



Article

# Keras/TensorFlow in Drug Design for Immunity Disorders

Paulina Dragan <sup>1</sup>, Kavita Joshi <sup>1</sup>, Alessandro Atzei <sup>1,2</sup> and Dorota Latek <sup>1,\*</sup>

<sup>1</sup> Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-903 Warsaw, Poland; p.dragan@uw.edu.pl (P.D.); alessandro.atzei@unica.it (A.A.)

<sup>2</sup> Department of Life and Environmental Science, Food Toxicology Unit, University of Cagliari, University Campus of Monserrato, SS 554, 09042 Cagliari, Italy

\* Correspondence: dlatek@chem.uw.edu.pl

**Abstract:** Homeostasis of the host immune system is regulated by white blood cells with a variety of cell surface receptors for cytokines. Chemotactic cytokines (chemokines) activate their receptors to evoke the chemotaxis of immune cells in homeostatic migrations or inflammatory conditions towards inflamed tissue or pathogens. Dysregulation of the immune system leading to disorders such as allergies, autoimmune diseases, or cancer requires efficient, fast-acting drugs to minimize the long-term effects of chronic inflammation. Here, we performed structure-based virtual screening (SBVS) assisted by the Keras/TensorFlow neural network (NN) to find novel compound scaffolds acting on three chemokine receptors: CCR2, CCR3, and one CXC receptor, CXCR3. Keras/TensorFlow NN was used here not as a typically used binary classifier but as an efficient multi-class classifier that can discard not only inactive compounds but also low- or medium-activity compounds. Several compounds proposed by SBVS and NN were tested in 100 ns all-atom molecular dynamics simulations to confirm their binding affinity. To improve the basic binding affinity of the compounds, new chemical modifications were proposed. The modified compounds were compared with known antagonists of these three chemokine receptors. Known CXCR3 compounds were among the top predicted compounds; thus, the benefits of using Keras/TensorFlow in drug discovery have been shown in addition to structure-based approaches. Furthermore, we showed that Keras/TensorFlow NN can accurately predict the receptor subtype selectivity of compounds, for which SBVS often fails. We cross-tested chemokine receptor datasets retrieved from ChEMBL and curated datasets for cannabinoid receptors. The NN model trained on the cannabinoid receptor datasets retrieved from ChEMBL was the most accurate in the receptor subtype selectivity prediction. Among NN models trained on the chemokine receptor datasets, the CXCR3 model showed the highest accuracy in differentiating the receptor subtype for a given compound dataset.

**Keywords:** Keras; TensorFlow; neural network; G protein-coupled receptors; chemokine receptors; CCR2; CCR3; CXCR3; structure-based virtual screening; molecular dynamics; immunity disorders; inflammation; cancer



**Citation:** Dragan, P.; Joshi, K.; Atzei, A.; Latek, D. Keras/TensorFlow in Drug Design for Immunity Disorders.

*Int. J. Mol. Sci.* **2023**, *24*, 15009.

<https://doi.org/10.3390/ijms241915009>

ijms241915009

Academic Editor: Masashi Tanaka

Received: 8 September 2023

Revised: 21 September 2023

Accepted: 29 September 2023

Published: 9 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Chemokines, or chemotactic cytokines, are a group of highly conserved small proteins that participate in the immune response via the chemotaxis of cells, either in response to tissue damage or infection (inflammatory chemokines) or to ensure homeostasis (homeostatic chemokines) [1,2]. Inflammatory chemokines (including CCL1-13, CCL23, CCL24, CCL26, CXCL1-3, and CXCL5-11) are expressed under inflammatory conditions and cause an increase in leukocyte trafficking towards the inflamed tissue. Homeostatic chemokines (including CCL14, CCL15, CCL16, CCL19, CCL21, CCL25, CCL27, CCL28, CXCL4, CXCL12, and CXCL13) are expressed constitutively and induce constant homeostatic migrations to lymph nodes throughout the body and as well as the homing of immune cells, such as lymphocytes [3]. However, dual-function chemokines, e.g., CCL11, CCL17, CCL22, or XCL1 (lymphotactin) and CXCL1 (fractalkine), also exist.

Due to the chemokine system being central to physiological processes such as homeostasis and immune responses by leukocyte transferring [4,5], the expression of these proteins is a promising prognostic method for various malignancies [6]. Furthermore, a dysregulation of the chemokine system is implicated not only in cancer pathogenesis [7–9] but also in the progression of inflammatory and immune diseases, making chemokine receptors an emerging target in the development of new drugs.

The pro-inflammatory or homeostatic effects of chemokines are exerted through the activation of their receptors—a family of rhodopsin-like G protein-coupled receptors (GPCRs) that, based on the arrangement of cysteine residues in the N-terminal of the chemokines they bind, can be divided into four subfamilies: XCR, CCR, CXCR, and CX3CR [10]. To date, roughly 19 standard chemokine receptors and four atypical chemokine receptors (ACKRs) have been characterized in humans [11,12]. The latter group is less known and lacks a full-length structural characterization in the PDB but represents a promising group of drug targets. For example, ACKR3 modulates the CXCR4 signaling by acting as a decoy receptor and scavenging of CXCL12 [13]. The CXCL12 chemokine that binds both ACKR3 and CXCR4 is classified as a homeostatic chemokine and is over-expressed in autoimmune and inflammatory diseases [13]. Among homeostatic receptors, CXCR4, CCR7, and CCR9 are the most well-known [14], but many others are still being investigated.

CCR2 is a conventional chemokine receptor responding to chemokines with the cysteine CC motif in their N-termini. CCR2 is expressed largely in T cells and monocytes [15] and is specifically involved in monocyte mobilization [16]. Similarly to other chemokine receptors, CCR2 can be activated non-selectively by many different chemokines, including CCL2, CCL7, CCL8, CCL12, CCL13, and CCL16 [16]. A recently discovered chemokine PSMP—PC3-secreted microprotein (microseminoprotein, prostate-associated MSMP), which is over-expressed in cancer and promotes hepatic fibrosis, has an affinity for CCR2 on a level similar to the most potent CCL2 [17]. This has implications for the importance of CCR2 in drug discovery for a variety of pathologies, e.g., inflammatory and autoimmune diseases, such as rheumatoid arthritis [15], multiple sclerosis [16], and autoimmunity-driven type-1 diabetes [18], but also ischemic stroke [16], liver disease [19], asthma, atherosclerosis, transplant rejection [20], diabetic nephropathy, neuropathic pain, and the promotion of cancer cell metastasis [18]. CCR2 is the target of multiple clinical candidates—according to ChEMBL (accessed on 3 July 2023) [21,22], nine are already in the 2nd phase of clinical trials, and one more is in the 3rd phase, but none has been approved for clinical use so far [23].

CCR3 belongs to the same subfamily of chemokine receptors as CCR2 and is expressed predominately on the surface of eosinophils [24] and basophils [25]. Although it is also over-expressed in certain types of cancer, it is connected with a rather poor prognosis (except in prostate and ovarian cancers) in contrast to a generally better prognosis associated with a high expression of CCR2 (except in glioma, testicular, and renal cancers) due to the CCR3-mediated migration of cancer cells [26]. CCR3 is known to bind chemokines CCL5, CCL7, CCL13, CCL15, eotaxin-1 (CCL11), eotaxin-2 (CCL24), and eotaxin-3 (CCL26) [25]. The activation of CCR3 by eotaxins (eosinophil chemotactic proteins) induces inflammation and thus is involved in asthma and allergies [24], including allergic skin diseases [27]. According to ChEMBL, there are currently no drugs or clinical candidates targeting this receptor. Only recently the active-state structures of CCR2 (bound to CCL2) and CCR3 (CCL11 not visible in electron density maps, but an active-like receptor structure) have been solved using cryo-EM [23] to provide a basis for the rational design of novel immunomodulators acting on these two receptors.

CXCR3 is a chemokine receptor that is expressed mainly on immune cells, such as natural killer cells and activated T lymphocytes. In humans, it can exist in three different isoforms: CXCR3-A, CXCR3-B, and CXCR3-alt, which are a result of gene splicing. While CXCR3-A and CXCR3-B sequences display a large overlap, there is a difference in their N-terminal, which is longer in the case of CXCR3-B due to the insertion of an additional sequence fragment from exon-2 of the CXCR3 gene. CXCR3-alt, however, consists of five transmembrane domains rather than seven as a result of the deletion of 337 base pairs from exon-3 [28]. Different isoforms are known to bind different chemokines—while CXCR3-A binds CXCL9, CXCL10, and CXCL11, CXCR3-B additionally binds CXCL4 [28]; CXCR3-alt is known to bind CXCL11 [29]. Similarly to CCR3, CXCR3 has been implicated in the progression of numerous diseases, including but not limited to multiple sclerosis, rheumatoid arthritis, transplant rejection [20], systemic lupus erythematosus [30], and allergies [31]. CXCR3 knockout mice are reported to be more resistant to autoimmune diseases [18]. A clinical candidate for acute lung inflammation targeting CXCR3 has been suggested by Meyer et al., but it has not yet been tested in clinical trials [32]. Biased ligands of CXCR3 (biaryl-type VUF10661 and VUF11418) have also been discovered in addition to the biased signaling observed for endogenous agonists of CXCR3 (CXCL11 bias towards  $\beta$ -arrestin). Recently, these three agonists of CXCR3 have been shown to activate the formation of the  $G\alpha i:\beta$ -arrestin complex in non-canonical GPCR signaling [33]. This emphasizes the importance of drug design for CXCR3 in numerous diseases, such as cancer, inflammatory diseases, and autoimmune disorders.

It is worth mentioning that except for the ACKRs, the structure of chemokine receptors is vastly conserved around the DRYLAIV motif in TM3 and the ICL2 loop [34,35]. However, only 3 drugs out of 45 in trials have so far been clinically approved [36]. Mogamulizumab was first approved in 2012 as a CCR4 antibody antagonist for cancer treatment. Maraviroc was approved in 2007 as an antiviral by acting as a CCR5 antagonist, while in 2008, Plerixafor was approved as a CXCR4 partial agonist for cancer therapies. To our knowledge, no drugs have been clinically approved so far for their action on CCR2, CCR3, or CXCR3.

As mentioned above, CCR2 is involved in a wide range of diseases; however, most of the clinical trials aiming to find new CCR2-binding drugs have failed in Phase II [36,37]. CCR3 seems to be a target for asthma and allergy, but ongoing studies present a potential role of CCR3 antagonism in two disorders associated with the aging population, such as macular degeneration (MAD) and cognitive dysfunction in mice models [37]. CXCR3, mostly expressed on the surface of activated T cells, B cells, and natural killer cells, plays a crucial role in infection, autoimmune diseases, and tumor immunity by binding to specific receptors on target cell membranes to induce targeted cell migration and resulting immune responses. CXCR3 and its main ligands (i.e., CXCL9, CXCL10, and CXCL11) have been linked to the development of many tumors (Table 1). Interestingly, the CXCR3 ligands CXCL9, CXCL10, and CXCL11 demonstrate a dichotomous activity in cancer ranging from inhibition to the promotion of tumor growth [38]. This can be explained by the varied expression patterns of CXCR3 in many tumor tissues. Therefore, it is necessary to better investigate the mode of action(s) (MoAs) and related signaling pathways for CXCR3 given its potential role as a new target for clinical tumor immunotherapy. Known antagonists and agonists of CCR2, CCR3, and CXCR3 receptors are given in Table 1.

**Table 1.** Known active ligands of CCR2, CCR3, and CXCR3 receptors.

Chemokine Receptor	Active Ligands		Mechanism of Action	Indications	Status	References
	Ligand	Drugs				
CCR2		CCX-140	Antagonist	Type 2 diabetes and diabetic nephropathy	In trial, II clinical phase	[36,39]
		Plozalizumab (MLN-1202)	Antagonist	Anti-inflammatory	In trial, II clinical phase	[36]
		Plozalizumab, MLN-1202	Antagonist	Antineoplastic	In trial, II clinical phase	[36]
		CNTX-6970	Antagonist	Analgesic	In trial, II clinical phase	[36]
		Incb3284	Antagonist	Anti-inflammatory	In trial, II clinical phase	[36]
		azd2423	Antagonist	Chronic obstructive pulmonary disorder	In trial, II clinical phase	[36]
		Ccx872	Antagonist		In trial, II clinical phase	[36]
		Cenicriviroc	Antagonist	Antiviral, HIV	In trial, II clinical phase	[36]
		Ccl2-lpm	Antagonist	Anti-inflammatory	In trial, II clinical phase	[36]
CCR3		Tpi-asm8	Antagonist	Anti-asthmatic	In trial	[36]
CXCR3	CXCL9/10		Promotes lymph node metastasis, lymph node and lung metastasis, malignant ascites production, and tumor growth and metastasis	Colorectal cancer, breast cancer, ovarian cancer, lung cancer, and stomach cancer		[40]
	CXCL9/10/11		Promotes proliferation and metastasis of cancer cells and distant metastasis and inhibits tumor growth and metastasis	Esophageal cancer, kidney cancer, and osteosarcoma		[40]
	CXCL10		Promotes tumor growth and metastasis and inhibits tumor growth	Prostate cancer and glioma and myeloma		[40]

Drug discovery is a long and costly process, but it can be enhanced by computational methods, including both structure- and ligand-based virtual screening (SBVS and LBVS, respectively). Of these two, SBVS is more time-consuming, requiring the use of a target structure or a homology model to compute the approximate free energy of the ligand binding [41]. The aim of SBVS is to screen a library of compounds using a receptor homology model or its cryo-EM/X-ray structure using scoring functions based on simplified force fields. The computed free energy of binding, approximated by a scoring function (SF), enables the selection of compounds that will likely evoke the highest response in vitro. Many different programs are available to perform such library screening, including Glide [42], AutoDock, AutoDock Vina [43,44], DOCK [45], MOE [46], and GOLD [47]. Comparative studies performed using AutoDock and AutoDock Vina indicate that the latter is better at predicting binding poses, though it cannot be said that one program is inherently superior to the other—they were found to be better fitted for different drug targets. The computational time is also crucial in deciding which docking program to choose for virtual screening purposes. Recently, AutoDock Vina has been modified to fit the GPU architecture, which significantly accelerates the computations and adds more advantages in comparison to other molecular docking software [48].

In recent years, scoring functions (SFs) used in SBVS have also begun to be based on machine learning. In contrast to classical SFs, ML-based SFs do not make use of a fixed

functional form (usually linear) that is based on the relationship between the characteristics of the protein–ligand complex and the binding affinity. Instead, in their case, the functional form is based purely on the information obtained by ML from the training data [49]. This way, it is possible to reflect non-linear relationships between the protein–ligand complex structure and the ligand-binding affinity, e.g., by using neural networks (NNs), random forest (RF), or support vector machines (SVMs) [50]. Deep learning methods—especially, convolutional neural networks (CNNs)—have been applied in SBVS in order to obtain more reliable results from docking calculations [50]. Such approaches include DeepVS [51], DenseFS [52], and Gonczarek et al.’s fingerprinting method involving learnable atomic convolution [53]. Furthermore, CNNs have also been used in the prediction of binding poses and affinities, and more robust models can be built by combining them with transfer and multitask learning [50]. Noteworthy, deep learning methods are not always better than those based on classical machine learning [54]. Classical ML methods are typically used for rescoring or ranking the output from popular molecular docking programs rather than being directly integrated into them, and their results are not easily interpretable [50]. The interpretability of the results of a deep learning method can be important, especially when it comes to medical applications [55], e.g., for finding gene–drug associations [56]. One of the common methods to explain ML results is SHapley Additive exPlanations (SHAP) [57]. Shapley values allow the importance of specific features to be assessed by computing three properties: consistency, missingness, and local accuracy. This method demonstrates a high consistency with human intuition [57]. Other interpretability methods include DeepLIFT [58], especially used for deep NNs, or Grad-CAM++ [59], which is used to visually explain the predictions of CNN models [57]. There are also frameworks joining various methods to uncover global feature importance in contrast to the local interpretation of each feature, e.g., SAGE [60,61].

In principle, LBVS is much faster, based solely on the structure and physicochemical properties of ligands known to interact with the molecular target in order to predict the affinities of yet untested compounds [62]. This makes it possible to use when the structure of a receptor is unavailable, which is often the case with GPCRs. The applicability of machine learning methods in ligand-based virtual screening has been widely discussed so far, e.g., in [63]. Constantly increasing the number of available ligand datasets for various drug targets and improving the quality and quantity of such datasets enhance the accuracy of computational drug discovery despite minor problems with the integration and optimization of used ML methods [64]. In supervised ML, feature selection is used to recognize relevant molecular (in case of drug discovery) or genomic (in case of genomic analysis) features that inform about drug responses or drug–gene associations. These techniques, however, require the labeling of the used training dataset, e.g., prior knowledge about drug–target associations [65]. Thus, their use may inhibit the discovery of potential new actives, as compounds not possessing the preselected features could be discarded.

Data-driven concepts, e.g., the discovery of new drugs or drug targets, require efficient and accurate algorithms, in advance, to process massive data from large biomedical repositories and to reflect subtle differences in compounds that have a huge impact on the observed biological response, respectively. Although conventional ML algorithms belonging to a supervised learners’ group, such as gradient boosting or support vector machines, seem to be the most accurate in tasks including predictions of compound activity or binding affinity [54,63,66–68], NNs constantly draw attention [64]. Learning of NNs can be carried out as supervised learning (first NNs, also including backpropagation NNs) or unsupervised learning (e.g., deep belief networks) where layers can detect relevant features. Other types of learning, e.g., reinforcement learning, can also be used to train NNs [69]. In principle, if no labeled input data are used to train NN, it has many more possibilities of finding new active-like scaffolds compared to supervised-learning methods. This is the basis for the popularity of NNs and deep learning NNs in various tasks, ranging from image recognition, natural language processing, and engineering applications [70], to the retrieval of relevant information from databases, e.g., in protein structure prediction [71]

or in drug design [72]. The two, most used systems for machine learning and especially for deep learning tasks are TensorFlow, developed by Google [73], and PyTorch [74], co-developed by A. Paszke. Keras API [75] makes it possible to define and train ML models implemented on TensorFlow or PyTorch platforms to easily release open-source projects and construct pipelines joining various libraries, e.g., RDKit [76], for compound fingerprints [77]. TensorFlow with or without high-level Keras API is widely used due to the easy implementation of algorithms that are otherwise difficult to optimize flawlessly, such as convolutional neural networks [78]. One of the key concepts recently introduced in TensorFlow2.0 and PyTorch is a “define-by-run” paradigm [79,80], in which connections in NNs are defined during the training, not before. This backpropagation allows for a more efficient automatic differentiation scheme compared to “define-and-run” in TensorFlow1.0. Recently, TensorFlow has been used for rapid screening for GPCR ligands [81].

Keras/TensorFlow NNs used in this study were based on a sequential NN model, meaning that it is built layer by layer. In such a model, each layer has one input tensor and one output tensor, and the resulting NN topology is linear [82]. As an input, it received a set of compounds represented as SMILES [83] and their activity was measured by pIC50 standardized to pChEMBL values. Here, extended connectivity fingerprints (ECFPs) based on Morgan fingerprints were used to describe compounds, together with pChEMBL values and an additional parameter referring to the compound activity class. As a result of using such a set of molecular descriptors, the NN model learned on both categorical and numerical data. Developed Keras/TensorFlow NN predicted the most probable activity class for a given compound in a multi-class classification task instead of binary classification (active vs. inactive) for which NNs are commonly applied.

Keras/TensorFlow or any other ML system used in LBVS, if additionally combined with SBVS, allows for a more precise and reliable assessment of the ligand-binding affinity and its detailed binding mode [54]. In the final step, molecular dynamics (MD) can be used to validate the molecular docking-based binding affinity and binding modes of discovered compounds and thus reduce the number of false positives before the bioassay studies [82]. This combined computational approach significantly reduces both the time and cost required to find novel chemotypes.

Here, we performed MD simulations for previously obtained novel CCR2 and CCR3 antagonists and used a combination of AutoDock Vina for SBVS, Keras/Tensorflow sequential model of neural network (NN) for LBVS, and MD simulations in order in order to find and validate novel small-molecule antagonists for CCR2, CCR3, and CXCR3 chemokine receptors. While previously [81] the impact of the ligand dataset composition on the ML results was discussed, here, we focused on the ability of ML to reflect slight structural differences between ligands matching the certain receptor subtype which account for their receptor subtype selectivity. In [83], we assessed gradient-boosting decision trees (LightGBM) in the recognition of the receptor subtype-selective and non-selective ligands of cannabinoid receptors. Here, we assess NNs (Keras/TensorFlow) in such a task using not only the curated compound datasets for CB1/CB2 cannabinoid receptors (<http://db-gpcr.chem.uw.edu.pl> (accessed on 20 August 2023)) but also for CCR2/CCR3/CXCR3 chemokine receptors.

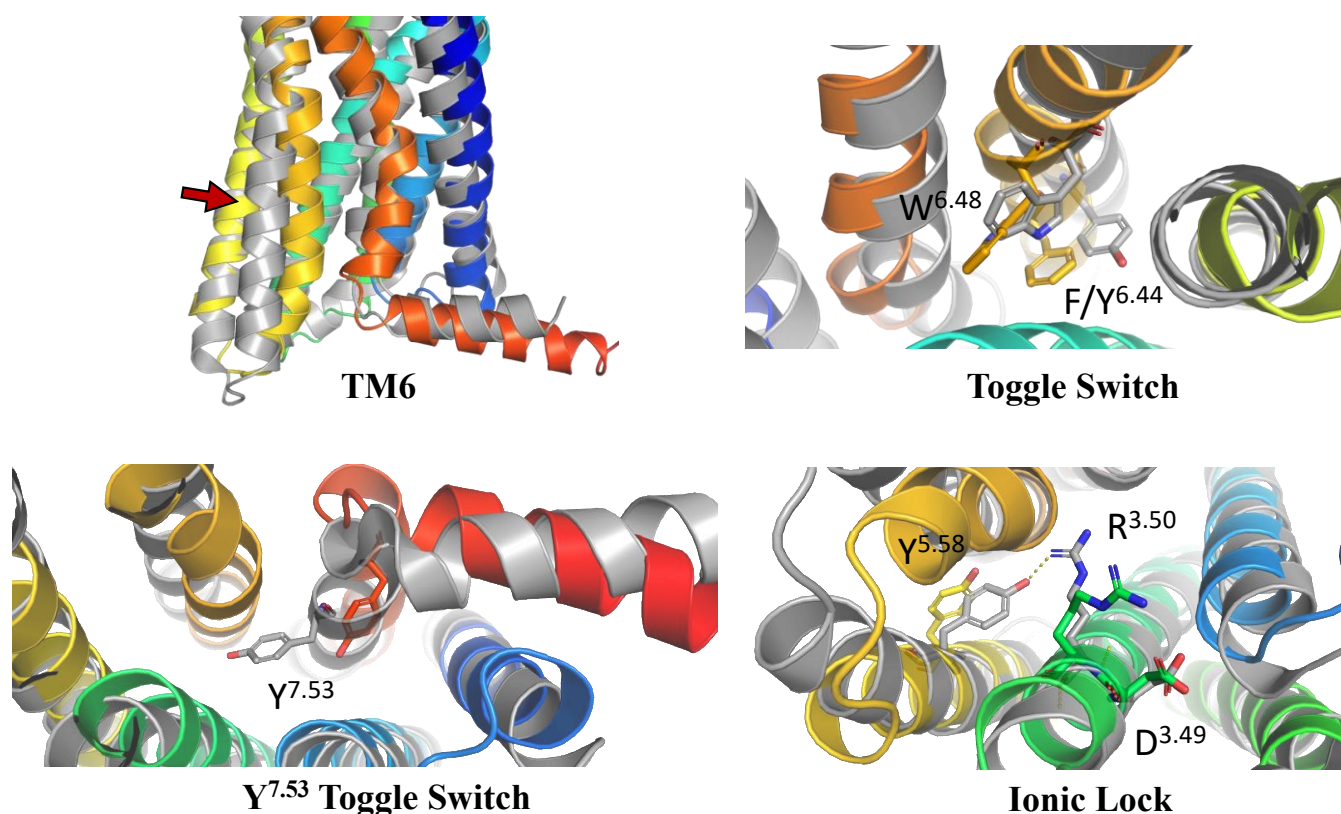
## 2. Results

### 2.1. CXCR3 Model Validation

The positions of transmembrane helices 5 and 6 (TM5 and TM6) were the first to be analyzed. It is well-known that during the activation of the GPCR receptor, the extracellular region of its TM5 helix moves inwards, while the intracellular region of TM6 moves outwards [84]. These differences in the inactive and active-state conformations are shown in Figure 1. Indeed, the location and shape of TM5 and TM6 in our CXCR3 model are similar to those observed for inactive-state structures of chemokine receptors.

Similarly, the conformations of selected residues were analyzed (see Figure 1 and Supplementary S1 Figure S1 [85]): the W<sup>6.48</sup> toggle switch with F<sup>6.44</sup> from the PIF motif,

$Y^{7.53}$  toggle switch, and  $D^{3.49}R^{3.50}Y^{3.51}$  ionic lock [84,86,87].  $W^{6.48}$  with  $F^{6.44}$  was rotated counter clockwise (from the extracellular side) as compared to the active-state structure, the  $Y^{7.53}$  toggle switch was rotated counter clockwise (from the intracellular side), and  $R^{3.50}$  in the ionic lock was closed to  $D^{3.49}$  instead of interacting with  $Y^{4.58}$  as in active-state structures of chemokine receptors. On this basis, it was concluded that the model was suitable for performing structure-based virtual screening, i.e., the molecular switches present in the model were in the conformations expected to be present in an inactive-state chemokine receptor.

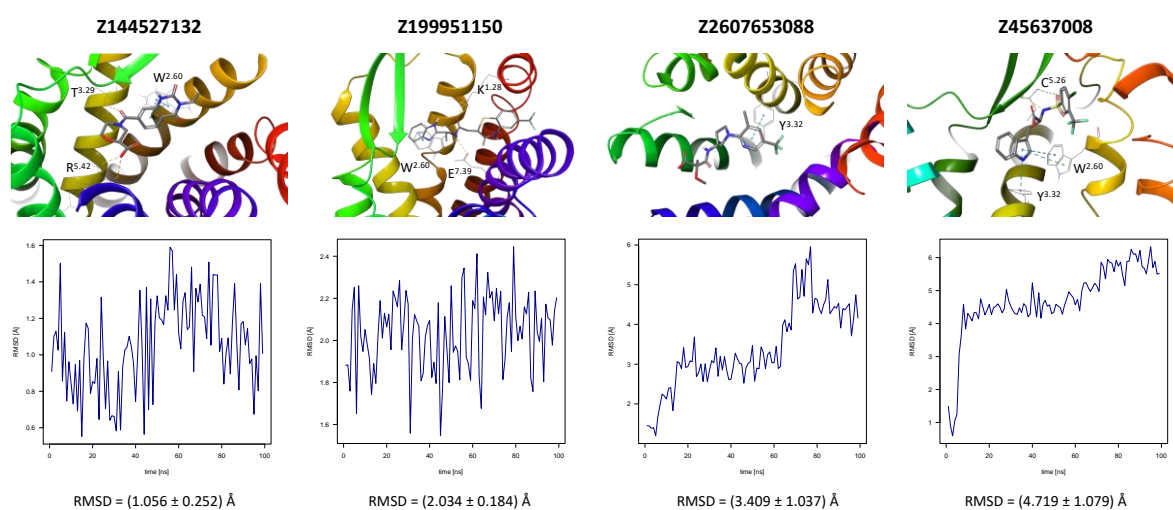


**Figure 1.** The validation of the inactive-state CXCR3 model through an analysis of micro- and macroswitches. The inactive-state model of CXCR3 (blue-to-red) was superposed on active-state chemokine receptor structures (gray): 7O7F (CCR5) [88] for TM6 and the tryptophan toggle switch W6.48 with F6.44 from the PIF motif [89], and 6WWZ (CCR6) for comparison of the tyrosine toggle switch Y7.53 and the ionic lock including R3.50 from the DRY motif. The residues have been labeled using the Ballesteros–Weinstein numbering system [90].

## 2.2. MD-Based Validation of Ligand Binding Modes

CCR2 and CCR3 actives proposed in a recent study by Dragan et al. were docked to the same receptor structures as before but with a different algorithm (AutoDock Vina) to confirm their binding modes. Based on these results, 6 out of 10 CCR2 actives and 7 out of 12 CCR3 actives were discarded as Glide and AutoDock Vina provided significantly different binding modes for them. For CXCR3, only AutoDock Vina was used for molecular docking prior to MD simulations. Notably, molecular docking algorithms, extremely useful for virtual screening, have limitations regarding their reproducibility of ligand binding modes. This is due to simplified force fields, in which some molecular interactions are approximated to decrease the computational time and to efficiently screen large libraries of compounds [63]. For this reason, all-atom MD simulations were used to validate the binding modes of proposed active compounds (see Supplementary S1 Tables S1–S3), following a previous study [91,92]. Of the four ligands tested for CCR2, only three remained stable throughout the course of the simulation (Figure 2). Both Z144527132 and Z199951150

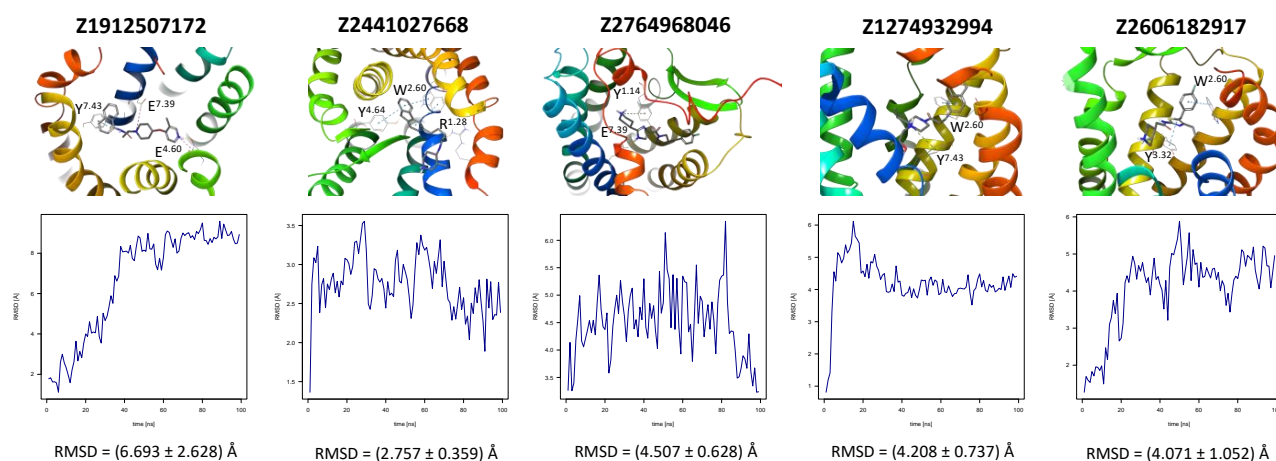
displayed high stability from the very beginning of the production run—both the RMSD values and their standard deviations were low. On the other hand, Z2607653088 was stable initially, but its hydroxyl group began moving upwards after 60 ns, as if it were leaving the receptor. At the 100 ns cutoff, the only interaction noted by Maestro for this ligand was pi-pi stacking with Y<sup>3.32</sup>; at this point in the simulation, the other tested molecules all displayed more binding interactions, further undermining the Z2607653088 binding mode obtained from molecular docking. The final compound, Z45637008, stabilized after 10 ns of the simulation, but after a noticeable relocation of its trifluoromethyl end. This change made it possible for the sulfonyl group and a nearby nitrogen atom to form hydrogen bonds with C<sup>5.26</sup>. Overall, pi-pi stacking interactions with W<sup>2.60</sup> and Y<sup>3.32</sup> appeared in multiple cases, suggesting these residues may play a significant role in the binding of CCR2 ligands.



**Figure 2.** Results of the MD simulations for CCR2 for four different compounds proposed by virtual screening. **(Top)** The interactions between the receptor and the ligand obtained after 100 ns of the simulation. The receptor was shown in the red-to-blue color scheme; yellow dashed lines—hydrogen bonds; blue dashed lines—pi-pi stacking. The residues have been labeled using Ballesteros–Weinstein numbering system [90]. **(Bottom)** The RMSD plots obtained for each of the ligands over the 100 ns simulation, as well as the average RMSD with its fluctuation range.

In the same way, five molecules were tested for the inactive-state CCR3 Robetta model (Figure 3). Z1912507172 was highly unstable over the first ca. 40 ns of the simulation. The ligand was observed to move much deeper into the binding site than according to the molecular docking results. After 40 ns, its location stabilized and remained that way until the end of the simulation. A comparison of its binding mode, both obtained in molecular docking and refined in the MD simulations, is presented in Supplementary S1 Table S2 [93]. Of the tested CCR3 compounds, Z2441027668 was the most stable, barely changing its position with respect to the results from molecular docking. Slightly larger, though still low, RMSD fluctuations were observed for the remaining three compounds: Z2764968046, Z1274732994, and Z2606182917. In general, the RMSD fluctuations obtained for the CCR3 complexes were higher than those obtained for CCR2 complexes, likely because a receptor model was used here rather than a high-quality structure. Similarly, as in the case of CCR2, W<sup>2.60</sup> and Y<sup>3.32</sup> were shown to frequently participate in ligand binding; in addition, interactions with E<sup>4.60</sup> were noted in two separate cases.





**Figure 3.** Results of the MD simulations for CCR3 for five proposed compounds. (**Top**) The interactions between the receptor and the ligand obtained after 100 ns of the simulation. The receptor was shown in the red-to-blue color scheme; yellow dashed lines—hydrogen bonds; blue dashed lines—pi-pi stacking; green dashed lines—pi-cation; purple dashed lines—salt bridges. The residues have been labeled using Ballesteros–Weinstein numbering system [90]. (**Bottom**) The RMSD plots obtained for each of the ligands over the 100 ns simulation, as well as the average RMSD with its fluctuation range.

A further five compounds were tested for the CXCR3 model—they all demonstrated a high stability of their binding mode. The most interactions with the receptor were observed for Z2233592864—mainly, pi-pi stacking. This compound, along with Z107207944 and Z1903257002, demonstrated the most stable binding mode. The two final compounds, Z1167188972 and Z1510954688, fluctuated to a much greater extent, and no specific interactions with the receptor were observed for the latter one, while for the former interactions with residues N<sup>3.33</sup> and Y<sup>3.37</sup> were rarely formed during the simulations.

Interestingly, the best compounds selected by the Keras/TensorFlow NN for CXCR3 (in the range of 9 and above of pChEMBL predicted values) did not include any of the compounds proposed for CCR2 or CCR3 (Figures 2 and 3). It means that like previously [54], predictions made by the NN model are selective for the receptor subtype because they are based on known active ligands only and not on the receptor structures which could be too similar, e.g., for SBVS. Thus, diverse CCR2, CCR3, or CXCR3 ligand training sets will provide diverse novel chemotypes for each of these receptors, while the similar structures of these receptors could only provide similar compounds in SBVS. This is further discussed in Section 2.5.

### 2.3. Chemical Modifications of Proposed Ligands

Chemical modifications of the structures of the proposed compounds were suggested in order to increase their affinity towards the receptors. The modified structures as well as their contacts with the receptor were presented in Supplementary S1 Tables S4–S6. In the case of the proposed CCR2 ligand Z144527132, the addition of the alkyl substituent with the hydroxyl group to the hydrogenated quinazoline ring was proposed in order to enable the formation of a hydrogen bond with E<sup>7.39</sup>, a residue that participated in the binding of another ligand for this receptor. A second alkyl group was added in order to fill out the hydrophobic pocket. For Z199951150, a shortening of the molecule (specifically, of the end with the trifluoride group) could be suggested, as it did not appear to greatly contribute to the binding. More significant modifications should be introduced to the structure of Z2607653088 in order to prevent it from immediately leaving the receptor. A phenyl ring in the place of the ligand part with carbonyl and hydroxyl groups might support the formation of pi-pi stacking interactions with the nearby F<sup>5.30</sup> or Y<sup>5.29</sup>. Furthermore, an introduction of a cyclopentene ring between one of the carbons and an oxygen could facilitate the

formation of a hydrogen bond with H<sup>5.38</sup>. No modifications were suggested for Z45637008. However, only one modified compound based on Z2607653088 (together with Z199951150) was in the 8–9 predicted activity range by Keras/TensorFlow NN, while the other two (see Supplementary S1 Table S4) were predicted as inactive (below 5).

For CCR3, the introduction of a hydroxyl group into the structure of Z1274732994 was suggested to facilitate the formation of a hydrogen bond with E<sup>7.39</sup>. In the case of Z1912507172, the addition of two separate alkyl groups were suggested in order to better fill out the hydrophobic region of the binding pocket of the receptor, as well as a hydroxyl group that could interact with Y<sup>3.32</sup> to form a hydrogen bond. For Z2441027668, it was suggested that the methylpiperidine ring could be transformed into methylpyridine in order to allow for potential pi-pi stacking interactions with Y<sup>1.14</sup>. For Z2606182917, the addition of an alkyl chain is suggested in order to better fill the binding cavity, as well as an oxygen that could form a hydrogen bond with H<sup>5.38</sup>. In the case of Z2764968046, a cyclopentane ring was added to the structure in order to better fill out the binding pocket. All modified compounds were in the highest predicted activity range (above 9 or in the 8–9 range) except for Z2606182917 which fell into the medium predicted activity range (7–8).

For CXCR3, an additional double bond to introduce aromaticity was added to the indane ring of Z107207944. Thus, the formation of pi-pi stacking interactions with the nearby F<sup>3.32</sup>, W<sup>6.48</sup>, or Y<sup>6.51</sup> could be facilitated. For Z1167188972, a benzene ring could be a replacement for the cyclohexane ring to facilitate the pi-pi stacking interactions with F<sup>4.63</sup>. Furthermore, an alkene chain was added to fill out the binding pocket. In the case of Z1510954688, the tetrahydropyran ring can be replaced with a benzene ring, and one of the methyl groups was removed. This would allow for pi-pi stacking interactions with Y<sup>3.37</sup>. In addition, a transformation of one of the other methyl groups present in the molecule into a hydroxyethyl group would allow for the formation of a hydrogen bond with D<sup>4.60</sup>. For Z1903257002, the cyclohexane ring could be replaced with a benzene ring to facilitate interactions with Y<sup>6.51</sup>. A subsequent relocation of one of the methyl groups would help fill out the binding pocket. No modifications were suggested for Z2233592864. Interestingly, this compound (Z2233592864) together with a modified Z107207944 was the best among all modified compounds according to Keras/TensorFlow NN (the 8–9 predicted activity range).

#### 2.4. Comparison of Proposed Compounds with Known CXCR3 Ligands

Recently, Meyer et al. [32] published a novel CXCR3 antagonist. A comparison of the described structures and Enamine's Hit Locator Library (HLL) provided three hits. Two of the molecules, Z2755039307 and Z2755039304, proved most similar to ACT-7779991 (the clinical candidate) with Tanimoto similarities equal to 0.255 and 0.253, respectively, as well as ACT-672125, with similarities equal to 0.141 for both. A third compound, Z1695828968, was most similar to ACT-660602, with a similarity of 0.207. Here, the previous two compounds had similarities equal to 0.196 and 0.195, respectively. AutoDock Vina-approximated binding affinities for these similar compounds in Enamine HLL were rather low to medium, ranging from 6.5 to 7.5. For two compounds, Z2755039307 and Z2755039304, the NN results were also unsatisfactory (see Table 2). However, the third compound Z1695828968 was assessed by the Keras/TensorFlow NN as highly active (the activity range above 9—the highest one), and it was included in less than 20% of the best compounds for CXCR3 in Enamine HLL. It shows that the Keras/TensorFlow NN does not reproduce molecular docking results but indeed may provide substantial new information on the compound activity not accessible to physics-based force fields. For comparison, the NN results for CXCR3 antagonists proposed and tested in MD simulations in this study fell into 18.6%, 3.7%, 13.5%, 19.5%, and 14.0% of top NN predictions (see Figure 4) and 72.5%, 14.4%, 52.6%, 76.1%, and 54.9% of top predictions of the 9 and above activity range, respectively.

In addition to the above-mentioned CXCR3 antagonists, we also conducted a search for compounds similar to the CXCR3-biased ligand VUF10661. Here, the results of NN

were even better (see Table 2). All three similar compounds were in the activity range of predicted pChEMBL values of 8 and above, meaning they were predicted as highly active for CXCR3. All three of these compounds were also among the best compounds found in Enamine HLL. Despite these results, in our opinion, Keras/TensorFlow NN is a method to be used in combination with classical virtual screening methods such as SBVS rather than be used solely in VS.

All four known CXCR3 ligands were docked with AutoDock Vina to five receptor conformations obtained at the end of MD simulations of five HLL compounds (see Figure 4). The best binding affinities predicted by AutoDock Vina are given in Table 2, while the binding modes are given in Supplementary S1 Table S7. There was no common receptor conformation that proved the best fit for all four compounds, but the Z1903257002-fitted conformation was discarded by all compounds possibly because of the steric hindrance caused by Y<sup>3.37</sup> forming interactions with Z1903257002 (see Figure 4).

**Table 2.** A comparison of three known CXCR3 active compounds proposed by Meyer et al. and a biased CXCR3 ligand with the most similar compounds present in Enamine's Hit Locator Library. The common substructure of all six compounds is presented in the last row.

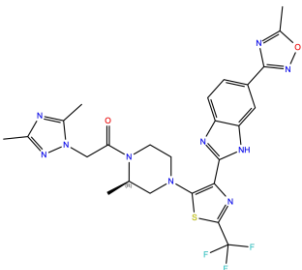
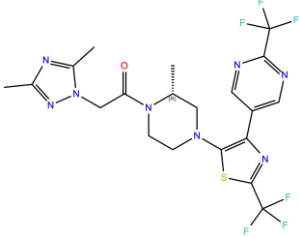
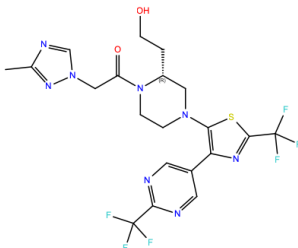
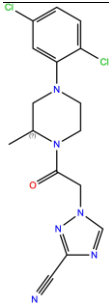
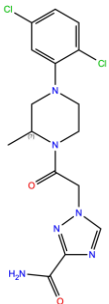
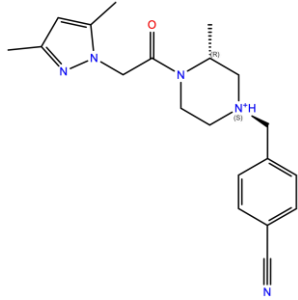
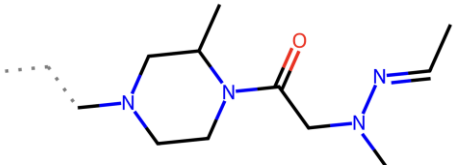
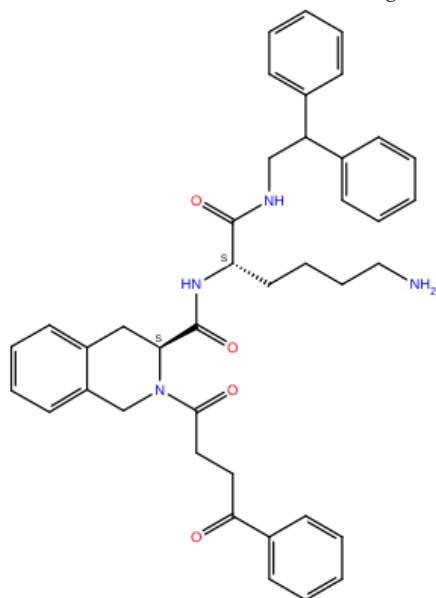
Known CXCR3 Antagonists		
 <p>ACT-672125 Predicted activity range: 7–8 Predicted binding affinity: −11.570 (receptor conformation—Z1167188972)</p>	 <p>ACT-660602 Predicted activity range: 8–10 Predicted binding affinity: −10.930 (receptor conformation—Z1510954688)</p>	 <p>ACT-777991 Predicted activity range: 7–8 Predicted binding affinity: −9.065 (receptor conformation—Z107207944)</p>
Enamine HLL		
 <p>Z2755039307 Predicted activity range: 7–8 Predicted binding affinity: −6.503 In 63.3% of top NN predictions (20.9% top predictions in the 7–8 activity range)</p>	 <p>Z2755039304 Predicted activity range: 7–8 Predicted binding affinity: −7.403 In 66.5% of top NN predictions (29.9% top predictions in the 7–8 activity range)</p>	 <p>Z1695828968 Predicted activity range: 9–above Predicted binding affinity: −7.081 In 18.6% of top NN predictions (72.5% top predictions in the 9–above activity range)</p>
Common Substructure		
		

Table 2. Cont.

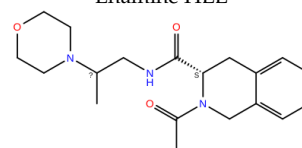
Known biased CXCR3 small-molecule ligand



VUF10661

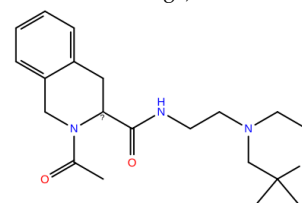
Predicted activity range: 8–9  
 Predicted binding affinity:  $-10.690$  (the best receptor conformation—Z2233592864)

Enamine HLL



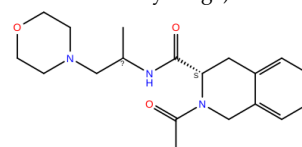
Z364673996

Predicted activity range: 8–9  
 In 26.2% of top NN predictions (2.0% top predictions in the 8–9 activity range)



Z1128741153

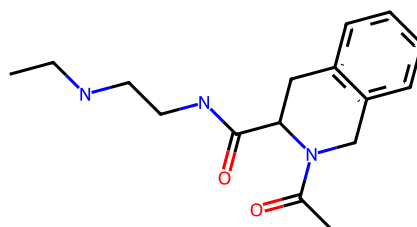
Predicted activity range: 9–above  
 In 19.0% of top NN predictions (74.4% top predictions in the 9–above activity range)



Z422692598

Predicted activity range: 8–9  
 In 36.7% of top NN predictions (36.5% top predictions in the 8–9 activity range)

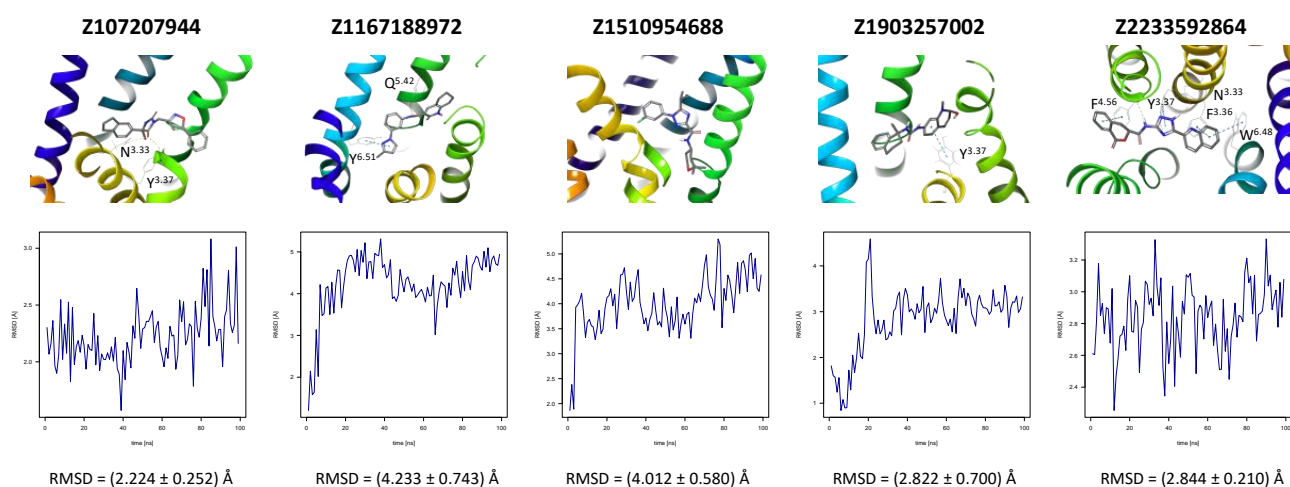
Common Substructure



All four CXCR3 ligands were additionally assessed by NN trained on the CXCR3 dataset and the CCR2 and CCR3 datasets to check if NN is sensitive to the receptor subtype selectivity. The CXCR3 model assessed these ligands rather highly, similarly to the CCR3 model, while the CCR2 model assessed them as inactive compounds. A similar observation was made for the modified compounds. They were among the top-scored compounds assessed by the CXCR3 and CCR3 models, but not by the CCR2 model. This means that the CCR3 and CXCR3 NN models were less selective with respect to each other in their activity predictions for these compounds in comparison to the CCR2 model. This again suggests the dependency of NN on the training dataset composition [54], yet in this case with the desired outcome.

The NN and SBVS predictions were not fully consistent for Meyer's compounds, meaning that the best compound proposed by NN was not the best compound proposed by SBVS. However, both NN and SBVS assessed VUF10661 as the best compound out of these four actives. This could be due to the fact that VUF10661 consists of much more functional groups than Meyer's compounds. More functional groups decrease the energy of interactions computed in molecular docking as we observed previously for statins [94]. On the other hand, the presence of more functional groups ensures that the compound

resembles at least any subset of active compounds used for training of NN; thus, NN will select it as an active compound.



**Figure 4.** Results of the MD simulations for CXCR3 for five proposed compounds. (**Top**) The interactions between the receptor and the ligand obtained for 100 ns of the simulation. The receptor was shown in the red-to-blue color scheme; yellow dashed lines—hydrogen bonds; blue dashed lines—pi-pi stacking. The residues have been labeled using Ballesteros–Weinstein numbering system [90]. (**Bottom**) The RMSD computed for each of the ligands over the 100 ns simulation, as well as the average RMSD with its fluctuation range.

#### 2.5. Performance of Keras/TensorFlow NN in the Receptor Subtype Selectivity Prediction Tasks

To compare with the previous ML study on cannabinoid receptors (LightGBM, CB1/CB2 selectivity), we also used CB1 and CB2 datasets for the NN training. This time, we included as many ChEMBL-retrieved compounds as possible (>5000) in contrast to previously limited datasets for these two receptors [83] available at <https://db-gpcr.chem.uw.edu.pl> (accessed on 20 August 2023). The average Tanimoto coefficients between the current datasets and the previous datasets were equal to 0.138 (mode: 0.17) and 0.141 (mode: 0.17) for CB1 and CB2, respectively. Both datasets included small-molecule compounds only. However, the previous datasets included data from assays that provided pKi values, while the current datasets included only data from assays that provided pIC50 (standardized to pChEMBL values). This means that the current datasets include only CB1 or CB2 small-molecule inhibitors and not all CB1 and CB2 actives, as previously observed. Furthermore, the previous datasets did not include any inactive or weakly active compounds (pChEMBL < 4), the addition of which to training sets was recently discussed in [54]. In the current datasets, nearly 40% and 25% (CB1 and CB2, respectively) of compounds were inactive compounds (pChEMBL equal to 0). Among active compounds in the current datasets, 1% and 2% were weakly active compounds (pChEMBL less than 5, CB1 and CB2, respectively). Histograms showing the distribution of the activity classes in the current and previous datasets were provided in Supplementary S1 Table S8. Despite these differences, the results of the receptor subtype selectivity prediction tasks were similar for the current and previous datasets, with only a slight improvement in comparison to the previous ones. The accuracy of the prediction for validation datasets was ca. 0.5 for the same receptor subtype, 0.2 for the other receptor subtype, less than 0.02 for CB2 selective compounds with the inconsistent receptor subtype, and nearly 1 for the consistent receptor subtype (see Tables 3 and 4). In the latter case of the CB2 selective compounds with the matching receptor subtype, the CB2 model trained on the previous dataset performed much better than that trained on the current dataset (accuracy: 0.876 vs. 0.521, see Table 4). However, this could be due to the higher average similarity between the previous training sets and the CB2 selective set (0.153 and 0.141 for the previous training set and the current one, respectively, see Table 4).

**Table 3.** Performance of Keras/TensorFlow NN in the chemokine receptor subtype selectivity tasks.

Training Set	Number of Datapoints	Validation Set	Number of Datapoints	Loss	Accuracy (Change)	Average Tanimoto Coefficient Training vs. Validation Set	Mode Tanimoto Coefficient Training vs. Validation Set
CCR2	1995	CCR2	399	5.406	0.190	0.139	0.16
		CCR3	121	9.621	0.231 (+0.041)	0.141	0.16
		CXCR3	199	17.878	0.126 (−0.064)	0.125	0.13
CCR3	603	CCR3	121	5.332	0.223	0.243	0.15
		CCR2	399	14.013	0.115 (−0.108)	0.142	0.10
		CXCR3	199	14.853	0.191 (−0.032)	0.143	0.13
CXCR3	994	CXCR3	199	4.727	0.322	0.198	0.15
		CCR2	399	13.681	0.125 (−0.197)	0.124	0.13
		CCR3	121	8.029	0.182 (−0.140)	0.142	0.13

**Table 4.** Performance of Keras/TensorFlow NN in the cannabinoid receptor subtype selectivity tasks.

Training Set	Number of Datapoints	Validation Set	Number of Datapoints	Loss	Accuracy (Change)	Average Tanimoto Coefficient Training vs. Validation Set
2023 ChEMBL datasets						
CB1	4509	CB1	902	2.455	0.503	0.129
		CB2	818	7.174	0.246 (−0.257)	0.133
		CB2 selective	35	6.499	0.0146	0.139
CB2	4087	CB2	818	2.987	0.418	0.137
		CB1	902	5.354	0.291 (−0.127)	0.131
		CB2 selective	35	2.450	0.521	0.141
2020 ChEMBL datasets [83] from <a href="https://db-gpcr.chem.uw.edu.pl">https://db-gpcr.chem.uw.edu.pl</a> (accessed on 20 August 2023)						
CB1	1566	CB1	314	1.943	0.464	0.152
		CB2	418	4.417	0.203 (−0.261)	0.148
		CB2 selective	35	9.187	0.0135	0.150
CB2	2093	CB2	418	1.919	0.509	0.152
		CB1	314	4.263	0.210 (−0.299)	0.147
		CB2 selective	35	0.487	0.876	0.153

These results confirmed that although the composition of the training dataset has a noticeable impact on the classification results [63], neural networks are still able to classify correctly despite an increase in noise in the training sets. Here, noise in datasets was introduced by adding inactive compounds to the current dataset. This advantage of NNs over supervised methods like gradient-boosting decision trees (LightGBM) is mostly due to the fact that NNs can also act as unsupervised learners using unlabeled datasets for training. What is more, adding inactive or weakly active compounds to training sets only slightly worsened the accuracy of the activity prediction, which was also expected based

on [54]. Adding inactive compounds to training sets could improve the binary classification (active vs. inactive compounds) but not the activity value prediction, which is a multiclass classification task [54].

If we compare the results presented in Tables 3 and 4, the NN model trained on the cannabinoid receptor datasets seems to be more accurate in the selectivity prediction than models trained on the chemokine receptor datasets. In the case of the CB1 model, the prediction accuracy dropped by more than 0.2 when the validation set with the inconsistent receptor subtype was tested. In the case of the CB2 model, the accuracy changed by 0.3. In the case of the chemokine receptor models, the most significant change in the accuracy was for the CXCR3 model (nearly 0.2 for the CCR2 validation set), but the remaining models showed only ca. 0.1 or less change in the accuracy. The worst model regarding the selectivity prediction was the CCR2 model, which is consistent with the fact that the CCR2 ligands from the training set were almost as similar to ligands from the CCR2 validation set as from the CCR3 or CXCR3 validation sets. The CXCR3 model performed the best in the receptor subtype selectivity prediction task for the same reason. The CXCR3 ligands retrieved from ChEMBL were the most dissimilar to both CCR2 and CCR3 ligands. In all cases, the prediction accuracy of NNs correlated with values of the Tanimoto coefficient between the training and validation sets.

### 3. Discussion and Conclusions

Due to the role they play in numerous diseases, chemokine receptors represent promising drug targets—however, drug design is hindered by the unavailability of many of their structures. In such cases, homology modeling makes it possible to create models of receptors based on their similarity to other receptors with solved structures. Though this can be completed using web servers, standalone programs, such as Modeller, give researchers the opportunity to take a more hands-on approach and adjust the modeling process to suit their own needs. The created models can then be used in SBVS in order to search for novel active compounds for the receptors in question, and the results validated through the use of properly trained machine learning algorithms. Regarding virtual screening, the comparison to known CXCR3 ligands showed that our recently developed machine learning approach to ligand-based virtual screening provides substantially new information on the compound activity, different to predictions made by molecular docking in SBVS. Nevertheless, Keras/TensorFlow NN or LightGBM cannot be used solely but rather as a filtering method to decrease the number of compounds tested in SBVS for their precise binding modes and affinities. Machine learning also allows the screening of much larger compound libraries than those accessible to SBVS. Such novel algorithms offer better accuracy and better computational time efficiency than classical QSAR methods. The pre-filtering of large compound libraries before the SBVS step requires accurate but fast computational methods, which can be easily fulfilled by ML. In our opinion, the only limitation of ML remains in its dependency on the composition of the used training datasets [54]. This seems to be more crucial for LightGBM, while NNs encounter problems arising from the limited size of the assay-derived datasets.

Molecular dynamics, though more computationally expensive than LBVS or SBVS, is much more reliable than these methods in the validation of the ligand–receptor interactions, as it provides a dynamic image of the protein system in a time-dependent manner. Here, MD simulations allowed us to decide which of the previously selected compounds could serve as novel scaffolds for each of the studied receptors and which would require modifications to improve their binding affinity. As a result, we obtained four novel chemotypes for CCR2, five for CCR3, and five for CXCR3. These molecules can serve as a basis for further drug design involving ligand-binding assays and bioassays to confirm their ability to enhance the biological response of the receptor.

The combination of various computational methods allows us to overcome the limitations of each method. For example, SBVS does not use any prior knowledge about known active ligands of a given target and encounters problems arising from a simplification of

used force fields. Nanosecond MD simulations do not allow for the scanning of all possible receptor binding sites and all possible ligand conformations. Machine learning used in LBVS does not use any explicit information about the receptor and its interactions with ligands. On the other hand, SBVS allows performing an exhaustive search through all possible ligand conformations and ligand–receptor interactions to find the global free energy minimum. Nanosecond MD simulations allow unstable ligand–receptor interactions to be discarded and ligand-binding modes to be corrected using detailed all-atom force fields. ML can perform an extremely fast search for active ligands among huge datasets of compounds and thus significantly limits the number of ligands to be tested in SBVS. GPU-accelerated neural networks designed in Keras/TensorFlow or using GPUs for LighGBM offer the next level of processing cheminformatic data.

Among ML methods, NNs or deep learning NNs built on the Keras/TensorFlow platform have been used so far mainly in binary classification tasks in drug design [95]. Here, we showed that NNs can also be used in drug design as efficient multi-class classifiers when trained on datasets with discrete compound activity values [54]. To our knowledge, this is the first such application of Keras/TensorFlow NNs. Keras/TensorFlow NN multi-class classifier allows discarding not only inactive compounds from active ones but also low-active compounds from highly active compounds. This is especially important for drug design referring to large datasets, in which the number of low-active compounds is so high and they are so diverse that they would introduce nothing but noise when used as training sets for binary classification.

Another important application of NN models is the prediction of the receptor subtype selectivity of a compound. As already discussed, Keras/TensorFlow NNs can accurately distinguish ligand datasets matching different receptor subtypes. The only requirement is a sufficient dissimilarity between such ligand datasets, which was met in the case of CB1/CB2 datasets. Structural differences between ligands of different chemokine receptor subtypes were hardly sufficient, except for the CXCR3 dataset. Thus, based on the datasets currently available in ChEMBL, we could only develop the CB1/CB2 selective NN model and CCR/CXCR selective model of an accuracy sufficient for drug design purposes.

## 4. Materials and Methods

### 4.1. Ligand-Based Virtual Screening

In the LBVS step, we used a method described in detail elsewhere [54]. The method uses the Keras/TensorFlow library version 2.11.0 for constructing, training, and evaluating the currently used sequential model of neural network. NN was trained on the ChEMBL datasets of CCR2, CCR3, and CXCR3 compounds following the procedure described in [54]. Extended connectivity fingerprints with bond diameter 4 (ECFP4) [77] based on Morgan fingerprints were used to describe compound features with RDKit version 2022.9.2 [76]. To emphasize, Keras/TensorFlow NN was used here not as a typical binary NN classifier but as a multi-class classifier that is able to distinguish not only active and inactive compounds but also low- and medium-active from highly active compounds. This was carried out by labeling the datasets with seven activity categories based on logarithmic pChEMBL values: 1 (below 4), 2 (4–5), 3 (5–6), 4 (6–7), 5 (7–8), 6 (8–9), and 7 (above 9). Categories 5, 6, and 7 referred to highly active compounds, while 3 and 4 to medium-active, and 1 and 2 to inactive or low-active compounds. An NN was built and trained using the categorical cross-entropy loss function and stochastic gradient descent to minimize the loss function (Adaptive Moment Estimation Optimizer). Due to the multi-class application of NN, Softmax conversion leading to a probability distribution was used as the activation function for the last layer, instead of the sigmoid function that is used typically for binary classification. The Rectified Linear Unit activation function (ReLU) was used for hidden layers for quick convergence. One thousand epochs were used to ensure the sufficient minimization of the model, although a much smaller number could be also used, e.g., 200, as cross-entropy loss and accuracy stabilized after 200 epochs (see Supplementary S1, Figures S3 and S4).



In principle, in the case of neural networks fitted to solve big data problems, increasing the training set from 40% to 80% (see Supplementary S1 Figures S5 and S6) should improve both the model accuracy and the model training efficiency. This improvement was indeed visible in the case of CCR2 and CCR3 (Supplementary S1 Figure S5). Nevertheless, the bootstrapping analysis should be performed to undoubtedly confirm this.

For the receptor subtype selectivity tests, the following curated datasets were used for training (80% randomly selected compounds from the ChEMBL-retrieved datasets): 1995 (CCR2), 603 (CCR3), 994 (CXCR3), 4509 (CB1), and 4087 (CB2), and for validation, the remaining compounds were used. For cannabinoid receptors, two additional training sets [63,83] from <https://db-gpcr.chem.uw.edu.pl> (accessed on 20 August 2023) were used, consisting of 1566 and 2093 compounds (CB1 and CB2, respectively). A further 35 CB2-selective compounds (from <https://db-gpcr.chem.uw.edu.pl> (accessed on 20 August 2023)) were used as one of the validation sets included in Table 4. To generate the results presented in Table 4, the number of epochs was set to 100, and the average loss and accuracy were computed for 100 independent training runs of NN.

Python scripts with imported modules from the latest versions of RDKit (v2022.9.2), scikit-learn (v1.0.2), Keras (v2.11.0), and TensorFlow (v2.11.0) used for data processing.

#### 4.2. Preparation of CCR2, CCR3, and CXCR3 Structures

The 6GPX structure [96] of the inactive-state CCR2 receptor was downloaded from the Protein Data Bank (PDB) [97], and a model of CCR3 was generated using the Robetta webserver (accessed on 2 February 2022) [98]. Both the structure and the model were preprocessed using Maestro v2023-1 [99] and evaluated as described in a previous study [54].

The amino acid sequence of CXCR3-A was obtained from UniProt [100] (see Supplementary S1 Figure S2). Protein BLAST (<https://blast.ncbi.nlm.nih.gov> accessed on 9 May 2023) was used to perform a search against PDB to find solved GPCR structures with a high sequence similarity to CXCR3-A [101]. Of these, the inactive-state 5LWE [102] and 6MEO [103] PDB entries, with a 33.44% and 36.49% sequence similarity to the target, respectively, were selected as templates for homology modeling. Modeller v10.4 [104] was used to generate 5000 models of CXCR3-A. The lowest energy models were analyzed in PyMol v2.4.0 [105] and validated by comparing their structures to those of other GPCRs with well-described molecular switches, including PDB entries 7O7F [88] and 6WWZ [106].

#### 4.3. Structure-Based Virtual Screening

SBVS assisted by machine learning was performed using Glide v2021-4 for CCR2 and CCR3, as described previously [54]. Here, to confirm their binding modes and to test if two molecular docking programs based on completely different force fields (OPLS and Amber for Glide and AutoDock, respectively) provide similar results, we used AutoDock Vina v1.2.3 [43,44,107]. Although the compound ranking proposed by AutoDock Vina was very similar to the one obtained previously by Glide, a few compounds were discarded due to significant differences in their binding modes provided with AutoDock Vina in comparison to Glide results. The remaining CCR2 and CCR3 compounds were subjected to validation with MD simulations.

The validated model of CXCR3 was used for structure-based virtual screening (SBVS) with AutoDock Vina, using the Enamine Hit Locator Library (HLL) (accessed on 17 November 2022) [108], consisting of over 460,000 compounds. The position of the grid box for AutoDock Vina was determined based on the positions of the ligands in the corresponding template structures, and its size was  $31.19 \times 29.17 \times 38.56$ . Ten binding modes were generated for each ligand, and the energy cut-off for selecting ligand poses was equal to  $-10.5$ . The results were analyzed using the vs-analysis.py script [109], and 31 compounds with the best binding affinities were selected for further investigation.

A set of known CXCR3 inhibitors—the IC50 subset—was downloaded from the ChEMBL (accessed on 15 May 2023). After the data were curated and compounds with no

specified activity values (pChEMBL values) were removed, the CXCR3 dataset was used as a training set for a neural network implemented in Keras/TensorFlow according to a procedure described elsewhere [54]. The algorithm was then used to predict the activity values of the molecules in the HLL compound library. The compounds with the highest predicted activity values (above 9) were mapped against those obtained via SBVS, and as a result, nine potential CXCR3 actives were obtained. Out of these, five of the best-assessed compounds were selected for further MD simulations.

#### 4.4. Molecular Dynamics Simulations

For the selected compounds, their complexes with receptors for the MD simulations were prepared using CHARMM-GUI's v3.7 [110–112] Membrane Builder [113–116]. Information about the disulphide bonds in the receptor structures was provided based on known structures of chemokine receptors in the PDB, and the ligand parameterization was performed using CGenFF [117] and 3D structural files generated by Maestro. The ligand–receptor complexes were inserted into a lipid bilayer consisting of a 3:1 ratio of POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) to cholesterol. The periodic rectangular water box (TIP3P) was fitted to the complex, and each simulation system was neutralized by adding Na<sup>+</sup> and Cl<sup>−</sup> ions at a concentration of 0.15 M. The number of atoms in each simulation system was equal to between 135,000 and 148,000 atoms, depending on the system. The Charmm36 force field was used in each simulation.

The equilibration step included 10,000 steps of the steepest descent minimization, then 25,000 steps of the conjugated gradients minimization. The equilibration simulation was performed in NVT using the Langevin dynamics (303.15 K). The time integration step in the equilibration and production runs was set to 2 fs. The production run in NPT was performed using the Langevin piston Nose–Hoover method (1 bar, 303.15 K) and lasted for 100 ns for each system. The GPU-accelerated version of NAMD v3.0 [118] was used for all MD simulations. The obtained trajectories were analyzed using VMD v1.9.3 [119].

#### 4.5. Suggested Structural Modification of Active Compounds

Chemical modifications of functional groups of the proposed active compounds for each receptor were suggested in order to improve their binding affinities. Maestro was used to analyze the interactions between the modified ligands and the receptor in the final frame of the MD simulation and to suggest possible changes. Modified structures of proposed compounds were minimized in Maestro (OPLS4 force field), in order to prevent clashes.

#### 4.6. Structural Comparison of CXCR3 Antagonists

Compounds described by Meyer et al. [32] were reproduced in Maestro in order to perform a search for similar structures in the HLL compound library. The Fingerprint Similarity tool was used with the Tanimoto similarity metric. The docking scores and predicted activities as well as their ranks provided by Keras/TensorFlow NN were extracted for compounds with the highest Tanimoto coefficients.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms241915009/s1>, References [63,83,85,90,93] are cited in the supplementary materials.

**Author Contributions:** Conceptualization, P.D. and D.L.; methodology, P.D., K.J. and D.L.; software, P.D., K.J. and D.L.; validation, P.D., K.J. and D.L.; formal analysis, P.D., K.J. and D.L.; investigation, P.D., K.J., A.A. and D.L.; resources, D.L.; data curation, P.D. and D.L.; writing—original draft preparation, P.D., K.J., A.A. and D.L.; writing—review and editing, P.D., A.A. and D.L.; visualization, P.D. and D.L.; supervision, D.L.; project administration, D.L.; funding acquisition, D.L. Detailed contributions: Introduction—P.D. and D.L., with contributions from A.A. (Table 1) and K.J.; Results—P.D. (Figures 1–4, Table 2, Sections 2.1–2.4), D.L. (Tables 2–4, Sections 2.2, 2.4 and 2.5); Discussion and Conclusions—P.D. and D.L.; Materials and Methods—P.D. (Sections 4.1–4.6), D.L. (Section 4.1) with contributions from K.J. (performing AutoDock Vina computations for CCR2, CCR3, and CXCR3, and ML computations for CXCR3); Supplementary S1—P.D. (Figures S1 and S2,

Tables S1–S6), D.L. (Figures S3–S7, Tables S7–S10); Graphical Abstract: D.L. and P.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Science Centre in Poland, grant number 2020/39/B/NZ2/00584. Computational resources were provided by Poland’s high-performance Infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH), grant number PLG/2023/016255.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Datasets used for NN training are publicly available on <https://www.ebi.ac.uk/chembl/> (accessed on 24 May 2023) and <https://db-gpcr.chem.uw.edu.pl> (accessed on 20 August 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

GPCR	G protein-coupled receptor
CCL	C-C motif chemokine ligand
CXCL	C-X-C motif chemokine ligand
CCL	C-C motif chemokine ligand
CCR	C-C motif conventional chemokine receptor
CXCR	C-X-C motif chemokine receptor
ACKR	Atypical chemokine receptor
MSMP	Prostate-associated microseminoprotein
SBVS	Structure-based virtual screening
LBVS	Ligand-based virtual screening
MD	Molecular dynamics
HLL	Hit Locator Library
NN	Neural network
ML	Machine learning
Cryo-EM	Cryogenic electron microscopy

## References

- Borsig, L.; Wolf, M.J.; Roblek, M.; Lorentzen, A.; Heikenwalder, M. Inflammatory Chemokines and Metastasis—Tracing the Accessory. *Oncogene* **2014**, *33*, 3217–3224. [[CrossRef](#)] [[PubMed](#)]
- Zlotnik, A.; Yoshie, O. The Chemokine Superfamily Revisited. *Immunity* **2012**, *36*, 705–716. [[CrossRef](#)] [[PubMed](#)]
- Jaeger, K.; Bruenle, S.; Weinert, T.; Guba, W.; Muehle, J.; Miyazaki, T.; Weber, M.; Furrer, A.; Haenggi, N.; Tetaz, T.; et al. Structural Basis for Allosteric Ligand Recognition in the Human CC Chemokine Receptor 7. *Cell* **2019**, *178*, 1222–1230.e10. [[CrossRef](#)] [[PubMed](#)]
- Hughes, C.E.; Nibbs, R.J.B. A Guide to Chemokines and Their Receptors. *FEBS J.* **2018**, *285*, 2944–2971. [[CrossRef](#)] [[PubMed](#)]
- Sokol, C.L.; Luster, A.D. The Chemokine System in Innate Immunity. *Cold Spring Harb. Perspect. Biol.* **2015**, *7*, a016303. [[CrossRef](#)] [[PubMed](#)]
- Vilgelm, A.E.; Richmond, A. Chemokines Modulate Immune Surveillance in Tumorigenesis, Metastasis, and Response to Immunotherapy. *Front. Immunol.* **2019**, *10*, 333. [[CrossRef](#)] [[PubMed](#)]
- Balkwill, F. Cancer and the Chemokine Network. *Nat. Rev. Cancer* **2004**, *4*, 540–550. [[CrossRef](#)]
- Mukaida, N.; Baba, T. Chemokines in Tumor Development and Progression. *Exp. Cell Res.* **2012**, *318*, 95–102. [[CrossRef](#)]
- Ono, S.J.; Nakamura, T.; Miyazaki, D.; Ohbayashi, M.; Dawson, M.; Toda, M. Chemokines: Roles in Leukocyte Development, Trafficking, and Effector Function. *J. Allergy Clin. Immunol.* **2003**, *111*, 1185–1199. [[CrossRef](#)]
- Fernandez, E.J.; Lolis, E. Structure, Function, and Inhibition of Chemokines. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 469–499. [[CrossRef](#)]
- Kufareva, I.; Salanga, C.L.; Handel, T.M. Chemokine and Chemokine Receptor Structure and Interactions: Implications for Therapeutic Strategies. *Immunol. Cell Biol.* **2015**, *93*, 372–383. [[CrossRef](#)] [[PubMed](#)]
- Bachelier, F.; Graham, G.J.; Locati, M.; Mantovani, A.; Murphy, P.M.; Nibbs, R.; Rot, A.; Sozzani, S.; Thelen, M. New Nomenclature for Atypical Chemokine Receptors. *Nat. Immunol.* **2014**, *15*, 207–208. [[CrossRef](#)] [[PubMed](#)]
- García-Cuesta, E.M.; Santiago, C.A.; Vallejo-Díaz, J.; Juarranz, Y.; Rodríguez-Frade, J.M.; Mellado, M. The Role of the CXCL12/CXCR4/ACKR3 Axis in Autoimmune Diseases. *Front. Endocrinol.* **2019**, *10*, 585. [[CrossRef](#)] [[PubMed](#)]
- Zlotnik, A.; Burkhardt, A.M.; Homey, B. Homeostatic Chemokine Receptors and Organ-Specific Metastasis. *Nat. Rev. Immunol.* **2011**, *11*, 597–606. [[CrossRef](#)] [[PubMed](#)]

15. Chen, W.; Fang, Y.; Wang, H.; Tan, X.; Zhu, X.; Xu, Z.; Jiang, H.; Wu, X.; Hong, W.; Wang, X.; et al. Role of Chemokine Receptor 2 in Rheumatoid Arthritis: A Research Update. *Int. Immunopharmacol.* **2023**, *116*, 109755. [CrossRef] [PubMed]
16. Chu, H.X.; Arumugam, T.V.; Gelderblom, M.; Magnus, T.; Drummond, G.R.; Sobey, C.G. Role of CCR2 in Inflammatory Conditions of the Central Nervous System. *J. Cereb. Blood Flow Metab.* **2014**, *34*, 1425–1429. [CrossRef] [PubMed]
17. She, S.; Wu, X.; Zheng, D.; Pei, X.; Ma, J.; Sun, Y.; Zhou, J.; Nong, L.; Guo, C.; Lv, P.; et al. PSMP/MSMP Promotes Hepatic Fibrosis through CCR2 and Represents a Novel Therapeutic Target. *J. Hepatol.* **2020**, *72*, 506–518. [CrossRef] [PubMed]
18. Elemam, N.M.; Hannawi, S.; Maghazachi, A.A. Role of Chemokines and Chemokine Receptors in Rheumatoid Arthritis. *ImmunoTargets Ther.* **2020**, *9*, 43–56. [CrossRef]
19. She, S.; Ren, L.; Chen, P.; Wang, M.; Chen, D.; Wang, Y.; Chen, H. Functional Roles of Chemokine Receptor CCR2 and Its Ligands in Liver Disease. *Front. Immunol.* **2022**, *13*, 812431. [CrossRef]
20. Proudfoot, A.E.I. Chemokine Receptors: Multifaceted Therapeutic Targets. *Nat. Rev. Immunol.* **2002**, *2*, 106–115. [CrossRef]
21. Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J.P. ChEMBL Web Services: Streamlining Access to Drug Discovery Data and Utilities. *Nucleic Acids Res.* **2015**, *43*, W612–W620. [CrossRef] [PubMed]
22. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [CrossRef] [PubMed]
23. Shao, Z.; Tan, Y.; Shen, Q.; Hou, L.; Yao, B.; Qin, J.; Xu, P.; Mao, C.; Chen, L.-N.; Zhang, H.; et al. Molecular Insights into Ligand Recognition and Activation of Chemokine Receptors CCR2 and CCR3. *Cell Discov.* **2022**, *8*, 44. [CrossRef] [PubMed]
24. Erin, E.M.; Williams, T.J.; Barnes, P.J.; Hansel, T.T. Eotaxin Receptor (CCR3) Antagonism in Asthma and Allergic Disease. *Curr. Drug Targets-Inflamm. Allergy* **2002**, *1*, 201–214. [CrossRef] [PubMed]
25. CCL3 C-C Motif Chemokine Ligand 3 [Homo Sapiens (Human)]. Gene—NCBI. Available online: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=6348> (accessed on 9 November 2022).
26. Korbecki, J.; Kojder, K.; Simińska, D.; Bohatyrewicz, R.; Gutowska, I.; Chlubek, D.; Baranowska-Bosiacka, I. CC Chemokines in a Tumor: A Review of Pro-Cancer and Anti-Cancer Properties of the Ligands of Receptors CCR1, CCR2, CCR3, and CCR4. *Int. J. Mol. Sci.* **2020**, *21*, 8412. [CrossRef] [PubMed]
27. Amerio, P.; Frezzolini, A.; Feliciani, C.; Verdolini, R.; Teofoli, P.; Pità, O.; Puddu, P. Eotaxins and CCR3 Receptor in Inflammatory and Allergic Skin Diseases: Therapeutical Implications. *Curr. Drug Targets-Inflamm. Allergy* **2003**, *2*, 81–94. [CrossRef] [PubMed]
28. Satarkar, D.; Patra, C. Evolution, Expression and Functional Analysis of CXCR3 in Neuronal and Cardiovascular Diseases: A Narrative Review. *Front. Cell Dev. Biol.* **2022**, *10*, 882017. [CrossRef] [PubMed]
29. Ehlert, J.E.; Addison, C.A.; Burdick, M.D.; Kunkel, S.L.; Strieter, R.M. Identification and Partial Characterization of a Variant of Human CXCR3 Generated by Posttranscriptional Exon Skipping. *J. Immunol.* **2004**, *173*, 6234–6240. [CrossRef]
30. Lacotte, S.; Brun, S.; Muller, S.; Dumortier, H. CXCR3, Inflammation, and Autoimmune Diseases. *Ann. N. Y. Acad. Sci.* **2009**, *1173*, 310–317. [CrossRef]
31. Gangur, V.; Birmingham, N.; Thanavorakul, S.; Joseph, S. CCR3 and CXCR3 as Drug Targets for Allergy: Principles and Potential. *Curr. Drug Targets-Inflamm. Allergy* **2003**, *2*, 53–62. [CrossRef]
32. Meyer, E.A.; Äänismaa, P.; Ertel, E.A.; Hühn, E.; Strasser, D.S.; Rey, M.; Murphy, M.J.; Martinic, M.M.; Pouzol, L.; Froidevaux, S.; et al. Discovery of Clinical Candidate ACT-777991, a Potent CXCR3 Antagonist for Antigen-Driven and Inflammatory Pathologies. *J. Med. Chem.* **2023**, *66*, 4179–4196. [CrossRef] [PubMed]
33. Zheng, K.; Smith, J.S.; Eiger, D.S.; Warman, A.; Choi, I.; Honeycutt, C.C.; Boldizar, N.; Gundry, J.N.; Pack, T.F.; Inoue, A.; et al. Biased Agonists of the Chemokine Receptor CXCR3 Differentially Signal through G $\alpha$ <sub>i</sub>; $\beta$ -Arrestin Complexes. *Sci. Signal.* **2022**, *15*, eabg5203. [CrossRef] [PubMed]
34. Bachelier, F.; Ben-Baruch, A.; Burkhardt, A.M.; Combadiere, C.; Farber, J.M.; Graham, G.J.; Horuk, R.; Sparre-Ulrich, A.H.; Locati, M.; Luster, A.D.; et al. International Union of Basic and Clinical Pharmacology. LXXXIX. Update on the Extended Family of Chemokine Receptors and Introducing a New Nomenclature for Atypical Chemokine Receptors. *Pharmacol. Rev.* **2014**, *66*, 1–79. [CrossRef] [PubMed]
35. Arimont, M.; Hoffmann, C.; De Graaf, C.; Leurs, R. Chemokine Receptor Crystal Structures: What Can Be Learned from Them? *Mol. Pharmacol.* **2019**, *96*, 765–777. [CrossRef] [PubMed]
36. Zhao, S.; Wu, B.; Stevens, R.C. Advancing Chemokine GPCR Structure Based Drug Discovery. *Structure* **2019**, *27*, 405–408. [CrossRef] [PubMed]
37. Solari, R.; Pease, J.E.; Begg, M. Chemokine Receptors as Therapeutic Targets: Why Aren't There More Drugs? *Eur. J. Pharmacol.* **2015**, *746*, 363–367. [CrossRef] [PubMed]
38. Russo, E.; Santoni, A.; Bernardini, G. Tumor Inhibition or Tumor Promotion? The Duplicity of CXCR3 in Cancer. *J. Leukoc. Biol.* **2020**, *108*, 673–685. [CrossRef]
39. Hanefeld, M.; Schell, E.; Gouni-Berthold, I.; Melichar, M.; Vesela, I.; Johnson, D.; Miao, S.; Sullivan, T.J.; Jaen, J.C.; Schall, T.J.; et al. Orally-Administered Chemokine Receptor CCR2 Antagonist CCX140-B in Type 2 Diabetes: A Pilot Double-Blind, Randomized Clinical Trial. *J. Diabetes Metab.* **2012**, *3*, 225. [CrossRef]
40. Wang, X.; Zhang, Y.; Wang, S.; Ni, H.; Zhao, P.; Chen, G.; Xu, B.; Yuan, L. The Role of CXCR3 and Its Ligands in Cancer. *Front. Oncol.* **2022**, *12*, 1022688. [CrossRef]

41. Li, Q.; Shah, S. Structure-Based Virtual Screening. In *Protein Bioinformatics: From Protein Modifications and Networks to Proteomics*; Wu, C.H., Arighi, C.N., Ross, K.E., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2017; pp. 111–124, ISBN 978-1-4939-6783-4.
42. *Schrödinger Release 2021-4: Glide*; Schrödinger, LLC: New York, NY, USA, 2021. Available online: <https://www.schrodinger.com/products/glide> (accessed on 8 February 2022).
43. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
44. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [[CrossRef](#)] [[PubMed](#)]
45. Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288. [[CrossRef](#)] [[PubMed](#)]
46. Molecular Operating Environment (MOE). Chemical Computing Group ULC, Montreal, Canada, 2023. Available online: <https://www.chemcomp.com/Products.htm> (accessed on 8 February 2022).
47. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking 1 Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267*, 727–748. [[CrossRef](#)] [[PubMed](#)]
48. Ding, J.; Tang, S.; Mei, Z.; Wang, L.; Huang, Q.; Hu, H.; Ling, M.; Wu, J. Vina-GPU 2.0: Further Accelerating AutoDock Vina and Its Derivatives with Graphics Processing Units. *J. Chem. Inf. Model.* **2023**, *63*, 1982–1998. [[CrossRef](#)]
49. Li, H.; Sze, K.; Lu, G.; Ballester, P.J. Machine-learning Scoring Functions for Structure-based Virtual Screening. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1478. [[CrossRef](#)]
50. Shen, C.; Ding, J.; Wang, Z.; Cao, D.; Ding, X.; Hou, T. From Machine Learning to Deep Learning: Advances in Scoring Functions for Protein–Ligand Docking. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1429. [[CrossRef](#)]
51. Pereira, J.C.; Caffarena, E.R.; Dos Santos, C.N. Boosting Docking-Based Virtual Screening with Deep Learning. *J. Chem. Inf. Model.* **2016**, *56*, 2495–2506. [[CrossRef](#)] [[PubMed](#)]
52. Imrie, F.; Bradley, A.R.; van der Schaar, M.; Deane, C.M. Protein Family-Specific Models Using Deep Neural Networks and Transfer Learning Improve Virtual Screening and Highlight the Need for More Data. *J. Chem. Inf. Model.* **2018**, *58*, 2319–2330. [[CrossRef](#)]
53. Gonczarek, A.; Tomczak, J.M.; Zareba, S.; Kaczmar, J.; Dąbrowski, P.; Walczak, M.J. Interaction Prediction in Structure-Based Virtual Screening Using Deep Learning. *Comput. Biol. Med.* **2018**, *100*, 253–258. [[CrossRef](#)]
54. Dragan, P.; Merski, M.; Wiśniewski, S.; Sanmukh, S.G.; Latek, D. Chemokine Receptors—Structure-Based Virtual Screening Assisted by Machine Learning. *Pharmaceutics* **2023**, *15*, 516. [[CrossRef](#)]
55. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2020**, *23*, 18. [[CrossRef](#)] [[PubMed](#)]
56. Lee, S.-I.; Celik, S.; Logsdon, B.A.; Lundberg, S.M.; Martins, T.J.; Oehler, V.G.; Estey, E.H.; Miller, C.P.; Chien, S.; Dai, J.; et al. A Machine Learning Approach to Integrate Big Data for Precision Medicine in Acute Myeloid Leukemia. *Nat. Commun.* **2018**, *9*, 42. [[CrossRef](#)] [[PubMed](#)]
57. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874. [[CrossRef](#)]
58. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning Important Features through Propagating Activation Differences. *arXiv* **2017**, arXiv:1704.02685. [[CrossRef](#)]
59. Chattopadhyay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018.
60. Covert, I.; Lundberg, S.; Lee, S.-I. Understanding Global Feature Contributions with Additive Importance Measures. *arXiv* **2020**, arXiv:2004.00668. [[CrossRef](#)]
61. Covert, I.C.; Lundberg, S.; Lee, S.-I. Explaining by Removing: A Unified Framework for Model Explanation. *J. Mach. Learn. Res.* **2021**, *22*, 1–90.
62. Vázquez, J.; López, M.; Gibert, E.; Herrero, E.; Luque, F.J. Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. *Molecules* **2020**, *25*, 4723. [[CrossRef](#)] [[PubMed](#)]
63. Mizera, M.; Latek, D. Ligand-Receptor Interactions and Machine Learning in GCGR and GLP-1R Drug Discovery. *Int. J. Mol. Sci.* **2021**, *22*, 4060. [[CrossRef](#)]
64. Ekins, S.; Puhl, A.C.; Zorn, K.M.; Lane, T.R.; Russo, D.P.; Klein, J.J.; Hickey, A.J.; Clark, A.M. Exploiting Machine Learning for End-to-End Drug Discovery and Development. *Nat. Mater.* **2019**, *18*, 435–441. [[CrossRef](#)]
65. Ali, M.; Aittokallio, T. Machine Learning and Feature Selection for Drug Response Prediction in Precision Oncology Applications. *Biophys. Rev.* **2019**, *11*, 31–39. [[CrossRef](#)]
66. Czarnecki, W.M.; Podlowska, S.; Bojarski, A.J. Robust Optimization of SVM Hyperparameters in the Classification of Bioactive Compounds. *J. Cheminform.* **2015**, *7*, 38. [[CrossRef](#)] [[PubMed](#)]
67. Rodríguez-Pérez, R.; Bajorath, J. Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *J. Med. Chem.* **2020**, *63*, 8761–8777. [[CrossRef](#)] [[PubMed](#)]
68. Siemers, F.M.; Feldmann, C.; Bajorath, J. Minimal Data Requirements for Accurate Compound Activity Prediction Using Machine Learning Methods of Different Complexity. *Cell Rep. Phys. Sci.* **2022**, *3*, 101113. [[CrossRef](#)]

69. Mahmud, M.; Kaiser, M.S.; Hussain, A.; Vassanelli, S. Applications of Deep Learning and Reinforcement Learning to Biological Data. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2063–2079. [CrossRef]
70. Haghighat, E.; Juanes, R. SciANN: A Keras/TensorFlow Wrapper for Scientific Computations and Physics-Informed Deep Learning Using Artificial Neural Networks. *Comput. Methods Appl. Mech. Eng.* **2021**, *373*, 113552. [CrossRef]
71. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]
72. Bai, Q.; Liu, S.; Tian, Y.; Xu, T.; Banegas-Luna, A.J.; Pérez-Sánchez, H.; Huang, J.; Liu, H.; Yao, X. Application Advances of Deep Learning Methods for de Novo Drug Design and Molecular Dynamics Simulation. *WIREs Comput. Mol. Sci.* **2022**, *12*, e1581. [CrossRef]
73. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv* **2016**, arXiv:1603.04467. [CrossRef]
74. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, arXiv:1912.01703. [CrossRef]
75. Chollet, F. Keras: The Python Deep Learning Library. *Astrophys. Source Code Libr.* **2018**, ascl:1806.022.
76. Landrum, G.; Tosco, P.; Kelley, B.; Ric, C.; Cosgrove, D.; Sriniker; Gedeck; Vianello, R.; NadineSchneider; Kawashima, E.; et al. Rdkit/Rdkit: 2023\_03\_2 (Q1 2023) Release 2023. Available online: [https://github.com/rdkit/rdkit/releases/tag/Release\\_2023\\_03\\_2](https://github.com/rdkit/rdkit/releases/tag/Release_2023_03_2) (accessed on 8 February 2022).
77. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef] [PubMed]
78. Pang, B.; Nijkamp, E.; Wu, Y.N. Deep Learning with TensorFlow: A Review. *J. Educ. Behav. Stat.* **2020**, *45*, 227–248. [CrossRef]
79. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. *arXiv* **2019**, arXiv:1907.10902. [CrossRef]
80. Tokui, S.; Okuta, R.; Akiba, T.; Niitani, Y.; Ogawa, T.; Saito, S.; Suzuki, S.; Uenishi, K.; Vogel, B.; Vincent, H.Y. Chainer: A Deep Learning Framework for Accelerating the Research Cycle. *arXiv* **2019**, arXiv:1908.00213. [CrossRef]
81. Remington, J.M.; McKay, K.T.; Beckage, N.B.; Ferrell, J.B.; Schneckel, S.T.; Li, J. GPCRLigNet: Rapid Screening for GPCR Active Ligands Using Machine Learning. *J. Comput. Aided Mol. Des.* **2023**, *37*, 147–156. [CrossRef] [PubMed]
82. Ge, H.; Wang, Y.; Li, C.; Chen, N.; Xie, Y.; Xu, M.; He, Y.; Gu, X.; Wu, R.; Gu, Q.; et al. Molecular Dynamics-Based Virtual Screening: Accelerating the Drug Discovery Process by High-Performance Computing. *J. Chem. Inf. Model.* **2013**, *53*, 2757–2764. [CrossRef] [PubMed]
83. Mizera, M.; Latek, D.; Cielecka-Piontek, J. Virtual Screening of C. Sativa Constituents for the Identification of Selective Ligands for Cannabinoid Receptor 2. *Int. J. Mol. Sci.* **2020**, *21*, 5308. [CrossRef] [PubMed]
84. Liu, K.; Shen, L.; Wu, M.; Liu, Z.-J.; Hua, T. Structural Insights into the Activation of Chemokine Receptor CXCR2. *FEBS J.* **2022**, *289*, 386–393. [CrossRef]
85. *Schrödinger Release 2022-3: Maestro*; Schrödinger, LLC: New York, NY, USA, 2021. Available online: <https://www.schrodinger.com/products/maestro> (accessed on 15 March 2023).
86. Trzaskowski, B.; Latek, D.; Yuan, S.; Ghoshdastider, U.; Debinski, A.; Filipek, S. Action of Molecular Switches in GPCRs-Theoretical and Experimental Studies. *Curr. Med. Chem.* **2012**, *19*, 1090–1109. [CrossRef]
87. Filipek, S. Molecular Switches in GPCRs. *Curr. Opin. Struct. Biol.* **2019**, *55*, 114–120. [CrossRef]
88. Isaikina, P.; Tsai, C.J.; Dietz, N.; Pamula, F.; Grahl, A.; Goldie, K.N.; Guixà-González, R.; Branco, C.; Paolini-Bertrand, M.; Calo, N.; et al. Structural basis of the activation of the CC chemokine receptor 5 by a chemokine agonist. *Sci. Adv.* **2021**, *16*, eabg8685. [CrossRef] [PubMed]
89. Ishimoto, N.; Park, J.H.; Kawakami, K.; Tajiri, M.; Mizutani, K.; Akashi, S.; Tame, J.R.H.; Inoue, A.; Park, S.Y. Structural basis of CXC chemokine receptor 1 ligand binding and activation. *Nat. Commun.* **2023**, *14*, 4107. [CrossRef] [PubMed]
90. Ballesteros, J.A.; Weinstein, H. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Methods Neurosci.* **1995**, *25*, 366–428. [CrossRef]
91. Langer, I.; Latek, D. Drug Repositioning For Allosteric Modulation of VIP and PACAP Receptors. *Front. Endocrinol.* **2021**, *12*, 711906. [CrossRef] [PubMed]
92. Latek, D.; Langer, I.; Krzysko, K.; Charzewski, L. A Molecular Dynamics Study of Vasoactive Intestinal Peptide Receptor 1 and the Basis of Its Therapeutic Antagonism. *Int. J. Mol. Sci.* **2019**, *20*, 4348. [CrossRef] [PubMed]
93. Madeira, F.; Pearce, M.; Tivey, A.R.N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50*, W276–W279. [CrossRef]
94. Latek, D.; Rutkowska, E.; Niewieczeral, S.; Cielecka-Piontek, J. Drug-Induced Diabetes Type 2: In Silico Study Involving Class B GPCRs. *PLoS ONE* **2019**, *14*, e0208892. [CrossRef] [PubMed]
95. Sid, K.; Zertal, S.; Mezioud, C. DeepD\_DrugC: Deep and Distributed Workflow to Predict Drug- Candidates. In Proceedings of the 2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS), Oum El Bouaghi, Algeria, 12–13 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–8.
96. Apel, A.-K.; Cheng, R.K.Y.; Tautermann, C.S.; Brauchle, M.; Huang, C.-Y.; Pautsch, A.; Hennig, M.; Nar, H.; Schnapp, G. Crystal Structure of CC Chemokine Receptor 2A in Complex with an Orthosteric Antagonist Provides Insights for the Design of Selective Antagonists. *Structure* **2019**, *27*, 427–438.e5. [CrossRef]

97. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
98. Robetta. Available online: <https://robeta.bakerlab.org/> (accessed on 8 February 2022).
99. *Schrödinger Release 2023-1: Maestro*; Schrödinger, LLC: New York, NY, USA, 2021. Available online: <https://www.schrodinger.com/products/maestro> (accessed on 15 March 2023).
100. The UniProt Consortium UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531. [[CrossRef](#)]
101. Boratyn, G.M.; Camacho, C.; Cooper, P.S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T.L.; Matten, W.T.; McGinnis, S.D.; Merezuk, Y.; et al. BLAST: A More Efficient Report with Usability Improvements. *Nucleic Acids Res.* **2013**, *41*, W29–W33. [[CrossRef](#)]
102. Oswald, C.; Rappas, M.; Kean, J.; Doré, A.S.; Errey, J.C.; Bennett, K.; Deflorian, F.; Christopher, J.A.; Jazayeri, A.; Mason, J.S.; et al. Intracellular Allosteric Antagonism of the CCR9 Receptor. *Nature* **2016**, *540*, 462–465. [[CrossRef](#)]
103. Shaik, M.M.; Peng, H.; Lu, J.; Rits-Volloch, S.; Xu, C.; Liao, M.; Chen, B. Structural Basis of Coreceptor Recognition by HIV-1 Envelope Spike. *Nature* **2019**, *565*, 318–323. [[CrossRef](#)]
104. Šali, A.; Blundell, T.L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [[CrossRef](#)]
105. The PyMOL Molecular Graphics System, Schrödinger. Available online: <https://pymol.org/2/> (accessed on 8 February 2022).
106. Wasilko, D.J.; Johnson, Z.L.; Ammirati, M.; Che, Y.; Griffor, M.C.; Han, S.; Wu, H. Structural Basis for Chemokine Receptor CCR6 Activation by the Endogenous Protein Ligand CCL20. *Nat. Commun.* **2020**, *11*, 3031. [[CrossRef](#)]
107. Nguyen, N.T.; Nguyen, T.H.; Pham, T.N.H.; Huy, N.T.; Bay, M.V.; Pham, M.Q.; Nam, P.C.; Vu, V.V.; Ngo, S.T. Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *J. Chem. Inf. Model.* **2020**, *60*, 204–211. [[CrossRef](#)]
108. Enamine. Available online: <https://enamine.net/> (accessed on 21 November 2022).
109. Muniba, F. Vs\_Analysis.Py: A Python Script to Analyze Virtual Screening Results of Autodock Vina. 2021, 8, pp. 12–16. Available online: <https://bioinformaticsreview.com/20210509/vs-analysis-a-python-script-to-analyze-virtual-screening-results-of-autodock-vina/> (accessed on 8 February 2022).
110. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865. [[CrossRef](#)]
111. Brooks, B.R.; Brooks III, C.L.; Mackerell, A.D., Jr.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [[CrossRef](#)]
112. Lee, J.; Cheng, X.; Swails, J.M.; Yeom, M.S.; Eastman, P.K.; Lemkul, J.A.; Wei, S.; Buckner, J.; Jeong, J.C.; Qi, Y.; et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413. [[CrossRef](#)]
113. Wu, E.L.; Cheng, X.; Jo, S.; Rui, H.; Song, K.C.; Dávila-Contreras, E.M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R.M.; et al. CHARMM-GUI Membrane Builder toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014**, *35*, 1997–2004. [[CrossRef](#)]
114. Jo, S.; Lim, J.B.; Klauda, J.B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009**, *97*, 50–58. [[CrossRef](#)]
115. Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE* **2007**, *2*, e880. [[CrossRef](#)]
116. Lee, J.; Patel, D.S.; Stähle, J.; Park, S.-J.; Kern, N.R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M.A.; Holst, O.; et al. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **2019**, *15*, 775–786. [[CrossRef](#)]
117. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31*, 671–690. [[CrossRef](#)]
118. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hémin, J.; Jiang, W.; et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130. [[CrossRef](#)]
119. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.