



HAL
open science

Applying in silico Tools to the Discovery of Novel CXCR4 Inhibitors

Violeta I. Pérez-Nueno, David Ritchie

► **To cite this version:**

Violeta I. Pérez-Nueno, David Ritchie. Applying in silico Tools to the Discovery of Novel CXCR4 Inhibitors. Drug Development Research, 2010, 10.1002/ddr.20406 . inria-00550645

HAL Id: inria-00550645

<https://inria.hal.science/inria-00550645v1>

Submitted on 29 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Applying *in silico* Tools to the Discovery of Novel CXCR4 Inhibitors

Violeta I. Pérez-Nueno and David W. Ritchie

INRIA Nancy – Grand Est, LORIA, 615 rue du Jardin Botanique, 54506 Vandoeuvre-lès-Nancy,
France

Tel: + +33-3-83593045. Fax: + +33-3-83413079. E-mail: violeta.pereznueno@loria.fr;
dave.ritchie@loria.fr

Key words: virtual screening; HIV; CXCR4/CCR5 inhibitors; GPCRs; homology modelling;
docking; shape matching; pharmacophore; QSAR; similarity search; *de novo* design

37 Text Pages, 8 Figures and 1 Table

ABSTRACT

The process of HIV entry begins with the binding of the viral envelope glycoprotein gp120 to both the CD4 receptor and one of the CXCR4 or CCR5 chemokine coreceptors. There is currently considerable interest in developing novel ligands which can bind to these coreceptors and hence block virus-cell fusion. This article reviews the use of different *in silico* structure-based and ligand-based virtual screening (VS) tools for the discovery of potential HIV entry inhibitors for the CXCR4 receptor. More specifically, it discusses homology modelling, *de novo* design, docking, QSAR analyses, pharmacophore modelling, and similarity searches. Results from retrospective VS of a library of known CXCR4 inhibitors taken from the literature and from prospective VS of a combinatorial virtual library are reviewed. The structures of active compounds found by these approaches, as well as CXCR4 inhibitors currently in development are also discussed.

INTRODUCTION

Acquired Immune Deficiency Syndrome (AIDS) has become a global disease. According to the World Health Organization, an estimated 33.4 million people worldwide have AIDS. AIDS is caused by infection of host cells by the Human Immunodeficiency Virus (HIV). The process of HIV entry begins with the binding of the gp120 viral envelope glycoprotein to both the CD4 cell surface receptor and one of CXCR4 or CCR5 chemokine coreceptors, which leads to fusion of the viral capsid with the cell membrane. Current antiretroviral therapies (ARTs) against AIDS are generally based on reverse transcriptase inhibitors and protease inhibitors. Despite advances in the development of these potent agents which block HIV transcription and assembly, there remain problems regarding drug resistance, latent viral reservoirs, and drug induced toxic effects, which can all compromise effective control of the virus. Hence there is a need to develop new classes of anti-HIV drugs with different modes of action.

HIV entry inhibitors have emerged as a new generation of antiretroviral drugs that block viral fusion with the CXCR4 and CCR5 membrane co-receptors [De Clercq, 2005; Bean, 2005]. Several small molecule antagonists for these co-receptors have been developed, e.g. AMD3100 (Plerixafor) which is a CXCR4 antagonist, and Maraviroc (Selzentry®) which is a CCR5 inhibitor. Several further entry inhibitors are currently in clinical trials, e.g. the CXCR4 inhibitors AMD3465, AMD11070, CS-3955, KRH-2731, the CCR5 inhibitors SCH-C (SCH-351125), SCH-D (SCH-417690), GW-873140, ONO-4128, AK-602, TAK 779, TAK 220, TAK 652, and also the dual inhibitor AMD3451 [Tauzin, 2009; ClinicalTrials.gov]. However, because no crystal structures for the co-receptor proteins are available, the binding modes of the known inhibitors within the co-receptor extracellular pockets must be analyzed using site-directed mutagenesis (SDM) and homology modelling. Generally, the objective of these computational approaches is to screen large numbers of candidate drug compounds rapidly. Virtual screening (VS) has recently become a useful complement to laboratory-based high-throughput screening methods for large libraries of compounds. Several receptor-based and ligand-based VS approaches have been used to find CXCR4 and CCR5 antagonists which could potentially serve as HIV entry inhibitors [Pérez-Nueno et al., 2008a; Pérez-Nueno et al., 2008b; Pérez-Nueno et al., 2009; Carrieri et al., 2009]. *De novo design* methods may also be applied to identify new HIV entry blockers although this is difficult because only limited structural information is available. This

paper reviews the use of *in silico* ligand-based and structure-based VS tools to discover novel CXCR4 inhibitors.

CXCR4 is a seven transmembrane G-protein coupled receptor (GPCR) which acts as the co-receptor for T-trophic strains of HIV. GPCRs are the largest family of transmembrane (TM) receptor proteins and are involved in many physiological processes in humans. They are also extremely important pharmaceutical targets [Klabunde and Hessler, 2002]. Currently, 50% of marketed drugs target the GPCR family [Howard et al., 2001]. However, the full potential of GPCR therapeutics has not been achieved due to the difficulties in determining their structures by X-ray crystallography [Howard et al., 2001]. At the time of writing this review, the crystal structures of only six GPCRs have been published in the Protein Data Bank (PDB) [Mustafi et al., 2009]. Just recently, the first crystal structure of CXCR4 has been solved. Five independent crystal structures of CXCR4 bound to an antagonist small molecule IT1t and a cyclic peptide CVX15 at 2.5 to 3.2 angstrom resolution have been reported [Wu et al., 2010].

Despite the lack of structural data until now, CXCR4 homology models are useful for receptor-based VS, and binding mode analyses have been carried out using such models and known high affinity ligands [Wong et al., 2008; Pérez-Nueno et al., 2008a; Pettersson et al., 2008; Pettersson et al., 2010]. Furthermore, the performance of docking-based and ligand-based VS of CXCR4 inhibitors has been compared [Pérez-Nueno et al., 2008a]. Pharmacophore modelling and several shape-based and property-based approaches using high-affinity ligands as query molecules have been shown to be successful in finding CXCR4 inhibitors [Pérez-Nueno et al., 2008a; Pérez-Nueno et al., 2009]. However, it is worth mentioning that although several groups have performed CXCR4 homology modelling and binding mode analyses, only a few groups have done VS. For this reason, much of this review refers to our own CXCR4 VS results using our own purpose-built libraries.

Receptor-based and ligand-based methods have been used for both retrospective and prospective VS. The former used a large database of known CXCR4 inhibitors and similar presumed inactive molecules assembled from the literature [Pérez-Nueno et al., 2008a], and the later used a combinatorial library derived from the AMD3100 reference ligand [Pettersson et al., 2008; Pérez-Nueno et al., 2009]. Our VS library of CXCR4 ligands and decoys is essentially the largest, and this should become a useful resource for the community. Both libraries were

queried using known binders, and the enrichment factors and diversity of the resulting virtual hit lists were analyzed. Receiver-operator-characteristic (ROC) analyses were used to compare the performance of all the screening approaches. Retrospective analyses are useful to validate the different VS approaches for a given target and to set the best parameters for the screening process. Consensus ranking of ligand and structure based approaches were then used to obtain prospective virtual hit lists. This approach has successfully identified novel CXCR4 inhibitors with activity values in the range of 20 to 0.008 $\mu\text{g/ml}$ [Pérez-Nueno et al., 2009].

G Protein Coupled Receptors

GPCRs are located at the cell surface and are responsible for the transduction of an endogenous signal into an intracellular response. GPCRs comprise a large protein super-family sharing a conserved structure of seven transmembrane (TM) helices. These helices surround a small molecule or peptide binding site within the TM domain. The helices are linked by three extracellular loops (ELs) and three intracellular loops (ILs), as illustrated schematically in Figure 1. Extracellular loops vary significantly and generally play an important role in ligand recognition.

Figure 1

Until recently, bovine rhodopsin was the only GPCR structure to have been solved by crystallography [Palczewski et al., 2000]. Knowledge of how drugs interact with GPCRs was therefore limited to models based on homology and SDM experiments. In the last three years, three new GPCR structures have been solved, namely the $\beta 1$ and $\beta 2$ adrenergic receptors, and the adenosine A_{2a} receptor [Cherezov et al., 2007; Rasmussen et al., 2007; Jaakola et al., 2008; Warne et al., 2008]. These structures have provided further insights into the three-dimensional (3D) topology of GPCRs, allowing more accurate GPCR homology models to be built [Michino et al., 2009].

Although the GPCR signalling mechanism is not clearly understood, it is believed that when an endogenous agonist binds to the TM binding site, it causes conformational changes to be propagated through highly conserved “conformational switches” in the TM helices, leading to G-protein signalling [Nygaard et al., 2009]. This implies that all GPCRs possess a common

conformationally flexible TM region which allows them to move from an unbound or antagonist-bound state to an agonist-bound signalling state.

For the CXCR4 and CCR5 co-receptors, it is now thought that chemokines activate the receptor at least in part via its N-terminal residues near the trans-membrane region, and that CCR5 and CXCR4 antagonists allosterically prevent switching to the agonist-bound state [Kondru et al., 2008; Wong et al. 2008].

As well as the inevitable inaccuracies inherent in homology models, an additional complication when modelling GPCR structures by homology has been that until recently all of the available structures are in the “unactivated” conformation. Bissantz et al. [2003] propose that homology models based on the unactivated state of rhodopsin could be suitable for structure-based VS of antagonists but not of agonists. On the other hand, Shim et al. [2003] and Evers et al. [2003a,b] have developed knowledge-based and pharmacophore-based approaches to predict activated agonist-bound forms of the receptor which could extract known agonists from a VS library. An important recent development has been the solution of the structure of opsin, the ligand-free form of rhodopsin [Park et al., 2008; Scheerer et al. 2008], which is found to be quite different from the known GPCR structures. Opsin's most prominent features are the activating movement of TM6, rearrangements in regions of TM5 and TM6, and restructuring of the helix 8 kink of TM7, which correspond to an active or partially active conformation. This suggests that future structure-based modelling exercises should use an activated or unactivated GPCR template for screening antagonists or agonist inhibitors, respectively.

CXCR4 HIV Entry Inhibitors

The CXCR4 entry inhibitors can be divided into three main groups: small antagonist molecules [Mosley, et al., 2009], large and small peptide antagonists [Tamamura et al., 2001; Tamamura et al., 2005; Cluzeau et al., 2007], and chemokine-based agents [Shaeen et al., 2004; Kazmierski et al., 2005]. Among the small molecule antagonists, there are at least ten families, consisting of tetrahydroquinolinamine derivatives (with two main categories: the benzimidazoles and imidazopyrimidines) [Bridger et al., 2002; Bridger et al., 2003; Bridger et al., 2004], phenanthroline derivatives [Catalano et al., 2010; Jenkinson et al., 2010], KRH derivatives [Ichiyama et al., 2003; Murakami et al., 2004; Yamazaki et al., 2003; Yamazaki et al., 2004],

macrocycles [Bridger et al., 1999; Bridger et al., 2010], AMD bicyclam derivatives [Bridger et al., 1999; De Clercq, 2000; Esté et al., 1999; Egberink et al., 1999; Hatse et al., 2005; Princen et al., 2004], dipicolil amine zinc(II) complexes [Tamamura et al., 2006], peptidomimetic compounds consisting of a common peptidic backbone with various alkyl nitrogenated branches [Kureha, 2004], isothiourea derivatives [Novartis, 2007], pyrimidine derivatives [Ono, 2007; Taigen, 2006], and amine-based derivatives, either noncyclam amines, such as ethanediamine polyamine compounds [Taigen, 2004; Moyle et al., 2007] and diamine compounds [Teixidó et al., 2006; Pettersson et al., 2008], or monocyclam amines [Pettersson et al., 2010; Rosenkilde et al., 2007]. Figure 2 shows some representative molecules of each family. To date, VS tools applied to CXCR4 inhibitors have focused on finding new antagonist [Pérez-Nueno et al., 2008a; Pérez-Nueno et al., 2009].

Figure 2

Pre-filtering and Library Generation

When applying *in silico* tools in drug discovery, it is advisable to start with a pre-filtered library of compounds to eliminate structures which do not have drug-like functional groups or physical properties. Typical drug-like criteria include the Lipinski rule of five, the presence of functional groups in common established drugs, the absence of toxic functional or unstable groups, and good ADMET properties (absorption, distribution, metabolism, excretion, and toxicity). It is also possible to filter on metabolic parameters such as *e.g.* blood brain barrier (BBB) permeation, cytochrome P450 metabolism, albumin binding, and aqueous and DMSO solubility. Using these pre-filters allows the search to be reduced from hundreds of thousands or millions of compounds to a more reasonable number.

Virtual libraries can also be generated by using combinatorial chemistry techniques to preserve the main features of a lead drug compound. This ensures that chemical functional groups and physical properties are similar to those of the lead drug structure. For example, Pettersson et al. [2008] and Zhan et al. [2007] designed two different combinatorial libraries based on AMD3100 to find novel CXCR4 antagonists. The first library was more restrictive, whereas the second library had greater scaffold diversity.

Receptor-based *In Silico* Tools Applied to CXCR4

Homology modelling

In general, structural homology models are considered to be reasonably reliable if the sequence identity is >40% [Schaffner et al., 2001]. However, like other GPCR targets, the sequence similarity of CXCR4 and CCR5 with the template is generally much lower than this (20%) [Vaidehi et al., 2002]. For example, several homology models have been built for CXCR4 [Zhou et al., 2001; Huang et al., 2003; Trent et al., 2003; Gerlach et al. 2001; Rosenkilde et al., 2004; Gerlach et al. 2003; Pérez-Nueno et al., 2008a; Singh et al., 2007], all of which have approximately 21% identity between CXCR4 and the GPCR template. Homology modelling protocols often begin by predicting TM segments using different methods such as HMMTOP, TMHMM, MEMSAT or DAS. Secondly, the TM segments are aligned with those of the template. Third, the loop segments are aligned in the same way. Finally, the 3D structure of CXCR4 is built by transferring equivalent conserved regions from the template and by using *ab initio* techniques to model any poorly conserved loop regions. The loops are especially difficult to model because their sequences are very variable. Hence it is common to refine them after homology modelling by *ab initio* methods. For example, we used MODELLER to align the TM and loop segments and build the 3D structure, and CONGEN to refine the loop regions *ab initio* [Pérez-Nueno et al., 2008a]. It is worth mentioning that the conformation of CXCR4 extracellular loop two (EL2) is important for ligand recognition. Ligands can bind in the extracellular pocket only when EL2 has an open conformation. Hence, it requires careful attention when modelling to be sufficiently open to allow ligand binding.

CXCR4 small antagonists binding site and binding mode Analyses

AMD3100 is often considered as an archetypical non-peptidic CXCR4 antagonist inhibitor. Both experimental and computational binding mode studies have been carried out on the AMD3100-CXCR4 complex. Mutagenic substitutions of 16 amino acid residues in TM3, TM4, TM5, TM6 and TM7 helices have identified three acidic residues: Asp171, Asp262 and Glu288 as key interaction points for AMD3100. Two of these are at one side of the extracellular pocket

(Asp262 in TM6 and Glu288 in TM7) and the third is at the opposite side (Asp171 in TM4). Both Asp171 and Asp262 have been identified as essential for recognition by the viral gp120 protein [Brelot et al., 2000]. Furthermore, SDM results show that when metal species are bound to the AMD3100 cyclam rings, the affinity to CXCR4 is increased. Although all three residues are very important for binding, it seems that the metal effect is mediated preferentially by only Asp262 [Gerlach et al., 2003].

According to previous molecular modelling work it seems that the cyclam rings in a bicycam compound bind preferentially to certain carboxylic acid groups of TM helices 4, 6, and 7 in the pocket. These interactions may be in at least three different ways: a) two nitrogens of one cyclam ring interact with the two oxygens of Asp171 (TM 4), and two nitrogens of the other cyclam ring interact with the two oxygens of Asp262 (TM 6) [Gerlach et al. 2001]; b) two nitrogens of one cyclam ring interact with the two oxygens of Asp262, and two nitrogens of the other cyclam ring interact with the two oxygens of Glu288 (TM 7) [Trent et al., 2003]; c) two nitrogens of one cyclam ring interact with the two oxygens of Asp171, and the other cyclam ring is sandwiched between Asp262 and Glu288 [Liang et al., 2002; Rosenkilde et al., 2004].

Regarding the monocyclam compounds, Rosenkilde et al. [2007] suggested that AMD3465 seems to mimic the AMD3100 cyclam interactions whereby the noncyclam moiety has additional favourable interactions with residues near the extracellular end of TM 6 and TM 7 (*e.g.* His281).

It is worth mentioning that two types of docking analyses can be useful. Binding site and binding mode analyses allow validation of homology modelled receptors. For example, if a good 3D model of a receptor can be built, then an automatic docking protocol should be able to locate correctly a high affinity ligand within the binding site. Hence, binding site analyses can confirm the location of the receptor binding site by blind docking [Hetényi and Van der Spoel, 2002], whereas binding mode analyses allow the orientation and conformation of the ligand in the binding site to be identified by conventional docking techniques. Perez-Nueno et al. [2008a] performed both binding site and binding mode analyses of AMD3100 on a CXCR4 homology model using AUTODOCK, obtaining results which agree with experimental SDM data. The binding site analyses used a large grid enclosing the entire protein structure with the ligand initially placed far from the protein to avoid excluding the possibility of finding other binding sites.

Binding mode analyses used a localized grid with the ligand already placed on the SDM-defined ligand binding site. Figure 3 shows a close-up view of the lowest energy AMD3100-CXCR4 binding conformation obtained in docking binding mode analyses. Wong et al. [2008] and Pettersson et al. [2010] have also reported binding mode analyses for bicyclam, monocyclam and noncyclam compounds using MOE and AUTODOCK, respectively. Both agree that none of the dockings into CXCR4 can explain all mutant results by a direct ligand-receptor interaction. That is to say, none of the predicted binding modes involved all of the residues known to affect the binding of the well-known CXCR4 bicyclam, monocyclam or noncyclam inhibitors. In other words, it seems that different ligands bind in different ways within the SDM-defined pocket. Furthermore, molecular dynamics (MD) simulations are commonly used to refine docked binding poses by allowing flexible motions of both receptor and ligand, and by taking into account solvation effects. CXCR4 antagonist binding modes refined by molecular dynamics have also been reported in the literature [Zhou et al. 2001; Huang et al., 2003; Singh et al., 2007; Pettersson et al., 2010].

Figure 3

Wong et al. [2008] performed both experimental binding studies and molecular modelling/docking experiments comparing bicyclam, monocyclam, and noncyclam binding modes. Their results suggest that these CXCR4 inhibitors bind to overlapping but not identical residues in the TM regions of the receptor. Similar results were obtained by Pettersson et al. [2010]. Moreover, Wong et al. identified residues involved in unique interactions with the AMD11070 noncyclam inhibitor and the AMD3465 monocyclam, which suggests an extended binding pocket in the transmembrane regions close to EL2. Furthermore, taking into account the proposed new interaction sites, Wong et al. suggested the possibility of multiple binding modes for the bicyclam (AMD3100), monocyclam (AMD3465), and noncyclam (AMD11070). It could be conceivable that two compounds might bind simultaneously to the inhibitor binding pocket, but their heterologous competitive binding curves indicate that only one inhibitor binds to one receptor molecule. Nevertheless, as Wong et al. suggest, further investigation of the stereochemistry of inhibitor binding would be interesting.

Docking-based VS

Docking-based VS approaches aim to predict how compounds in a database might bind to the active site of a receptor. A docking VS protocol is often characterized by two aspects: the method used to sample the ligand-receptor conformational space, and the scoring function used to evaluate the binding affinity. Perez-Nueno et al. [2008a] performed retrospective docking-based VS on CXCR4 antagonist inhibitors using AUTODOCK, GOLD, FRED, and HEX, and constraining the ligands to form a hydrogen bond to one of the Glu288, Asp171, or Asp262 oxygens in GOLD. All of the ligands were initially flexibly aligned in MOE to the docked AMD3100 obtained in binding mode analyses in order to be rigidly docked using FRED. Figure 4 shows the enrichment curves obtained using these docking-based scoring functions. This shows that our docking-based VS results are comparable to those of other docking-based VS exercises on different model-built GPCR targets [Kellenberger et al. 2007; Bissantz et al., 2000; Evers et al., 2005]. Figure 4 also shows screening curves obtained when applying the retrospective protocol to the combinatorial virtual library designed by Pettersson et al. [2008]. These are discussed in further detail below.

Figure 4

Structure-based de novo design

Structure-based *de novo* design allows new molecules to be designed using knowledge of the active receptor site or a structure-based 3D pharmacophore. *De novo* design methods mainly fall into two categories. Energy-based approaches place fragments in the active site by energetically exploring the space or by means of simulation techniques. On the other hand, knowledge-based methods use information about geometry and preferential interactions observed in crystal structures to identify the active site regions where the binding of fragments is favourable by hydrogen bond or hydrophobic interactions. After fragments have been placed in the active site, they need to be connected. This is the most problematic step in *de novo* design given the high quantity of interconnection possibilities and the difficulty in evaluation synthetic

feasibility. Retro synthetic software has been developed to help identify compounds obtained from *de novo* techniques which are synthetically accessible [Gasteiger et al., 1994a].

In principle, a well designed homology model coupled with SDM information about key binding site residues should be able to deliver a reliable pharmacophoric representation of the target. However, *de novo* design of GPCR ligands is especially challenging because of the limited structural information available for the target proteins. In particular, structure-based *de novo* design for CXCR4 inhibitors has not been explored thoroughly. Nonetheless, Pérez-Nueno [2009] performed structure-based *de novo* design of CXCR4 inhibitors using the LUDI approach implemented in Discovery Studio. In this study, receptor interaction sites were generated on CXCR4 binding pocket according to key SDM-defined binding residues (Asp171, Asp262, and Glu288). The LUDI approach yielded a total of 200 new ligands based on the 20 highest-ranked scaffolds from standard calculations. The range of LUDI scores were from 1040 (corresponding with a predicted K_i of about 0.01 μM to 280 μM). Due to the difficult synthetic feasibility of some of the proposed molecules, a virtual combinatorial library of synthetically accessible compounds was designed. This used commercially available fragments to virtually build the compounds focusing on *de novo* selected fragments, distances between heteroatoms and chemistry of *de novo* ligands. The final library was virtually screened in order to select the best compounds to be synthesized and tested. Their synthesis and biological test is under current development.

Ligand-Based *In Silico* Tools Applied to CXCR4

Ligand-based approaches exploit the fact that molecules which are structurally related often show similar biological activities. Ligand-based VS techniques are fast and efficient, and they are especially useful when the structure of the receptor is not available. However, ligand-based design can become complicated when different classes of agonists or antagonists can interact with different sets of receptor site points, which usually corresponds to ligands with very different chemistry and consequently difficult to superpose, or when the ligand sets are particularly flexible. It is also worth noting that small structural differences can sometimes cause large differences in activity.

2D ligand-based similarity searches

Ligand-based similarity approaches start from the description of one or several lead structures using one or more structural descriptors, along with similar descriptions of the compounds in a virtual library. These descriptors are often termed “1D” if they describe only extensive quantities such as molecular weight, “2D” if they include topological connectivity, and “3D” if they encode the 3D structure of a molecule. Many metrics exist to measure similarity and distance using these attributes. Their behaviour depends on the set of descriptors used and molecules to compare. Chemical similarity searches for CXCR4 and CCR5 inhibitors have been carried out by Pérez-Nueno [2009] and Carrieri et al. [2009] using QIKPROP to calculate 1D, 2D, and 3D descriptors for all compounds. The Tanimoto similarity coefficient was calculated using QIKSIM for all database compounds with respect to a given lead compound and the average of physico-chemical and biological properties of a set of actives.

3D pharmacophore screening

A pharmacophore can be defined as a set of structural characteristics of one or more ligands which are directly related to the recognition of a ligand in the binding site of the receptor. Often, a pharmacophore model starts from an alignment of active molecules to identify common pharmacophoric groups and thus to obtain the spatial configuration of the key chemical features responsible for binding. The pharmacophoric groups commonly used are atoms with positive and negative charges, hydrogen bond acceptors and donors, and atoms with hydrophobic character. Once a pharmacophore model has been built, it can be used to search for compounds which contain the same pharmacophore in a database. When designing the pharmacophore model, it is possible to consider the flexibility of the starting actives (templates) and their superposition, as well as the number and conservation of the pharmacophoric features that must be present in the aligned molecules. Programs have been developed to derive hypotheses automatically based on superpositions and multiple alignments, although user intervention is still necessary to select the best proposal.

Ligand-based pharmacophore approaches have been widely used in the GPCR area using both small molecules and peptides to generate the pharmacophore model. Regarding CXCR4 entry inhibitors, Pérez-Nueno et al. [2009] built a pharmacophore model for CXCR4 antagonist inhibitors using MOE and Discovery Studio software suites with four families of known actives,

namely the AMD3100 derivatives, KRH1636 derivatives, dipicolil amine zinc(II) complexes, and the diamine derivatives. Figure 5 shows the pharmacophore model which gave the best balance between sensitivity and specificity (giving no false positives and only nine false negatives) in a retrospective VS of 151 known CXCR4 antagonists. This study also found that even better VS results could be obtained by using a consensus of five models. This consensus was applied prospectively to the combinatorial virtual library designed previously by Pettersson et al. [2008] obtaining interesting hits (see Table 1).

Figure 5

Shape matching

Shape matching approaches are based on the superposition and comparison of the 3D shapes of a set of molecules against a known active molecule. Compounds that have a shape similar to that of a well-known active will have a good probability of fitting in the biological receptor and may potentially have greater activity. The biggest problem with shape-matching techniques is the selection of the initial query conformation.

Often, the crystallographic conformation of the complexed ligand is used as the query. However, if there is no crystallographic information available for a target, computational methods have to be used. Usually, the lowest energy computed conformation is used in these cases, or different query conformations are calculated and the one most similar to superposed compounds is used in each case. Another limitation of shape-matching approaches is that if a target has an active site which can fit ligands in different ways, the query selected will represent only a single binding mode. Therefore, only those compounds in the database with this single shape will superpose correctly onto the query. Hence, other potentially active compounds may be missed. Pérez-Nueno et al. [2008b] proposed a spherical harmonic (SH) consensus shape matching algorithm to help solve this problem. In this approach, the shape of a consensus (or average) pseudo-molecule is calculated from the SH representation of each active. Hence, consensus shape can capture the essential features of several known high-affinity ligands and encode these in the form of a single representative pseudo-molecule. Figure 6 shows the consensus shape of the three most active compounds in our dataset of 248 high affinity CXCR4 antagonists.

Figure 6

Perez-Nueno et al. [2008a] performed retrospective shape matching VS on CXCR4 antagonist inhibitors using ROCS shape-only and shape plus chemistry (combo), PARAFIT consensus shape and shape-only, and HEX shape-only. Figure 7 shows the enrichment curves obtained using these shape matching approaches for both retrospective experiment and when applied to the virtual library designed by Pettersson et al. [2008].

Figure 7

3D-QSAR

3D-QSAR methods aim to correlate chemical structures with biological activities of a series of related compounds. They use a number of location-dependent measures that describe molecular properties without any explicit calculation of their interaction with the targets. 3D-QSAR models are derived from multiple structural and spatial descriptors combined with linear and non-linear (genetic algorithms or neuronal networks) optimization techniques and multivariate analysis techniques PLS (partial least squares). Generally, the derivation of descriptors for 3D-QSAR models takes into account the calculation of autocorrelation vectors of molecular electrostatic potentials (MEPs) mapped over the molecular surfaces of the ligands [Gasteiger et al., 1994b], as well as molecular field descriptors based on the description of receptor-ligand interactions by means of molecular interaction potentials (MIPs). Among the various 3D-QSAR approaches, comparative molecular field analysis (CoMFA) is one of the earliest examples and is still in widespread use. CoMFA relies on a 3D description of molecular properties based on analysis of molecular fields detected by a probe atom and mapped on a grid [Cramer et al., 1998]. Later, CoMSIA (comparative molecular similarity indices analyses) were developed, in which molecular fields are expressed in the form of Gaussian-type functions [Klebe et al., 1994]. The first account of using CoMFA to GPCRs is probably given by Greco et al. [1991]. The later publication of the rhodopsin crystal structure gave impetus to performing 3D-QSAR techniques, as receptor docking became a practical way of superimposing ligands in

the 3D space. As an alternative to CoMFA, GRID and GOLPE programs [Goodford, 1985] may be used to generate predictive GPCR 3D-QSAR models [Brea et al., 2002; Broer et al., 2003; Audouze et al., 2004].

Regarding QSAR analysis of CXCR4 inhibitors, Pérez-Nueno et al. [2009] carried out a QSAR study on a dataset of 39 compounds with known EC₅₀ activity values consisting of AMD3100 plus 38 diamine derivatives synthesized by Pettersson et al. [2008]. This dataset was divided into a training subset of 30 compounds and an external test set of 9 compounds. A total of 194 descriptors were calculated with MOE, including 2D and 3D descriptors. These descriptors were then pruned using correlation analysis, forward-selection and backward-elimination methods.

Partial Least Squares (PLS) regression was used to build the QSAR models using the above descriptors as independent variables and using the biological activities as the dependent variables. A Grubbs test was used to detect significant outliers which were discarded. The model was then validated using leave-one-out (LOO) cross-validation and an external test set of 9 compounds. Several statistical parameters were used to evaluate the quality of the model, namely the R² correlation coefficient, cross-validated R² and test set validation R² against an external dataset, the root mean squared error (RMSE) for the model, the cross-validation, and the external test set validation, and the Fisher test. After descriptor pruning had been applied, five descriptors were selected to build the QSAR models, namely: VAdjEq, Q_VSA_HYD, dipoleY, SlogP_VSA8 and FASA+. Three models were built which all showed broadly similar statistical results. One model used only three descriptors, whereas the other two required four descriptors. Hence, the model with less descriptors and better statistics was selected for prediction of activity values for compounds in Pettersson et al. [2008] combinatorial library. The use of *dipoleY*, an external 3D descriptor, as independent variable in the three models enhanced the importance of a correct alignment of the molecules in order to obtain a reliable predicted activity value. The final selected model was calculated as:

$$pEC_{50} = 2.52586 + 0.00940 \cdot (Q_VSA_HYD) + 0.00507 \cdot (SlogP_VSA8) + 0.10611 \cdot (dipoleY)$$
$$N=29, R^2=0.81, RMSE=0.42, F=36.45, R^2_{LOO}=0.75, RMSE_{LOO}=0.49, R^2_{test}=0.69, RMSE_{test}=0.57,$$
$$n=9, R_o^2=0.77, (R^2 - R_o^2)/R^2=0.049, k=0.99, PRESS=5.20, SSY=27.93, PRESS/SSY=0.19.$$

Figure 8 shows the resulting correlation between experimental and predicted pEC₅₀.

Figure 8

Ligand-based de novo design

Ligand-based *de novo* design allows new molecules to be built from aligned known actives. It uses pharmacophoric features to place fragments and hence assemble compounds that have similar features known to be critical/important for active ligands binding to a specific target. It uses structural and biological information of the bound conformation of superimposed known active ligands to generate a 'hypothetical receptor' within novel chemically diverse inhibitors can be built through the hybridisation and/or linking of the input structures [Schmidt et al., 2003; Pierce et al., 2004]. This approach has been implemented in several software [Murcko et al., 1997; Schneider et al., 2005], and molecular modelling suites such as MOE (BREED) and Discovery Studio (DS De Novo Ligand Builder). One particularly interesting development of *de novo* ligand design is the 'scaffold-hopping' approach described in Lloyd et al. [2004]. On the other hand, structure-based *de novo* design uses an interaction map created from the receptor active site, which give rise to lists of compounds built solely to fit into a specific target binding site. This can also be carried out using suites as MOE (MultiFragment Search) or Discovery Studio (De Novo Receptor, De Novo Link, and De Novo Evolution). Similarly, it is possible to integrate features derived from the receptor structure with those derived from known ligands to create a more complete model of the characteristics needed for ligand-receptor binding. This is especially useful in the context of designing CXCR4 inhibitors, as well as other GPCR inhibitors, because the receptor structure is not yet known in detail.

Combining Ligand and Structure-Based Methodologies

As well as *de novo* design, another way to combine receptor and ligand information is the direct guided docking approach of Fradera et al. [2004] which introduces receptor binding mode information in docking extracted from complexes co-crystallized with other ligands. It is normally assumed that the binding mode is conserved between different ligands, although this is not always the case. Another approach is the construction of pharmacophore models considering both the characteristics of active ligands and the structure of the active site of the receptor, especially if co-crystallized ligand-protein complexes are available. The pharmacophoric features extracted from the alignment of known active ligands are combined with those

extracted from the generation of a map of interactions of the active site of the receptor (hydrogen bond donors, hydrogen bond acceptors, and hydrophobic regions). Related to this, 3D-QSAR models can also use a combination of the ligand and receptor information.

A crucial element in the construction of 3D-QSAR models is the structural alignment of the ligands, which can be especially challenging when dealing with structurally diverse or highly flexible compounds. Hence, the utilization of alignment techniques based on binding site geometries and minimizations within binding sites can be an effective alternative to conformational searches and superpositions solely based on the molecular properties of the ligands. Another type of approaches is based on encoding the 3D protein-ligand contacts in bit string "fingerprints" derived from the number of residues/atoms in the binding cavity of the protein. Each bit denotes the presence (1) or absence (0) of a particular interaction: hydrogen bond, hydrophobic contact, or van der Waals. The implementation of these "interaction" fingerprints can vary depending on the definition of the bit string and the type of interactions considered. The interactions of a ligand co-crystallized with a receptor are translated to an interaction fingerprint, which is used as a reference for the comparison with the interaction fingerprints extracted from the best configurations obtained by the docking of all the molecules of the database to screen.

Receptor-based 3D-QSAR, in which receptor-ligand complexes are not used to directly calculate binding interactions, but only to generate the 3D alignment of the ligands, has been applied successfully to GPCRs, by means of docking studies conducted at homology models derived using rhodopsin as structural template [Constanzi et al., 2007; Moro et al., 2005; Moro et al., 2006; Jojart and Marki, 2007; Iskander et al., 2006].

Consensus Scoring

Another way to combine ligand and structure-based scoring information is consensus scoring. Consensus scoring combines the scores from different programs to try to find "the common ground" between compounds which rank highly across different algorithmic approaches. It has been demonstrated that consensus scoring reduces false positives when post-processing VS hit ranking lists. It can also be applied to choose ligands with the lowest binding energies from a given a set of ligands or to choose the best conformations from a set of docking solutions

[Ferrara et al., 2004]. There are different strategies for performing consensus scoring [Charifson et al., 1999; Wang et al., 2001]. For example, rank-by-number orders candidates according to the average values of all the scoring functions. Rank-by-rank orders compounds firstly by each individual scoring function, an average of the ranks for each of the compounds taking into account the different scoring functions is performed, and finally compounds are re-ranked according to their average ranking. Finally, rank-by-vote orders compounds present in the x% of re-ranked lists for all the individual scoring functions used in the combination.

Consensus ranking has been applied to CXCR4 inhibitors VS. Pérez-Nueno et al. [2009] applied a combination of ligand-based and receptor-based VS tools, and generated a ranked hit list using consensus rank-by-vote. The first ranked compounds obtained by the consensus rank-by-vote were selected and synthesized. Table 1 shows selected compounds using rank-by-vote consensus scoring for the prospective screening of Pettersson et al. [2008] virtual combinatorial library. Molecules selected had EC₅₀ activity values between 0.008 and 4.1 µg/ml. In that study, the ligand-based shape matching and pharmacophore tools performed better than the docking tools. However, looking at the first percentages of the ranked hit lists obtained, the compounds selected by the ligand-based methods were generally similar to those found by docking. Compounds selected by ligand-based VS tools (pharmacophore, shape matching, QSAR) were practically the same, whereas those selected by structure-based docking tools (AUTODOCK, GOLD, HEX, and FRED) also included some others. However, although ligand-based searches gave better results than structure-based docking for both retrospective and prospective VS analyses of the two databases of Pérez-Nueno et al. [2008a] and Pettersson et al. [2008], the pharmacophore models, compounds selected by QSAR and AUTODOCK docked energy gave the best correlation with experimental data.

Table 1

Future Trends and Challenges

There is no doubt that a major limitation of some current structure-based *in silico* tools is the use of rigid receptor models. It is clear that treating the receptor as a rigid entity must give way to methods which explicitly deal with receptor flexibility, and can thus address induced fit effects.

Of course, MD simulations can do this (LIE [Åqvist et al., 1994], MM-PBSA [Srinivasan et al., 1998]), but because of the high computational cost it is generally impractical to include an explicit solvation model or to cover the full conformational space of thousands of putative receptor-ligand pairs. It is also becoming increasingly apparent that using docking programs in high throughput modes do not sufficiently or rigorously sample the conformational space of the ligands as well [Erickson et al., 2004; Cecchini et al. 2004; Good et al., 2003]. Further current limitations of structure-based methods are the inadequate treatment of tautomerisation [Pospisil et al., 2003], and of water as both the biological solvent and as an explicit partner in ligand binding [Klebe, 2005; Gunther et al., 2003; Rush et al., 2005].

On the other hand, the general opinion in the field suggests that the docking component of the VS protocol is working quite well, but that the scoring techniques still need to be improved in order to be truly predictive [Ha et al., 2000; Warren et al., 2005]. Although docking approaches do not provide the level of accuracy necessary to rank properly the affinity of a series of analogues, they do have sufficient accuracy to allow discrimination between binders and non-binders, which make them very useful in VS. It is also worth mentioning that docking applied to homology models such as GPCRs is also promising. However, advances in solving more GPCR structures, coupled with the introduction of more sophisticated modelling algorithms and the increase in computer power, will open up further opportunities to target CXCR4 and other GPCRs using VS techniques. In this way, the crystal structure of CXCR4 recently solved [Wu et al., 2010] provides new clues about the interactions between CXCR4 and its natural ligand CXCL12 and with the HIV-1 glycoprotein gp120. This information is crucial for the improvement of VS techniques applied to GPCRs and, particularly, the discovery of novel CXCR4 inhibitors.

Regarding ligand-based *in silico* tools, these have been shown to be able to identify valuable starting points for GPCR lead generation. However, processes that depend on existing (possibly patented) structures to generate alternative ligands often run the risk of finding similar compounds. Hence, there is a need to develop new ligand-based tools to identify novel chemotypes [Lloyd et al., 2004; Schneider et al., 2000]. Ligand-based *de novo* design tries to solve this problem by searching for fragments which fit specific features needed for binding without using knowledge of the actives. However, these approaches still need to be improved, especially with respect to fragment linking and synthetic feasibility of the final assembled

molecules. With regard to QSAR approaches, they made possible the discovery of GPCR ligands before the publication of the experimentally elucidated rhodopsin structure that opened the way to homology modelling. Although QSAR approaches can generate accurate and robust models leading to precise ranking of the activity of a series of analogue compounds, these models are highly dependent on the training sets. Thus, they are not applicable to situations in which comparable activity data for a set of ligands of a specific target are not available, and they often fail to predict the activity of molecules whose structures differ significantly from those of training set.

CONCLUSIONS

CXCR4 is a GPCR, which today implies that homology modelling is needed to apply receptor-based *in silico* tools. Several homology models have been built using bovine rhodopsin because it was the first GPCR structure to be solved crystallographically. The availability of this structure facilitated receptor-based VS endeavours. Currently, three more GPCRs have been solved, which will allow better CXCR4 models to be built. Structure-based *in silico* tools applied to the discovery of novel CXCR4 inhibitors have been shown to give respectable results, selecting compounds similar to the ones selected by ligand-based *in silico* tools. Although structure-based *de novo* design is challenging for CXCR4, if a good homology model can be built, it can be used to search for *de novo* structures. Ligand-based *de novo* design can be used in combination with structure-based to build a more complete model. Regarding ligand-based approaches, QSAR models have been obtained which give accurate quantitative predictions. Screening large libraries of compounds using pharmacophore and shape matching approaches have also been shown to be useful for finding interesting hits relatively quickly. However, careful consideration of the shape-based and pharmacophoric features is necessary to obtain satisfactory results. Consensus scoring is also a useful way to combine receptor-based and ligand-based VS information.

We have used all of the above approaches in both retrospective and prospective VS experiments against CXCR4, and we have obtained novel hits with high activity values. Clearly, many *in silico* drug discovery tools are available nowadays, but our results show that one should

apply and combine them in appropriate ways for the particular target under study. For GPCR targets such as CXCR4, progress can also be made by making the best use of the available templates and by exploiting existing knowledge of known actives.

REFERENCES

Åqvist J, Medina C, Samuelsson JE. 1994. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng* 7:385-391.

Audouze K, Nielsen EO, Peters D. 2004. New series of morpholine and 1,4-oxazepane derivatives as dopamine D-4 receptor ligands: synthesis and 3D-QSAR model. *J Med Chem* 47:3089-3104.

Bean P. New Drugs Targets for HIV. 2005. *Clin Infect Dis* 41:96-100.

Bissantz C, Folkers G; Rognan D. 2000. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* 43:4759-4767.

Bissantz C, Bernard P, Hibert M, Rognan D. 2003. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets?. *Proteins: Struct Funct Gen* 50:5-25.

Brea J, Rodrigo J, Carrieri A, Sanz F, Cadavid MI, Enguix MJ, Villazon M, Mengod G, Caro Y, Masaguer CF, Ravina E, Centeno NB, Carotti A, Loza MI. 2002. New serotonin 5-HT(2A), 5-HT(2B), and 5-HT(2C) receptor antagonists: Synthesis, pharmacology, 3D-QSAR, and molecular modeling of (aminoalkyl)benzo and heterocycloalkanones. *J Med Chem* 45:54-71.

Brelot A, Heveker N, Montes M, Alizon M. 2000. Identification of Residues of CXCR4 Critical for Human Immunodeficiency Virus Coreceptor and Chemokine Receptor Activities. *J. Biol. Chem.* 275:23736-23744.

Bridger GJ, Skerlj RT, Padmanabhan S, Martellucci SA, Henson GW, Struyf S, Witvrouw M, Schols D, De Clercq E. 1999. Synthesis and Structure-Activity Relationships of Phenylenebis(methylene)-Linked Bis-azamacrocycles That Inhibit HIV-1 and HIV-2 Replication by Antagonism of the Chemokine Receptor CXCR4. *J Med Chem* 42:3971-3981.

Bridger G, Skerlj R, Kaller A, Harwing C, Bogucki D, Wilson TR, Crawford J, McEachern EJ, Atsma B, Nan S, Zhou Y. World Patent WO 0022600, 2002.

Bridger G, Skerlj R, Kaller A, Harwing C, Bogucki D, Wilson TR, Crawford J, McEachern EJ, Atsma B, Nan S, Zhou Y. World Patent WO 0022599, 2002.

Bridger G, Skerlj R, Kaller A, Harwing C, Bogucki D, Wilson TR, Crawford J, McEachern EJ, Atsma B, Nan S, Zhou Y. World Patent WO 00234745, 2002.

Bridger G, Skerlj R, Kaller A, Harwing C, Bogucki D, Wilson TR, Crawford J, McEachern EJ, Atsma B, Nan S, Zhou Y. World Patent WO 055876, 2003.

Bridger G, Skerlj R, Kaller A, Harwig C, Bogucki D, Wilson TR, Crawford J, McEachern EJ, Atsma B, Nan S, Zhou Y, Schools D, Smith CD, Di Fluir RM. U.S. Patent 7091217, 2004.

Bridger G, Skerlj RT, Hernandez-Abad PE, Bogucki DE, Wang Z, Zhou Y, Nan S, Boehringer EM, Wilson T, Crawford J, Metz M, Hatse S, Princen K, De Clercq E, Schols D. 2010. Synthesis and Structure-Activity Relationships of Azamacrocyclic C-X-C Chemokine Receptor 4 Antagonists: Analogues Containing a Single Azamacrocyclic Ring are Potent Inhibitors of T-Cell Tropic (X4) HIV-1 Replication. *J Med Chem* 53:1250-1260.

Broer BM, Gurrath M, Holtje HD. 2003. Molecular modeling studies on the ORL1-receptor and ORL1-agonists. *J Comp-Aided Mol Des* 17:739-754.

Carrieri A, Pérez-Nueno VI, Fano A, Pistone C, Ritchie DW, Teixido J. 2009. Biological profiling of anti-HIV agents and insights into CCR5 antagonist binding. *ChemMedChem* 4:1153-1163.

Catalano JG, Gudmundsson KS, Svolto A, Boggs SD, Miller JF, Spaltenstein A, Thomson M, Wheelan P, Minick DJ, Phelps DP, Jenkinson S. 2010. *Bioorg Med Chem Lett* 20:2186-2190.

Cecchini M, Kolb P, Majeux N, Caflich A. 2004. Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J Comput Chem* 25:412-422.

Charifson PS, Corkery JJ, Murcko MA, Walters PW. 1999. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J Med Chem* 42:5100-5109.

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS et al. 2007. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318:1258-1265.

ClinicalTrials.gov. United States National Library of Medicine National Institutes of Health, Health & Human Services. <http://clinicaltrials.gov/>

Cluzeau J, Oishi S, Ohno H, et al. 2007. Design and synthesis of all diastereomers of cyclic pseudo-dipeptides as mimics of cyclic CXCR4 pentapeptide antagonists. *Org Biomol Chem* 5:1915 -23.

Costanzi S, Tikhonova IG, Ohno M, Roh EJ, Joshi BV, Colson AO, Houston D, Maddileti S, Harden TK, Jacobson KA. 2007. P2Y1 antagonists: combining receptor-based modeling and QSAR for a quantitative prediction of the biological activity based on consensus scoring. *J Med Chem* 50:3229-3241.

Cramer RD, Patterson DE, Bunce JD. 1988. Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110:5959-5967.

De Clercq E. 2000. Inhibition of HIV Infection by Bicyclams, Highly Potent and Specific CXCR4 Antagonists. *Mol Pharmacol* 57:833-839.

De Clercq E. 2005. Emerging anti-HIV drugs. *Expert Opin Emerg Drugs* 10:241-274.

Egberink HF, De Clercq E, Van Vliet ALW, Balzarini J, Bridger GJ, Henson G, Horzinek MC, Schols D. 1999. Bicyclams, selective antagonists of the human chemokine receptor CXCR4, potently inhibit feline immunodeficiency virus replication. *J Virol* 73:346-6352.

Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. 2004. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J Med Chem* 47:45-55.

Esté JA, Cabrera C, De Clercq E, Struyf S, Van Damme J, Bridger G, Skerlj RT, Abrams MJ, Henson G, Gutierrez A, Clotet B, Schols D. 1999. Activity of Different Bicyclam Derivatives against Human Immunodeficiency Virus Depends on Their Interaction with the CXCR4 Chemokine Receptor. *Mol Pharmacol* 55:67-73.

Evers A, Klebe G. 2003a. Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew Chem Int Ed Engl* 43:248-251.

Evers A, Gohlke H, Klebe G. 2003b. Ligand-supported Homology Modelling of Protein Binding-sites using Knowledge-based Potentials. *J Mol Biol* 334:327-345.

Evers A, Hessler G, Matter H, Klabunde T. 2005. Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J Med Chem* 48:5448-5465.

Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. 2004. Assessing Scoring Functions for Protein-Ligand Interactions. *J Med Chem* 47:3032-3047.

Fradera X, Mestres J. 2004. Guided Docking Approaches to Structure-Based Design and Screening. *Curr Top Med Chem* 4:687-700.

Gasteiger J, Ihlenfeldt WD, Fick R, Rose JR. 1994a. Similarity Concepts for the Planning of Organic Reactions and Synthesis. *J Chem Inf Comput Sci* 15:793-813.

Gasteiger J, Li X, Rudolph C, Sadowsky J, Zupan J. 1994b. Representation of Molecular Electrostatic potentials by Topological Feature Maps. *J Am Chem Soc* 116: 4608-4620.

Gerlach LO, Skerlj RT, Bridger GJ, Schwartz TW. 2001. Molecular interactions of cyclam and bicyclam non-peptide antagonists with the CXCR4 chemokine receptor. *J Biol Chem* 276:14153-14160.

Gerlach LO, Jakobsen JS, Jensen KP, Rosenkilde MR, Skerlj RT, Ryde U, Bridger GJ et al. 2003. Metal ion enhanced binding of AMD3100 to Asp262 in the CXCR4 receptor. *Biochemistry* 42:710-717.

Good AC, Cheney DL. 2003. Analysis and optimization of structure-based virtual screening protocols (1): exploration of ligand conformational sampling techniques. *J Mol Graphics Modell* 22:23-30.

Goodford PJ. 1985. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J Med Chem* 28:849-857.

Greco G, Novellino E, Silipo C, Vittoria A. 1991. Comparative molecular field analysis on a set of muscarinic agonists. *Quant Struct-Act Relat* 10:289-299.

Gunther J, Bergner A, Hendlich M, Klebe G. 2003. Utilising structural knowledge in drug design strategies: applications using Relibase. *J Mol Biol* 326:621-36.

Ha S, Andreani R, Robbins A, Muegge I. 2000. Evaluation of docking/scoring approaches: a comparative study based on MMP3 inhibitors. *J Comput Aided Mol Des* 14:435-448.

Hatse S, Princen K, De Clercq E, Rosenkilde MM, Schwartz TW, Hernandez-Abad PE, Skerlj RT, Bridger GJ, Schols D. 2005. AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem Pharmacol* 70:752-761.

Hetényi C, Van der Spoel D. 2002. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Prot Sci* 11:1729-1737.

Howard AD, McAllister G, Feighner SD, Liu Q, Nargund RP, Van der Ploeg LH, Patchett AA. 2001. Orphan G-protein-coupled receptors and natural ligand discovery. *Trends Pharmacol Sci* 22:132-140.

Huang X, Shen J, Cui M, Shen L, Luo X, Ling K, Pei G, et al. 2003. Molecular dynamics simulations on SDF-1a: binding with CXCR4 receptor. *Biophys J* 84:171-184.

Ichiyama K, Yokohama-Kumakura S, Tanaka Y, Tanaka R, Hirose K, Bannai K, Edamatsu T, Yanaka M, Niitani Y, Miyako- Kurosaki N, Takaku H, Koyanagi Y, Yamamoto N. 2003. A duodenally absorbable CXC chemokine receptor 4 antagonist, KRH-1636, exhibits a potent and selective anti-HIV-1 activity. *Proc Natl Acad Sci USA* 100:4185-4190.

Iskander MN, Leung LM, Buley T, Ayad F, Di Iulio J, Tan YY, Coupar IM. 2006. Optimization of a pharmacophore model for 5-HT₄ agonists using CoMFA and receptor based alignment. *Eur J Med Chem* 41:16-26.

Jaakola V, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR et al. 2008. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322:1211-217.

Jenkinson S, Thomson M, McCoy D, Edelstein M, Danehower S, Lawrence W, Wheelan P, Spaltenstein A, Gudmundsson K. 2010. Blockade of X4-Tropic HIV-1 Cellular Entry by

GSK812397, a Potent Noncompetitive CXCR4 Receptor Antagonist. *Antimicrobial Agents and Chemotherapy* 54:817-824.

Jojar B, Marki A. 2007. Receptor-based QSAR studies of non-peptide human oxytocin receptor antagonists. *J Mol Graph Model* 25:711-720.

Kazmierski WM, Peckman JP, Duan M, Kenakin TP, Jenkinson S, Gudmundsson KS, Piscitelli SC, Feldman PL. 2005. Recent Progress in the Discovery of New CCR5 and CXCR4 Chemokine Receptor Antagonists as Inhibitors of HIV-1 Entry. Part 2. *Curr Med Chem Anti-Infect Agents* 4:133-152.

Kellenberger E, Springael JY, Parmentier M, Hachet-Haas M, Galzi JL, Rognan D. 2007. Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening. *J Med Chem* 50:1294-1303.

Klabunde T, Hessler G. 2002. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* 3:928-944.

Klebe G, Abraham U, Mietzner T. 1994. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 37:4130.

Klebe G. 2005. In *Virtual Screening in Drug Discovery*, Alvarez JC, Shoichet B. Eds.; CRC Press: Boca Raton. 3-24pp.

Kondru R, Zhang J, Ji C, Mirzadegan T, Rotstein D, Sankuratri S et al. 2008. Molecular interactions of CCR5 with major classes of small molecule anti-HIV CCR5 antagonists. *Mol Pharmacol* 73:789-800.

Kureha Corporation. Novel nitrogenous compounds and use thereof. US20040254221; 2004.

Liang X, Parkinson JA, Weishaüpl M, Gould RO, Paisey SJ, Park H, Hunter TM, Blindauer CA, Parsons S, Sadler PJ. 2002. Structure and Dynamics of Metallomacrocycles: Recognition of Zinc Xylyl-Bicyclam by an HIV Coreceptor. *J Am Chem Soc* 124:9105-9112.

Lloyd DG, Buenemann CL, Todorov NP, Manallack DT, Dean PM: Scaffold hopping in de novo design. Ligand generation in the absence of receptor information. *JMed Chem* 2004, 47:493-496.

Michino M, Abola E, Participants G, Brooks Cr, Dixon J, Moulton J, et al. 2009. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat Rev Drug Discov* 8:455-63.

Moro S, Braiuca P, Deflorian F, Ferrari C, Pastorin G, Cacciari B, Baraldi PG, Varani K, Borea PA, Spalluto G. 2005. Combined target-based and ligand-based drug design approach as a tool to define a novel 3D-pharmacophore model of human A3 adenosine receptor antagonists: pyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine derivatives as a key study. *J Med Chem* 48:152-162.

Moro S, Deflorian F, Bacilieri M, Spalluto G. 2006. Novel strategies for the design of new potent and selective human A3 receptor antagonists: an update. *Curr Med Chem* 13:639-645.

Mosley CA, Wilson LJ, Wiseman JM, Skudlarek JW, Liotta DC. 2009. Recent patents regarding the discovery of small molecule CXCR4 antagonists. *Expert Opin Ther Patents* 19:23-38.

Moyle G, DeJesus E, Boffito M, Wong R, Coakley E, Gibney C, Badel K, Calandra G, Bridger G, and Becker S. 2007. CXCR4 antagonism: proof of activity with AMD11070, in Proceedings of the 14th Conference on Retroviruses and Opportunistic Infections. Feb 25-28; Los Angeles, CA. Abstract No. 511.

Murakami T, Yoshida A, Tanaka R, Mitsuhashi S, Hirose K, Yanaka M, Yamamoto N, Tanaka Y. 2004. KRH-2731: An Orally Bioavailable CXCR4 Antagonist Is a Potent Inhibitor of HIV-1 Infection. In 2004 AntiVirals Pipeline Report; Camp, R., Ed.; Proceedings of the 11th Conference on Retroviruses and Opportunistic Infection, San Francisco CA, Feb. 8-11, 2004; Treatment Action Group: San Francisco, CA; Abstract No. 541.

Mustafi D, Palczewski K. 2009. Topology of class A G protein-coupled receptors: insights gained from crystal structures of rhodopsins, adrenergic and adenosine receptors. *Mol Pharmacol* 75:1-12.

Novartis. Isothiourea derivatives. US20070161618; 2007.

Nygaard R, Frimurer TM, Holst B, Rosenkilde MM, Schwartz TW. 2009. Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol Sci* 30:249-259.

Ono Pharmaceutical Co. Nitrogen-containing heterocyclic compounds and medicinal use thereof. US20070167459; 2007.

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA et al. 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289:739-745.

Park JH, Scheerer P, Hofmann KP, Choe H, Ernst OP. 2008. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 454:183-187.

Pérez-Nueno VI, Ritchie DW, Rabal O, Pascual R, Borrell JI, Teixidó J. 2008a. Comparison of Ligand-Based and Receptor-Based Virtual Screening of HIV Entry Inhibitors for the CXCR4 and CCR5 Receptors Using 3D Ligand Shape Matching and Ligand-Receptor Docking. *J Chem Inf Model* 48:509-533.

Pérez-Nueno VI, Ritchie DW, Borrell JI, Teixidó J. 2008b. Clustering and Classifying Diverse HIV Entry Inhibitors Using a Novel Consensus Shape-Based Virtual Screening Approach: Further Evidence for Multiple Binding Sites within the CCR5 Extracellular Pocket. *J Chem Inf Model* 48:2146-2165.

Pérez-Nueno VI, Pettersson S, Ritchie DW, Borrell JI, Teixidó J. 2009. Discovery of Novel HIV Entry Inhibitors for the CXCR4 Receptor by Prospective Virtual Screening. *J Chem Inf Model* 49: 810-823.

Pérez-Nueno VI. 2009. PhD Thesis. Universitat Ramon Llull (Barcelona, Spain), <http://www.tesisenxarxa.net/TDX-0526109-130043/index.html>

Pettersson S, Pérez-Nueno VI, Ros-Blanco L, Puig de la Bellacasa R, Rabal O, Batllori X, Clotet B, Clotet-Codina I, Armand-Ugón M, Esté J, Borrell JI, Teixidó J. 2008. Discovery of novel non-cyclam polynitrogenated CXCR4 coreceptor inhibitors. *ChemMedChem* 3:1549-1557.

Pettersson S, Pérez-Nueno VI, Mena MP, Clotet B, Esté J, Borrell JI, Teixidó J. 2010. Novel Monocyclam Derivatives as HIV Entry Inhibitors: Design, Synthesis, Anti-HIV Evaluation, and Their Interaction with the CXCR4 Coreceptor. *ChemMedChem*. DOI: 10.1002/cmdc.201000124.

Pospisil P, Ballmer P, Scapozza L, Folkers G. 2003. Tautomerism in Computer Aided Drug Design. *J Recept Signal Transduction* 23:361–371.

Princen K, Hatse S, Vermeire K, Aquaro S, De Clercq E, Gerlach LO, Rosenkilde M, Schwartz TW, Skerlj R, Bridger G, Schols D. 2004. Inhibition of Human Immunodeficiency Virus Replication by a Dual CCR5/CXCR4 Antagonist. *J Virol* 78:12996-13006.

Rasmussen SGF, Choi H, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC et al. 2007. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450:383-387.

Rosenkilde MM, Gerlach LO, Jakobsen JS, Skerlj RT, Bridger GJ, Schwartz TW. 2004. Molecular mechanism of AMD3100 antagonism in the CXCR4 receptor: transfer of binding site to the CXCR3 receptor. *J Biol Chem* 279:3033-3041.

Rosenkilde MM, Gerlach LO, Hatse S, Skerlj RL, Schols D, Bridger G, Schwartz TW. 2007. Molecular mechanism of action of monociclam versus biciclam non-peptide antagonist in the CXCR4 chemokine receptor. *J Biol Chem* 282:27354-27365.

Rush TS, Manas ES, Tawa GJ, Alvarez JC. 2005. In *Virtual Screening in Drug Discovery*, Alvarez JC, Shoichet B. Eds.; CRC Press: Boca Raton. 249-277pp.

Schafferhans A, Klebe G. 2001. Docking ligands onto binding site representations derived from proteins built by homology modelling. *J Mol Biol* 307:407-27.

Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe H et al. 2008. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* 455:497-502.

Schneider G, Clément-Chomienne O, Hilfiger L, Schneider P, Kirsch S, Böhm H-J, Neidhart W. 2000. Virtual screening for bioactive molecules by evolutionary de novo design. *Angew Chem Int Ed Engl* 39:4130-4133.

Shaheen F, Collman RG. Co-receptor antagonist as HIV-1 entry inhibitors. 2004. *Curr Opin Infect Dis* 17:7-16.

Shim JY, Welsh WJ, Howlett AC. 2003. Homology model of the CB1 cannabinoid receptor: sites critical for nonclassical cannabinoid agonist interaction. *Biopolymers* 71:169-189.

Singh S, Malik BK, Sharma DK. 2007. Targeting HIV-1 Through Molecular Modeling and Docking Studies of CXCR4: Leads for Therapeutic Development. *Chem Biol Drug Des* 69:191-203.

Srinivasan J, Cheatham III, Kollman P, Case DA. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. 1998. *J Am Chem Soc* 120:9401-9409.

Taigen Biotechnology. Polyamine compounds for treating chemokine receptor mediated diseases. WO04089360; 2004.

Taigen Biotechnology. Pyrimidine compounds. US20060281712; 2006.

Tamamua H, Omagari A, Hiramatsu K, et al. 2001. Development of specific CXCR4 inhibitors possessing high selectivity indexes as well as complete stability in serum based on an anti-HIV peptide T140. *Bioorg Med Chem Lett* 11: 1897-902.

Tamamura H, Araki T, Ueda S, Wang Z, Oishi S, Esaka A, Trent JO, Nakashima H, Yamamoto N, Peiper SC, Otaka A, Fujii N. 2005. Identification of novel low molecular weight CXCR4 antagonists by structural tuning of cyclic tetrapeptide scaffolds. *J Med Chem* 48:3280-3289.

Tamamura H, Ojida A, Ogawa T, Tsutsumi H, Masuno H, Nakashima H, Yamamoto N, Hamachi I, Fujii N. 2006. Identification of a new class of low molecular weight antagonists against the chemokine receptor CXCR4 having the dipicolylamine-zinc(II) complex structure. *J Med Chem* 49:3412–3415.

Tauzin B. Report 2009. Medicines in development for HIV/AIDS. Presented by America's Pharmaceutical Research Companies.
http://www.phrma.org/files/meds_in_dev/AIDS2007SP.pdf

Teixidó J, Borrell JI, Nonell S, Pettersson S, Ros L, Puig de la Bellacasa R, Rabal MO, Pérez-Nueno VI, Esté J, Clotet-Codina I, Armand-Ugón M. ES Patent ES200602764, 2006 (filing date: October 26, 2006).

Trent JO, Wang ZX, Murray JL, Shao W, Tamamura H, Fujii N, Peiper SC. 2003. Lipid bilayer simulations of CXCR4 with inverse agonists and weak partial agonists. *J Biol Chem* 278:47136-47144.

Vaidehi N, Floriano WB, Trabianino R, Hall SE, Freddolino P, Choi EJ, Zamanakos GA, Goddard W. 2002. Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci USA* 99:12622-12627.

Wang R, Wang S. 2001. How does consensus scoring work for virtual library screening? An idealized computer experiment. *J Chem Inf Comput Sci* 41:1422-1426.

Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards C, Henderson R et al. 2008. Structure of a beta(1)-adrenergic G-protein-coupled receptor. *Nature* 454:486-491.

Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. 2006. A Critical Assessment of Docking Programs and Scoring Functions. *J Med Chem* 49:5912-5931.

Wong RSY, Bodart V, Metz M, Labrecque J, Bridger G, Fricker SP. 2008. Comparison of the Potential Multiple Binding Modes of Bicyclam, Monocyclam, and Noncyclam Small-Molecule CXCR4 Chemokine Receptor 4 Inhibitors. *Mol Pharmacol* 74:1485-1495.

World Health Organisation AIDS Update 2009,
<http://www.who.int/hiv/pub/9789241599450/en/index.html>

Wu B, Chien ETY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC. 2010. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. Science. DOI: 10.1126/science.1194396

Yamazaki T, Saitou A, Ono M, Yokohama S, Bannai K, Hirose K, Yanaka M. World Patent WO 029218, 2003.

Yamazaki T, Kikumoto S, Ono M, Saitou A, Takahashi H, Kumakura S, Hirose K. World Patent WO 024697, 2004.

Zhan W, Liang Z, Zhu A, Kurtkaya S, Shim H, Snyder JP, Liotta DC. 2007. J Med Chem 50: 5655-5664.

Zhou N, Luo Z, Luo J, Liu D, Hall JW, Pomerantz RJ, Huang Z. 2001. 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 276:42826-42833.

The following pages contain 8 Figures and 1 Table for the article.

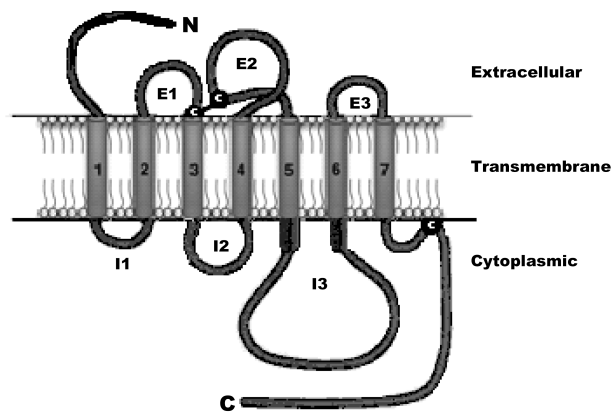


Fig 1. Topology of rhodopsin-like G protein coupled receptors. The three extracellular loops are labelled E1, E2, E3 and the three intracellular loops are labelled I1, I2, I3. Reproduced from [Pérez-Nueno et al., 2008a] with permission from ACS.

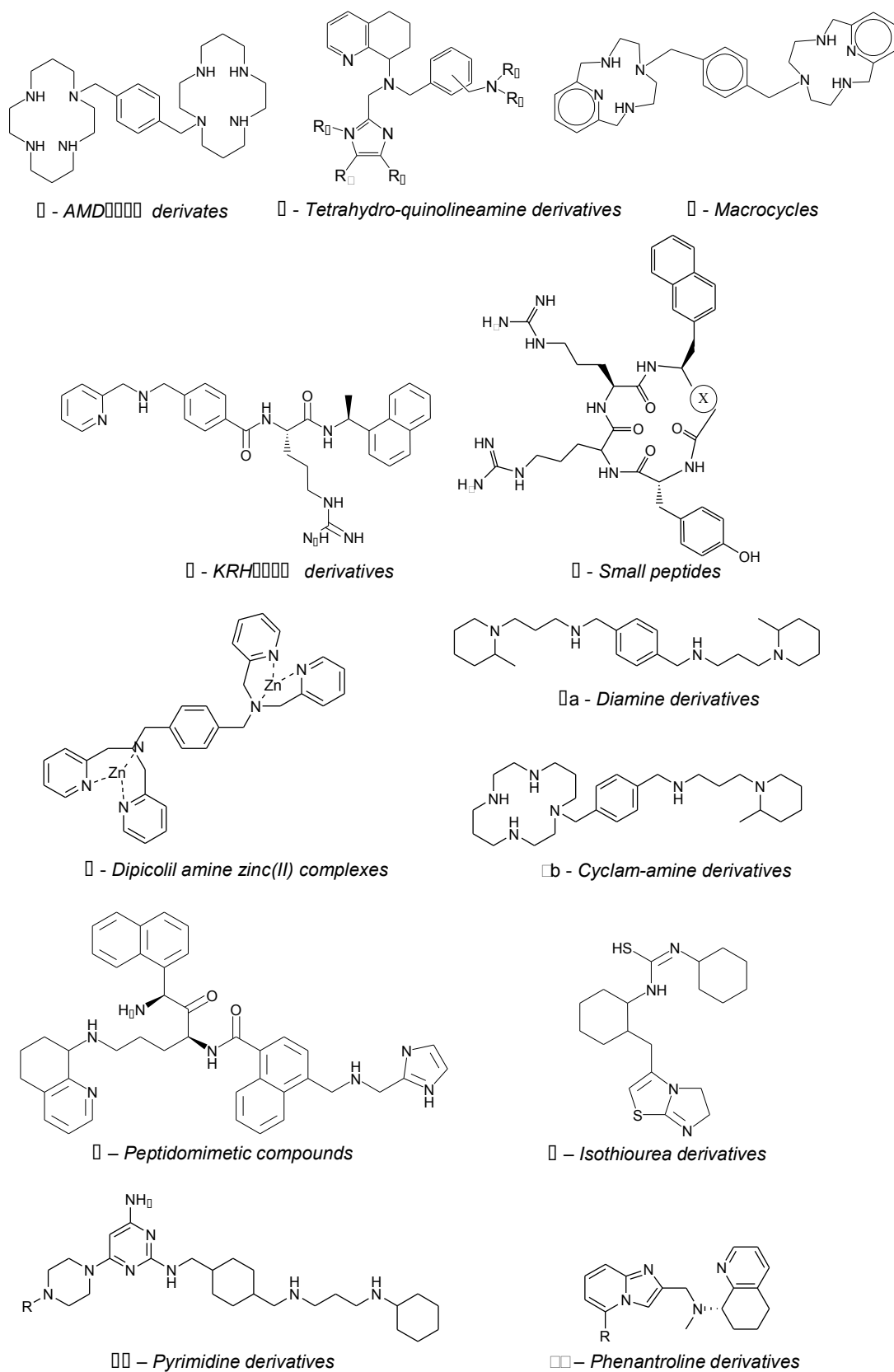


Fig. 2. Representative structures of eleven families of CXCR4 antagonist inhibitors.

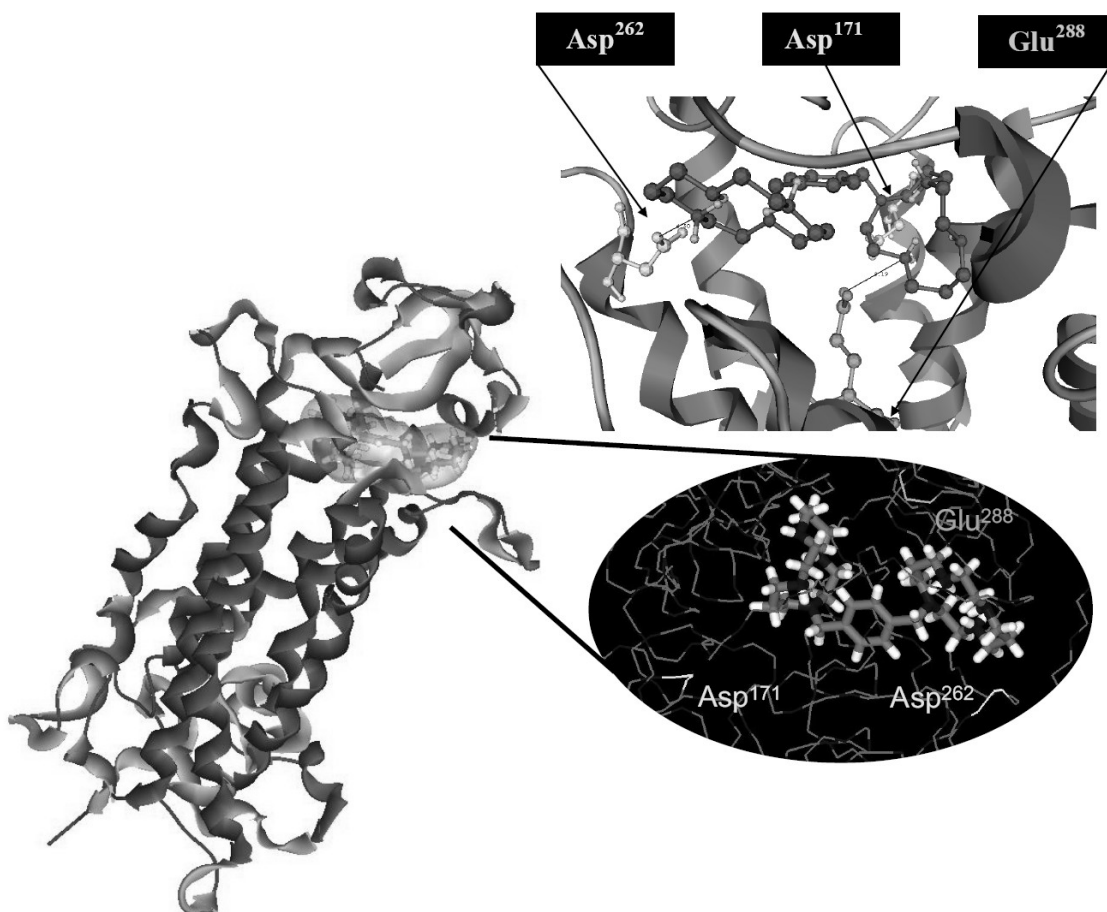


Fig. 3. Close-up view of the lowest energy AMD3100-CXCR4 binding conformation. The view on the left shows AMD3100 docked within the CXCR4 pocket. The AMD3100 molecular volume is depicted using a spherical harmonic surface. The view on the right shows in detail the calculated binding conformation. In this docking prediction, two nitrogens of one cyclam ring interact with the two carboxylic oxygens of Asp262, and two nitrogens of the other cyclam ring interact with the two carboxylic oxygens of Glu288. Reproduced from [Pérez-Nueno et al., 2008a] with permission from ACS.

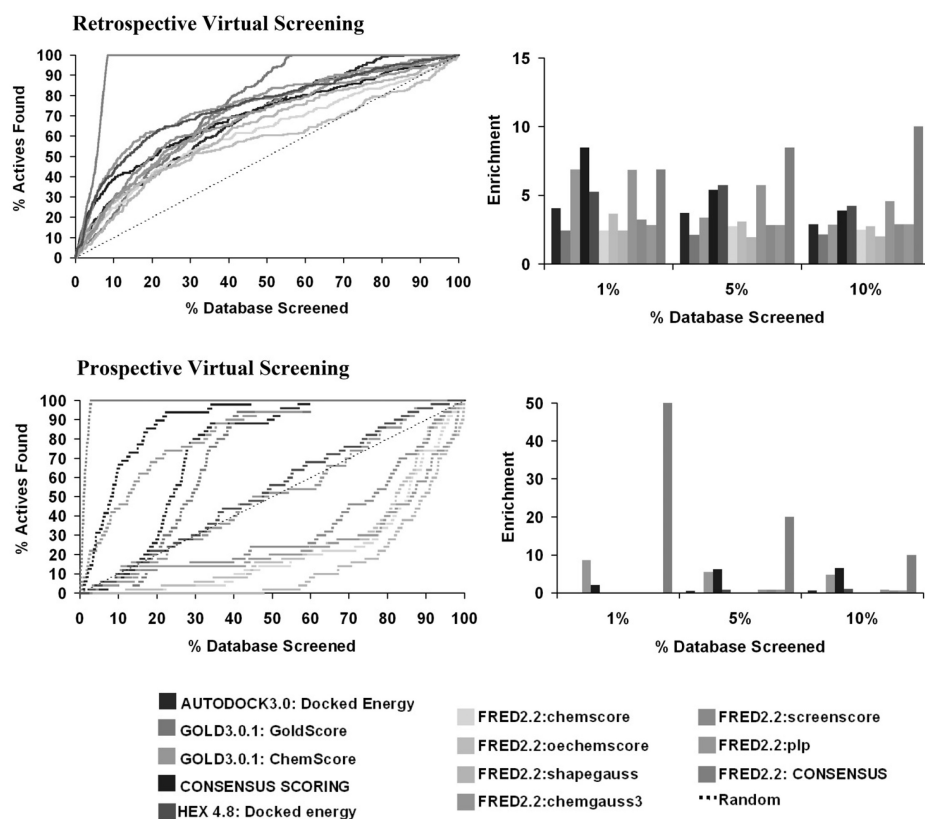


Fig. 4. CXCR4 docking-based enrichment plots. On the left, enrichment results for several docking protocols for retrospective (top) and prospective (bottom) virtual screening analyses. The dotted black line represents the expected values if actives are selected at random. On the right, enrichment factors for actives found within the top-ranking 1%, 5%, and 10% of the screened known-inhibitor database (top) and screened virtual combinatorial library (bottom). Consensus scoring (Autodock, ChemScore, GoldScore), FRED Consensus, FRED Chemgauss3, ChemScore and Hex docked energy performed the best for both retrospective and prospective VS analyses. Reproduced from [Pérez-Nueno et al., 2009] with permission from ACS.

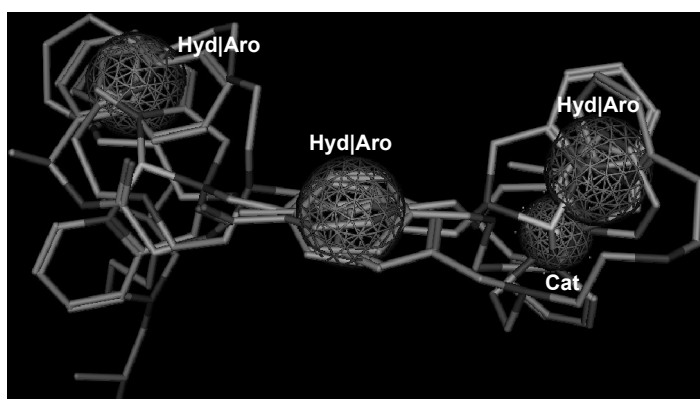


Fig 5. Alignment of the compounds in the training set and the pharmacophore model obtained using MOE Pharmacophore Elucidate module. The pharmacophore comprises three hydrophobic and aromatic features (Hyd|Aro) and one cationic (Cat) feature. Reproduced from [Pérez-Nueno et al., 2009] with permission from ACS.

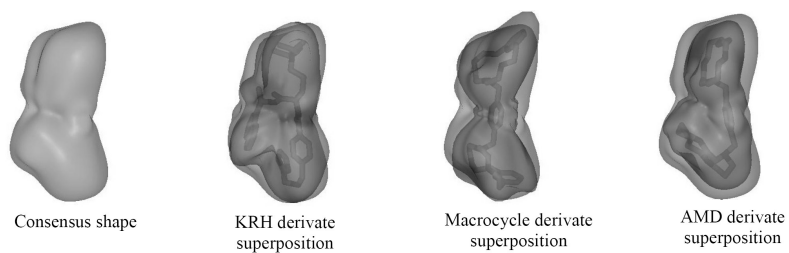


Fig. 6. CXCR4 antagonist consensus shapes. The image on the left shows the consensus shape calculated from the three most active compounds of different scaffolds families in a CXCR4 inhibitor database compiled from the literature for retrospective screening: an AMD derivative, a macrocycle derivative, and a KRH derivative. The following three images show the superpositions of these compounds onto the consensus. Reproduced from [Pérez-Nueno et al., 2008b] with permission from ACS.

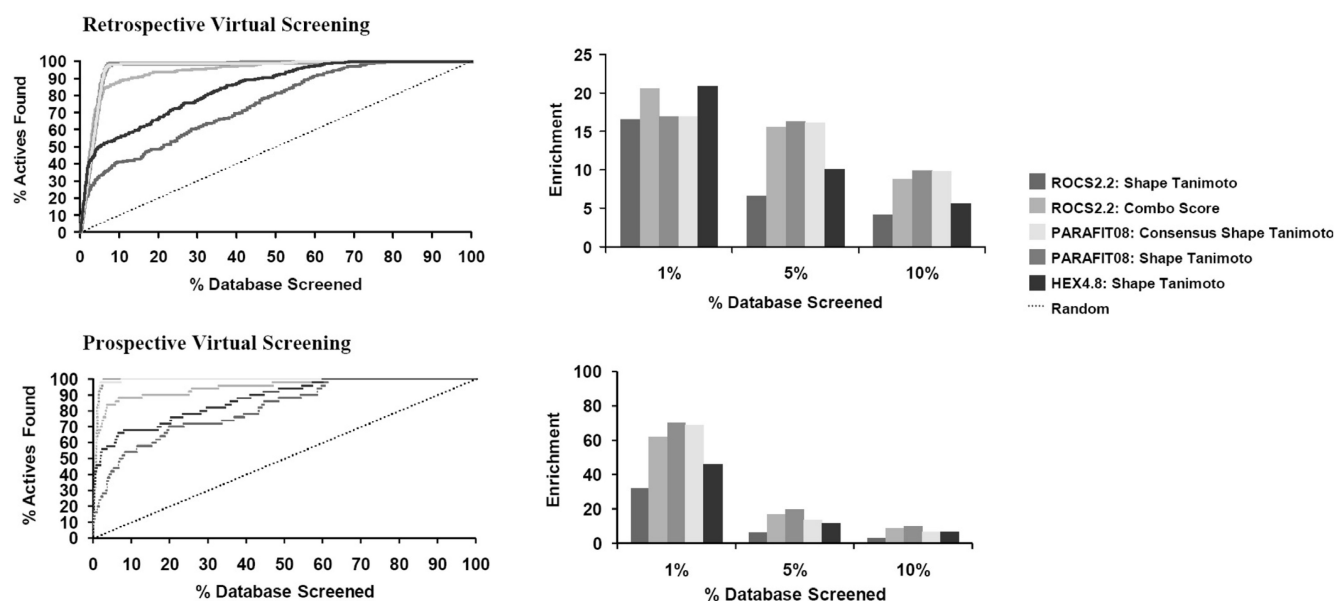


Fig. 7. CXCR4 shape matching-based enrichments. On the left, enrichment curves obtained for various shape matching protocols on the known inhibitor database (top) and compounds from a designed virtual combinatorial library to prospectively screen (bottom). The dotted line represents the expected enrichment if actives are selected at random. On the right, enrichment values for actives found within the top-ranking 1%, 5%, and 10% of the screened database (top) and screened virtual combinatorial library (bottom). ROCS combo score, PARAFIT consensus shape and PARAFIT shape performed the best for both retrospective and prospective VS analyses. Reproduced from [Pérez-Nueno et al., 2009] with permission from ACS.

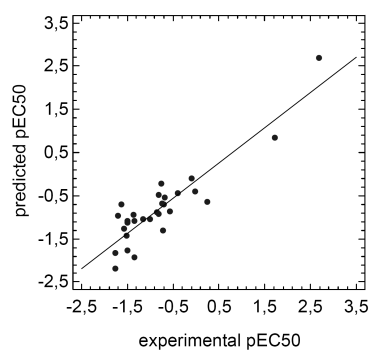


Fig. 8. Correlation of experimental versus predicted pEC_{50} for Pérez-Nueno et al. [2009] CXCR4 inhibitors QSAR model. Reproduced from [Pérez-Nueno et al., 2009] with permission from ACS.

Compound	Name	EC ₅₀ / µg/ml	CC ₅₀ / µg/ml
	1{8,8}	0.008	>25
	1{8,9}	0.03	>25
	1{5,6}	0.2	>25
	1{8,10}	0.4	>25
	8{1,8}	> 4.1	4.1
	8{2,8}	0.6	14.6
	1{7,8}	0.4	> 25
	10{8}	0.022	> 25
	10{11}	0.058	> 25

Table 1. Summary of prospective VS-selected hits. EC₅₀ denotes anti-HIV activity, and CC₅₀ is the cytotoxicity value (µg/ml). Reproduced from [Pérez-Nueno et al., 2009] with permission from ACS.