, M. Parrinello, Flexible docking in solution using metadynamics, J. Am. Chem. Soc. 127 (2005) 2600–2607.

[59] M. Mangoni, D. Roccatano, A. Di Nola, Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation, Proteins Struct. Funct. Bioinforma. 35 (1999) 153–162.

[60] Z. Huang, C.F. Wong, Docking flexible peptide to flexible protein by molecular dynamics using two implicit-solvent models: an evaluation in protein kinase and phosphatase systems, J. Phys. Chem. B. 113 (2009) 14343–14354.

[61] O. Dagliyan, E.A. Proctor, K.M. D'Auria, F. Ding, N. V Dokholyan, Structural and dynamic determinants of protein-peptide recognition, Structure. 19 (2011) 1837–1845.

[62] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct. Biol. 9 (2002) 646–652. [63]A.H. Mahmoud, M.S.A. Elsayed, M. ElHefnawi, Structure-based predictive model for some benzimidazole inhibitors of hepatitis C virus NS5B polymerase, Med. Chem. Res. 22 (2012) 1866–1883.

[64] P.-C. Shih, M.-S. Yang, S.-C. Lin, Y. Ho, J.-C. Hsiao, D.-R. Wang, et al., A turn-like structure "KKPE" segment mediates the specific binding of viral protein A27 to heparin and heparan sulfate on cell surfaces, J. Biol. Chem. 284 (2009) 36535–36546.

[65] Q.A.T. Le, J.C. Joo, Y.J. Yoo, Y.H. Kim, Development of thermostable Candida antarctica lipase B through novel in silico design of disulfide bridge, Biotechnol. Bioeng. 109 (2012) 867–876.

[66] W. Im, M. Lee, C. Brooks, Generalized born model with a simple smoothing function, J. Comput. Chem. 24 (2003) 1691–1702.

[67] J. Chen, C.B. III, J. Khandogin, Recent advances in implicit solvent-based methods for biomolecular simulations, Curr. Opin. Struct. Biol. 18 (2008) 140–148.

[68] O. Yuzlenko, T. Lazaridis, Membrane protein native state discrimination by implicit membrane models, J. Comput. Chem. 34 (2013) 731–738.

[69] LigPrep, version 2.4, Schrödinger, LLC, New York, NY, (2010).

[70] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.

[71] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites, J. Mol. Graph. Model. 21 (2003) 289–307.

[72] D. Zhang, Z.-G. Gao, K. Zhang, E. Kiselev, S. Crane, J. Wang, et al., Two disparate ligand-binding sites in the human P2Y1 receptor, Nature. 520 (2015) 317–321.

[73] B.D. Cox, A.R. Prosser, B.M. Katzman, A.A. Alcaraz, D.C. Liotta, L.J. Wilson, et al., Anti-HIV small-molecule binding in the peptide subpocket of the CXCR4:CVX15 crystal structure, Chembiochem. 15 (2014) 1614–1620.

[74] A. Srivastava, J. Yano, Y. Hirozane, G. Kefala, F. Gruswitz, G. Snell, et al., High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875., Nature. 513 (2014) 124–7.

[75] J. Garcia-Perez, P. Rueda, J. Alcami, Allosteric model of maraviroc binding to CC chemokine receptor 5 (CCR5), J. Biol. Chem. 286 (2011) 33409–33421.

[76] R.S.Y. Wong, V. Bodart, M. Metz, J. Labrecque, G. Bridger, S.P. Fricker, Comparison of the potential multiple binding modes of bicyclam, monocylam, and noncyclam small-molecule CXC chemokine receptor 4 inhibitors, Mol. Pharmacol. 74 (2008) 1485–1495.

[77] S. Tian, W. Choi, D. Liu, J. Pesavento, Distinct functional sites for human immunodeficiency virus type 1 and stromal cell-derived factor 1α on CXCR4 transmembrane helical domains, J. Virol. 79 (2005) 12667–12673.

[78] N. Zhou, Z. Luo, J. Luo, D. Liu, J.W. Hall, R.J. Pomerantz, et al., Structural and functional characterization of human CXCR4 as a chemokine receptor and HIV-1 co-receptor by mutagenesis and molecular modeling studies, J. Biol. Chem. 276 (2001) 42826–42833.

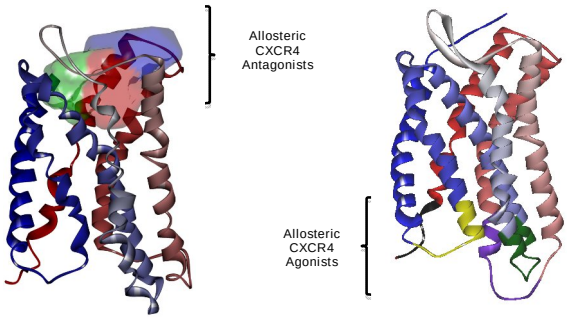
[79] S. Thiele, J. Mungalpara, Determination of the binding mode for the cyclopentapeptide CXCR4 antagonist FC131 using a dual approach of ligand modifications and receptor mutagenesis, Br. J. Pharmacol. 171 (2014) 5313–5329.

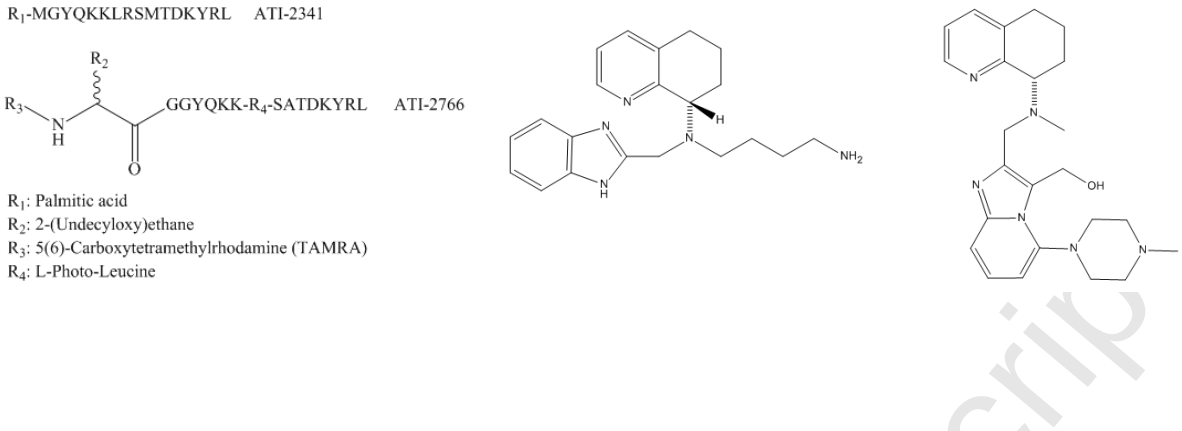
[80] J.-S. Surgand, J. Rodrigo, E. Kellenberger, D. Rognan, A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors, Proteins. 62 (2006) 509–538.

[81] D.J. Scholten, M. Canals, D. Maussang, L. Roumen, M.J. Smit, M. Wijtmans, et al., Pharmacological modulation of chemokine receptor function, Br. J. Pharmacol. 165 (2012) 1617–1643.

[82] B. Lagane, J. Garcia-Perez, E. Kellenberger, Modeling the allosteric modulation of CCR5 function by Maraviroc, Drug Discov. Today. Technol. 10 (2013) e297–e305.

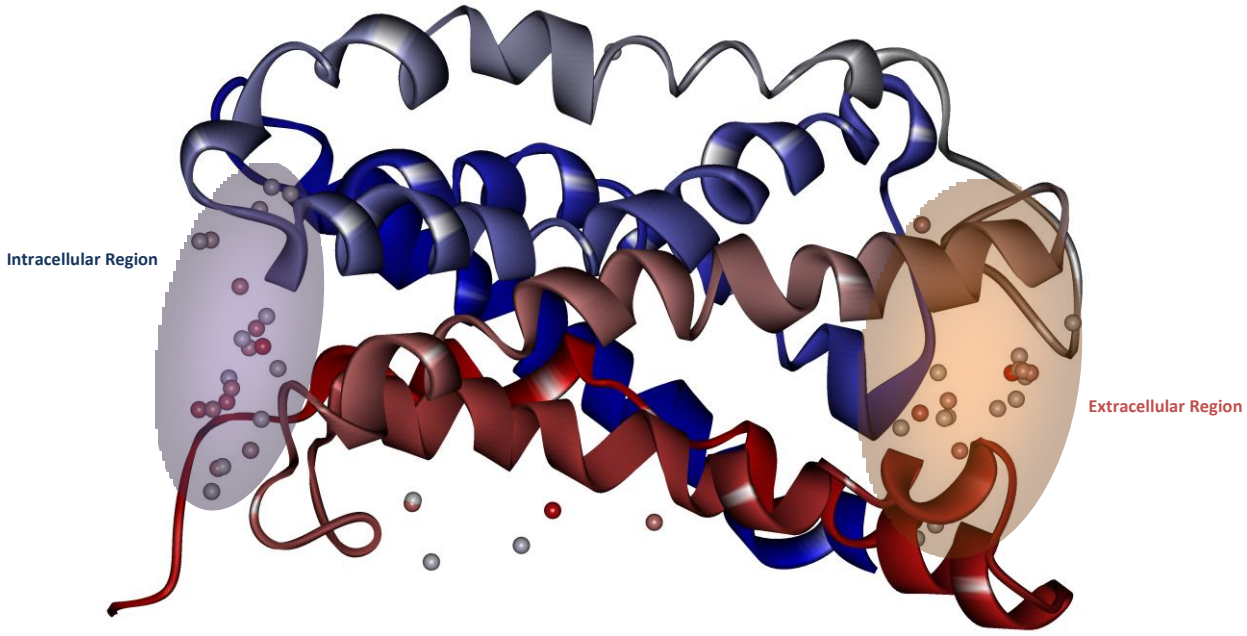
GRAPHICAL ABSTRACT





A **B** **C**

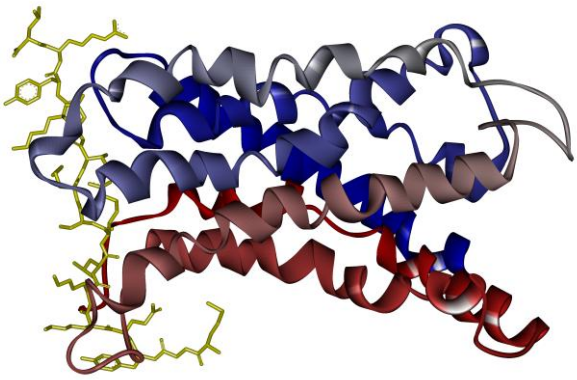
Figure 1 Structures of CXCR4 allosteric inhibitors. **A.** Allosteric peptidomimetic agonists pepducins ATI-2341 and ATI-2766 with their corresponding substituents **B**. Small organic antagonist AMD11070 **C.** Small organic antagonist GSK812397.

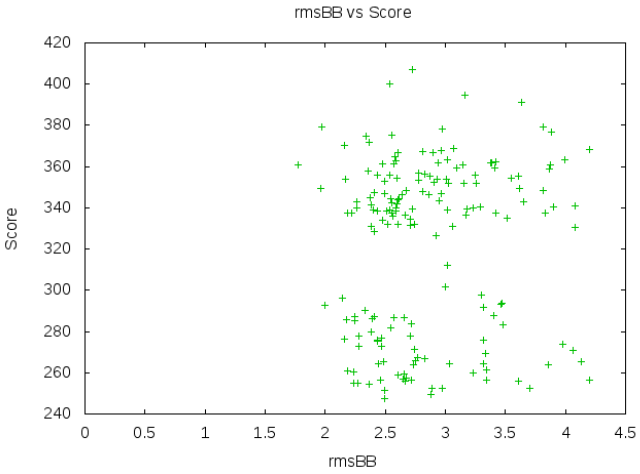


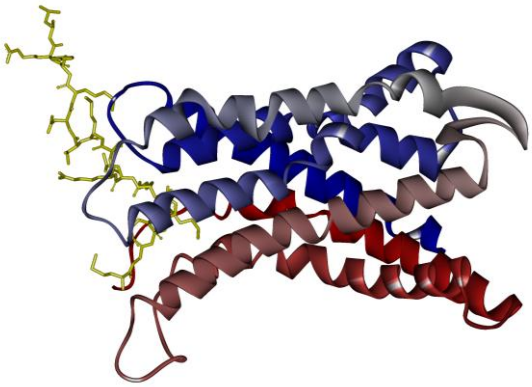
PDB	Extracellular Region	Intracellular Region
3OE6	18 poses	23 poses

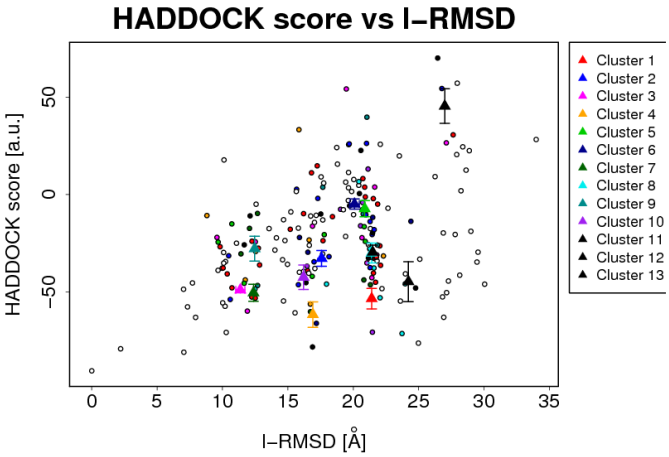
Figure 2 3OE6 regions where the best 50 ATI-2341 ZDock poses were distributed. Pepducin poses are represented in balls in colors ranging from red to blue according to ZDock Score (red=best and blue=poor ZDock Scores)

A









B

C

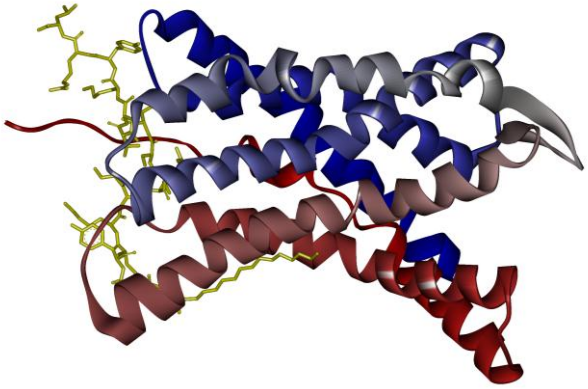
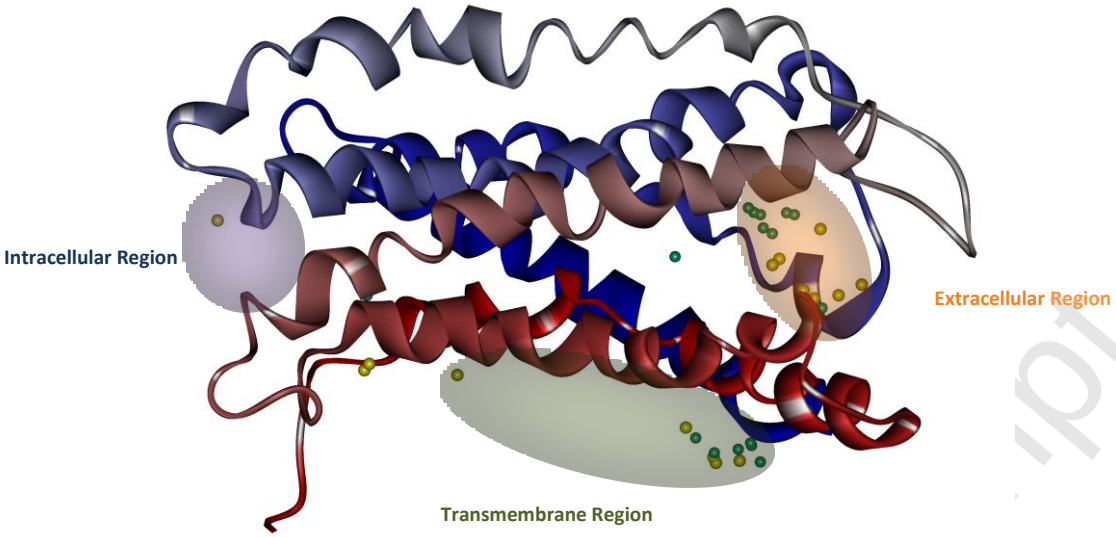
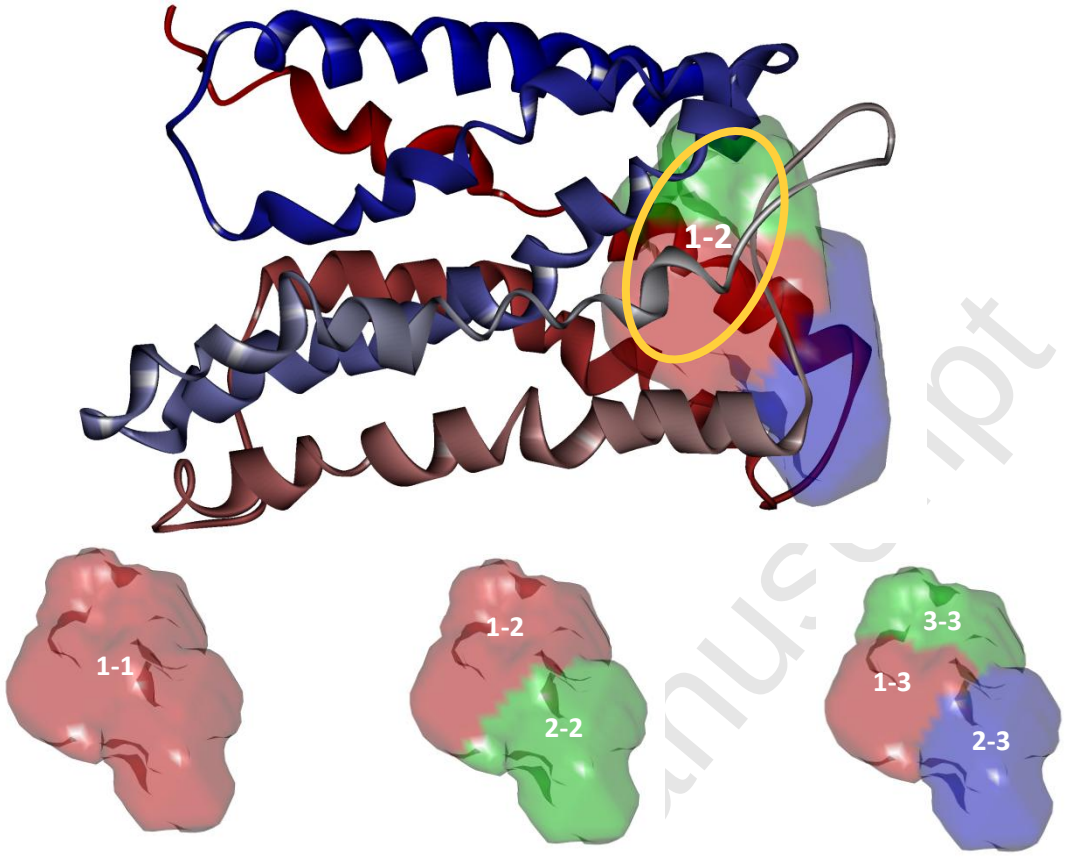


Figure 3 **A** Best ATI-2341 pose obtained by FlexPepDock, located in ICL1 of PDB 3OE6. On the right, the global flexible docking results. Score values are represented versus ligand RMSD over peptide backbone heavy atoms (*rmsBB*) with respect to the initial conformation. **B** HADDOCK best scored pose (left) for 3OE6 and the global results grouped in clusters (right). Score values are represented versus the ligand RMSD (l-RMSD)[56]. **C** Final and stabilized conformation of ATI-2341-3OE6 obtained by MD simulation.



PDB	Extracellular Region	Transmembrane Region	Intracellular Region
	AMD11070/GSK812397	*AMD11070/GSK812397*	*AMD11070/GSK812397*
3OE6	**8**/7 poses	7/**4** poses	0/1 poses

Figure 4 Scheme of the docking regions containing the majority of AMD11070 and GSK812397 blind-docked poses. Transmembrane Region is located on the surface of helix 6 and 7 in the cytoplasmic membrane domain. Intracellular Region is located between ICL1 and ICL3 domains. Extracellular Region covers the main extracellular pocket of CXCR4. The table below shows the number of best poses of AMD11070 and GSK812397 found by region. Values in bold indicate the best energy pose and the region where it was placed. AMD11070 and GSK812397 poses are represented in balls green and yellow, respectively.



Subsite

PDB		*1-1*	*1-2*	*1-3*	*2-2*	*2-3*	*3-3*
3OE6	AMD11070	-	5	11	0	0	0
	GSK812397		**4**	11	0	1	0

Figure 5 Representation of the 3 partition levels applied to CXCR4 extracellular pocket in order to dock AMD11070 and GSK812397 by subsites. Subsite 1-1 is generated when only one partition level is applied to the binding site. Subsites 1-2 and 2-2 are generated when two partition levels are applied. Subsites 1-3, 2-3 and 3-3 are created when three partition levels are applied. Applying 3 partition levels creates a total of 6 subsites. Hence, an exhaustive site exploration can be performed by docking AMD11070 and GSK81239 into each sub-site. The table below shows the number of best-docked poses located in each subsite for structures AMD11070 and GSK812397. Values in bold indicate the subsite and the structure that returned the best-scored pose.

ATI-2341	Tyr65	Gln66	Lys67	Lys68	Leu69	Arg70	Ser71	Met72	Thr73	Asp74	Tyr76	Arg77
FlexPepDock (3OE6)				H				H				
Haddock (3OE6)					H				H			
MD (3OE6)									2H			

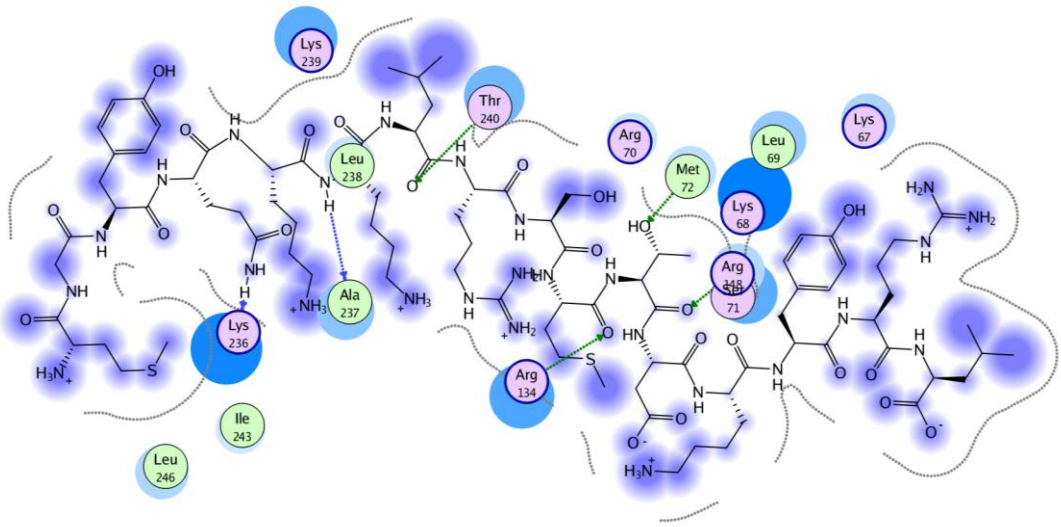
ATI-2341	Ile130	Asp133	Arg134	Arg148	Lys149
FlexPepDock (3OE6)			H	H	
Haddock (3OE6)	H	H	H		
MD (3OE6)		H	H	2H	

ATI-2341	Gln233	Lys234	Arg235	Lys236	Ala237	Leu238	Lys239	Thr240	Thr241	Val242	Ile243	Leu244	Leu246	Phe249
FlexPepDock (3OE6)				H	H			H						
Haddock (3OE6)					H			H			H			
MD (3OE6)				H	H	H								

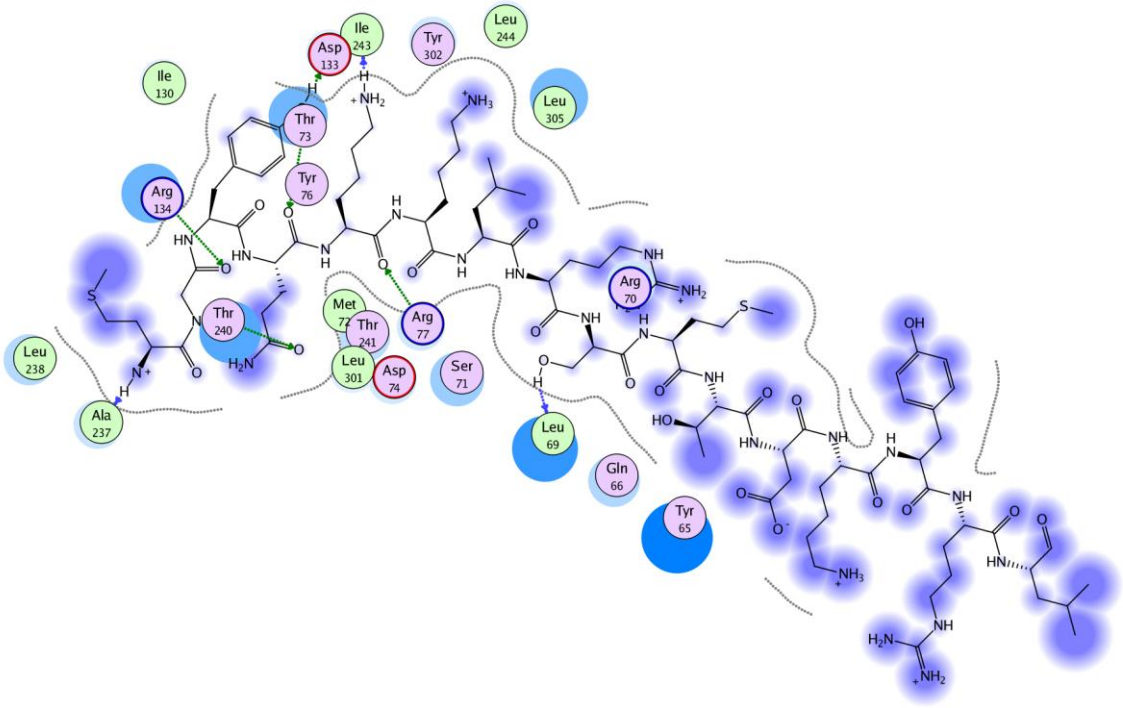
ATI-2341	Leu301	Leu302	Ala303	Phe304	Leu305
FlexPepDock (3OE6)					
Haddock (3OE6)					
MD (3OE6)					

Figure 6 Illustration of the interactions established between pepducin ATI-2341 and residues of CXCR4 according to each computational method. CXCR4 structure is shown with its 7-helixes colored by N-to-C terminal in a continuous gradient from blue at the N-terminus through white to red at the C-terminus. Each table is related to the CXCR4 fragments colored in yellow for ICL1, in dark green for ICL2, in purple for ICL3 and in black for the C-terminal domain. Colored domains correspond to residues of CXCR4 that establish interactions with ATI-2341 according to FlexPepDock, HADDOCK, or MD results. Tables show polar interactions in magenta boxes and hydrophobic interactions in light green boxes. Side chain hydrogen bonds (*H*) are indicated in red letters and backbone hydrogen bonds in black letters.

A



B



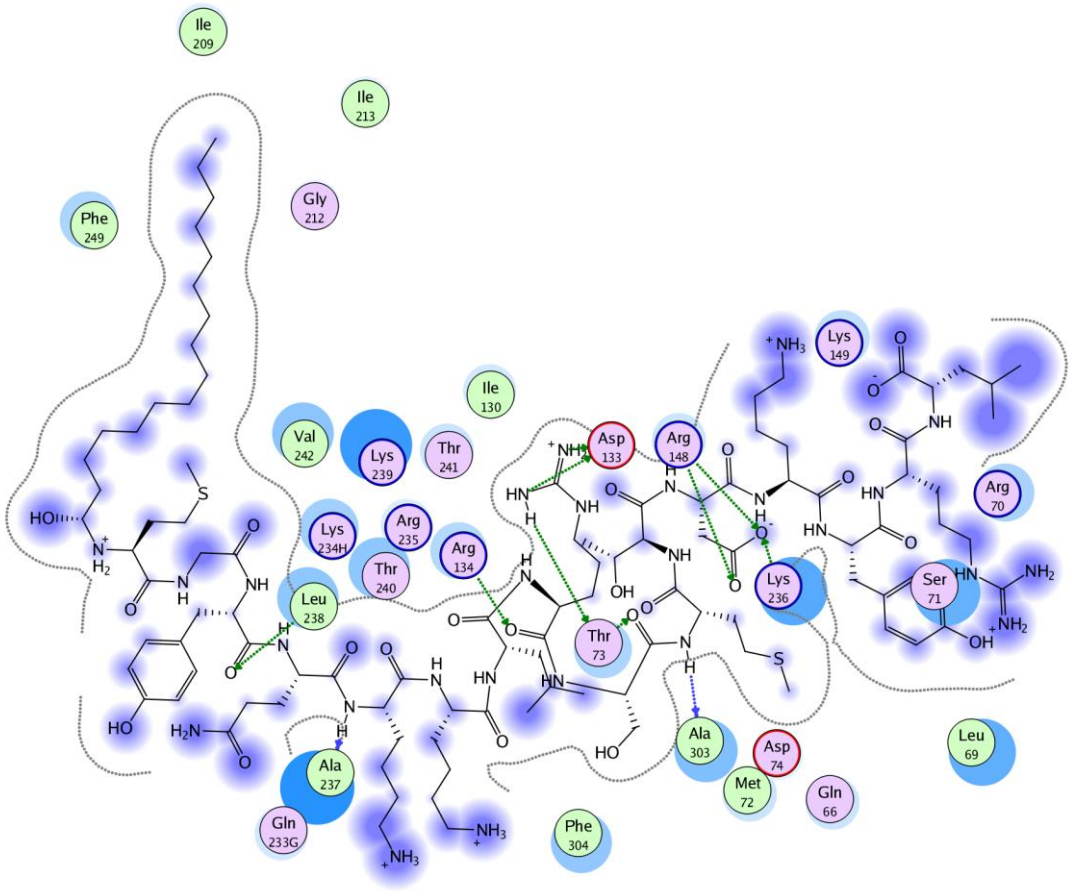


Figure 7 Protein-ligand interactions established for the complexes CXCR4 - ATI-2341 generated by FlexPepDock (**A**), HADDOCK (**B**) and MD (**C**). Circles in magenta represent residues involved in acidic, basic or polar interactions. Circles in light green represent residues involved in non-polar interactions. A blue halo around the residue represents a solvent accessible surface; this halo is proportional to the solvent accessible surface. In blue dashed lines with arrow heads are represented hydrogen bonds with amino acid backbone chains and green dashed lines with arrow heads represent hydrogen bonds with amino acid side chains.

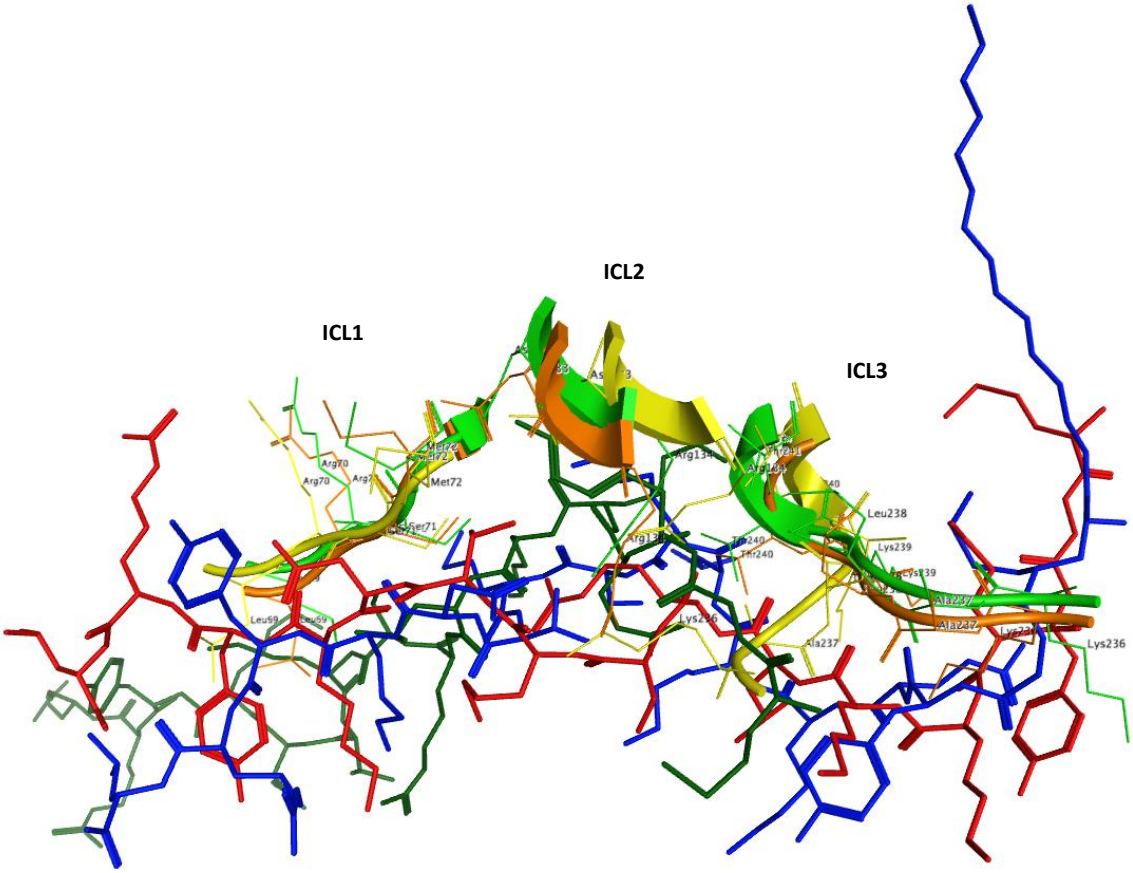
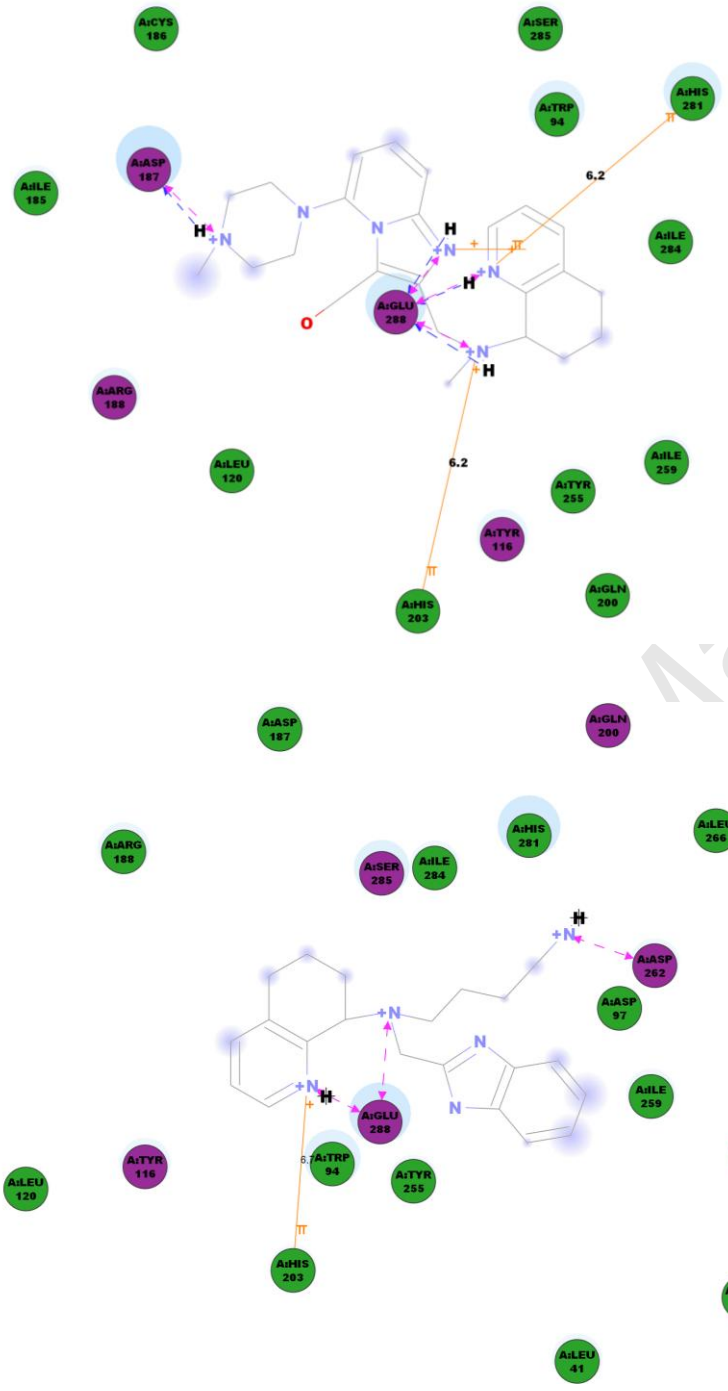
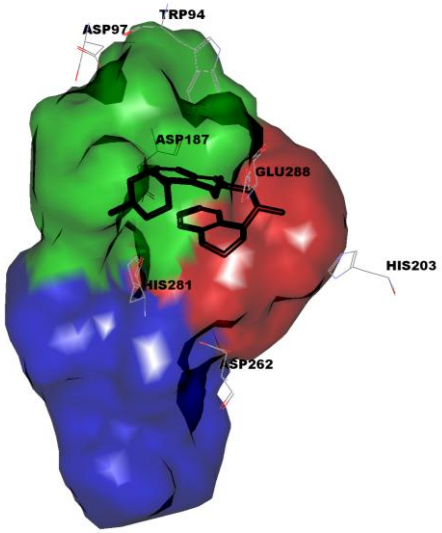


Figure 8 Superposed pepducin ATI-2341 conformations returned by FlexPepDock in red, by Haddock in green and by MD in blue. Representation generated by protein superposition.

A



B



C

D

Figure 9 Protein ligand interaction charts for the best-scored LigandFit-DockScore pose in binding subsite 1-2. The best pose corresponds to structure GSK812397 (**A**) and AMD11070 (**C**). The chart shows π interactions in orange lines. Charge interactions are shown in pink dashed lines with arrowheads on both sides. In blue dashed lines with arrow heads towards the electron donor are represented hydrogen bonds. Circles in magenta represent residues involved in hydrogen bond, charge or polar interactions. Circles in green represent residues involved in van der Waals interactions. A blue halo around the residue represents a solvent accessible surface; this halo is proportional to the solvent accessible surface. GSK812397 (**B**) and AMD11070 (**D**) best docked pose appeared docked in the same binding subsite 1-2.