Published in final edited form as:

J Chem Inf Model. 2018 May 29; 58(5): 1104–1120. doi:10.1021/acs.jcim.8b00004.

Maximal Unbiased Benchmarking Data Sets for Human Chemokine Receptors and Its Comparative Analysis

Jie Xia^{△, ‡}, Terry-Elinor Reid[§], Song Wu[△], Liangren Zhang^{‡, *}, and Xiang Simon Wang^{§, *}

[△]State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Department of New Drug Research and Development, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050

[‡]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China

[§]Molecular Modeling and Drug Discovery Core Laboratory for District of Columbia Center for AIDS Research (DC CFAR); Department of Pharmaceutical Sciences, College of Pharmacy, Howard University, Washington DC 20059, U.S.A

Abstract

Chemokine receptors (CRs) have long been druggable targets for the treatment of inflammatory diseases and HIV 1 infection. As a powerful technique, virtual screening (VS) has been widely applied to identifying small molecule leads for modern drug targets including CRs. For rational selection of a wide variety of VS approaches, ligand enrichment assessment based on a benchmarking data set has become an indispensable practice. However, the lack of versatile benchmarking sets for the whole CRs family that are able to unbiasedly evaluate every single approaches including both structure and ligand based VS, somewhat hinders modern drug discovery efforts. To address this issue, we constructed Maximal Unbiased Benchmarking Data sets for human Chemokine Receptors (MUBD-hCRs) using our recently developed tools of MUBD-DecoyMaker. The MUBD-hCRs encompasses 13 subtypes out of 20 chemokine receptors, composed of 404 ligands and 15756 decoys so far and are readily expandable in the future. It had been thoroughly validated that MUBD-hCRs ligands are chemically diverse while its decoys are maximal unbiased in terms of “artificial enrichment”, “analogue bias”. In addition, we studied the performance of MUBD-hCRs, in particular CXCR4 and CCR5 data sets, in ligand enrichment assessments of both structure and ligand based VS approaches in comparison with other benchmarking data sets available in public domain and demonstrated that MUBD-hCRs is much capable of designating the optimal VS approach. Taken together, MUBD-hCRs is a unique and maximal-unbiased benchmarking set that covers major CRs subtypes so far.

*Correspondence should be addressed to L.R.Z. (liangren@bjmu.edu.cn) or X.S.W. (x.simon.wang@gmail.com): Liangren Zhang, Ph.D., School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Rd, Beijing 100191, China, Xiang Simon Wang, Ph.D., Howard University College of Pharmacy, 2300 4th St. NW, Washington, DC 20059, U.S.A.

Supporting Information

The whole data sets of MUBD hCRs are available at ResearchGate: https://www.researchgate.net/profile/Xiang_Wang36. The Murcko scaffold analysis for CCR1 ligand set (Table S1), corresponding ROC curves for data in Table 4 and Figure 5 are available free of charge via the Internet at <http://pubs.acs.org>.

The authors declare no competing financial interest.

Graphical Abstract

Keywords

Chemokine receptor; virtual screening; ligand enrichment assessment; benchmarking data set; MUBD DecoyMaker

INTRODUCTION

Chemokines receptors (CRs) are a class of rhodopsin like G protein coupled receptors (GPCRs) that transduce cellular signals triggered by chemokines and mediate immune defense.^{1, 2} They are classified into four families, comprised of seven CXC receptors (CXCRs, CXCR1–7), ten CC receptors (CCRs, CCR1–10), one CX3C receptor (CX3CR1) and one XC receptor (XCR1).³ Excessive expression of chemokines and/or CRs are associated with various inflammatory diseases such as chronic inflammation,⁴ chronic obstructive pulmonary disease⁵, tumor progression and metastasis.^{6, 7} In addition, CCR5 and CXCR4 had been long identified as co receptors for HIV-1 infection.^{8–10} As a result of their pivotal roles in immune system, CRs have gained much popularity as druggable targets and drug discovery efforts targeting CRs had successfully led to therapeutics for inflammatory diseases and HIV-1 infection.^{1, 2, 11}

In modern drug discovery, virtual screening (VS) has become a favored and powerful technique to identify novel hits from large scale chemical libraries.¹² There are two well-known classes of VS approaches, i.e. ligand based virtual screening (LBVS) and structure based virtual screening (SBVS).¹³ LBVS is normally applied when three-dimensional structures of biological targets had not been solved while the information of known ligands is readily available. Examples of LBVS approaches include pharmacophore modeling, quantitative structure activity relationship (QSAR) modeling and fingerprint based similarity search.^{14–20} SBVS refers to molecular docking, i.e. a large amount of compounds are docked into the binding site of the three dimensional target structure (e.g. X-ray crystal structures or homology models) and ranked according to their estimated binding affinities by scoring functions.^{21–23} Thus far, both classes of VS approaches have been applied to assist drug discovery efforts targeting chemokine receptors, including the well-studied CCR5^{24–28}, CXCR4^{29–34} as well as other subtypes such as CCR1,³⁵ CCR2,³⁶ CCR3,³⁷ CCR4,^{38, 39}

CXCR2,⁴⁰ CXCR3,⁴¹ CXCR7.⁴² Among them, there were limited successes which performed VS approaches directly without any method evaluation or parameter optimization, e.g. for CCR4^{38, 39}. In most cases, benchmarking evaluation prior to screening of large-scale chemical libraries was an indispensable practice. To be specific, benchmarking study is the assessment of ligand enrichment of different VS approaches in the form of retrospective small-scale VS based on a compiled data set that consists of known ligands and structurally close inactives (i.e. decoys), named benchmarking data set.⁴³ Unfortunately, the benchmarking data sets for chemokine receptors used in the prior studies were generated in a less strict manner. Also the lack of comprehensive and uniform benchmarking sets for chemokine receptors made the reported ligand enrichments of different VS approaches non comparable across those studies. The prior benchmarking studies were thus not able to suggest the most effective approach.

In fact, uniform benchmarking sets date back to 2000 when Rognan *et al.* created the first pioneering benchmarking sets.⁴⁴ Over the ten years, there were continuous efforts on creation and optimization of benchmarking sets by reducing three main types of benchmarking biases, i.e. “artificial enrichment” (specifically for docking), “analogue bias” and “false negatives”.^{43, 45} Both “artificial enrichment” and “analogue bias” make ligand enrichment of VS approaches unrealistically easy thus cause performance overestimation. “Artificial enrichment” bias is mainly caused by significant mismatching of low-dimensional physicochemical properties between designed decoys and ligands. A benchmarking set of “analogue bias” is normally characterized by highly similar chemical structures (i.e. analogues) in the ligand set. By contrast, “false negative” bias reduces ligand enrichment and it occurs if presumed inactives in the decoy set turn out to be actives. Those efforts had brought forth a variety of uniform and bias-corrected benchmarks.⁴³ Among them, the series that were started from directory of useful decoys (DUD)⁴⁶ and designed for benchmarking SBVS approaches are the most widely used, including DUD clusters,⁴⁷ charge-matched DUD,⁴⁸ the recent DUD-Enhanced (DUD-E)⁴⁹ and its extended version of nuclear receptors ligands and structures benchmarking database (NRLiSt BDB).⁵⁰ Other benchmarking sets for SBVS cover world of molecular bioactivity (WOMBAT)⁴⁷, virtual decoy sets (VDS),⁵¹ G protein coupled receptor (GPCR) ligand library (GLL) and GPCR Decoy Database (GDD),⁵² demanding evaluation kits for objective *in silico* screening (DEKOIS)⁵³ and DEKOIS 2.0.⁵⁴ In the meantime, data sets such as DUD LIB VS 1.0,⁵⁵ database of reproducible virtual screens (REPROVIS-DB)⁵⁶ and maximum unbiased validation (MUV)⁵⁷ were developed specifically for benchmarking LBVS approaches. Along with the benchmarking sets, a few standalone or on line tools to build benchmarking sets were furnished to public, e.g. DecoyFinder⁵⁸, MUV and DUD-E decoy makers. In spite of such a large quantity of benchmarking sets and tools, there are still perplexing problems when they are applied to ligand enrichment study of VS approaches against chemokine receptors: (1) All the ready-to-apply data sets except for DUD-E and GLL/GDD do not cover every chemokine receptors; (2) The only available data sets for chemokine receptors, i.e. CXCR4 in DUD-E and CCR5 in GLL/GDD were initially designed for benchmarking SBVS (i.e. molecular docking) approaches; (3) The DecoyFinder and DUD-E decoy-maker currently available were also developed for building SBVS-specific benchmarking sets. The only one applicable for both SBVS and LBVS approaches, the MUV tool, is very limited to

its available chemokine receptors as well as its number of decoys which are experimental inactives from PubChem.⁵⁷

Recently, we developed a method of building maximal-unbiased benchmarking datasets (MUBD).⁴⁵ It has been implemented successfully in Pipeline Pilot (version 7.5, Accelrys Software, Inc) and named as MUBD-DecoyMaker. It provides a viable opportunity to customize MUBD for the whole panel of chemokine receptors with an aim of benchmarking both LBVS and SBVS approaches. In this study, we applied our tools to build MUBD for human chemokine receptors (MUBD-hCRs) and conduct thorough validation of the data sets generated by measuring potential benchmarking bias. In addition, we studied the performance of MUBD-hCRs in benchmarking studies (i.e. evaluation of both LBVS and SBVS approaches) by comparing CXCR4 data sets and CCR5 data sets with other databases of DUD-E and GLL/GDD. Based on the above comparative studies, we discussed strengths and weaknesses of MUBD-hCRs, DUD-E and GLL/GDD for ligand enrichment assessment and proposed the fairest benchmarking outcomes, i.e. the optimal VS approaches for CXCR4 and CCR5. (cf. Scheme 1) We anticipate the wide application of MUBD-hCRs in future benchmarking studies of both LBVS and SBVS approaches which can aid greatly the drug discovery efforts toward chemokine receptors.

METHODS

Ligand Collection and Curation.

ChEMBL (<https://www.ebi.ac.uk/chembl/>) is a publicly accessible bioactivity database that contains a large number of drug-like bioactive compounds for a broad range of drug targets.^{59, 60} It has been the major source of ligands for different subtypes of chemokine receptors.⁶⁰

The following steps were applied to ligand collection for each subtype of chemokine receptors. At the first step, those ligands whose ChEMBL confidence scores (cf. <https://www.ebi.ac.uk/chembl/faq#faq24>) were either greater than or equal to 4 were retained for further process. This criterion was originally defined by the Shoichet group for the compilation of DUD-E ligand set.⁴⁹ At the second step, ligands whose IC₅₀ values are 1 μ M or better were retained. In a few cases that there were not enough ligands to meet this criterion, the following restraint-relaxing measures were gradually applied: (1) lift of the cutoff for IC₅₀, (2) use of activity data of Ki or EC₅₀ and (3) inclusion of ligands reported in the recent literature but not in ChEMBL. Every measure above was aimed to ensure a sufficient number of diverse ligands for each benchmarking set, as Jahn *et al.* suggested the use of an acceptable number of chemotypes for ligand enrichment assessment.⁵⁵ At the last step, all the data records for the retained ligands were merged according to their unique ligand IDs. In this way, the scenario where an individual ligand had multiple data records, i.e. activity data from various bioassays or publications can be addressed. Those unique ligands constitute the “raw ligand” data sets.

Data curation was performed using Pipeline Pilot (version 7.5, Accelrys Software, Inc), including essential steps such as stripping salts, standardizing molecules, filtering “raw

ligands” by the criteria of “RBs > 20 or MW < 600”,⁴⁹ and protonation at pH range of 7.3–7.5. The resulting protonated ligands constitute the so called “curated ligand” data set.

Construction of MUBD-hCRs.

Computational Tool.—Our in house MUBD DecoyMaker (https://www.researchgate.net/publication/271531936_Codes_of_Unbiased_Benchmarking_Method) implemented in Pipeline Pilot was used to construct each data set of MUBD hCRs. It consists of three main consecutive modules, i.e. ligand processor (selection of diverse ligands and property calculation), preliminary filter and precise filter. The input of MUBD-DecoyMaker was the “curated ligand” data set from the last section. In the current study, the source database of decoys was the “All Purchasable Molecules” subset of ca. 18 million compounds from ZINC (<http://zinc.docking.org/>).^{61, 62}

Ligand Processor.—Firstly, this module coded each ligand in “curated ligand” data set with MACCS structural keys⁶³ and calculated pairwise “*similarity in structure*” (“*sims*”), i.e. Tanimoto coefficient (*Tc*)⁶³. It then selected diverse ligands from the “curated ligand” set according to the following algorithm. Let *n* be the total number of the curated ligands and *i* be the index of each curated ligand (*i*=1,2,..., *n*). Each ligand *i* had an array of *n-1* *Tc* values that measured the pairwise similarity of *n-1* other ligands to this reference ligand. The *n-1* *Tc* values of the reference ligand was automatically checked and any one that met the criterion of *Tc* < 0.75 was defined as the analogue. These analogues and their related *Tc* values were not considered in the next round of check. The automatic check and exclusion was not performed till all the curated ligands except for those already excluded ones were used as references. The remaining ligands were then determined as diverse ligands. Secondly, the module calculated six physicochemical properties for each diverse ligand, i.e. *AlogP*, Molecular Weight (MW), and number of hydrogen bond acceptors (HBAs), number of Hydrogen Bond Donors (HBDs), number of Rotatable Bonds (RBs) and Net (formal) Charge (NC). In the end, the output of this section is a “diverse ligand” data set that contains diverse ligands annotated with physicochemical properties.

Preliminary Filter.—Preliminary filter is able to narrow down the number of compounds in the source database of ZINC so as to speed up the calculation of decoys. The inputs are the “diverse ligand” data set and the large ZINC database. The latter was filtered by applying a preliminary target specific filter defined by the maximum and minimum values of each physiochemical property of diverse ligands, and a topology (“*sims*”-based) filter defined by the range of MACCS “*sims*”, i.e. from the minimum value (or lower bound) of MACCS “*sims*” of the diverse ligands to a constant 0.75. The remaining compounds were rendered as potential decoys (PDs). It should be noted that pairwise MACCS “*sims*” values as well as physicochemical properties of PDs to diverse ligands are within the ranges determined by the “diverse ligand” data set.

Precise Filter.—Precise Filter is able to accurately select 39 unique final decoys (FDs) for each diverse ligand from the above PDs. Its inputs are the “diverse ligand” data set and PDs data set. Two formulas were defined, i.e. (1) similarity in physiochemical properties (“*simp*”) between the PDs and the query ligand (eq. 1); (2) average difference (“*simsdiff*”)

between two structural similarities, i.e. MACCS “*sims*” between the query ligand and other ligands, MACCS “*sims*” between the query ligand and PDs. (eq. 2)

$$simp_{T,R} = 1 - \sqrt{\frac{1}{n} \sum_{i=1}^n (p_{i,T} - p_{i,R})^2} \quad (1)$$

$$\overline{simsdiff}_{i,k} = \frac{1}{m-1} \sum_{j=1}^{m-1} |sims_{k,j} - sims_{i,j}| \quad (2)$$

In eq. 1, T is a PDs molecule while R is a query ligand. n is the number of physicochemical properties ($n=6$) and i is the index of physicochemical properties ($i=1,2,\dots,n$). In eq.2, m , i and k are all constants, representing the total number of diverse ligands, the index of a query ligand, a potential decoy of ligand i . $\overline{simsdiff}_{i,k}$ is the average of *simsdiffs* for a potential decoy k of ligand i . $sims_{k,j}$ represents MACCS “*sims*” between decoy k and each remaining ligand j ($j=1,2,\dots,m-1$). $sims_{i,j}$ represents MACCS “*sims*” between query i and each remaining ligand j ($j=1,2,\dots,m-1$).

The precise filters were based on the above two formulas. For “*simp*”, the initial cutoff was 0.95 and the passing decoys were then sorted by $\overline{simsdiff}$ values. The top-ranked 39 decoys were selected as FDs. For certain diverse ligands, cutoffs for “*simp*” values were adjusted from 0.95 to 0.5 stepwise by 0.05 to make sure that enough (i.e. 39) decoys were obtained. Also, non duplicate decoys for each individual ligand were ensured. These FDs constitute the decoy set of MUBD hCRs.

Thorough Validation of MUBD-hCRs.

Every data set in MUBD-hCRs had been validated by measuring potential benchmarking bias. Since “false negative” bias was not able to be quantified in real practice, only the other two types of potential benchmarking bias, i.e. “artificial enrichment” and “analogue bias”⁴³ were quantified. The way of quantification was to applying Leave-One-Out Cross-Validation (LOO CV) to similarity search based on “*simp*” to evaluate “artificial enrichment” or MACCS “*sims*” to evaluate “analogue bias”.⁴⁵ Similarity search based on “*simp*” was such a validation in which each ligand was left out as a query and coded by six physicochemical properties, followed by “*simp*” calculation against their corresponding decoys. Similarity search based on MACCS “*sims*” follows almost the same protocol as the “*simp*” calculation. Both “*simp*” and MACCS “*sims*” calculations generate compound lists with pairwise “*simp*” or “*sims*” scores. Based on the lists and true classes (i.e. 1 for ligand and 0 for decoy) of each compound, the receiver operator characteristic (ROC) curve was plotted and the area under curve (AUC) was calculated. Because LOO CV output multiple ROC curves and their corresponding AUCs, the average value of all AUCs, i.e. mean(AUCs), was calculated to measure the overall enrichment by the above similarity search approaches based on MUBD-hCRs. For a data set free of “artificial enrichment” and “analogue bias”, its ligands and decoys are supposed to be randomly distributed in chemical space. According to the

definition of ROC analysis, the diagonal line (AUC=0.5) in the plot happens to represent random assignment of two classes (e.g. ligands and decoys).⁶⁴ Therefore, that mean(AUC) equal to 0.5 in LOO CV is used as an indicator for “optimal embedding” of ligands in decoys in this series of benchmarking studies.⁴⁵ A value greater than 0.5 indicates the existence of “artificial enrichment” or “analogue bias”, while a value less than 0.5 suggests the “anti-screening” phenomenon. In the case that a number of decoys are more similar to the query than other true ligands in terms of physicochemical properties or MACCS structural keys, they rank at the top of the compound list which makes ligand enrichment rather difficult (i.e. AUC < 0.5) using the above similarity search approaches.⁴⁵

For “artificial enrichment”, the distribution curve is a classical way of evaluation thus has been widely used.^{46, 49, 52} Since the distribution curves of ligands and decoys can help localize the mismatching property, they were plotted as a supplement to mean(AUCs) from “simp”-based similarity search. For this metric, the ideal situation is the perfect matching of every property between ligands and decoys, for which the value of mean(AUCs) is 0.5. Once again, it should be noted that property mismatching can be a sign of either “benchmarking bias” or “antiscreeing” phenomenon.

Comparative Benchmarking Study.

Benchmarking Data Sets.—Because CXCR4 data set in DUD-E⁴⁹ (<http://dude.docking.org>) and CCR5 data set in GLL/GDD⁵² (<http://cavasotto-lab.net/Databases/GDD/Download/>) are the only ready to apply data sets for chemokine receptors, our current comparative study is limited to CXCR4 ligand enrichment based on MUBD-hCRs & DUD-E as well as CCR5 ligand enrichment based on MUBD-hCRs & GLL/GDD. Firstly, CXCR4 data set in DUD-E⁴⁹ (<http://dude.docking.org>) and CCR5 data set in GLL/GDD⁵² (<http://www.cavasotto-lab.net/Databases/GDD/>) were obtained.

Retrospective Small-scale VS.—Retrospective small scale-virtual screenings using both molecular docking and similarity search were conducted based on each individual benchmarking data set of target subtypes, i.e. CXCR4/MUBD-hCRs, CXCR4/DUD-E, CCR5/MUBD-hCRs and CCR5/GLL/GDD. The detailed procedures for retrospective small-scale virtual screening are listed below.

For molecular docking, two classic programs named GOLD (version 3.0.1) and FRED⁶⁵ (now OEDocking, version 3.0.1) were applied. The details as well as parameters of performing docking by these two programs are as follows. For docking by GOLD, ligands and decoys in each benchmarking set were directly submitted for docking. The screening mode of “7–8 times speed up” was adopted herein. The protein structures for both CXCR4 (PDB code: 3ODU) and CCR5 (PDB code: 4MBS) were prepared using the “Clean Protein” module of Discovery Studio (version 2.5, Accelrys Software, Inc) and defined as receptors. The binding site residues were defined by a sphere centering on the coordinates of their cognate ligands with a radius of 8 Å (CXCR4) or 10 Å (CCR5). All docking poses for each compound were scored using ChemScore and the pose of the highest score was retained. For docking studies using FRED, a multi conformer database for all ligands and decoys in each benchmarking data set was prepared by Omega⁶⁶ (version 2.5.1.4) prior to molecular

docking. Next, the crystal structure of CXCR4 or CCR5 was converted to a receptor, whose active site was defined by its cognate ligand. After that, FRED docked molecules from the multi-conformer database into the well prepared receptor. Chemgauss4, a default scoring function in FRED, scored all docking poses and the pose with the top-ranked score was selected.

For similarity search, a powerful type of circular fingerprints, i.e. function class fingerprints of maximum diameter 6 (FCFP_6), was employed to code the compounds in addition to MACCS structural keys. The application of such a different kind of fingerprints from the latter is aimed to avoid the potentially beneficial effect for MUBD-hCRs, since MACCS structural keys were the fingerprints we adopted during the method development. The similarity search with FCFP_6 was conducted in the same way as that based on MACCS “*sims*”.

Ligand enrichments from the above virtual screenings were calculated and incorporated into comparison. The overall ligand enrichment was presented using ROC AUC generated from the ranked scores (i.e. Chemscore, Chemgauss4 or FCFP_6 “*sims*”) and/or true classes of the compounds. Since early recognition is more of practical significance than overall enrichment in real world screening campaign,⁶⁷ the ROC enrichment (ROCE) value at the top 1% of screening was also calculated.⁵⁵ The ROCE value is quotient of true positive rate divided by the false positive rate at a given percentage of binding decoys, e.g. 1%. In addition, all the ROC curves were plotted and presented in this article.

Metrics for Comparison.—For comparison purpose, our focus was on benchmarking outcomes from different data sets, i.e. the rank of three VS approaches by ligand enrichment. To be more specific, the approaches were ranked according to their overall enrichment (e.g. ROC AUC) and early enrichment (e.g. ROCE@1%), respectively. The rankings from CXCR4/DUD-E vs. CXCR4/MUBD-hCRs or CCR5/GLL/GDD vs. CCR5/MUBD-hCRs were compared for their consistency. To fully explore the associated factors that contribute to the assessment outcome, we applied multiple metrics in order to uncover the differences in those data sets.

For ligand enrichment by molecular docking, physicochemical property matching between ligands and decoys is the major correlated factor. Similar to the validation of MUBD-hCRs, mean(AUCs) from similarity search based on “*simp*” in the form of LOO CV was also conducted in other benchmarking sets to quantify the overall property matching. Distribution curves were plotted to show the wellness of individual property matching graphically. For ligand enrichment by 2D-fingerprint based similarity search, “2D bias” is a typical and measurable “analogue bias” which results in the renowned “LBVS-favorable” outcome, i.e. the enrichment for 2D-fingerprint based similarity search is artificially easier than it is supposed to be in real world VS.⁶⁸ In our previous study, we had defined *NLBScore* (*Nearer Ligands Bias Score*) to measure “2D bias” in the benchmarking data set.⁴³ It was designed based on LOO CV and defined as the average percentage of Nearer Ligands (NL) from all iterations in LOO CV (cf. eq. 3 and 4). δ_k is a parameter to determine the status of a ligand k (1 for NL and 0 for non NL) in each iteration by comparing $sims_{Q,L}$ (i.e. *sims* between the query ligand and ligand k) and $sims_{Q,D_{max}}$ (i.e. query ligand and the nearest or most similar

decoy). Herein, “*sims*” can be calculated based on any 2D-fingerprints but its performance normally depends on the type of fingerprints during the similarity search.

$$NLBScore = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n-1} \sum_{k=1}^{n-1} \delta_k \right) \quad (3)$$

$$\delta_k = \begin{cases} 1, & \text{if } sims_{Q,k} > sims_{Q,D_{\max}} \\ 0, & \text{if } sims_{Q,k} \leq sims_{Q,D_{\max}} \end{cases} \quad (4)$$

Distribution curves of *Tc* for different benchmarking sets are able to reveal the overall difference in structural similarity between all ligands and decoys. Chemical diversity has been widely used as an indicator for structural similarity within ligands. Since both factors are highly associated with *NLBScore*, *Tc* distribution and scaffold analysis were also conducted for DUD-E and GLL/GDD and compared with MUBD-hCRs.

RESULTS AND DISCUSSION

Overview of MUBD-hCRs.

Target Coverage and Activity Types/Cutoffs.—As shown in Table 1, MUBD-hCRs contains 13 members out of 20 subtypes of chemokine receptors, including families of CC chemokine receptors (CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, and CCR8) and CXC chemokine receptors (CXCR1, CXCR2, CXCR3, CXCR4, CXCR5, and CXCR7). For the other 7 subtypes, the number of diverse ligands after analogue exclusion was too small even after the restraint relaxing measures were applied, thus they were not included. In essence, the target coverage of MUBD-hCRs highly depends on the chemical diversity of raw ligands obtained from ChEMBL19. From another perspective, the target coverage reflects the popularity of chemokine receptors subtypes and their progress of drug discovery pipeline.

Though the non-strict criterion (i.e. ChEMBL confidence score = 4) was applied, it turned out that all the ChEMBL confidence scores of the diverse ligands in MUBD-hCRs were either 8 or 9 (cf. Table 1). It indicates the ligands assigned to each subtype are targeting a homologous or a direct single protein. Though the cell lines expressing that protein are different, a homologous single protein and a direct one belong to the same subtype/target. Thus, those ligands in MUBD-hCRs can be used for ligand enrichment assessment of VS approaches targeting that protein.

As for the activity type, IC₅₀ was commonly adopted by all the ligand sets in MUBD-hCRs, though both binding data of IC₅₀ and Ki were applicable for benchmarking sets. When the ligands were being collected, no Ki data point or rather limited Ki data can be retrieved from ChEMBL for a few subtypes (e.g. CCR6, CCR8, CXCR5 and CXCR7), while IC₅₀ data points were accessible for all targets. Because of this, we gave priority to IC₅₀ instead of Ki in data curation, which was aimed to keep the criterion as consistent across multiple subtypes as possible.

Ideally, the binding affinity of all ligands (IC_{50} or K_i) should be better than $1\mu M$. However, it turned out that the cutoffs of activity were lifted for a few subtypes, e.g. $20\mu M$ for CCR6 and $60\mu M$ for CCR8. This measure was adopted to obtain a sufficient number of compounds to maximize the computational accuracy.

Size of Ligand and Decoy Sets.—Table 1 lists the numbers of raw ligands, curated (protonated) ligands, diverse ligands (i.e. MUBD-hCRs ligand sets) and decoys (i.e. MUBD-hCRs decoy sets). We initially collected a total of 5112 raw ligands. This number was then reduced to 4426 (i.e. for curated ligands) by data curation. Following that, the amount went down further to 404 as a result of analogue exclusion. Generally, the number of ligands for each subtype shows the same decreasing trend. In the data sets of “raw ligand”, the number of ligands ranges from 25 to 1520. In “curated ligand” data sets, the number is within the range of 19 to 1228. In “diverse ligand” data sets, the maximum number is 72 while the minimum is 12. For each “diverse ligand” data set, we identified Murcko scaffolds⁶⁹ using Pipeline Pilot and calculated the ratio of ligands per scaffold.⁴⁵ As shown in Table 1, the average of the ratios of ligands per scaffold across all subtypes is 1.08, with a maximum of 1.22 and the minimum of 1.00. Therefore, almost each ligand represents one unique Murcko scaffold. To give examples, the chemical structures of CCR1 diverse ligands and their Murcko scaffolds are shown in Table S1. All the above data demonstrates that analogue exclusion not only greatly scales down ligand entries but also ensures the chemical diversity. These data sets are advantageous for ligand enrichment assessment because they can evaluate the performance of different methods in terms of screening accuracy and scaffold hopping while keeping the computing cost low. As for decoy sets, we kept using 39 as the ratio of decoys per ligand which had been rationalized well in our prior studies.⁴⁵ In summary, the decoy sets include a total of 15756 decoys, with the minimum of 468 for CCR8 and the maximum of 2808 for CCR5.

Benchmarking Bias Corrections for MUBD-hCRs.

Artificial Enrichment Correction.—Plots of ROC curves from LOO CV using similarity search based on “*simp*” for all 13 subtypes of chemokine receptors are shown in Figure 1. In each plot, most of ROC curves are in proximity to the diagonal line, i.e. the random distribution curve of ligands and decoys. The average of AUCs calculated based on all ROC curves in each plot, i.e. $\text{mean}(\text{AUCs})$, is listed in Table 2 and shown in Figure 2. The value of $\text{mean}(\text{AUCs})$ for each subtype is approximately 0.5, with a minimum value of 0.432 and a maximum of 0.495. In addition, the $\text{mean}(\text{AUCs})$ values for 10 subtypes are greater than 0.45. That every $\text{mean}(\text{AUCs})$ is less than 0.5 suggests that (1) the data set is free of “artificial enrichment”; (2) the existence of “anti-screening” phenomenon, for which it is rather challenging to distinguish ligands from maximal unbiased decoys using physicochemical properties.⁴⁵ Besides, the fact that every $\text{mean}(\text{AUCs})$ is still close to 0.5 denotes that no significant difference exists in physicochemical properties between ligands and maximal-unbiased decoys, though the overall physicochemical property matching is not perfect.

Figure 3 shows the physicochemical properties distributions of ligands and decoys in each data set of MUBD-hCRs. According to those property matching plots in Figure 3, we

learned that (1) for each data set, there are always many physicochemical properties that match well, e.g. AlogP, RBs, and NC; (2) no data set achieves perfect matching for all physicochemical properties. These observations are consistent with that the value of mean(AUCs) is less than but close to 0.5. Also, it appears that the wellness of overall property matching is closely associated with std. of AUCs. For instance, ligands and decoys in CCR6 data set show the finest matching in every physicochemical property. Meanwhile, the std. of AUCs for this data set is the smallest (i.e. 0.007 in Table 2). To our knowledge, the combined usage of mean(AUCs) and property distribution curves is not only an excellent metric to measure “artificial enrichment”, but also able to show details of each property matching.

Analogue Bias Correction.—ROC curves from LOO CV using similarity search based on MACCS “*sims*” for each subtype were generated and shown in Figure 4. In general, all ROC curves for every subtype are close to the diagonal line, i.e. random distribution curve. Table 2 lists the average value of all AUCs, i.e. mean(AUCs) calculated from those ROC curves for each subtype while Figure 2 shows those mean(AUCs)s graphically. To be noted, the values of mean(AUC)s for 12 out of 13 subtypes are less than 0.6 and the mean(AUCs) for the only one remaining (CXCR4) is merely 0.614. The fact that mean(AUCs) is close to 0.5 implies that ligands and decoys in each data set are difficult to be discriminated by applying similarity search based on MACCS structural keys. The potential benchmarking bias due to the existence of inherent analogues, i.e. “analogue bias”, had been significantly minimized for all data sets in the MUBD-hCRs.

Representative Ligand and Decoy Structure for Each Data Set.—For each data set, we list chemical structures of a representative ligand and a representative decoy as well as values of “*simp*”, “*simsdiff*” and “*sims*” for that decoy, aiming to assist users to understand MUBD-hCRs in depth (cf. Table 3). From the table, we noticed that (1) the values of “*simp*” for all representative decoys are greater than or equal to 0.9. Since “*simp*” represents similarity in physiochemical properties, the large value of “*simp*” (0.9) of each representative decoy^{52, 58} properties; (2) the value of *simsdiff* for each representative decoy is approximately 0 (cf. Table 3), demonstrating that the decoy is difficult to be distinguished from the representative ligand using MACCS “*sims*”-based similarity search with any other ligand in the “diverse ligand” data set as a query; (3) the pairwise similarity (i.e. MACCS “*sims*”, or *Tc*) is less than 0.75 between every listed ligands and decoys. In summary, as shown by these representative ligands and decoys, the benchmarking sets are made certain to be maximal unbiased against “artificial enrichment”, “analogue bias”. Since *Tc*=0.75 based on MACCS is normally defined as a cutoff for binders and non-binders, the decoys in MUBD-hCRs are presumed not to be binders, i.e. not to be “false negatives”.

MUBD-hCRs for CXCR4 Ligand Enrichment Assessment.

General Outcomes.—Figure 5 (A) shows both the overall (i.e. AUC) and early enrichments (i.e. ROCE@1%) by FRED are greater than those by GOLD, no matter whether the assessment is based on CXCR4/MUBD-hCRs or CXCR4/DUD-E. Table 4 lists the exact values of AUC and ROCE@1%. Based on CXCR4/MUBD-hCRs, the overall enrichment by FRED is 0.563 while it is 0.541 by GOLD. Moreover, the early enrichment by FRED is

much larger than that by GOLD, i.e. 11.115 vs. 0. Consistently, based on CXCR4/DUD E the overall enrichment by FRED is also greater than that by GOLD, i.e. 0.812 vs. 0.694. Remarkably, FRED also shows much greater early enrichment for CXCR4 ligands than GOLD, i.e. 14.756 vs. 2.459. All these data indicates that the benchmarking outcome for specific docking programs based on CXCR4/MUBD-hCRs is consistent with that based on CXCR4/DUD-E. Interestingly, the ranks of FCFP_6 based similarity search are inconsistent based on those two benchmarking sets (cf. Figure 5(A) and Table 4). Based on CXCR4/MUBD-hCRs, FCFP_6 based similarity search ranks at the 3rd place for overall enrichment and the 2nd place for early enrichment. Its AUC equals to 0.518 and ROCE@1% equals to 9.201. However, based on CXCR4/DUD E FCFP_6-based similarity search ranks the 1st place for both overall enrichment (i.e. 0.843) and early enrichment (i.e. 56.265).

Benchmarking Bias.—With thorough validations, we have demonstrated that MUBD-hCRs is a benchmarking data set free of “artificial enrichment”. For CXCR4/MUBD hCRs, the value of mean(AUCs) from LOO CV for “*simp*”-based similarity search is 0.432 with a std. of 0.074 (cf. Table 2 or 5). In addition, most physicochemical properties of ligands and decoys in this data set, e.g. AlogP, MW and RBs, match well. (cf. Figure 3) We also investigated CXCR4/DUD E for “artificial enrichment” in the same way as for CXCR4/MUBD hCRs. The value of mean(AUCs) for CXCR4/DUD-E is 0.621 with a std. of 0.081 (cf. Table 5), which implies the existence of minor “artificial enrichment”. The matching of physicochemical property between ligands and decoys in CXCR4/DUD E is not as good either as that for CXCR4/MUBD-hCRs, in particular for MW and RBs. (cf. Figure 6)

For “analogue bias” in benchmarking LBVS approaches, we have validated MUBD-hCRs as maximal-unbiased based on LOO CV for similarity search using MACCS structural keys. Since the current ligand enrichment assessment applied similarity search based on FCFP_6 fingerprints rather than MACCS structural keys as a LBVS approach. It is thus ideal to employ the same metrics based on FCFP_6 fingerprints to measure potential “analogue bias” in CXCR4/MUBD hCRs and CXCR4/DUD-E. At first, we calculated *NLBScores* based on FCFP_6 fingerprints. As shown in Table 5, the *NLBScore* for CXCR4/DUD-E (0.459) is much greater than that for CXCR4/MUBD hCRs (0.069), indicating that CXCR4/DUD-E is highly “2D-biased” while the latter is almost unbiased for similarity search based on FCFP_6 fingerprints. In order to locate the underlined causes of the differences related to *NLBScore*, we plotted distribution curves of *Tc* values between ligands and decoys based on FCFP_6 fingerprints, and conducted scaffold analysis for ligands. In Figure 7, we observed that the distribution curve of *Tc* values for CXCR4/DUD-E shifts to the left side with reference to the corresponding curve for CXCR4/MUBD-hCRs. It illustrates that ligands and decoys in CXCR4/DUD-E are generally more structurally dissimilar than those in CXCR4/MUBD-hCRs. In Table 5, the ratios of ligands per Murcko scaffold for CXCR4/DUD-E and CXCR4/MUBD-hCRs are listed. For all protonated forms of ligands, the ratio in CXCR4/DUD-E is 1.968. Even after multiple protonated forms are counted as one chemical entity, the ratio is still as high as 1.481. In comparison, the ratio in CXCR4/MUBD hCRs is much lower at the value of 1.000. Based on the above information, it is likely that the low pairwise similarity between structures of ligands and decoys as well as the high pairwise similarity within ligands cause high *NLBScore*, i.e. “2D bias” in CXCR4/DUD-E.

As a matter of fact, DUD-E was specifically designed for benchmarking SBVS approaches which are not sensitive to “2D bias”.⁶⁸ The authors chose the most highly dissimilar compounds as decoys based on ECFP4 fingerprints and did not optimize the ligand set accordingly.⁴⁹ Because of the concept of this design, ligands in DUD-E tend to be apparent against decoys when DUD-E is used to benchmark 2D fingerprints based similarity search. Unlike DUD-E, MUBD-hCRs was designed for both SBVS and LBVS approaches, thus it maintains a reasonable degree of structural similarity between decoys and ligands as well as dissimilarity in order to balance “2D bias” and “false negative” bias.^{43, 45} Therefore, the different design concepts lead to the significant difference in “2D bias” for CXCR4/DUD-E and CXCR4/MUBD-hCRs. In addition, Shoichet, B. K. *et al.* mentioned the design concept of DUD-E may also contribute to artificial boost in docking enrichment.⁴⁹ Theoretically, high structural dissimilarity between ligands and decoys can make it easy to discriminate themselves by molecular docking. We have observed that both AUCs and ROCE@1% by multiple docking programs based on CXCR4/DUD-E are greater than those based on CXCR4/MUBD-hCRs. In addition to the validated “artificial enrichment” due to property mismatching, we deem that the increased enrichment for CXCR4/DUD-E may result from its unique design.

Concluding Remarks.—According to the above analysis, CXCR4/MUBD-hCRs is unbiased for “artificial enrichment” and “2D bias” while CXCR4/DUD-E appears to be biased for both. Because the evaluation outcome for FRED and GOLD from CXCR4/DUD-E is consistent with that from CXCR4/MUBD-hCRs, the minor “artificial enrichment” in CXCR4/DUD-E may not affect ligand enrichment assessment greatly for docking programs. However, it seems that the “2D bias” in CXCR4/DUD-E significantly boosts ligand enrichment of FCFP_6 based similarity search, thus rendering the evaluation outcome “LBVS-favorable”. In other words, the assessment based on CXCR4/DUD-E may overestimate the performance of FCFP_6 based similarity search in real world VS. Due to its unbiased feature, it appears to be much fair to perform CXCR4 ligand enrichment assessment for FRED, GOLD and FCFP_6 based similarity search using CXCR4/MUBD-hCRs. Since early recognition is more critical in practical application, we deem that FRED would perform finely for early recognition of CXCR4 ligands in real-world screening.

MUBD-hCRs for CCR5 Ligand Enrichment Assessment.

General Outcomes.—Based on CCR5/MUBD hCRs, FRED performs approximately the same as GOLD for the overall enrichment (0.732 vs. 0.724) but worse for early enrichment (11.112 vs. 16.668). Based on CCR5/GLL/GDD, FRED performs better than GOLD for the overall enrichment (0.720 vs. 0.566). However, the early enrichments (ROCE@1%) for both of them are 0. For FCFP_6-based similarity search, it performs worse than docking approaches based on CCR5/MUBD-hCRs since both its overall and early enrichment are lower (0.571 for AUC and 10.723 for ROCE@1%). Based on CCR5/GLL/GDD, FCFP_6-based similarity search shows the lowest overall enrichment (0.556 for AUC), though it provides higher enrichment at 1% of binding decoys (6.825) in comparison to FRED and GOLD. (cf. Table 4 and Figure 5(B))

Benchmarking Bias.—We measured “artificial enrichment” and “2D bias” in CCR5/MUBD-hCRs and CCR5/GLL/GDD using the same metrics for CXCR4 data sets. From LOO CV for “*simp*”-based similarity search, the value of mean(AUCs) is 0.456 for CCR5/MUBD-hCRs, with a std. of 0.075. And the value of mean(AUCs) for CCR5/GLL/GDD is 0.480 with a std. of 0.126. (cf. Table 5) The facts that both values are less than 0.5 indicates (1) no “artificial enrichment” bias exists in docking enrichment by FRED or GOLD from either CCR5/MUBD-hCRs or CCR5/GLL/GDD; (2) “antiscreeing” phenomenon does exist during the “*simp*”-based similarity search. The property distribution curves of ligands and decoys in Figure 6 further validate the above observations. Apparently, the curves from CCR5/MUBD-hCRs match superiorly to those for CCR5/GLL/GDD though neither one had achieved perfect matching for every property. As mentioned before, the *NLBScore* was calculated to measure potential “2D bias” (cf. Table 5). Both *NLBScores* from CCR5/MUBD hCRs and CCR5/GLL/GDD are approximately 0, indicating that both CCR5 data sets are almost free of the “2D bias”. In Figure 7, the overall shape of *Tc* distribution curve for CCR5/GLL/GDD appears to be similar to that of CCR5/MUBD-hCRs. Besides, as shown in Table 5, the ratios of ligands per scaffold are approximately 1.00 for both data sets. These data indicate that CCR5/GLL/GDD and CCR5/MUBD-hCRs are equivalent in general.

Concluding Remarks.—Through the analysis of benchmarking bias, we demonstrated that both CCR5/MUBD-hCRs and CCR5/GLL/GDD are almost unbiased in terms of “artificial enrichment” and “2D bias”. To our knowledge, the result is expected for MUBD-hCRs since the MUBD DecoyMaker tool was specifically designed to reduce these benchmarking biases for every data set. In comparison, GLL/GDD chooses property matching decoys based on the “first come, first served” principle and does not require the resemblance of decoys to ligands.⁵² Based on this principle, it is inferred that their method normally cannot reduce “2D bias. Our earlier study that identified “analogue bias” from the 17 representative data sets of GLL/GDD validated our insights.⁴⁵ Therefore, CCR5/GLL/GDD seems to be a unique data set different from those previously studied data sets. The only difference we can identify was its rather limited number of ligands, i.e. 6.

Jahn, A. *et al.* mentioned that a limited number of chemotypes may bring forth bias, thus the benchmarking outcome based on CCR5/GLL/GDD is somewhat ambiguous in that we are not sure whether FCFP_6-based similarity search truly performs better than the docking approaches. It is even more suspicious that all 6 ligands in CCR5/GLL/GDD are so challenging that no molecular docking program, e.g. FRED or GOLD, is able to enrich any of them at the top 1% of their corresponding decoy sets. To our knowledge, such an extreme case may mask the inherent differences of various docking programs. On the contrary, from the benchmarking studies based on our CCR5/MUBD hCRs (i.e. 72 ligands) we are able to clearly identify that: (1) molecular docking undoubtedly performs greater than FCFP_6-based similarity search; (2) GOLD performs preferably than FRED in this case. CCR5/MUBD-hCRs appears to be more effective for benchmarking studies to weigh over the optimal approaches for future use in real world VS.

CONCLUSIONS

In this study we had built maximal unbiased benchmarking data sets for the whole family of human chemokine receptors, i.e. MUBD-hCRs, by applying our recently developed tools of MUBD-DecoyMaker.⁴⁵ MUBD-hCRs covers 13 out of 20 chemokine receptors at the current phase with a total of 404 diverse ligands and 15756 decoys, and is readily expandable to more subtypes in the future when more binders becomes available. The ready-to-use property of MUBD hCRs makes the data sets advantageous as it avoids users' subjective collection of ligands, a seemingly unavoidable step for both standalone and on line decoy generators. Therefore, MUBD-hCRs represents a uniform benchmark for human chemokine receptors, based on which the benchmarking outcomes can be comparable across different VS studies.

We thoroughly validated every data sets in MUBD-hCRs by (1) measuring the “artificial enrichment” through the application of LOO CV using similarity search based on “*simp*” and distribution curves of multiple physicochemical properties; (2) quantifying “analogue bias” by the means of LOO CV using similarity search based on MACCS “*sims*”. ROC curves and mean(AUCs) from LOO CV for similarity search based on “*simp*” have proved that decoys in MUBD hCRs matched well to ligands in overall physicochemical properties. Distribution curves further validated the wellness of matching for individual physicochemical properties. Meanwhile, ROC curves and mean(AUCs) from LOO CV using similarity search based on MACCS “*sims*” have demonstrated that MUBD hCRs decoys are rather challenging to be distinguished from ligands by 2D similarity search. Therefore, every data sets in MUBD-hCRs are almost free of “artificial enrichment” and “analogue bias”. Although the criterion of “ $T_c < 0.75$ ” was used to reduce “false negative” bias in MUBD hCRs, it should be noted that currently no computational method is able to preclude the possibility of the decoys being active. Therefore, this type of bias may still exist in MUBD-hCRs.

We also studied the performance of MUBD-hCRs in the assessments of FRED, GOLD and FCFP_6-based similarity search with the reference of DUD-E and GLL/GDD. For the target of CXCR4, we identified potential benchmarking bias in particular “2D bias” in CXCR4/DUD-E while CXCR4/MUBD-hCRs appears to be much fair in benchmarking LBVS approaches, e.g. FCFP_6 based similarity search. As the benchmarking outcome from MUBD hCRs was consistent with that from DUD E in benchmarking docking programs, we conclude that DUD-E remains a gold-standard benchmark for SBVS approaches. For the target of CCR5, both CCR5/GLL/GDD and CCR5/MUBD-hCRs are unbiased for “analogue bias” (or “2D bias”). Nevertheless, the latter appeared to be more effective to assess the optimal molecular docking program, e.g. FRED or GOLD, vs. FCFP_6 based similarity search in that it contains much larger chemical diversity in its ligand set. Both the minimal “2D” bias and the strong capacity to distinguish between docking programs contributed to a fair assessment for the above three VS approaches. Accordingly, we proposed that FRED was the optimal approach for VS against CXCR4, while GOLD was likely to lead to high hit rate in real world screening for CCR5 ligands. Based on the case studies for CXCR4 and CCR5, we have demonstrated our MUBD hCRs was specifically designed to make up the

weakness of other currently available benchmark, e.g. DUD E in ligand enrichment assessment of LBVS approaches.

In summary, MUBD-hCRs is a maximal unbiased benchmarking sets collection and can be applied to comparing ligand enrichments of all VS approaches in a fair way so as to suggest the optimal method(s). The most useful case in practice would be rational design of a pipeline that integrates both SBVS and LBVS approaches for real world VS screening against chemokine receptors. However, the users need to beware that the construction of MUBD-hCRs was highly restricted by the accessible chemical data. For instance, the ligands were annotated with activity status (i.e. active) but not with functional effects (e.g. agonists and antagonists). To construct a high-quality benchmarking sets, we will keep updating MUBD-hCRs by the inclusion of information such as binding sites as well as functional effects of ligands in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported in part by District of Columbia Developmental Center for AIDS Research (P30AI087714), National Institutes of Health Administrative Supplements for U.S.-China Biomedical Collaborative Research (5P30AI087714-02). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are also grateful to National Natural Science of China (91213302, 81161120404, and 81603027), CAMS Innovation Fund for Medical Sciences (2017-I2M-2-004), Universities and Colleges Key Programs for Foreign Talent of State Administration of Foreign Experts Affairs P.R. China (T2017069).

ABBREVIATIONS USED

CRs	chemokines receptors
GPCRs	G protein-coupled receptors
CXCRs	CXC receptors
CCRs	CC receptors
CX3CR	CX3C receptor
XCR	XC receptor
VS	virtual screening
LBVS	ligand-based virtual screening
SBVS	structure-based virtual screening
QSAR	quantitative structure activity relationship
DUD	directory of useful decoys
DUD-E	DUD-Enhanced

NRLiSt BDB	nuclear receptors ligands and structures benchmarking database
WOMBAT	world of molecular bioactivity
VDS	virtual decoy sets
GLL	GPCR ligand library
GDD	GPCR Decoy Database
DEKOIS	demanding evaluation kits for objective in silico screening
REPROVIS-DB	database of <i>re</i> producible <i>vir</i> tual screens
MUV	maximum unbiased validation
MUBD	maximal-unbiased benchmarking datasets
hCRs	human chemokine receptors
sims	similarity in structure
<i>Tc</i>	Tanimoto coefficient
MW	molecular weight
HBA	hydrogen bond acceptor
HBD	hydrogen bond donor
RB	rotatable bond
NC	net charge
PDs	potential decoys
FDs	final decoys
<i>simp</i>	similarity in physiochemical properties
LOO CV	leave-one-out cross validation
ROC	receiver operator characteristic
AUC	area under curve
FCFP_6	function class fingerprints of maximum diameter 6
ROCE	ROC enrichment
NLBScore	nearer ligands bias score
NL	nearer ligand

REFERENCES

1. Solari R; Pease JE; Begg M, "Chemokine Receptors as Therapeutic Targets: Why Aren't There More Drugs?". *Eur. J. Pharmacol* 2015, 746, 363–367. [PubMed: 25016087]
2. Allegretti M; Cesta MC; Garin A; Proudfoot AE, Current Status of Chemokine Receptor Inhibitors in Development. *Immunol. Lett* 2012, 145, 68–78. [PubMed: 22698186]
3. Scholten DJ; Canals M; Maussang D; Roumen L; Smit MJ; Wijtmans M; de Graaf C; Vischer HF; Leurs R, Pharmacological Modulation of Chemokine Receptor Function. *Br. J. Pharmacol* 2012, 165, 1617–1643. [PubMed: 21699506]
4. White GE; Iqbal AJ; Greaves DR, Cc Chemokine Receptors and Chronic Inflammation Therapeutic Opportunities and Pharmacological Challenges. *Pharmacol. Rev* 2013, 65, 47–89. [PubMed: 23300131]
5. Caramori G; Di Stefano A; Casolari P; Kirkham PA; Padovani A; Chung KF; Papi A; Adcock IM, Chemokines and Chemokine Receptors Blockers as New Drugs for the Treatment of Chronic Obstructive Pulmonary Disease. *Curr. Med. Chem* 2013, 20, 4317–4349. [PubMed: 24059236]
6. Balkwill F, Cancer and the Chemokine Network. *Nat. Rev. Cancer* 2004, 4, 540–550. [PubMed: 15229479]
7. Palacios Arreola MI; Nava Castro KE; Castro JI; Garcia Zepeda E; Carrero JC; Morales Montor J, The Role of Chemokines in Breast Cancer Pathology and Its Possible Use as Therapeutic Targets. *J. Immunol. Res* 2014, 2014, 849720. [PubMed: 25165728]
8. Deng H; Liu R; Ellmeier W; Choe S; Unutmaz D; Burkhart M; Di Marzio P; Marmon S; Sutton RE; Hill CM; Davis CB; Peiper SC; Schall TJ; Littman DR; Landau NR, Identification of a Major Co Receptor for Primary Isolates of Hiv 1. *Nature* 1996, 381, 661–666. [PubMed: 8649511]
9. Feng Y; Broder CC; Kennedy PE; Berger EA, Hiv 1 Entry Cofactor: Functional Cdna Cloning of a Seven Transmembrane, G Protein-Coupled Receptor. *Science* 1996, 272, 872–877. [PubMed: 8629022]
10. Dragic T; Trkola A; Thompson DA; Cormier EG; Kajumo FA; Maxwell E; Lin SW; Ying W; Smith SO; Sakmar TP; Moore JP, A Binding Pocket for a Small Molecule Inhibitor of Hiv 1 Entry within the Transmembrane Helices of Ccr5. *Proc. Natl. Acad. Sci. U. S. A* 2000, 97, 5639–5644. [PubMed: 10779565]
11. Choi WT; Duggineni S; Xu Y; Huang Z; An J, Drug Discovery Research Targeting the Cxc Chemokine Receptor 4 (Cxcr4). *J. Med. Chem* 2012, 55, 977–994. [PubMed: 22085380]
12. Irwin JJ, Community Benchmarks for Virtual Screening. *J. Comput. Aided Mol. Des* 2008, 22, 193–199. [PubMed: 18273555]
13. Stahura FL; Bajorath J, New Methodologies for Ligand Based Virtual Screening. *Curr. Pharm. Des* 2005, 11, 1189–1202. [PubMed: 15853666]
14. Braga RC; Andrade CH, Assessing the Performance of 3d Pharmacophore Models in Virtual Screening: How Good Are They? *Curr. Top. Med. Chem* 2013, 13, 1127–1138. [PubMed: 23651486]
15. Schuster D; Wolber G, Identification of Bioactive Natural Products by Pharmacophore Based Virtual Screening. *Curr. Pharm. Des* 2010, 16, 1666–1681. [PubMed: 20222852]
16. Horvath D, Pharmacophore Based Virtual Screening. *Methods Mol. Biol* 2011, 672, 261–298. [PubMed: 20838973]
17. Kim KH; Kim ND; Seong BL, Pharmacophore Based Virtual Screening: A Review of Recent Applications. *Expert Opin. Drug Discovery* 2010, 5, 205–222.
18. Tropsha A; Golbraikh A, Predictive Qsar Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Curr. Pharm. Des* 2007, 13, 3494–3504. [PubMed: 18220786]
19. Willett P, Similarity Searching Using 2d Structural Fingerprints. *Methods Mol. Biol* 2011, 672, 133–158. [PubMed: 20838967]
20. Ripphausen P; Nisius B; Bajorath J, State of the Art in Ligand Based Virtual Screening. *Drug Discovery Today* 2011, 16, 372–376. [PubMed: 21349346]
21. Tuccinardi T, Docking Based Virtual Screening: Recent Developments. *Comb. Chem. High Throughput Screen* 2009, 12, 303–314. [PubMed: 19275536]

22. Cheng T; Li Q; Zhou Z; Wang Y; Bryant SH, Structure Based Virtual Screening for Drug Discovery: A Problem Centric Review. *AAPS J* 2012, 14, 133–141. [PubMed: 22281989]
23. Waszkowycz B; Clark DE; Gancia E, Outstanding Challenges in Protein Ligand Docking and Structure-Based Virtual Screening. *Wiley Interdiscip. Rev.-Comput. Mol. Sci* 2011, 1, 229–259.
24. Debnath AK, Generation of Predictive Pharmacophore Models for Ccr5 Antagonists: Study with Piperidine and Piperazine Based Compounds as a New Class of Hiv 1 Entry Inhibitors. *J. Med. Chem* 2003, 46, 4501–4515. [PubMed: 14521412]
25. Afantitis A; Melagraki G; Sarimveis H; Koutentis PA; Markopoulos J; Igglessi Markopoulou O, Investigation of Substituent Effect of 1-(3,3-Diphenylpropyl)-Piperidinyl Phenylacetamides on Ccr5 Binding Affinity Using Qsar and Virtual Screening Techniques. *J. Comput. Aided Mol. Des* 2006, 20, 83–95. [PubMed: 16783600]
26. Kellenberger E; Springael JY; Parmentier M; Hachet Haas M; Galzi JL; Rognan D, Identification of Nonpeptide Ccr5 Receptor Agonists by Structure Based Virtual Screening. *J. Med. Chem* 2007, 50, 1294–1303. [PubMed: 17311371]
27. Perez Nueno VI; Ritchie DW; Borrell JI; Teixido J, 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* 2008, 48, 2146–2165. [PubMed: 18942828]
28. Carrieri A; Perez Nueno VI; Fano A; Pistone C; Ritchie DW; Teixido J, Biological Profiling of Anti Hiv Agents and Insight into Ccr5 Antagonist Binding Using in Silico Techniques. *ChemMedChem* 2009, 4, 1153–1163. [PubMed: 19544518]
29. Perez Nueno VI; Ritchie DW; Rabal O; Pascual R; Borrell JI; Teixido J, 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* 2008, 48, 509–533. [PubMed: 18298095]
30. Perez Nueno VI; Pettersson S; Ritchie DW; Borrell JI; Teixido J, Discovery of Novel Hiv Entry Inhibitors for the Cxcr4 Receptor by Prospective Virtual Screening. *J. Chem. Inf. Model* 2009, 49, 810–823. [PubMed: 19358515]
31. Neves MA; Simoes S; Sa e Melo ML, Ligand Guided Optimization of Cxcr4 Homology Models for Virtual Screening Using a Multiple Chemotype Approach. *J. Comput. Aided Mol. Des* 2010, 24, 1023–1033. [PubMed: 20960031]
32. Planesas JM; Perez Nueno VI; Borrell JI; Teixido J, Impact of the Cxcr4 Structure on Docking Based Virtual Screening of Hiv Entry Inhibitors. *J. Mol. Graphics Modell* 2012, 38, 123–136.
33. Karaboga AS; Planesas JM; Petronin F; Teixido J; Souchet M; Perez Nueno VI, Highly Specific and Sensitive Pharmacophore Model for Identifying Cxcr4 Antagonists. Comparison with Docking and Shape Matching Virtual Screening Performance. *J. Chem. Inf. Model* 2013, 53, 1043–1056. [PubMed: 23577723]
34. Vitale RM; Gatti M; Carbone M; Barbieri F; Felicita V; Gavagnin M; Florio T; Amodeo P, Minimalist Hybrid Ligand/Receptor Based Pharmacophore Model for Cxcr4 Applied to a Small Library of Marine Natural Products Led to the Identification of Phidianidine a as a New Cxcr4 Ligand Exhibiting Antagonist Activity. *ACS Chem. Biol* 2013, 8, 2762–2770. [PubMed: 24102412]
35. Vaidehi N; Schlyer S; Trabanino RJ; Floriano WB; Abrol R; Sharma S; Kochanny M; Koovakat S; Dunning L; Liang M; Fox JM; de Mendonca FL; Pease JE; Goddard WA, 3rd; Horuk R, Predictions of Ccr1 Chemokine Receptor Structure and Bx 471 Antagonist Binding Followed by Experimental Validation. *J. Biol. Chem* 2006, 281, 27613–27620. [PubMed: 16837468]
36. Kim JH; Lim JW; Lee SW; Kim K; No KT, Ligand Supported Homology Modeling and Docking Evaluation of Ccr2: Docked Pose Selection by Consensus Scoring. *J. Mol. Model* 2011, 17, 2707–2716. [PubMed: 21213000]
37. Jain V; Saravanan P; Arvind A; Mohan CG, First Pharmacophore Model of Ccr3 Receptor Antagonists and Its Homology Model Assisted, Stepwise Virtual Screening. *Chem. Biol. Drug Des* 2011, 77, 373–387. [PubMed: 21284830]
38. Davies MN; Bayry J; Tchilian EZ; Vani J; Shaila MS; Forbes EK; Draper SJ; Beverley PC; Tough DF; Flower DR, Toward the Discovery of Vaccine Adjuvants: Coupling in Silico Screening and in

- Vitro Analysis of Antagonist Binding to Human and Mouse Ccr4 Receptors. PloS one 2009, 4, e8084. [PubMed: 20011659]
39. Bayry J; Tchilian EZ; Davies MN; Forbes EK; Draper SJ; Kaveri SV; Hill AV; Kazatchkine MD; Beverley PC; Flower DR; Tough DF, In Silico Identified Ccr4 Antagonists Target Regulatory T Cells and Exert Adjuvant Activity in Vaccination. Proc. Natl. Acad. Sci. U. S. A 2008, 105, 10221–10226. [PubMed: 18621704]
40. Asadollahi T; Dadfarnia S; Shabani AM; Ghasemi JB; Sarkhosh M, Qsar Models for Cxcr2 Receptor Antagonists Based on the Genetic Algorithm for Data Preprocessing Prior to Application of the Pls Linear Regression Method and Design of the New Compounds Using in Silico Virtual Screening. Molecules 2011, 16, 1928–1955. [PubMed: 21358586]
41. Huang D; Gu Q; Ge H; Ye J; Salam NK; Hagler A; Chen H; Xu J, On the Value of Homology Models for Virtual Screening: Discovering Hcxcr3 Antagonists by Pharmacophore Based and Structure Based Approaches. J. Chem. Inf. Model 2012, 52, 1356–1366. [PubMed: 22545675]
42. Yoshikawa Y; Oishi S; Kubo T; Tanahara N; Fujii N; Furuya T, Optimized Method of G Protein Coupled Receptor Homology Modeling: Its Application to the Discovery of Novel Cxcr7 Ligands. J. Med. Chem 2013, 56, 4236–4251. [PubMed: 23656360]
43. Xia J; Tilahun EL; Reid T-E; Zhang L; Wang XS, Benchmarking Methods and Data Sets for Ligand Enrichment Assessment in Virtual Screening. Methods 2015, 71, 146–157. [PubMed: 25481478]
44. Bissantz C; Folkers G; Rognan D, Protein Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. J. Med. Chem 2000, 43, 4759–4767. [PubMed: 11123984]
45. Xia J; Jin H; Liu Z; Zhang L; Wang XS, An Unbiased Method to Build Benchmarking Sets for Ligand Based Virtual Screening and Its Application to Gpcrs. J. Chem. Inf. Model 2014, 54, 1433–1450. [PubMed: 24749745]
46. Huang N; Shoichet BK; Irwin JJ, Benchmarking Sets for Molecular Docking. J. Med. Chem 2006, 49, 6789–6801. [PubMed: 17154509]
47. Good AC; Oprea TI, Optimization of Camd Techniques 3. Virtual Screening Enrichment Studies: A Help or Hindrance in Tool Selection? J. Comput. Aided Mol. Des 2008, 22, 169–178. [PubMed: 18188508]
48. Mysinger MM; Shoichet BK, Rapid Context Dependent Ligand Desolvation in Molecular Docking. J. Chem. Inf. Model 2010, 50, 1561–1573. [PubMed: 20735049]
49. Mysinger MM; Carchia M; Irwin JJ; Shoichet BK, Directory of Useful Decoys, Enhanced (Dud E): Better Ligands and Decoys for Better Benchmarking. J. Med. Chem 2012, 55, 6582–6594. [PubMed: 22716043]
50. Lagarde N; Ben Nasr N; Jeremie A; Guillemain H; Laville V; Labib T; Zagury JF; Montes M, Nrlst Bdb, the Manually Curated Nuclear Receptors Ligands and Structures Benchmarking Database. J. Med. Chem 2014, 57, 3117–3125. [PubMed: 24666037]
51. Wallach I; Lilien R, Virtual Decoy Sets for Molecular Docking Benchmarks. J. Chem. Inf. Model 2011, 51, 196–202. [PubMed: 21207928]
52. Gatica EA; Cavasotto CN, Ligand and Decoy Sets for Docking to G Protein Coupled Receptors. J. Chem. Inf. Model 2012, 52, 1–6. [PubMed: 22168315]
53. Vogel SM; Bauer MR; Boeckler FM, Dekois: Demanding Evaluation Kits for Objective in Silico Screening a Versatile Tool for Benchmarking Docking Programs and Scoring Functions. J. Chem. Inf. Model 2011, 51, 2650–2665. [PubMed: 21774552]
54. Bauer MR; Ibrahim TM; Vogel SM; Boeckler FM, Evaluation and Optimization of Virtual Screening Workflows with Dekois 2.0 a Public Library of Challenging Docking Benchmark Sets. J. Chem. Inf. Model 2013, 53, 1447–1462. [PubMed: 23705874]
55. Jahn A; Hinselmann G; Fechner N; Zell A, Optimal Assignment Methods for Ligand Based Virtual Screening. J. Cheminf 2009, 1, 14.
56. Ripphausen P; Wassermann AM; Bajorath J, Reprovis Db: A Benchmark System for Ligand Based Virtual Screening Derived from Reproducible Prospective Applications. J. Chem. Inf. Model 2011, 51, 2467–2473. [PubMed: 21902278]

57. Rohrer SG; Baumann K, Maximum Unbiased Validation (Muv) Data Sets for Virtual Screening Based on Pubchem Bioactivity Data. *J. Chem. Inf. Model* 2009, 49, 169–184. [PubMed: 19434821]
58. Cereto Massague A; Guasch L; Valls C; Mulero M; Pujadas G; Garcia Vallve S, Decoyfinder: An Easy to Use Python Gui Application for Building Target Specific Decoy Sets. *Bioinformatics* 2012, 28, 1661–1662. [PubMed: 22539671]
59. Gaulton A; Bellis LJ; Bento AP; Chambers J; Davies M; Hersey A; Light Y; McGlinchey S; Michalovich D; Al Lazikani B; Overington JP, ChEMBL: A Large Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res* 2012, 40, D1100–1107. [PubMed: 21948594]
60. Bento AP; Gaulton A; Hersey A; Bellis LJ; Chambers J; Davies M; Kruger FA; Light Y; Mak L; McGlinchey S; Nowotka M; Papadatos G; Santos R; Overington JP, The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res* 2014, 42, D1083–1090. [PubMed: 24214965]
61. Irwin JJ; Sterling T; Mysinger MM; Bolstad ES; Coleman RG, Zinc: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model* 2012, 52, 1757–1768. [PubMed: 22587354]
62. Irwin JJ; Shoichet BK, Zinc a Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model* 2005, 45, 177–182. [PubMed: 15667143]
63. Cabrera AC; Gil Redondo R; Perona A; Gago F; Morreale A, Vsdmp 1.5: An Automated Structure and Ligand Based Virtual Screening Platform with a Pymol Graphical User Interface. *J. Comput. Aided Mol. Des* 2011, 25, 813–824. [PubMed: 21826555]
64. Fawcett T, An Introduction to Roc Analysis. *Pattern Recog. Lett* 2006, 27, 861–874.
65. McGann M, Fred Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model* 2011, 51, 578–596. [PubMed: 21323318]
66. Hawkins PC; Skillman AG; Warren GL; Ellingson BA; Stahl MT, Conformer Generation with Omega: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model* 2010, 50, 572–584. [PubMed: 20235588]
67. Hsieh JH; Yin S; Wang XS; Liu S; Dokholyan NV; Tropsha A, Cheminformatics Meets Molecular Mechanics: A Combined Application of Knowledge Based Pose Scoring and Physical Force Field Based Hit Scoring Functions Improves the Accuracy of Structure Based Virtual Screening. *J. Chem. Inf. Model* 2012, 52, 16–28. [PubMed: 22017385]
68. Cleves AE; Jain AN, Effects of Inductive Bias on Computational Evaluations of Ligand Based Modeling and on Drug Discovery. *J. Comput. Aided Mol. Des* 2008, 22, 147–159. [PubMed: 18074107]
69. Bemis GW; Murcko MA, The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem* 1996, 39, 2887–2893. [PubMed: 8709122]

Figure 1. ROC curves from Leave-One-Out Cross Validation using “*simp*”-based similarity search.

Figure 2. Average values of AUCs of ROC curves from Leave-One-Out Cross-Validation using “*simp*”-based and MACCS “*sims*”-based similarity search. Color code: “*simp*”-based, blue; MACCS “*sims*”-based, red; random, black.

Figure 3. Distributions of physicochemical properties for ligands and decoys of all 13 data sets in MUBD-hCRs. Color code: ligands, blue; decoys, red.

Figure 4. ROC curves from Leave-One-Out Cross-Validation using MACCS “*sims*”-based similarity search.

Figure 5. Ligand enrichment assessments (AUC or ROCE@1%) of FRED, GOLD and FCFP_6-based similarity search for CXCR4 data sets from DUD-E and MUBD hCRs (A), and CCR5 data sets from GLL/GDD and MUBD hCRs (B).

Figure 6.

Distributions of physicochemical properties of ligands and decoys for CXCR4 data sets from MUBD-hCRs and DUD-E (upper panel), and CCR5 data sets from MUBD-hCRs and GLL/GDD (lower panel). Color code: MUBD-hCRs Ligands, blue; MUBD-hCRs Decoys, red; DUD-E or GLL/GDD Ligands, black; DUD-E or GLL/GDD Decoys, green.

Figure 7. Distributions of pairwise structural similarity (Tanimoto coefficient, Tc) based on FCFP₆ fingerprints between ligands and decoys for CXCR4 data sets (A) and CCR5 data sets (B). Color codes: MUBD-hCRs, green; DUD E, red; GLL/GDD, blue.

Scheme 1.

The construction of MUBD hCRs and comparative studies with DUD-E and GLL/GDD.

Table 1.

Summary of the Ligand Sets and Decoy Sets during the Construction of MUBD-hCRs.

class	subtype	no. of raw ligands	no. of curated ligands ^a	no. of diverse ligands	no. of scaffolds	ratio of ligands per scaffold	no. of decoys	activity type	cutoffs (μ M)	ChEMBL confidence score ^c
CCRs	CCR1	404	394	27	24	1.13	1053	IC ₅₀	1	8/9
	CCR2	987	891	60	49	1.22	2340	IC ₅₀	1	8/9
	CCR3	611	614	40	38	1.05	1560	IC ₅₀	1	8/9
	CCR4	158	138	18	17	1.06	702	IC ₅₀	1	8/9
	CCR5	1520	1228	72	63	1.14	2808	IC ₅₀	1	8/9
	CCR6	66	63	53	52	1.02	2067	IC ₅₀	20	9
	CCR8	127	122	12	12	1.00	468	IC ₅₀ /K _i	60	9
CXCRs	CXCR1	197	198	18	17	1.06	702	IC ₅₀ /K _i	10	8/9
	CXCR2	244	249	29	24	1.21	1131	IC ₅₀	1	8/9
	CXCR3	420	315	27	26	1.04	1053	IC ₅₀	1	8/9
	CXCR4	304	144	18	18	1.00	702	IC ₅₀ /K _i /EC ₅₀	50	8/9
	CXCR5	25	19	17	16	1.06	663	IC ₅₀	60	9
	CXCR7	49	51	13	13	1.00	507	IC ₅₀ /K _i	20	9
	max	1520	1228	72	63	1.22	2808			
	min	25	19	12	12	1	468			
	sum	5112	4426	404	369	1.08 ^b	15756			

^aCurated ligands may include multiple protonated forms for the same ligand.^bThe average of the ratios of ligands per scaffold.^cChEMBL confidence score equals to 8 and 9 refers to a homologous single protein target or a direct single protein target.

Table 2.

Mean(AUCs) and Standard Deviations (std) from Leave One Out Cross Validation Using “*simp*”-based Similarity Search and MACCS “*sims*”-based Similarity Search.

class	subtype	“ <i>simp</i> ”		MACCS “ <i>sims</i> ”	
		mean(AUCs)	std	mean(AUCs)	std
CCRs	CCR1	0.484	0.037	0.530	0.071
	CCR2	0.468	0.068	0.559	0.092
	CCR3	0.480	0.074	0.560	0.082
	CCR4	0.468	0.042	0.557	0.091
	CCR5	0.456	0.075	0.542	0.072
	CCR6	0.483	0.007	0.492	0.041
	CCR8	0.451	0.090	0.572	0.065
CXCRs	CXCR1	0.440	0.023	0.478	0.067
	CXCR2	0.449	0.030	0.508	0.093
	CXCR3	0.450	0.038	0.533	0.057
	CXCR4	0.432	0.074	0.561	0.130
	CXCR5	0.488	0.019	0.487	0.048
	CXCR7	0.495	0.137	0.614	0.111

Table 3.

Chemical Structures of Representative Ligand and Decoy from Each Benchmarking Set in MUBD-hCRs.

class	subtype	ligand ^a	decoy ^a	simp ^b	$\overline{simsdiff}^b$	sims ^b
CCR1				0.957	0.041	0.706
				0.975	0.048	0.630
				0.973	0.064	0.675
CCR2				0.958	0.019	0.747
				0.900	0.039	0.746
				0.963	0.030	0.739
CCR3				0.957	0.016	0.525
				0.957	0.016	0.525
				0.957	0.016	0.525
CCR4				0.957	0.016	0.525
				0.957	0.016	0.525
				0.957	0.016	0.525
CCR5				0.957	0.016	0.525
				0.957	0.016	0.525
				0.957	0.016	0.525
CCR6				0.957	0.016	0.525
				0.957	0.016	0.525
				0.957	0.016	0.525
CCR8				0.957	0.016	0.525
				0.957	0.016	0.525
				0.957	0.016	0.525
CXCRs	CXCR1			0.902	0.026	0.694
				0.902	0.026	0.694
				0.902	0.026	0.694

class	subtype	ligand ^a	decoy ^a	<i>simp</i> ^b	$\overline{simsdif}$ ^b	<i>sims</i> ^b
	CXCR2			0.960	0.047	0.609
	CXCR3			0.963	0.044	0.727
	CXCR4			0.901	0.044	0.746
	CXCR5			0.911	0.027	0.734
	CXCR7			0.918	0.046	0.659

^a All the structures are shown in the original (un protonated) form.

^b Three similarity values, i.e. “*simp*”, “ $\overline{simsdif}$ ” and “*sims*”, between the representative ligand and the decoy are also listed.

Table 4.

Benchmarking Performances using Three Virtual Screening Approaches for CXCR4 Data Sets from MUBD hCRs and DUD E, CCR5 Data Sets from MUBD hCRs and GLL/GDD.

data set	metric	virtual screening approaches		
		FRED	GOLD	FCFP_6 ^a
CXCR4/MUBD-hCRs	AUC	0.563	0.541	0.518
	ROCE@1%	11.115	0.000	9.201
CXCR4/DUD-E	AUC	0.812	0.694	0.843
	ROCE@1%	14.756	2.459	56.265
CCR5/MUBD-hCRs	AUC	0.732	0.724	0.571
	ROCE@1%	11.112	16.668	10.723
CCR5/GLL/GDD	AUC	0.720	0.566	0.556
	ROCE@1%	0.000	0.000	6.825

^aFor FCFP_6 based similarity search, the value is the average, i.e. mean(AUCs) or mean(ROCE@1%).

Table 5.

Metrics related to Benchmarking Bias for Different Data Sets, i.e. mean(AUCs) from LOO CV for “simp”-based Similarity Search, *NLBScore* based on FCFP_6 Fingerprints and Ratio of Ligands per Murcko Scaffold.

data set	mean(AUCs) $\pm$ std (“simp”)	<i>NLBScore</i> (FCFP_6)	ligand/scaffold(ratio)
CXCR4/MUBD-hCRs	0.432 $\pm$ 0.074	0.069	18/18 (1.000)
CXCR4/DUD-E	0.621 $\pm$ 0.081	0.459	122/62(1.968) 40/27(1.481) ^a
CCR5/MUBD-hCRs	0.456 $\pm$ 0.075	0.051	72/63 (1.140)
CCR5/GLL/GDD	0.480 $\pm$ 0.126	0.067	6/5(1.200)

^aIt is when different protonated forms of one ligand are considered as identical chemical entities.