Development of novel algorithms for data-driven prediction of interactions between proteins and small molecules

Maria Kadukova

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

préparée dans le cadre d'une cotutelle entre l'*Université Grenoble Alpes* et l'*Institut de physique et de technologie de Moscou*

Spécialité : **Mathematiques et informatique**

Arrêté ministériel : le 6 janvier 2005 – 25 mai 2016

Présentée par

Maria KADUKOVA

Thèse dirigée par **Stephane REDON** et **Sergei GRUDININ**
codirigée par **Vladimir CHUPIN**

préparée au sein du **Laboratoire Jean Kuntzmann et du Laboratoire de chimie et physique des lipides**

à **l'École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

Développement de nouveaux modèles, basés sur les données, pour prédire les interactions entre protéines et petites molécules

Thèse soutenue publiquement le **23 mars 2021**,
devant le jury composé de :

M. Dragos HORVATH
Président du jury, Directeur de recherche CNRS Delegation Alsace
M. Ruben ABAGYAN
Rapporteur, Professeur University of California, San Diego
Mme Maria Miteva
Rapportrice du jury, Directeur de recherche INSERM Ile-de-France (Paris 7)
M. Bogdan IORGA
Membre du jury, Directeur de recherche CNRS Delegation Ile-de-France Sud
M. Matthieu MONTES
Membre du jury, Professeur CNAM Paris
Mme Nathalie REUTER
Membre du jury, Professeur Universitetet I Bergen
M. Stephane REDON
Directeur de thèse, INGENIEUR DOCTEUR INRIA CENTRE GRENOBLE-RHONE-ALPES

Development of novel algorithms for data-driven prediction of interactions between proteins and small molecules

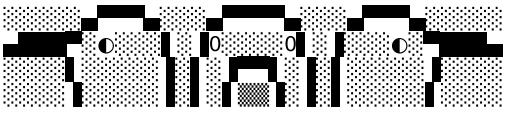
Author: Maria Kadukova[1,2]
Scientific advisors: Sergei Grudinin[2], Vladimir Chupin[1], Stephane Redon[3]

[1]Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, 141701 Dolgoprudniy, Russia
[2]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000 Grenoble, France
[3]OneAngstrom

May 26, 2021

Acknowledgements

First of all, I would like to thank my supervisors, Sergei Grudinin, Stephane Redon, and Vladimir Chupin. Without Sergei's help and guidance I wouldn't come to Grenoble to study machine learning and further apply it in structural bioinformatics. He has clearly shown me the beauty and depth of a good math model applied to some biological process that catches its underlying physics. Sergei also introduced me to our great collaborators: Elodie Laine, Karina dos Santos Machado, Pablo Chacón, and Česlovas Venclovas' team. I would like to particularly thank Pablo and Karina who have showed me the joy of working together and kind of pushed me beyond the borders of my understanding of protein-ligand interactions. Also many thanks to Kliment Olechnovič for the pleasant conversations and sharing his views on science. Besides the supervision, Sergei was always helping me with administrative issues, together with Imma Presseguer, Eugenia Chirkina, and Anna Roshchenko. So many thanks to them for their patient support with all the endless papers.
While being a student of a double degree program between MIPT and Inria, I was lucky to meet twice more colleagues as if I was working in a single lab. Many thanks to Alex Hoffmann and Krishna Kant Singh for introducing me to the world of French and Belgian beers, for the hikes, and vegetarian food. Thanks to Guillaume Pages, Francois Rousse, Ilya Igashov, Dmitry Marin, Fabian Gruber, Leonard Jaillet, Semeho Edorh and others for having lunches and occasionally spending time around Grenoble together.
Thanks to people from MIPT's CMM crystallography and MD teams. To Khalid Mustafin for sharing his understanding of the beauty of math problems. To Sergey Bukhdruker, Egor Marin, Anastasia Gusach, Ivan Gushchin, and Valentin Borshchevsky for fruitful conversations, in which they explained me a lot about experimental techniques and shook a bit my confidence in our ground-truth data. Although the double degree style of my PhD caused plenty of small administrative problems, without it I would never spend so much time with the people from room 609. Many thanks to Andrey, Nikolay, Andrey, Sergey, and Alexander. You've supported me a lot in the beginning of my thesis. Although we are not in the same room doing biology-related stuff anymore, Lovecraftian tabletop gaming with some of you is wonderful.
Thanks to all the friends I've met during studies in MIPT, UGA, and before. To people with whom I went hiking. To Karina Ashurbekova for spending wonderful time together, and for ultimately demonstrating me how cool are girls in math. Special thanks to the dude from MIPT, whose real name I still do not know. That discussions of dark music and decadent poetry have shifted my mind, showing how colorful and figurative can Russian and English languages be. Even in scientific writing.
Finally, I would like to thank my parents, and especially my mother Elena, for their encouragement to push my boundaries and for showing me the examples of truly altruistic deeds. And thanks to my boyfriend Mikhail for his endless support, and surely some help with C debugging. A single sentence is not enough to express all the things we did together that could be circumstantially related to this thesis.



Contents

Part I Introduction

Thesis description

Drug discovery is a very expensive process consisting of multiple phases. Computer simulations provide an opportunity to scale and speed up its early stages by performing the initial screening of drug candidates and modeling their interactions with the target receptors. Such modeling is generally performed with molecular docking methods. Three-dimensional (3D) models of the drug candidates are superposed with 3D receptor models by specific algorithms that are able to estimate the binding free energy change and detect the best configuration of the molecular complex. The binding free energy can be approximately evaluated with the so-called scoring functions. This thesis presents the development and assessment of several protein-ligand scoring functions with the associated docking and screening protocols.

Main contributions and the thesis outline

Research carried out in the scope of this thesis resulted in the development of three novel scoring functions described in Part II, namely Convex-PL, Convex-PL-R, and KORP-PL. From a practical point of view, the thesis author has participated in several docking challenges described in Part III that resulted in publications containing various docking protocols evaluation. The developed scoring functions are available at `https://team.inria.fr/nano-d/software/`.

1. Part I contains the thesis description and an overview of the current state of the art in the structure-based prediction of protein-ligand interactions. It also describes the methods used to obtain structural and kinetic data.

2. Part II Chapter 2 describes the development of Convex-PL – a knowledge-based pairwise distance-dependent scoring function for protein-ligand interactions, which is deduced by solving a quadratic optimization problem. Our motivation for its development was to prove that a knowledge-based scoring function can be derived by solving a classification convex optimization problem and also to demonstrate that the non-native ligand poses for the classification can be obtained with rigid constant-RMSD transformations of the native ones. Convex-PL is validated on several benchmarks and integrated into AutoDock Vina.

3. Part II Chapter 3 describes the development of Convex-PL^R – a machine learning-based scoring function that incorporates additional solvent and entropic terms. It demonstrates better affinity prediction and virtual screening performance if compared to Convex-PL. We developed Convex-PL^R to address the problem of a general scoring functions' preference of bigger protein-ligand interfaces, that partially happens due to insufficient consideration of the interactions with the solvent and mistreating the entropic contributions.

4. Part II Chapter 4 describes the development of KORP-PL – the first coarse-grained orientation-dependent knowledge-based scoring function for protein-ligand interactions. KORP-PL is based on the sidechain-free representation of a receptor and full-atom ligand representation. It was validated on several benchmarks and has proved a high virtual screening performance.

5. Part III Chapter 5 describes the participation in the pose prediction stage of D3R Grand Challenge 2. After the end of the Challenge, we have compared multiple docking protocols to understand the role of the receptor flexibility, and the choice of the docking input structures.

6. Part III Chapter 6 describes the participation in CAPRI round 41 that was focused on the docking of four oligosaccharide ligands. In this round we did scoring of docking poses with a combination of Convex-PL and protein conservation scores provided by our collaborators.

7. Part III Chapter 7 describes the participation in a cathepsin S sub-challenge of D3R Grand Challenge 3. In the pose prediction stage, we did not succeed to obtain near-native predictions for the majority of the ligands with our fully structure-based protocol. In the affinity prediction stage, we restricted the pose sampling with simplistic ligand-based constraints that produced more successful results.

8. Part III Chapter 8 describes the participation in the beta-secretase 1 sub-challenge of D3R Grand Challenge 4. The majority of the target ligands were macrocycles. In the first part of the pose prediction stage we suffered from an unnatural sampling of the macrocycles. After the analysis and improvement of the protocol, we obtained subangstrom results that were ranked 4th out of 70 in the second part of the pose prediction stage.

9. Part IV summarizes the results of the thesis.

This thesis contains purely computational research. It involved C++ and python programming, application of classical machine learning algorithms and convex optimization, and usage of bioinformatics software.
The thesis contributions were reported in 4 publications listed below and in the Conclusions, one more publication is in preparation.

1. Kadukova M., Grudinin S.. Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. J. Comp. Aid. Mol. Des.. 2017

2. Kadukova M., Grudinin S.. Docking of small molecules to farnesoid X receptors using AutoDock Vina with the Convex-PL potential: lessons learned from D3R Grand Challenge 2. J. Comp. Aid. Mol. Des.. 2018

3. Kadukova M., Chupin V., Grudinin S.. Docking rigid macrocycles using Convex-PL, AutoDock Vina, and RDKit in the D3R Grand Challenge 4. J. Comp. Aid. Mol. Des.. 2020

4. Kadukova M., dos Santos Machado K., Chacón P., Grudinin S.. KORP-PL: a coarse-grained knowledge-based scoring function for protein-ligand interactions. Bioinformatics. 2020

5. Kadukova M., Chupin V., Grudinin S.. Convex-PL^R – Revisiting affinity predictions and virtual screening using physics-informed machine learning. In preparation

Chapter 1. An overview of the protein-ligand interactions prediction techniques

This chapter contains a brief overview of experimental methods for protein-ligand structures and binding constants determination and an overview of those for protein-ligand interactions prediction. The latter can be generally classified into the *ligand-based* and *structure-based* methods. Ligand-based techniques mostly rely on ligand chemistry and are closely related to cheminformatics. Structure-based approaches utilize 3D structures of the protein-ligand complexes. In this chapter, I will focus presumably on the structure-based methods.

1.1 Binding thermodynamics and kinetics

Protein-ligand binding is a thermodynamic event that can be described with notions of statistical thermodynamics. The free energy of formation of a single complex in a solution can be written as [1, 2]

$$\Delta G = \mu^{PL} - \mu^{P} - \mu^{L}, \tag{1.1}$$

where μ is a chemical potential, PL, P, and L correspond to the protein-ligand complex, and unbound protein and ligand molecules. According to Gilson [2], the energy change of adding one protein molecule to a solution, or the protein chemical potential, in a canonical ensemble can be expressed as

$$\mu_P = -RT \ln \frac{Q_{N_P+1,N_S}}{Q_{N_P,N_S}} = -RT \ln \left(\frac{8\pi^2}{C_P} \int e^{-\frac{U(r_P)+W(r_P)}{RT}} dr_P \right), \tag{1.2}$$

where R is the ideal gas constant, T is the temperature, Q_{N_P+1,N_S} is a canonical partition function of $N_P + 1$ protein and N_S solvent molecules, C_P is the protein concentration, r_P are the internal protein coordinates, $U(r_P)$ and $W(r_P)$ are the potential and solvation energies. Here, solvation energy is equal to

$$W(r_P) = RT \ln \frac{\int e^{-\beta \Delta U(r_P, r_S)} e^{-\beta U(r_S)} dr_S}{\int e^{-\beta U(r_S)} dr_S}, \tag{1.3}$$

where $\Delta U(r_P, r_S)$ corresponds to the protein-solvent interactions, and the integral is taken over the solvent coordinates. After expressing the chemical potentials for the ligand molecule and the complex, the free energy of binding can be written as

$$\Delta G = -RT \ln \left(\frac{1}{8\pi^2} \frac{C_P C_L}{C_{PL}} \frac{\int e^{-\frac{U(r_{PL})+W(r_{PL})}{RT}} dr_{PL}}{\left(\int e^{-\frac{U(r_P)+W(r_P)}{RT}} dr_P \right) \left(\int e^{-\frac{U(r_L)+W(r_L)}{RT}} dr_L \right)} \right). \tag{1.4}$$

To derive the relation between the binding free energy and the protein-ligand concentrations ratio we must consider two relations. On the one hand, at equilibrium $\Delta G = 0$, and thus

$$\frac{C_{PL}}{C_P C_L} = \frac{1}{8\pi^2} \frac{\int e^{-\frac{U(r_{PL})+W(r_{PL})}{RT}} dr_{PL}}{\left(\int e^{-\frac{U(r_P)+W(r_P)}{RT}} dr_P \right) \left(\int e^{-\frac{U(r_L)+W(r_L)}{RT}} dr_L \right)}. \tag{1.5}$$

On the other hand, under the one-molar *standard concentration* C° the *standard free energy of binding*

$$\Delta G^\circ = -RT \ln \left(\frac{C^\circ}{8\pi^2} \frac{\int e^{-\frac{U(r_{PL})+W(r_{PL})}{RT}} dr_{PL}}{\left(\int e^{-\frac{U(r_P)+W(r_P)}{RT}} dr_P \right) \left(\int e^{-\frac{U(r_L)+W(r_L)}{RT}} dr_L \right)} \right). \tag{1.6}$$

A combination of these two equations yields the desired relation:

$$K_b = \frac{C_{PL} C^\circ}{C_P C_L} = e^{-\frac{\Delta G^\circ}{RT}}, \tag{1.7}$$

where K_b is the protein-ligand binding constant.
After some manipulations [1], the standard binding free energy can be also expressed as

$$\Delta G^\circ = \langle U^{PL} \rangle - \langle U^P \rangle - \langle U^L \rangle + \langle W^{PL} \rangle - \langle W^P \rangle - \langle W^L \rangle - T\Delta S^\circ_{config}, \tag{1.8}$$

where the P superscript refers to the interactions with the protein, L - with the ligand, $\langle U \rangle$ and $\langle W \rangle$ are the averaged potential and solvation energies, respectively, and ΔS_{config} is the entropy change related to protein and ligand motions upon complex formation. It is widely discussed that the ΔG° quantity is actually rather small in comparison with the separate terms of Eq. 1.8 that may counterbalance each other [3, 4, 5], however, this strongly depends on the way of these terms are computed [3].

1.2 Experimental methods to determine binding constants and protein-ligand structures

1.2.1 Binding constants

The binding constant K_b can be obtained from the experiments, or *assays* [6, 7, 8, 9]. Very often they are designed to measure some parameter, from which the relation between the concentrations of the protein, ligand, and complex can be deduced. For example, fluorescence polarization [10] of a ligand labeled with a fluorescent group changes with respect to its bound or unbound state. This allows to directly measure the rate of bound/unbound compounds. Binding to a chromophore-containing molecule, such as a protein with a heme co-factor or a tryptophan in the binding pocket, perturbs the UV/visible light absorbance spectrum of the chromophore. The differences between spectra can be measured with ultraviolet–visible absorption spectroscopy. In another spectroscopy-based method, surface plasmon resonance (SPR) [11], proteins are fixed on a metal surface. Ligand molecules are then injected to the solution. When they bind the immobilized proteins, the resonance of the sensitive surface plasmons change, thus changing the refraction of the surface that can be measured with an optical detector. After a while, ligands are washed out with a clean solution. From the resulting time-dependent curve of the SPR response, one can obtain the reaction rate constants and use them to compute the binding affinity. Separation of molecules in an electric field or with centrifugation may also provide information on binding kinetics. Another example of binding constants estimation is the isothermal titration calorimetry (ITC) [12] that can be used to determine an association constant. This method measures the reaction heat as the energy

that a thermostat should spend to maintain a constant temperature of the solution. Fitting the experimental data into a functional relation between the heat and molecule's concentration yields the association constant, stoichiometry, and the binding enthalpy change.
Binding data is often deposited in form of K_d, K_i, and $IC50$ constants. The dissociation constant K_d is inversely related to the association constant, or the binding constant K_b. Thus the binding free energy can be expressed as a value, proportional to the logarithm of a dissociation constant.

$$\log_{10} K_d = -\log_{10} K_b \propto \Delta G. \tag{1.9}$$

The inhibition constant K_i is analogous to K_d, but is defined solely for enzymes and their inhibitors. $IC50$ is also related to enzymes and corresponds to the inhibitor concentration that inhibits a given biological process by 50%. $IC50$ values are known to be dependent on the conditions of the experiment, especially if inhibition was competitive. For this case, Cheng-Prusoff equation can be used to convert $IC50$ to K_i [13].
Although there exist a lot of different approaches for binding constant determination, all of them can be subject to uncertainties [6]. Some methods are very condition-dependent and comparison of their results is rigorous only inside one assay. During the experiment, molecules may become damaged, aggregate into droplets or stick to the walls of the vessel. Such changes would influence the concentrations and the parameters used to analyze the experimental data. Assays that require immobilization and labeling may influence the binding free energy value since they change the molecules' entropy. Kinetic equations should be carefully chosen if a protein is known to have several binding sites.

1.2.2 Structural data

3D structures of macromolecules can be obtained with several experimental techniques with different resolutions and structure preparation requirements. These include X-ray diffraction [14, 15, 16], nuclear magnetic resonance (NMR) [17], cryogenic electron microscopy (Cryo-EM), small angle X-ray and neutron scattering (SAXS and SANS). Small angle scattering is currently unable to provide atomic resolution with trustworthy coordinates, especially for the small ligands, and is presumably used to study shapes and motions of proteins. Until recently, Cryo-EM suffered from similar low resolution problems. However, nowadays single-particle Cryo-EM is extensively evolving and starts being able to provide atomic resolution structures [18, 19].
X-ray diffraction is the most common method to obtain a molecular structure with atomic resolution. For example, almost 90% of protein structures deposited in the Protein Data Bank (PDB) [20] were determined with X-ray crystallography. X-ray waves diffract on the crystallographic planes of a protein crystal as on a 3D diffraction grating and thus form a diffraction pattern. The locations and intensities of the spots in the pattern represent the lattice parameters and the amplitudes of diffracted waves. The diffraction spots, generated by scattered waves, are interconnected with crystal electron densities by a Fourier transform. However, these densities can not be calculated directly from the spot's amplitudes, as the information about the phases is lost. This leads to the so-called *phase problem*. One of the most common approaches to solve the phase problem is molecular replacement (MR). If there exists a structure similar to the target protein, it can be superposed with the target cell and used for the phases computation. After the initial guess on phases

is obtained, the model is iteratively refined in specific software packages to better fit the electron density maps. The refinement process often requires manual intervention of the crystallographer, and relies on the knowledge-based stereochemical restraints. Ligand coordinates are usually identified after the preliminary refinement of the model is done. Unfortunately for the protein-ligand binding studies, ligand coordinates contain errors more often than the protein coordinates do [21]. To partially address this problem, the structural biology community has developed specific tools for ligand reliability evaluation on the basis of its electron density map [22, 23, 24]. There also exist initiatives for protein structure re-refinement, such as PDB-REDO [25].
One of the biggest limits of X-ray crystallography application is the protein crystallization. Prior to the diffraction step, protein molecules should be expressed in cells, purified, mixed with the ligand[1] and crystallized. Each of these steps may require a long and non-trivial search of specific conditions. Even if crystallization was successful, crystallized state and inter-molecular contacts that it causes are rather unnatural for a protein and may lead to the conformational shifts of its structure. Some macromolecules cannot be easily crystallized with classical methods. For example, crystallization of membrane proteins is hindered by their instability and often results in small crystalls that are unsuitable for classical X-ray crystallography. However, their structures can be determined with serial femtosecond crystallography based on the usage of X-ray free electron lasers [26].
Another popular method of atomic-level structure determination is multi-dimensional NMR [17] spectroscopy. Nuclei resonance frequencies depend on their adjacent functional groups. These small differences from the standard frequencies are called chemical shifts and can be used to determine the local environment of nuclei. In addition, nuclear Overhauser effect is used to obtain distance constraints between atoms. Unlike X-ray diffraction, NMR spectroscopy does not require crystallization and is able to catch the dynamics of a molecule. However, it is in general more expensive than X-ray crystallography and is restricted by the size of the protein, although several techniques were introduced to overcome this limit [27].

1.2.3 Important databases

Data obtained from the numerous assays and structural experiments is deposited and organized in specific databases. The most important archive of protein structures is the Protein Data Bank, which was founded in 1971 and is currently maintained by the Worldwide Protein Data Bank (wwPDB). wwPDB consists of four organizations located in USA, Europe, and Japan, namely RCSB PDB, PDBe, PDBj, and the NMR data bank BMRB. Each of them is a public database synchronized with each other. Small molecule crystallographic data is collected and maintained by the Cambridge Structural Database. However, the full access to its annotations and the possibility of data parsing is restricted by a subscription model. As an alternative, an open-access Crystallography Open Database [28] was founded in 2004.
CHEMBL [29] is one of the biggest databases storing the protein-ligand assay data. It contains and curates data from scientific publications, other public databases, and datasets shared by both non-profit and commercial organizations. Besides binding assays, it contains functional, toxicity, physicochemical, and other assays, as well as calculated properties of ligands. CHEMBL maintainers provide a python-based web interface for the database parsing [30]. BindingDB [31] specializes

[1] Proteins can be mixed with ligands before crystallization, or soaked in the ligand solution after crystallization

in binding assays and tries to include more details about the experimental conditions. It contains data from literature and public databases. PubChem BioAssay [32] contains assay data deposited by users and organizations.
In this work I am heavily relying on the PDBBind database [33] that stores both structural and binding information and was primarily created for the development of scoring functions. Starting from 2004, its maintainers collected structural data and annotated it with information on binding constants. Their workflow for the database compilation starts with the Protein Data Bank parsing. For a suitable complex, its primary reference indicated with the deposited structure is then checked for the binding data. If no binding data was found, PDBBind maintainers search this information in other publications related to the protein. When a binding affinity record is found, it is examined by two persons to minimize the human error. The current version of PDBBind, i.e. version 2019, contains 17 679 annotated protein-ligand complexes. One more database containing both structural and binding data is Binding MOAD [34]. Its system of annotations is more flexible that the one of PDBBind. For example, if a complex has several ligands and co-factors and the binding constant is measured for only one of these small molecules, PDBBind keeps only the ligand with the known binding constant. On the contrary, binding MOAD keeps all the small molecules present in the structure since they can be important for binding, and labels each of them with respect to the presence or absence of the binding data.

1.3 Structure-based prediction of protein-ligand interactions

Molecular docking methods can efficiently predict the binding energy and the correct 3D configuration of a protein-ligand complex. Generally speaking, molecular docking involves two algorithms. The first one is required to sample the possible configurations of the complex. The second one is designed to score these configurations and guide the sampling algorithm. The scoring is usually carried out with the so-called *scoring functions* that can also be applied at the very last step of docking for final re-scoring of the generated configurations. If the number of complexes of interest is relatively small, docking can be performed with accurate and rather slow methods that do not involve scoring functions. This can be molecular dynamics-based sampling in combination with thermodynamic integration for rigorous binding free energy prediction [35, 36, 37]. However, such methods are hardly applicable in computationally expensive tasks such as virtual screening, raising the demand for efficient sampling and scoring algorithms.

1.3.1 Scoring functions

A considerable number of scoring functions have been developed in the protein-ligand community throughout the past years. They are sometimes classified into four categories, although such classification is not very rigorous [38]. Knowledge-based potentials employ an assumption that statistical analysis of the protein-ligand complex structures' geometry may uncover the differences between native and non-native binding poses [39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49]. Typically, these potentials are given as a sum of pairwise terms that are derived from the inverse Boltzmann statistical distributions of distances (or, generally, geometric features) between atoms of protein-ligand complexes. They can be, however, augmented with terms that are calibrated using additional data [50, 51, 52, 47, 53], or derived in a different way if compared with classical statistical po-

tentials [54, 55]. In addition to the knowledge-based potentials that are most often derived in a statistical and unsupervised manner, a lot of scoring functions are based on other principles.
Physics-based scoring functions [56, 57, 58, 59, 60, 61] rely on direct simulations of the possible physical effects of protein-ligand interactions and are often related to force-fields, whose development requires very careful calibration. Empirical scoring functions [62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 53] were initially a linear combination of several terms that represent energy contributions of possible interactions at the protein-ligand interface. They are usually trained to fit binding constants obtained from experiments, and thus strongly depend on the quality of the experimental data. Many of them also utilize the force-fields for the computation of individual terms. The terms of empirical scoring functions may include the typical Lennard-Jones potential, terms for solvation and lipophilic effects, ligand features such as the number of rotatable bonds, orientation-based terms accounting for hydrogen bonding, or for the interactions with metals, terms for the ligand intramolecular energy, specific constraints and many others. One of the interesting examples of specific terms are the protein conservation scores [75]. Surprisingly enough, the classical empirical scoring functions such as AutoDock Vina and its modifications, while being more or less physically interpretable, still achieve stable state-of-the-art results in both pose and affinity predictions and are widely used.
The so-called machine learning-based scoring functions [76, 77, 78, 79, 80, 81, 82, 83, 84] are in some sense an extension of empirical scoring functions. They are also often trained to fit the binding constants, but usually rely on a bigger number of descriptors of different nature, and are based on non-linear machine learning models. Starting from 2015, a number of 3D deep learning architectures designed to either predict binding poses [85, 86], or fit the binding constants and classify binding and non-binding compounds appeared [87, 88, 89, 90]. The most recent architectures employ diverse training sets and objectives, and are able to predict both poses and affinities, as the classical all-purpose scoring functions do [91]. Although some of the recent machine-learning scoring functions often demonstrate high performance in affinity prediction and virtual screening, they are also subject to a number of flaws. While classical statistical potentials tend to be biased towards the number of contacts between the two molecules, learning on a relatively small number of available high-quality binding constants introduces biases towards experimental affinities. Very complex models, especially those from deep learning, may also introduce overfitting. For example, some recent architectures demonstrate excellent results on the DUD-E virtual screening benchmark if they are trained on a part of it. However, their performance is rather average if they are trained on other data sources [92]. These problems can be in principle solved by, for example, augmentation of the training set [93].
A conceptual difference between some of the recent ML-based scoring functions and more classical, all-purpose, scoring functions is in the understanding of the prediction task, as shown in Figure 1.1. Classical scoring functions, including some of those obtained with non-linear ML methods, are trained to predict energy, or a score that would be proportional to it, independently of the method and objective, with which the scoring function was trained. In contrast, some recent ML-based scoring functions are designed to provide the output, directly required by the task. For example, the multi-task scoring function by Ashtawy and Mahapatra [81] outputs three separate values for the pose, affinity and activity. The model developed by Kandemir and colleagues [86] computes a probability that a binding pose is a good one. Although the output values of these networks

could be trained to mimic energy, it is not needed in practice, at least because the models are less interpretable and the relation to energy would not make sense. This approach was not previously common for structure-based methods, unlike for the ligand-based ones.

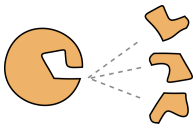
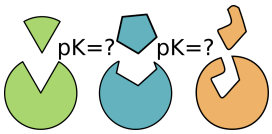
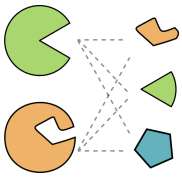
Task	Pose prediction	Affinity prediction	Virtual screening
			
	Find the correct 3D pose for the known ligand	Predict energy or affinity, rank by affinity	What ligands bind a protein and what do not
ML-based approach	Classification, ranking, quality scoring	Regression, ranking	Classification, positive unlabeled classification, ranking
Energy-based approach	Best pose – the lowest **energy**	**Energy** prediction	Ligand binds – **energy** is low

Figure 1.1: Machine learning-based and physics-based view on the classical scoring functions' tasks.

1.3.2 Solvent and entropic contributions, and the bias towards bigger ligands

As it was already mentioned in Eq.1.8 of Section 1.1, the binding Gibbs free energy can be written as [2, 1]

$$\Delta G = \langle U^{PL} \rangle - \langle U^{P} \rangle - \langle U^{L} \rangle + \langle W^{PL} \rangle - \langle W^{P} \rangle - \langle W^{L} \rangle - T\Delta S_{config}, \tag{1.10}$$

where the P superscript refers to the interactions with the protein, L - with the ligand, $\langle U \rangle$ and $\langle W \rangle$ are the averaged potential and solvation energies, respectively, and ΔS_{config} is the entropy change related to protein and ligand motions upon complex formation. However, many of the approaches would make very crude approximations of the entropic term and interactions with the solvent in the above equation. This causes the known flaw of many knowledge-based scoring functions preventing them from being used in screening tests. More precisely, many of them have a strong bias toward bigger and tighter protein-ligand interfaces. Conformations of a ligand inside a binding pocket that have a higher number of interactions with the protein, even weak ones, will often be preferred over the native ligand pose. However, in reality, some parts of the binding site and the ligand exposed to the solvent could be more favourable compared to the corresponding protein-ligand contacts.
The preference of larger interfaces can be illustrated by the publicly available results of scoring functions evaluation on the virtual screening test of CASF-2013 [94] and CASF-2016 [95] benchmarks shown in Figure 1.2 (a). Here one can see that the majority of the assessed scoring functions prefer binding with non-native ligands (decoys) with, on average, up to twice bigger buried solvent-accessible surface area (SASA) values than a native ligand has. Figure 1.2 (a) shows that, in fact, for some scoring functions this trend is even stronger if the total number of atoms is used instead of the SASA, buried upon binding. Notably, AutoDock Vina [70] and AutoDock Vina-based $\Delta_{\text{Vina}}\text{RF}_{20}$ [80] scoring functions do not suffer from this bias that much. This can be explained by the way AutoDock Vina scales its binding energies by the number of ligand's rotatable bonds.

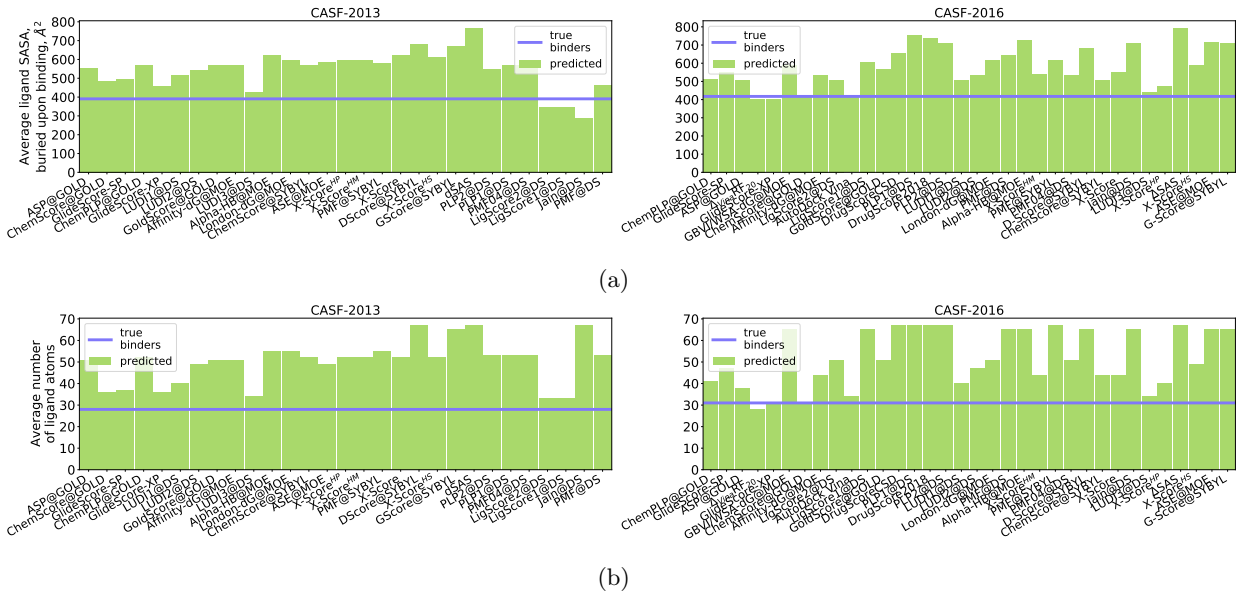


Figure 1.2: Statistics for the CASF 2013 and 2016 benchmarks. (a)-(b) The purple line represents the average values of buried SASAs and numbers of atoms computed for ligands that natively bind target proteins and should have been predicted as the most affine binders[1]. The green boxes correspond to decoys that were top-ranked by scoring functions assessed on the virtual screening test from the CASF benchmarks 2013 and 2016. The scoring functions are sorted by the ability to predict the highest affinity binder in the 5% of the top-ranked decoys. SASA values were computed with PyMOL's[96] *get_area()* function with *dot_solvent* set to 3.

[1] Or be among the most affine binders in several cases when the target protein was known to bind ligands with higher affinity but without co-crystal structure.

Interestingly, the AutoDock Vina scoring function was in some sense inspired by the X-Score scoring function and has similar terms. GlideScore-XP [68] also does not express any considerable bias toward the overall ligand molecule size neither. This is probably owing to its solvation term and the correct penalization of contacts between polar and hydrophobic groups.

Many other empirical [70, 66, 63, 72, 82] and some knowledge-based [51, 52, 53] scoring functions circumvent these problems by including additional entropic and solvation terms in their expressions. A classical approximation of ligand conformational entropy is the number of torsions or atoms involved in rotatable bonds [70, 52, 53, 66, 63]. Some scoring functions also include rigid-body contributions approximated with a logarithm of the ligand mass [53, 63], even though this approximation, as well as the involvement of the mass-dependent rigid-body entropy itself, is arguable [1, 97, 98]. Basic implicit representations of solvation include interaction terms proportional to the SASA [99, 52, 80], or solvent-accessible volume difference upon binding [100, 101, 102, 61]. Some algorithms utilize SASA in more sophisticated ways, such as calibration using the octanol-water partition coefficients alongside with separate hydrogen bonds description aiming at a better hydrophobic effects representation [103, 104], or integrating the surface curvature factor of the molecules over the solvent-accessible surface area [105]. Another way to compute solvation energy change with an implicit solvent model is to use the 3D-RISM [106], Poisson-Boltzmann, and generalized Born methods [2, 107, 108, 109]. They are, however, much more computationally demanding. Explicit solvent representation for molecular docking purposes [82] requires either high-quality X-ray structures with the hydration shell resolved or the hydration shell sampling performed by dedicated algorithms and molecular dynamics-based pipelines [110, 111, 112, 113]. Although these approaches look quite intuitive and generally improve the docking quality, they are mainly used

to predict the water sites for individual targets, which may require manual intervention, making it hard to apply them on a larger scale. However, this problem can be solved using statistical potentials for water molecules prediction derived from the crystallographic structures [114, 115]. Despite the variety of possible estimations of entropic and solvation terms, Figure 1.2 implies that some of these strategies are yet not sufficient.

1.3.3 Other components of the small molecules docking pipeline

Structure preparation

A good scoring function is crucial for correct predictions, but the docking success depends on multiple steps of the pipeline. In particular, accurate predictions require proper input structures. Receptor structures of interest can be found in a crystallographic database, mutated from homologous structures, or modeled from sequence. If a proper fully-homologous structure was found in the Protein Data Bank, structure preparation can be rather simplistic and, dependent on the docking algorithm, include hydrogens, partial charges and atomic types assignment. More complex structure preparation may involve water molecule sites prediction, modeling the gaps in the receptor's structure, minimization using a force-field, sampling of an ensemble of binding pocket conformations, etc. An important stage of the receptor preparation is the detection of the binding pocket. In some cases, binding pocket can be simply found by visual inspection of a receptor or from the comparison with known homologous structures, co-crystallized with ligands. However, in absence of homologous structures, this becomes quite challenging for proteins that do not have notable exposed cavities or those with vast binding sites located on the protein surface. Specific tools were developed for the binding site detection [116, 117]. In the worst scenario, docking can be inefficiently done to a set of binding regions defined to cover the whole protein structure, with subsequent clustering and selection of the best poses from the ensemble. In some cases, Normal Mode Analysis or molecular dynamics can be applied to open the pocket [118].
3D ligand structures are usually generated from the 2D interpretation with specific software, which is generally based on an optimized conformational sampling and an optional subsequent force-field based local optimization [119, 120, 121, 122]. Conformational sampling may extensively rely on knowledge-based information, such as pre-generated 3D fragment libraries and other constraints and heuristics, especially for ring conformations and torsion angles prediction [123, 124, 122, 125, 119, 126, 127, 120]. Some methods are based on the distance geometry computations [128, 129] that optimize a structure to fulfill a set of geometric inequalities for the bond lengths and thus utilize simplistic constraints for bond length and valence angle values. Overall, proper 3D ligand coordinates generation is extremely important, as many docking methods sample ligand conformations in only torsional coordinates. Thus, even small errors in local geometry may influence the docking performance. In this thesis, I was initially using OpenBabel[2] and then switched to RDKit's ETKDGv2. Both methods sometimes produced structures with local geometry different from the native co-crystal conformations.

[2]The version of *gen3d* function that was most likely developed prior to [126].

Pose sampling

A great variety of methods for the protein-ligand docking and pose sampling have been proposed in the recent 30 years [117, 130]. Here I will briefly list the principles that many of the classical structure-based approaches follow.

Some docking tools, especially the earliest ones, consider both ligand and receptor to be rigid. This requires generation of an ensemble of multiple ligand conformers, which are rotated and translated in the binding pocket, to find their best position. [130]. The search of an ideal ligand position can be then done with respect to the shape complementarity as in the earliest version of DOCK [131], with exhaustive-search methods [73], or even those accelerated with the fast Fourier transform [132, 133], brought from the protein-protein docking. Otherwise, ligand conformational states are explored inside the protein binding pocket. This can be achieved by very diverse stochastic optimization and search techniques, such as Monte Carlo methods [67, 134, 70] that can be biased by a priory known probability distribution [135], various genetic algorithms [136, 65, 137, 138, 139], swarm intelligence-based methods [140], or a combination of several optimization methods applied sequentially [72]. Some methods take an advantage of combining fast rigid-body docking of an internally generated library of the conformers with subsequent flexible fine-tuning of the best conformations [134]. The conformational search is most often done in the space of torsional coordinates of a ligand only. However, ICM [141] and Rosetta [142] are based on the search in the complete set of internal coordinates, i.e. bond lengths, bond angles, and dihedral angles. One more approach is the so-called *incremental construction* [143, 144], where a ligand is placed inside a pocket fragment by fragment. To speed-up the conformational search, many docking tools rely on the scoring function values precomputed on grids that can be then interpolated, or lookup tables. In this thesis I am using AutoDock Vina [70], as it is open-source and provides fast docking with minimal required structure preparation that can be easily automatized.

Here I have mentioned only a small portion of the existing approaches for molecular docking. In addition to the unrestricted sampling in a binding box, drug discovery often requires docking with search space restrictions, such as fragment-based [145], template-based, and covalent docking [146]. A number of specific tools have been developed for these purposes, and some of the above-mentioned methods have extensions to address such tasks. Docking can also be performed with molecular dynamics approaches that are not covered in this short overview.

While ligand sampling is a more or less solved problem, receptor flexibility remains a challenge [147, 148]. Sidechain conformations can be sampled simultaneously with the ligand conformational space exploration [138], but in many cases this increases the complexity [139] drastically. A popular approach is docking into an ensemble of the pre-sampled configurations of the binding pocket. If the exploration of the receptor conformations is optimized as a part of the docking process, this approach is called Multiple Receptor Conformation docking [149, 148], or the 4D-docking [150]. The conformations of a pocket can be explored prior to the docking with a molecular dynamics simulation, rotamer sampling, or more specific tools. One more strategy is to sample the sidechain positions for several rigid ligand conformations [134, 151] that can be done to resolve clashes between the molecules with a minimal number of sidechain rotations [152]. Nontheless, the problem of predicting huge backbone shifts, especially simultaneously with ligand docking, is still unsolved.

Post-processing

Some scoring functions cannot be used inside the sampling process because of their complexity. In such cases, docking is performed prior to re-scoring to explore the conformational landscape of the complex. Overall, after the sampling stage, protein-ligand poses can be re-scored with any desired scoring functions. These results can be then post-processed, which may involve analysis of the docking pose clusters, consensus scoring based on several scoring functions' decision [153, 154, 155] that is especially common in virtual screening, or human visual inspection.

1.3.4 Benchmarking of protein-ligand scoring functions

Multiple benchmarks were suggested in the protein-ligand modeling community to evaluate the performance of molecular docking algorithms and scoring functions. Some benchmarks contain co-crystal and docked structures and suggest a variety of tests to evaluate both pose, affinity, and activity prediction abilities [156, 157, 158, 159, 94, 160][3]. Other require a docking step performed on the user's side [161] and sometimes utilize exercises from the previous blind docking challenges [162, 163, 164]. There also exist a number of virtual screening benchmarks that can be used to assess both ligand-based methods and structure-based docking pipelines [165, 166, 167, 168, 169]. One of the common drawbacks of some virtual screening benchmarks is the unproved inactivity of generated decoys. The recent development of the machine learning-based scoring functions aroused many questions about the biases related to the benchmarks, as the latter are often used as training sets with subsequent cross-validation [170, 84]. The properties and biases of the popular virtual screening DUD-E benchmark were extensively discussed in this context [92, 170, 84].
In this thesis I am using the CASF-2013, CASF-2016, D3R-based, DUD, and DUD-E benchmarks that are described in more detail below.

CASF benchmarks

CASF-2013 and CASF-2016 [94, 160] benchmarks were created by the PDBBind maintainers and contain a set of respectively 195 and 285 complexes with known binding affinities. They also provide the results of assessment of 20 and 34 popular scoring functions. Both CASF benchmarks suggest four different metrics defined as *docking power*, *scoring power*, *ranking power*, and *screening power*. Docking power corresponds to the ability of a scoring function to predict the native or the best near-native docking pose among a set of computer-generated configurations. Scoring functions are evaluated by the number of the top-ranked predictions (top-1, top-2, and top-3) below a predefined cutoff distance from the crystal structure (1.0, 2.0, and 3.0 Å). Native structure may be either excluded or included to the comparison. Scoring and ranking powers measure the quality of affinity prediction of complexes with known co-crystal structures. Scoring power assesses the correlation of scoring function predictions with the experimental binding affinity data. Ranking power is related to the capability of a scoring function to correctly rank a set of known ligands for a target protein. In CASF-2016, where five known ligands are available for each target protein, it is measured by Spearman's correlation coefficient. However, in CASF-2013 only three ligands per protein are available and ranking power is represented with two numbers characterizing success rates of either correct ranking of all the given ligands, or finding the most affine one. Finally, screening power is

[3] A good overview of the older benchmarks can be found in the SI of [160]

related to the ability of a scoring function to identify true binders for a target protein among a set of small molecules. CASF benchmarks suggest two metrics to evaluate this ability. Enrichment Factor (EF) is calculated as a ratio between the total number of true binders observed among a fraction of top-ranked candidates (1%, 5%, and 10%) and the total number of true binders multiplied by this fraction. It represents the ability of a scoring function to correctly find active compounds compared to a random selection. 'Best binder success rate' is a success rate of identifying the highest-affinity binder among the 1%, 5%, or 10% of top-ranked ligands over all the test cases. In addition to this evaluation, CASF 2016 also introduces the *reverse screening* to identify potential target proteins for a given ligand.

D3R benchmarks

I have compiled our own benchmark based on the user submissions from the recent blind docking challenges. It contains pose prediction and scoring tests for three physiologically important proteins. The aim of blind challenges is the evaluation of docking protocols on previously unpublished structural data. After all participants have submitted their predictions, co-crystal structures become revealed and submissions get evaluated. A considerable effort was made by the D3R community to host data from the previous challenges. In particular, this resource contains all user submissions and answers, i.e. native structures and binding constants, from the recent three blind challenges, namely Grand Challenge 2 [171], Grand Challenge 3 [172], and Grand Challenge 4 [173]. Unfortunately, user submission data from the first D3R challenges is not publicly available.
Thus, I compiled a benchmark from the user submissions and published answers of the three blind challenges. Similar to the CASF benchmarks, it contains pose and affinity prediction exercises. However, this benchmark is different from CASF in several aspects. Unlike the CASF benchmarks, which were created from the data deposited in the Protein Data Bank, experimental data for each of the D3R challenge targets were provided by a single research group. Co-crystal structures were also visually inspected by the challenge organizers and participants. This allows us to expect higher quality and consistency of this data, especially for the binding constants, which are less trustworthy in the CASF benchmarks, and PDBBind in general. On the contrary, data from the D3R Challenges provides smaller diversity of both proteins and small molecules, since each of the three challenges was focused on one protein target binding with compounds of several chemical series. For example, the affinity prediction test made from the D3R Challenge data is closer to the CASF ranking test than to the scoring one.
For the pose prediction tests, I collected all available user submissions from the pose prediction stages of the three challenges. Root mean square deviations of atomic positions (RMSD) were obtained from the D3R website when possible, otherwise, I computed them using a modified version of symmetry-adapted RDKit's *GetBestRMS()* function, in which I disabled the ligand alignment, and PyMol's [96] *align* function to superpose each protein to its native structure. I excluded several submissions listed in Table A.2 because of various errors and clustered the rest of submissions with a 0.1 Å threshold without the binding pocket alignment. This clustering was mainly done to remove very similar or equivalent docking poses that were often present in submissions from the same users. Finally, I measured the pose prediction success rates on each test separately with and without the inclusion of the native structures. For the affinity prediction tests, I selected only the native structures and then measured the Spearman's correlation coefficients between predicted and

experimental binding constants for each of the Grand Challenges. When the ligand was present in several chains of the co-crystal structure, I scored all of the available complexes and took the average. The number of available submissions and binding constants is summarized in Table A.1.

DUD and DUD-E benchmarks

Directory of Useful Decoys (DUD)[165] and DUD-Enhanced (DUD-E)[167] datasets are specifically designed for virtual screening benchmarking. They contain protein targets with known 3D structures and binding pockets, a number of active compounds known to bind those proteins, and series of non-binding decoys generated for each active molecule. The goal is to discriminate active compounds and decoys. Please note that the word *decoy*, which is usually used in this thesis to describe the *generated non-native poses of ligands*, means *a computationally generated molecule without proved activity* in the scope of the DUD and DUD-E tests.
The DUD benchmark [165] contains 40 targets with a total of 2,950 active compounds. For each active compound, the dataset provides 36 decoys with similar physics, but various chemical topology. In my assessments, I have excluded nine targets containing co-factors listed in Table A.3, since our potentials Convex-PL and KORP-PL are not parametrized for predicting the ligand-ligand interactions. This resulted in 31 targets selected for benchmarking.
The DUD-E benchmark [167], the successor of DUD, is a very popular approach of assessing virtual screening abilities of various scoring functions and docking protocols. It consists of 102 targets, a set of active compounds per target known to bind it, and 50 inactive compounds, or decoys, per each active one. The total number of active compounds for all 102 targets equals to 22,886. For each target, one protein-ligand complex is provided and can be used for the identification of the binding pocket and molecular docking. The benchmark also contains 3D conformers of all the active and inactive compounds. As in the case with the DUD benchmark, DUD-E contains several targets with co-factors that seem to be crucial for binding. Again, I have excluded from the evaluation 12 complexes listed in Table A.4 that contained co-factors, namely HEM, NAD, NAP, FAD, ADP, and FMN.
Unlike the CASF benchmarks and the D3R-based benchmark that I have derived specifically for structure-based scoring functions assessment, evaluation on DUD and DUD-E requires a pose sampling stage. Therefore, I firstly performed molecular docking using either AutoDock Vina or VinaCPL (described in Chapter 3), and then re-scored the obtained poses with my scoring functions. For the evaluation metrics, I chose the ROC AUC value, the 5% enrichment factor, and the early enrichment BEDROC [174] metric.

1.3.5 Blind docking challenges

State-of-the-art benchmarks are useful for scoring functions validation and comparison, but the awareness of correct answers may lead to a bias towards the benchmark content. Blind docking challenges provide a way to assess docking protocols on a set of protein-ligand complexes with unknown crystallographic structures [162, 175, 164, 163, 176, 171, 172, 173]. They usually contain pose prediction exercises measuring the ability to find a near-native protein-ligand complex conformation, and scoring exercises that evaluate the quality of relative binding affinity predictions. After the end of the challenge, crystallographic structures of the complexes are revealed, as well as

the binding constants.

Blind challenges allow to summarize the results of various docking strategies. On the one hand, it provides an opportunity to compare protocols involving distinct methods such as Free Energy Perturbation and classical molecular docking on the same data. On the other hand, it is an opportunity to compare protocols based on the same methods and tools utilized by different teams. While the target-specific nature of the most challenges does not always allow to make general conclusions on docking protocols, an analysis of multiple submissions may provide a desirable degree of generalization. For example, the recent D3R Challenges have clearly shown the importance of receptor flexibility, demonstrated the success of template-based techniques, and provided some insights on the role of human intervention into the docking protocols.

Some of the past challenges were remarkable for the exercise design or specific features of the receptor or ligands. For example, the D3R Grand Challenge 4 contained docking of macrocycles, and in Phase 1 of CSAR 2013 exercise participants were asked to find the best protein sequence that binds with the same compound, which involved extensive homology modeling. In the D3R challenges, participants were encouraged to use existing crystallographic structures from the RCSB database, which could have been selected with ligand similarity-based methods. In the course of the latest CSAR and the first D3R competitions, various methods were used. Along with the classical docking methods [177, 178, 179, 180, 181, 182, 183, 73, 75, 184, 185], many teams relied on template-based approaches that incorporate search space restrictions with respect to known ligand structures crystallized with homologous proteins [186, 187, 188, 189, 190]. Other methods included QSAR models applied for affinity prediction [191, 192, 183], target-specific scoring functions [193, 192, 194], and sometimes combinations of those with more computationally expensive molecular dynamics-based methods [195, 196, 197, 198]. Participants reported rather classical problems related to the receptor modeling, such as the importance of taking into account the receptor flexibility, buried water molecules, presence of explicit solvent, lack of the homologous proteins present in databases, etc. In particular, several approaches were used to handle the receptor flexibility problem [199]. In case of a considerable number of complexes with homologous proteins available, their thorough inspection helped to choose either the proper docking target, or to identify a small number of residues suitable for flexible docking protocols [200, 197]. Side-chains repacking [188] and the induced-fit approach [201, 196] helped to optimize or widen the binding pockets. Flexible docking programs were also used [202].

The target protein of the D3R Grand Challenge 2 was a farnesoid X receptor (FXR). Its flexibility caused difficulties in pose predictions of several ligands, especially those of chemical series unrepresented in the crystallized homologous structures from the PDB. After the end of this Challenge, D3R organizers added a separate pose prediction stage, in which the native receptor conformations were revealed and available for docking. Again, the submitters relied on rather diverse structure-based strategies. A number of template-based approaches lead to relatively low RMSD values in the pose prediction stage [203, 204, 205], although their results in general were not outperforming the classical docking sampling [206, 207, 208, 209, 133, 210]. Molecular dynamics-based pipelines were widely used [206, 210, 211, 212, 213], sometimes in combination with the ligand-based restrictions [204]. The biases towards the ligand size that were discussed in the Introduction are reported in [204]. Despite the usage of ensemble docking, including the method proposed by the ICM developers [203], induced-fit docking [212], molecular dynamics and local optimization of the

sidechains [133], none of the submitters managed to find a near-native pose with RMSD less than 3 Å for the FXR_18 ligand, whose binding induced a considerable shift of the binding pocket's backbone. Only the group that was using SILCS [214] software and MD-based preparation of the receptor was able to obtain an acceptable RMSD of 3.06 Å.

Subchallenge 1 of D3R Grand Challenge 3 was focused on docking of chemically diverse ligand molecules to the cathepsin S receptor. Although the receptor itself was rather rigid, and a considerable number of homologous structures were available in the PDB, docking to its wide binding pocket exposed to the solvent turned out to be quite challenging for many classical structure-based approaches. The most successful strategies of ligand pose prediction for the CatS protein were the ligand template-based approaches [151, 215, 216, 217, 218]. Two of these submissions included 3D similarity-based ligand placement into the binding pocket with a subsequent optimization of the ligand and the receptor sidechains conformations [151, 215]. Knowledge of ligand locations in homologous proteins can also be directly included into the scoring function used in docking [218]. Participants also reported on additional molecular dynamics-based refinement that improved the pose prediction quality [151, 219]. Explicit water molecules might be very important for proper estimation of interactions with the wide binding pockets [216]. Other approaches included molecular dynamics-based sampling and thermodynamic averaging [220] and implicit ligand theory [221] for binding free energy predictions. Overall, the number of successful submitters was smaller than in other D3R challenges, and, unfortunately, not all of them have published an analysis of their protocols in the special journal issue.

Subchallenge 1 of D3R Grand Challenge 4 provided an interesting opportunity of macrocycle docking of beta secretase 1 ligands. Although sampling of the cycle conformations was a challenging task, the rigidity of the binding pocket and the availability of co-crystal ligand structures in the Protein Data Bank simplified the pose prediction stage for many participants. This time many teams used template-based methods [222, 223, 224, 225], some of which provided subangstrom average RMSD values. Classical structure-based docking and molecular dynamics methods were also successfully applied [226, 227, 228, 229]. One of these approaches yielded the best structure-based affinity predictions [230]. Although multiple structures of beta secretase 1 complexed with ligands were available, some teams focused on more realistic docking to the apo structure that required modifications of its binding pocket [231]. Several teams trained machine-learning-based models for target-specific affinity prediction [225, 232]. The three latest Grand Challenges are also remarkable for the first demonstrations of the 3D convolutional neural network-based scoring functions [233, 234]. In the Grand Challenges 3 and 4, Wei Guo-Wei with colleagues proposed several successful methods relying on quite novel graph-based features [217, 235].

Despite the amount of protocols' descriptions, both D3R and CSAR organizers report the difficulty of making conclusions about the best approaches, as the same methods used by different teams often lead to significantly different results. A notable example illustrating the difficulty of summarizing the results of the challenges is the visual inspection of predicted structures. Its importance was mentioned in the D3R 2015 results description. On the contrary, a year later visual inspection was not widely reported in the protocols of the top-performing methods. In the Grand Challenge 3, however, it was commonly applied again in the successful submissions. Finally, in the Grand Challenge 4, visual inspection was used in some of the successful submissions, but was not crucial for the methods' success in the majority of the cases.

The aim of our participation in blind challenges was mostly to evaluate our scoring functions, which was, however, dependent on the other parts of the docking protocol including initial structure preparation, docking, and post-processing of the results. Different versions of Convex-PL and its very early prototypes were assessed in the CSAR 2013 [163], CSAR 2014 [164], D3R 2015-2016 [176], and the D3R Grand Challenges 2 [171], 3 [172], and 4 [173]. Section 6 (D3R Grand Challenge) of my Master thesis is dedicated to our participation in the D3R Grand Challenge 2015. This thesis contains the description of our participation in the Grand Challenges 2-4, and in the 41th round of CAPRI.

Part II Development of scoring functions

Chapter 2. Convex-PL development

Several years ago, Sergei Grudinin and Petr Popov proposed a knowledge-based scoring function for protein-protein interactions [236] based on convex optimization. The idea was based on the earlier master theses of Georgy Derevyanko and Georgy Cheremovsky. A part of my master thesis project was to adapt this method to the protein-ligand interactions resulting in the development of a scoring function called Convex-PL. This thesis continues and further develops those studies. In particular, during the first year of my PhD program, I have improved the typization of ligand atoms and added the repulsion term that was especially necessary for the integration of Convex-PL into AutoDock Vina. I have also assessed the performance of Convex-PL in the case of exclusion of the complexes, homologous to the test set, from the training set. Subsections below describe the model of protein-ligand interactions used in Convex-PL and its derivation. The equations and theory from the parts describing the model of interactions and the optimization process are included in my master thesis. This section can also be found in our paper presenting Convex-PL [55].

2.1 Methods

2.1.1 Model of interactions

Let us consider P co-crystal (native) structures of protein-ligand complexes C_{i0}, $i = [1, P]$ obtained from a structural database. For each native configuration of the complex, we generate D non-native configurations (decoys) by applying rigid transformations to the ligand and obtain C_{ij} decoys with $j \in [1, D]$, where the first index indicates the protein-ligand complex and the second index indicates the generated decoys. Thus, for each complex we have $D + 1$ conformations, 1 native and D non-native. Our aim is to find a *scoring functional* E such that the following inequalities hold,

$$E(C_{i0}) < E(C_{ij}), \quad \forall i \in [1, P], \quad \forall j \in [1, D]. \tag{2.1}$$

This is a difficult problem in such a general formulation. In order to solve it, we need to make some simplifications. Thus, we represent the protein-ligand complex as a set of atoms, which are split into a finite number of types. This results in a total of $M_1 \times M_2$ pairs of different interactions, with M_1 being the total number of protein atom types, and M_2 – the total number of ligand atom types. Then, we assume that E depends only on the distribution of the distances between the atoms, with one atom located on the protein and the other on the ligand. We also assume these interactions to be short-ranged, which can be neglected if the distance between two interaction atoms is larger than a certain cutoff distance r_{max} that we set to 10 Å, as it gave good results in our earlier experiments [237]. Finally, we assume that E is a linear functional of the following form,

$$E(n(r)) = \sum_{k=1}^{M_1} \sum_{l=1}^{M_2} \int_0^{r_{max}} n^{kl}(r) f^{kl}(r)\, dr, \tag{2.2}$$

where $n^{kl}(r)$ are the *number densities of atom-atom pairs* at a distance r with the first atom of type k located on the protein, and the second atom of type l located on the ligand, and $f^{kl}(r)$ are the unknown *interaction potentials* between the atoms of types k and l. In our method, we use the

following functional form for the number densities $n^{kl}(r)$,

$$n^{kl}(r) = \sum_{ij} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r-r_{ij})^2}{2\sigma^2}}, \tag{2.3}$$

where each distance distribution is represented with a Gaussian function centered at r_{ij} with the standard deviation σ of 0.4 Å. This value was fixed and adapted from our previous studies [46]. The sum is taken over all pairs of atoms i of type k and j of type l separated by the distance r_{ij} smaller than r_{max}, with atom i located on the protein molecule and atom i located on the ligand molecule. Eq. 2.2 is very similar to the standard widely-used scoring formula, where individual protein-ligand distance-dependent interactions are summed up. Indeed, we can re-write functional E in the canonical way,

$$E = \sum_{ij} \sum_{k=1}^{M_1} \sum_{l=1}^{M_2} u^{kl}(r_{ij}), \tag{2.4}$$

with individual protein-ligand interactions $u^{kl}(r_{ij})$ given as a *convolution* of the interaction potentials $f^{kl}(r)$ with the Gaussians,

$$u^{kl}(r_{ij}) = \int_0^{r_{max}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r-r_{ij})^2}{2\sigma^2}} f^{kl}(r)\ dr. \tag{2.5}$$

In order to determine the unknown potentials $f^{kl}(r)$, we decompose them along with the number densities $n^{kl}(r)$ in a *polynomial basis*,

$$\begin{aligned} f^{kl}(r) &= \sum_{q=0}^{\infty} w_q^{kl} \psi_q(r), \quad r \in [0; r_{max}] \\ n^{kl}(r) &= \sum_{q=0}^{\infty} x_q^{kl} \psi_q(r), \quad r \in [0; r_{max}], \end{aligned} \tag{2.6}$$

where $\psi_q(r)$ are the basis functions orthogonal on $[0; r_{max}]$, and w_q^{kl} and x_q^{kl} are the expansion coefficients of $f^{kl}(r)$ and $n^{kl}(r)$, respectively. The orthogonality of the basis function implies that the following identity holds,

$$\int_0^{r_{max}} \psi_i(r)\psi_j(r)\Omega(r)\ dr = \delta_{ij}, \quad r \in [0; r_{max}], \tag{2.7}$$

where $\Omega(x)$ is a non-negative weight function with the support on $[0, r_{max}]$, and δ_{ij} is the Kronecker delta function. Without loss of generality, we can assume that the basis functions are always scaled in such a way that the weight function is unity. For this study, we have chosen simplistic rectangular basis functions. Thanks to the orthogonal basis functions, expansion coefficients w_q^{kl} and x_q^{kl} can

be determined from the orthogonality condition (2.7) as

$$\begin{aligned} w_q^{kl} &= \int_0^{r_{max}} f_{kl}(r)\psi_q(r)\ dr \\ x_q^{kl} &= \int_0^{r_{max}} n_{kl}(r)\psi_q(r)\ dr. \end{aligned} \tag{2.8}$$

Using expansions (2.6), the functional E can be rewritten as

$$E(n(r)) = \sum_{k=1}^{M_1}\sum_{l=1}^{M_2}\sum_{pq}^{\infty} w_q^{kl}x_p^{kl}\int_0^{r_{max}} \psi_q(r)\psi_p(r)\ dr. \tag{2.9}$$

Finally, to have a compact representation, and thanks to the orthogonality of the basis functions, the scoring functional E can be truncated up to the order Q as

$$E(n(r)) \approx \sum_{k=1}^{M_1}\sum_{l=1}^{M_2}\sum_{q=0}^{Q} w_q^{kl}x_q^{kl} = (\mathbf{w}\cdot\mathbf{x}), \qquad \mathbf{w},\mathbf{x} \in \mathbb{R}^{Q\times M_1\times M_2}. \tag{2.10}$$

We will refer to the vector $\mathbf{w}$ as to the *scoring vector*, whose value is to be determined, and to the vector $\mathbf{x}$ as to the *structure vector* that is computed from the structural data. We should note that vector $\mathbf{w}$ defines interatomic potentials $u^{kl}(r)$ for protein-ligand interactions. To conclude, Eqs. 2.8 provide a projection from a 3D structure into the *scoring space* on $\mathbb{R}^{Q\times M_1\times M_2}$, while Eq. 2.10 defines the scoring functional in this space.

2.1.2 Geometric interpretation

Using the expansion of the scoring functional E provided by Eq. 2.10, we can reformulate the scoring problem 2.1: given P native structure vectors $\mathbf{x}_i^{nat}$ and $P\times D$ nonnative structure vectors $\mathbf{x}_{ij}^{nonnat}$, find such a scoring vector $\mathbf{w}\in\mathbb{R}^{Q\times M_1\times M_2}$ that

$$\forall i = 1...P,\ \ \forall j = 1...D \quad (\mathbf{x}_i^{nat}\cdot\mathbf{w}) < (\mathbf{x}_{ij}^{nonnat}\cdot\mathbf{w}), \tag{2.11}$$

or, equivalently,

$$\forall i = 1...P,\ \ \forall j = 1...D \quad ([\mathbf{x}_{ij}^{nonnat} - \mathbf{x}_i^{nat}]\cdot\mathbf{w}) > 0, \tag{2.12}$$

which is a set of $P\times D$ *half-space equations* in $\mathbb{R}^{Q\times M_1\times M_2}$ with P parallel separation hyperplanes defined by the common normal $\mathbf{w}$. Figure 2.1.3 schematically shows three groups of structure vectors separated by three parallel hyperplanes with a common normal $\mathbf{w}$.
The set of inequalities (2.12) can have zero, one or infinite number of solutions [238]. Generally, this is an ill-posed problem. To obtain a single solution, we rewrite it as a *soft-margin quadratic optimization problem* [239] with an additional quadratic *regularization* term,

$$\begin{array}{ll} \text{Minimize (in } \mathbf{w},\ b_i,\ \xi_{ij}\text{):} & \frac{1}{2}\mathbf{w}\cdot\mathbf{w} + \sum_{ij} C_{ij}\xi_{ij} \\ \text{Subject to:} & y_{ij}\left[\mathbf{w}\cdot\mathbf{x}_{ij} + b_i\right] - 1 + \xi_{ij} \geq 0,\ i = 1..P,\ j = 0..D \\ & \xi_{ij} \geq 0 \end{array} \tag{2.13}$$

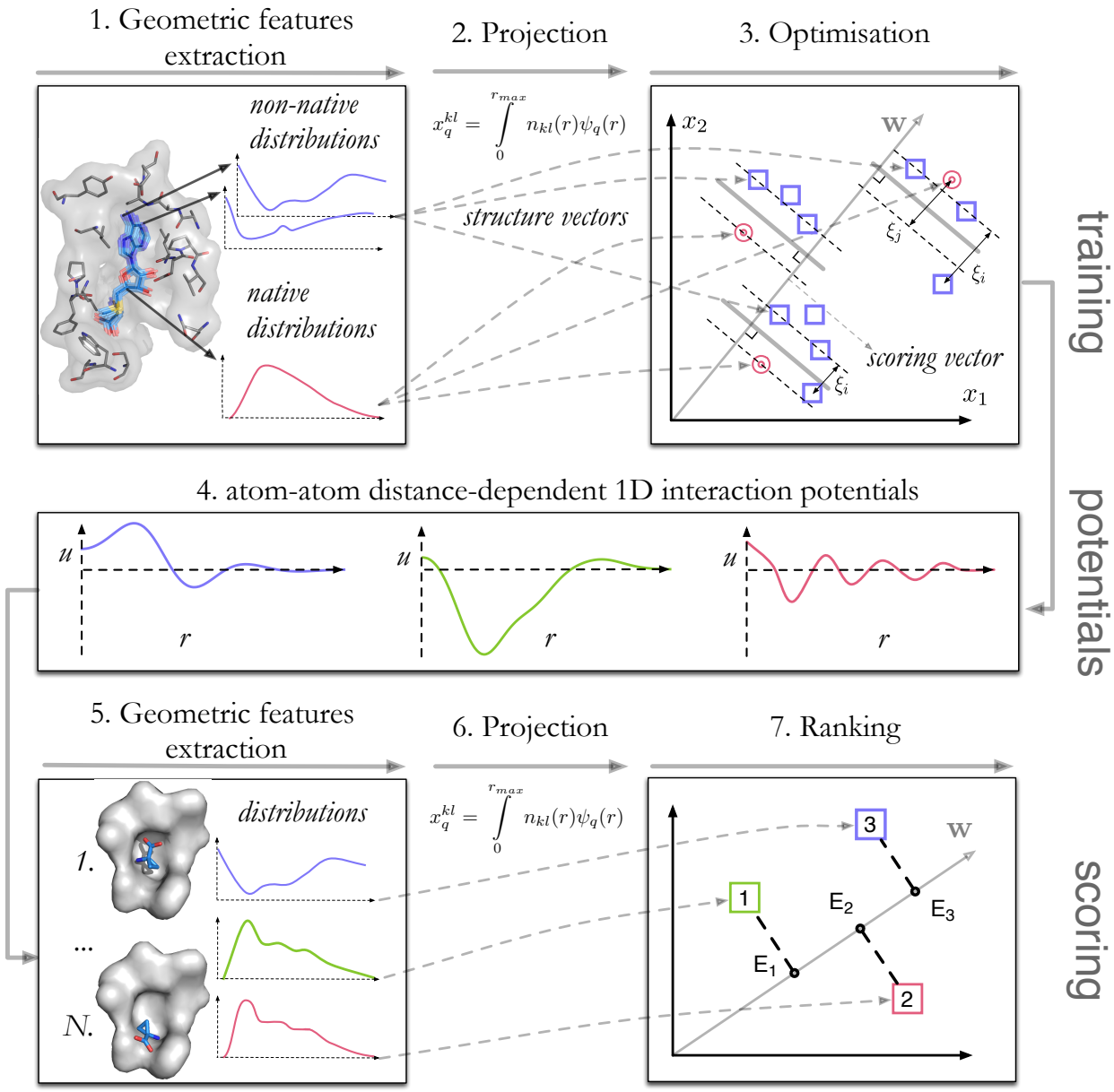


Figure 2.1: Schematic representation of the training stage (top) and the scoring stage (bottom) of the Convex-PL method.

Here, structure vectors $\mathbf{x}_{ij}$ are the same as in the above inequalities (2.11)-(2.12), index i runs over different protein-ligand complexes and index j runs over different conformations of the i-th protein-ligand complex. Particularly, protein conformations with $j = 0$ are native with the corresponding classifier $y_{i0} = +1$ and protein conformations with $j = 1..D$ are the decoys with the corresponding classifier $y_{ij} = -1$. Parameters C_{ij} can be regarded as regularization parameters, which control the importance of different structure vectors. To reduce the amount of adjustable parameters C_{ij} to a single regularization parameter C, we set values of C_{i0} for the native structure vectors to C, and the values of $C_{i1..D}$ for the non-native structure vectors to C/D. Then, we found the optimal value of the C parameter using the holdout cross-validation procedure [240]. The scoring vector $\mathbf{w}$, the offset vector $\mathbf{b}$ and the slack variables ξ_{ij} are the parameters to be optimized. Figure 2.1 shows a schematic workflow of the training and scoring stages of the presented method. We should note that the formulation 2.13 is, of course, not unique. We have also tried other regularization terms and other loss functions, but finally chose the quadratic regularization and the hinge-loss

misclassification penalties.

2.1.3 Optimization algorithm

Solutions and properties of quadratic optimization problems similar to the one stated above (2.13) have been extensively studied in theory of convex quadratic programming (QP) [241, 238]. Using the notion of *Lagrangian*, the optimization problem (2.13) can be converted into its dual form, which is a convex QP problem with the objective function solely depending on a set of *Lagrange multipliers* λ_{ij},

$$\begin{array}{ll} \text{Maximize in } \lambda_{ij}: & \mathcal{L}(\lambda_{ij}) = \sum_{ij} \lambda_{ij} - \frac{1}{2} \sum_{ij} \sum_{kl} y_{ij} y_{kl} \lambda_{ij} \lambda_{kl} \mathbf{x}_{ij} \cdot \mathbf{x}_{kl} \\ \text{Subject to:} & 0 \leq \lambda_{ij} \leq C_{ij} \\ & \sum_j y_{ij} \lambda_{ij} = 0, \quad \forall i \end{array} \tag{2.14}$$

The dual representation (2.14) of the original primal QP problem (2.13) allows us to break it into a series of smaller sub-problems. Here, we employ a *block-decomposition technique* and analytically iteratively maximize the Lagrangian with respect to pairs of multipliers according to *sequential minimal optimization* (SMO) algorithm [242]. Each block corresponds to one protein-ligand complex and its decoys.

Vectors $\mathbf{x}_{ij}$ for which $\lambda_{ij} > 0$ are called *support vectors*. Once the problem (2.14) is solved and the optimal Lagrange multipliers λ_{ij} are found, we can express the optimal solution of the original primal problem (2.13) (the scoring vector) as a linear combination of the support vectors,

$$\mathbf{w} = \sum_{\text{support vectors}} y_{ij} \lambda_{ij} \mathbf{x}_{ij}. \tag{2.15}$$

The initial values of our potentials were set to zero and no inverse Boltzmann statistics was used during the optimization.

2.1.4 Atom types

Convex-PL describes the ligand and protein atoms with 41 and 23 types, respectively. To make the typization of a ligand, I used our Knodle (KNowledge-Driven Ligand Extractor) library [243] that was developed as a part of my Master project at MIPT. The initial version of our potential described in my Master thesis contained 52 ligand atom types. In our participation in the first D3R Grand Challenge [237, 185], we used the version with 48 types. However, we later realized that these numbers are too large for the current training set, as the potentials for some specific types contained oscillations at large interaction distances. I merged them with more frequently occurring types and chose the typization with 41 atom types that provided the best cross-validation success rates on the control set. These are 8 carbon types, 14 nitrogen types, 7 oxygen types, 3 sulphur types, 2 phosphorus types, and 7 types describing halogens. These types are listed in Table A.5.

For the proteins I use a smaller typization set consisting of 23 atom types. Since hydrogens are rarely resolved experimentally and their inclusion would increase the dimensionality of the optimization problem, I did not include explicit hydrogens. Figure 2.2 shows the matrix of numbers of pairwise contacts between these 23 protein and 41 ligand atom types computed for the training set. As it can be expected, protein types that correspond to the protein backbone and carbon atoms are very

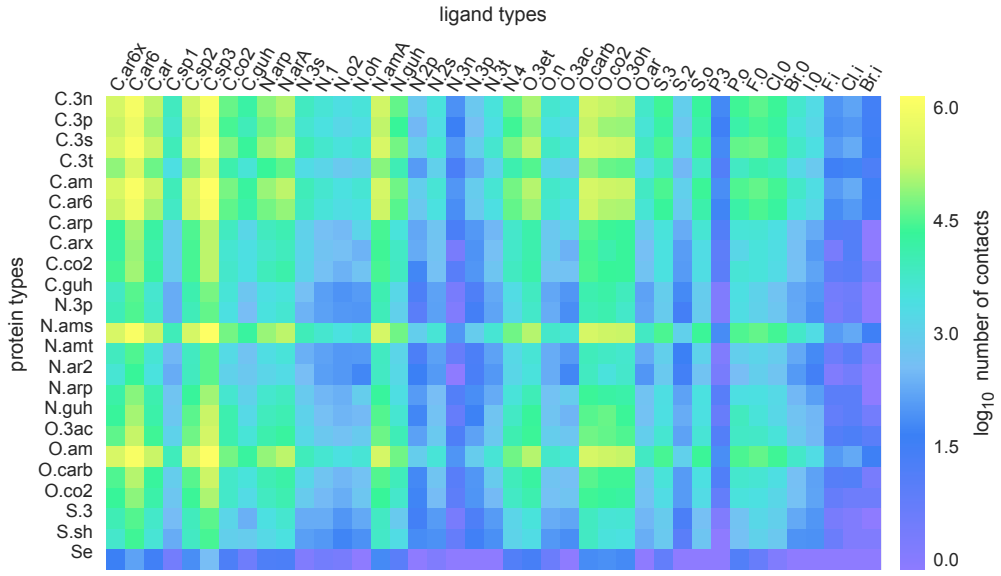


Figure 2.2: Numbers of pairwise contacts between the protein and the ligand atom types, as found in the training set. All numbers are shown in a log-10 scale.

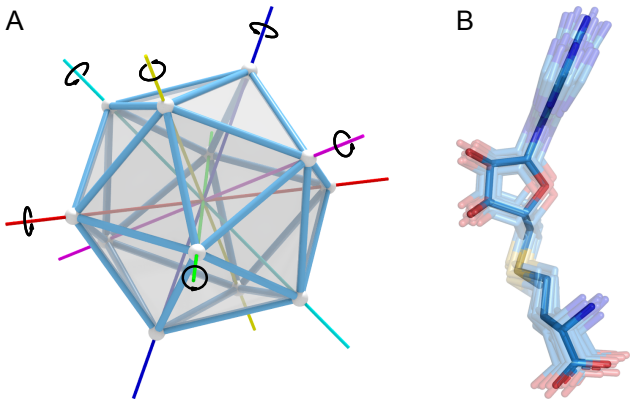


Figure 2.3: Decoys generation procedure. A : Six icosahedral axes about which we rotate the ligand. B : An example of a native ligand configuration with the corresponding 18 decoys generated with RMSD of 0.5 Å. These are 12 rotational decoys and 6 translational decoys.

frequent. However, we definitely lack statistical data for the selene atoms that occur in modified protein residues. The rarest ligand atom types are those that correspond to ionic halogens, one of the phosphorus types and one of the nitrogen types.

2.1.5 Training

Convex-PL was trained on the "general" set of the PDBBind 2015 database [244, 33] and validated on the two CASF benchmarks, benchmarks derived from blind docking challenges, and the blind docking challenges [237, 185, 207, 228], which is described in more detail below. Complexes from the test sets were excluded from the training set.

To construct the training set, I used randomly chosen 80% of protein-ligand complexes from the "general set" of PDBBind release 2015, excluding 195 complexes that intersect with the "core set" of the same database, as the "core set" forms the CASF 2013 benchmark. This resulted in 9,372 structures in the training set. The remaining 20% of protein-ligand complexes from the "general set" of the PDBBind release 2015, excluding 195 complexes from CASF 2013, formed the "control" set. I used this set to only adjust free parameters in our prediction model during its training. Both

the control and training sets included decoys generated by rigid-body rotations and translations that keep a constant RMSD value. After cross-validation that was done as a part of my Master thesis, I chose a value of 0.5 Å. Figure 2.3 illustrates the decoy generation procedure.

2.1.6 Repulsion term

The native complexes in the training set and also the generated decoys did not have atom-atom statistics at short interatomic distances. Consequently, the optimization procedure, or more precisely, its regularization term, left zero values of the scoring vector $\mathbf{w}$ and potentials $u^{kl}(r)$ at these distances, typically within two or three angstroms. To use our scoring function with the structures that contain atomic clashes, I manually filled these regions of $u^{kl}(r)$ potentials with artificial barriers of $\nu^{kl} r^{-2}$ shape. I adjusted the fitting coefficients ν^{kl} for each potential $u^{kl}(r)$ to match the first maximum of the $u^{kl}(r)$ curves. Also, to represent a soft repulsion at a zero separation distance, I replaced the barriers $\nu^{kl} r^{-2}$ at distances [0 Å, 0.4 Å] with a linear function, such that their values and the first derivatives match at a distance of 0.4 Å.

2.2 Results

2.2.1 Obtained potentials

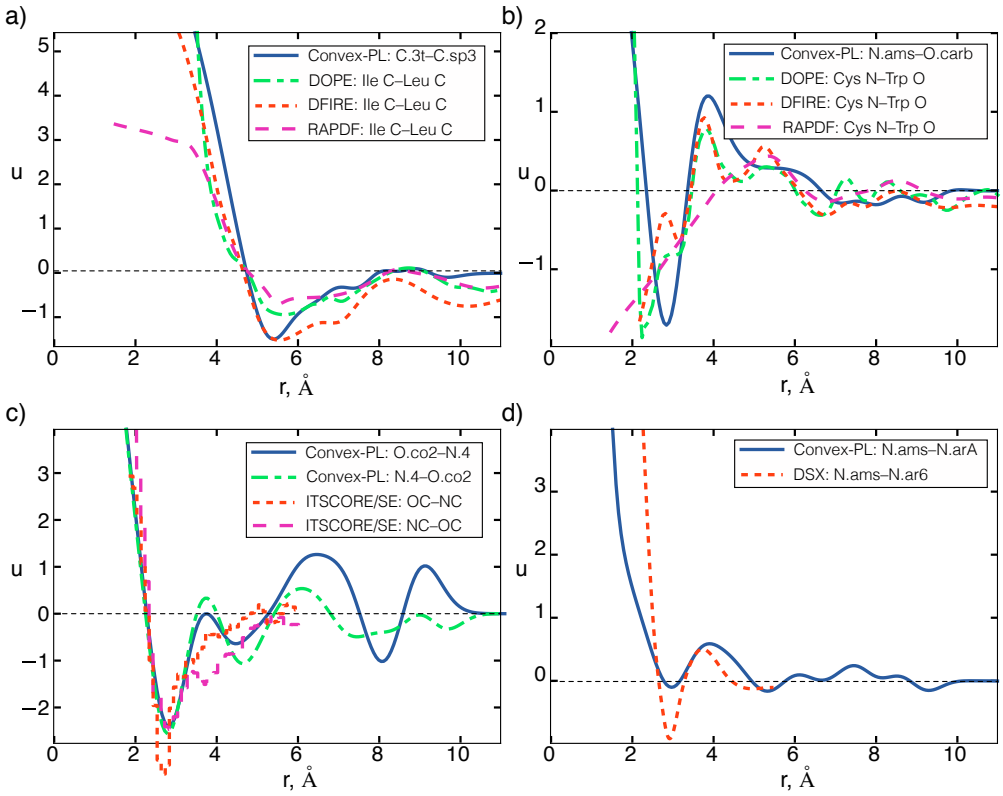


Figure 2.4: Comparison of the Convex-PL potentials with DOPE, DFIRE and RAPDF potentials for protein-protein interactions (a-b), with ITSCORE/SE (c) and with DSX (d) potentials for protein-ligand interactions. For the protein-ligand potentials, the two atom types in a pair correspond to the protein and ligand atoms, respectively. All the reference potentials' values were adapted from the plots found in the literature. The following interactions are plotted, a) sp3 carbon with sp3 carbon; b) secondary amide nitrogen with amide oxygen; c) negatively charged oxygen with positively charged nitrogen and vice versa; d) secondary amide nitrogen with aromatic nitrogen.

Starting from the initial vectors filled with zeros, after the optimization I obtained the scoring vector $\mathbf{w}$, which I also converted into 23×41 interatomic potentials $u^{kl}(r)$ for the sake of comparison with other methods. Some of these are shown in Fig. 2.4. In Figure 2.4 I plot our $u^{kl}(r)$ potentials together with several other potentials for relatively similar protein-protein and protein-ligand atom types that I found in the literature for DOPE [245], DFIRE [246], RAPDF [247], ITSCORE/SE [52] and DSX [45] scoring functions. Overall, it can be seen from this figure that Convex-PL predicts the first minimum and maximum peaks at similar locations compared to the other potentials. The difference between the first energy minimum locations in Fig. 2.4B may be caused by the fact that for protein-ligand interactions these interatomic separation distances are larger compared to similar protein-protein interactions. In Figure 2.4C we can also see two peaks for the Convex-PL potential with a protonated nitrogen on the ligand molecule that look unphysical. One of the possible explanations for this behavior is the difficulty of the correct type assignment for this type (N.4) on the ligand atoms. More precisely, this atom type can be easily mixed up with a sp3 nitrogen type, as the precise type assignment requires the presence of hydrogen atoms in the structure, while the corresponding lysine nitrogen on the protein molecule is considered to be protonated by default.

2.2.2 Assessment

Convex-PL demonstrated impressive pose prediction abilities, and was top-ranked in the CASF docking tests. For example, its success rates in CASF-2013 were 88.7%, 92.3% and 93.3% when predicting top-1, top-2 and top-3 poses, correspondingly, within RMSD of 2 Å. These results are slightly better than those reported in [55], as I have updated the clash term after the publication. The PDBBind "general set" contains a number of proteins homologous or even identical to those constituting the "core set", from which the CASF 2013 benchmark is constructed. To be confident that the presence of these proteins does not cause overfitting, we performed an accurate and computationally expensive leave-one-out cross-validation analysis. More precisely, for each of the 65 proteins from the "core set", I generated its own reduced dataset consisting of the "general set" without the protein's homologues. To do this, for each of the 65 proteins I detected its homologues in the "general set" using the 80% sequence identity criterion as computed by the BLASTP program of the BLAST+ package [248, 249]. After generating the 65 datasets, I divided them into the training and the control parts in the same manner as it was described above and ran 65 individual optimization processes. The resulting scoring functions were assessed on the docking test. All of these came to the same results as before, producing errors on a set of complexes that remains constant regardless the proteins excluded from the training set. Therefore, we can state that our scoring function is unbiased with respect to the proteins used in the CASF-2013 benchmark.
Convex-PL also performed very well in pose prediction in the benchmarks constructed from the D3R Grand Challenges 2 and 4. The results of Convex-PL's evaluation on these and the other tests from the CASF-2013 and CASF-2016, DUD, and D3R-based benchmarks can be seen in Figures 4.5, 4.6, 4.8, 4.9, and in Tables 4.2, A.11– A.13, A.23–A.22, A.28. These figures demonstrate less encouraging performance in the scoring and especially screening tests. Moreover, screening tests revealed the fact that Convex-PL is highly biased to the number of protein-ligand contacts and tighter interfaces. The statistics over the CASF screening test results, computed for Convex-PL similarly to the statistics reported in introduction, clearly illustrates this bias, as shown in Figure 2.5.

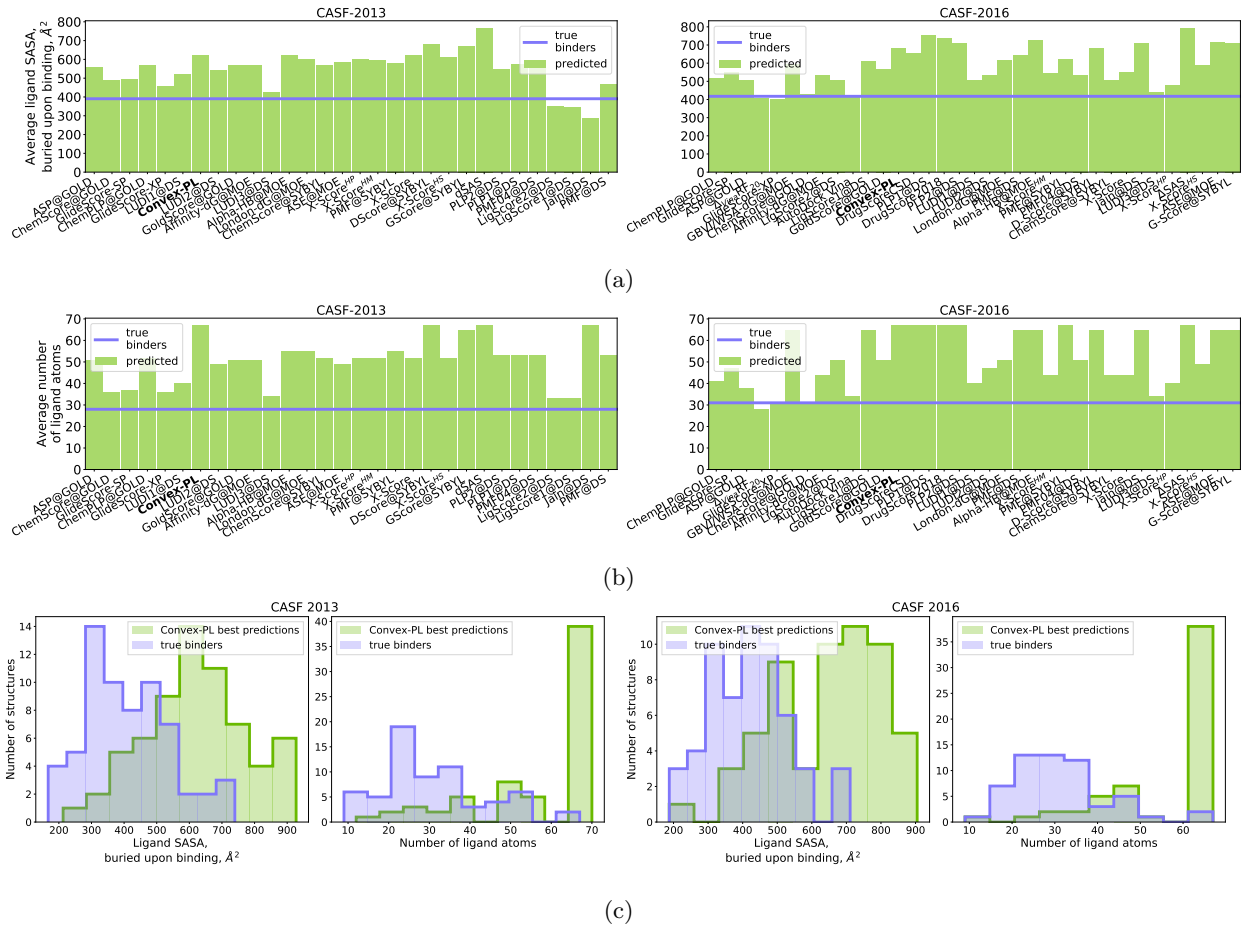


Figure 2.5: Statistics for the CASF 2013 and 2016 benchmarks, including the results computed for Convex-PL. (a)-(b) The purple line represents the average values of buried SASAs and numbers of atoms computed for ligands that natively bind target proteins and should have been predicted as the most affine binders[1]. The green boxes correspond to decoys that were top-ranked by scoring functions assessed on the virtual screening test from the CASF benchmarks 2013 and 2016. The scoring functions are sorted by the ability to predict the highest affinity binder in the 5% of the top-ranked decoys. (c) Histograms of average ligand buried SASA and number of atoms computed for the native "truly binding" ligands (purple) and decoy poses, top-ranked by Convex-PL (green) in the virtual screening tests from the CASF-2013 and CASF-2016 benchmarks. SASA values were computed with PyMOL's[96] *get_area()* function with *dot_solvent* set to 3.
[1] Or be among the most affine binders in several cases when the target protein was known to bind ligands with higher affinity but without co-crystal structure.

2.3 Conclusions

This section introduced Convex-PL – a knowledge-based distance-dependent scoring function that I started developing during my Master studies. During the training process, we do not impose any functional form of the scoring function. Instead, we decompose it into a polynomial basis and deduce the expansion coefficients from the knowledgebase by solving an optimization problem designed to classify native and non-native ligand poses. The decoys for the training set are generated with constant RMSD rigid-body deformations, making the obtained scoring function unbiased with respect to methods for the docking pose generation.

Convex-PL demonstrates very good results in pose prediction on the CASF benchmarks and in the pose prediction tests comprised of the D3R Grand Challenges 2 and 4 user submissions. However, its virtual screening abilities are rather average. One of the main reasons for such performance is the bias towards the number of protein-ligand contacts, from which Convex-PL definitely suffers.

Chapter 3. Development of Convex-PL modifications

The bias towards tighter protein-ligand interfaces that I have discovered for Convex-PL motivated us to develop a better model of protein-ligand interactions. Indeed, multiple even slightly favourable contacts of a huge ligand will sum up to lower energy compared to a smaller ligand with a small number of contacts. In reality, the bigger number of intermolecular contacts of bigger molecules is balanced by the interactions with solvent molecules and entropic contributions. We decided to revisit the derivation of empirical scoring functions and rework Convex-PL by supplementing it with several physics-inspired terms that aim to penalize the bias of large interfaces. This newer version of Convex-PL, trained with the *regression* model, is called Convex-PL^R. The development and assessment of Convex-PL^R were done by me, while the derivation of Convex-PL^R in terms of statistical physics was our joint effort with my scientific adviser. We have recently submitted a paper describing Convex-PL^R to the Journal of Chemical Information and Modeling [250].

3.1 Methods

3.1.1 Model of interactions

Following the protein-ligand binding free energy derivation in the canonical ensemble suggested in [1], the *standard binding free energy* of a protein A and a ligand molecule B in a solvent can be written as

$$\Delta G^o_{AB} = -RT \ln\left(\frac{C^o}{8\pi^2}\frac{\sigma_A\sigma_B}{\sigma_{AB}}\frac{Z_{N,AB}Z_{N,0}}{Z_{N,A}Z_{N,B}}\right) + P^o\Delta V, \tag{3.1}$$

where C^o and P^o are the standard concentration and pressure, σ_A, σ_B, σ_{AB} are the symmetry numbers of each molecule, $Z_{N,AB}$, $Z_{N,A}$, $Z_{N,B}$ are the configurational integrals of the protein-ligand complex, ligand, and protein in a solvent, respectively, $Z_{N,0}$ is a configurational integral of this solvent containing N atoms, and ΔV is a solute volume change upon binding. The integration in the partition functions is taken over the internal coordinates $\mathbf{r}_A$, $\mathbf{r}_B$, $\mathbf{r}_S$ of the receptor, ligand and solvent, respectively, and six external coordinates ζ_B of the ligand molecule B that are defined relative to the protein molecule A.

Our general aim is to approximate the configurational integrals ratios *using only a single conformation* of the protein-ligand complex instead of rigorously sampling the conformational ensemble. Firstly, we suppose that the ratio $\frac{Z_{N,AB}}{Z_{N,0}}$ of the complex and the solvent configurational integrals can be approximated with the original Convex-PL knowledge-based potential [55], which can be represented as a dot product between a *structure* vector $\mathbf{x}$ and a *scoring* vector $\mathbf{w}$,

$$E_{Convex-PL} = \sum_{k}^{M_1}\sum_{l}^{M_2}(\mathbf{x}_{kl}\cdot\mathbf{w}_{kl})\ . \tag{3.2}$$

It is important to note, however, that the given form of the Convex-PL function implicitly contains some additional interactions, especially the hydrophobic ones associated with the solvent entropy change.

The ligand configurational integral ratio can be rewritten as follows:

$$\frac{Z_{N,B}}{Z_{N,0}} = \frac{\int e^{-\beta(U(\mathbf{r}_B)+U(\mathbf{r}_B,\mathbf{r}_S)+U(\mathbf{r}_S))} d\mathbf{r}_B d\mathbf{r}_S}{\int e^{-\beta U(\mathbf{r}_S)} d\mathbf{r}_S} = \int e^{-\beta U(\mathbf{r}_B)} e^{-\beta W_B} d\mathbf{r}_B, \tag{3.3}$$

where $U(\mathbf{r}_B)$, $U(\mathbf{r}_S)$, $U(\mathbf{r}_B, \mathbf{r}_S)$ are the potential energy of the ligand, solvent, and the interactions between the ligand and solvent, respectively, and the integration is taken over all internal coordinates of the ligand B and solvent S. Also, $W(B)$ is the ligand *solvation energy* expressed as

$$W(B) = \frac{\int e^{-\beta(U(\mathbf{r}_B,\mathbf{r}_S)+U(\mathbf{r}_S))} d\mathbf{r}_S}{\int e^{-\beta U(\mathbf{r}_S)} d\mathbf{r}_S}. \tag{3.4}$$

We assume that for *sufficiently small ligands* we can neglect the intra-ligand interactions, $U(\mathbf{r}_B) = 0$, over the sampled ligand conformations $d\mathbf{r}_B$. We also presume that in this case, W_B is *almost constant* in the sampled volume. Thus, we obtain the following expression,

$$\ln \frac{Z_{N,B}}{Z_{N,0}} = -\beta W_B - \ln \int d\mathbf{r}_B. \tag{3.5}$$

We approximate the first term with a set of descriptors containing solvent-accessible surface areas of atoms, and grid-based descriptors representing the displaced solvent volume that are described in more detail in the section below. The second term corresponds to the volume of the ligand conformational space, which we approximate with the logarithm of the number of ligand conformational states that can be adopted by rotations about *rotatable bonds*,

$$\ln \int d\mathbf{r}_B \equiv S_{conf,B} \approx \ln(\prod_{i}^{\#\ bonds_B} w_i). \tag{3.6}$$

Here, the product is taken over the ligand rotatable bonds, and the weights $w_i = 4 - b_i$ are computed using ith bond order b_i. The conformational symmetry is partially taken into account by not counting bonds with the terminal atoms.
A similar procedure can be carried out for the receptor configurational integral, resulting in

$$\ln \frac{Z_{N,A}}{Z_{N,0}} = -\beta W_A - \ln \int d\mathbf{r}_A. \tag{3.7}$$

Following [1], we can split the integration coordinates $d\mathbf{r}_A$ to the *interface pocket* and *rigid* parts of the protein and neglect the integral over the rigid one. The interface integral can then be taken over the rotameric states of the pocket residues [97, 251, 252]. The proportion of pocket residue's bonds with allowed rotations in the unbound state can be estimated by the fraction of its solvent-accessible surface area $s_{i,unbound}$ in the unbound state, and the total surface area of the same residue, if it is extracted from the receptor, $s_{i,total}$. Thus, we obtain

$$\ln \int d\mathbf{r}_A \equiv S_{conf,A} \approx \sum_{i}^{\#residues} \frac{s_{i,unbound}}{s_{i,single}} \ln(\prod_{j}^{\#\ bonds_A} w_{ij}), \tag{3.8}$$

where the sum is taken over all the residues at the protein-ligand interface, the product is taken over the rotatable bonds in each residue, and the weights w_{ij} are computed the same way as for

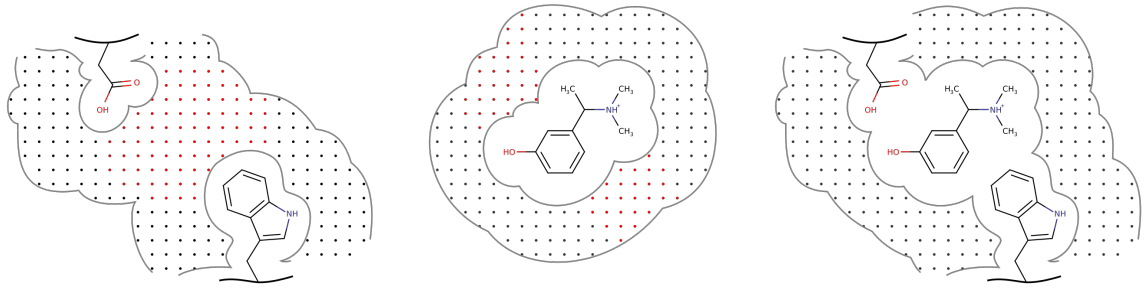
the $S_{conf,B}$. We should note that this expression is different from the one proposed by Sternberg and Chickos [252], which was estimated using the empirical scale of Pickett and Sternberg [253]. Their model assumes that solvent-accessible side chains (with relative SASA $> 60\%$) populate different rotamers, whereas buried side chains (with relative SASA $< 60\%$) are restricted to only one rotamer. Neglecting the displaced volume, constants, and some of the symmetry effects, we obtain our final equation,

$$\Delta G^o_{AB} \approx E_{Convex-PL} - W_A - W_B - S_{conf,A} - S_{conf,B}. \tag{3.9}$$

3.1.2 Interactions with the solvent

As I have mentioned above, Convex-PL implicitly counts some hydrophobic interactions in the corresponding pairwise potentials. To account for the underestimated interactions, I computed solvent-accessible surface areas of atoms of the polar, charged, and oriented types. I have tested two ways of calculating these descriptors using the POWERSASA library [254, 255] with a 1.4 Å probe atom radius. The first one was the computation of the usual ΔSASA buried upon binding. As an alternative, I computed SASAs only for those atoms that are considered to be interacting upon binding, i.e., are closer to each other than a cutoff distance that I set to 7 Å. The latter approach seems to penalize the preference for a larger number of contacts better.

For those solvation interactions that were possibly not taken into account by the combination of the Convex-PL term and SASA, we decided to use a grid representation of solvent following the ideas from the SBROD protein quality assessment scoring function [256]. We first constructed three grids corresponding to the ligand, protein, and the whole complex centered on the ligand molecule with the size equal to the ligand size plus a 5 Å padding. For each molecule of the ligand, protein, and the complex, we removed grid nodes intersecting with the molecule within the 1.4 Å margins. We then collected statistics of distance distributions between the grid points and the atoms of the complex, protein, and ligand of different types. We wrote it to the *structure vectors* $x^{sol}_{PL}, x^{sol}_{P}$, and x^{sol}_{L}, correspondingly, and then computed the feature representing the difference of the solvation geometry between the bound and unbound states, $x^{sol} = x^{sol}_{PL} - (x^{sol}_{P} + x^{sol}_{L})$. Figure 3.1 shows an example of the solvent grid for the 1gqs complex.



(a) (b) (c)

Figure 3.1: Schematic representation of solvent grids for a protein (a), ligand (b), and their complex (c). The difference between the complex grid and receptor and ligand grids is shown in red.

3.1.3 Regression model with solvent and entropic contributions

Convex-PL was initially trained to discriminate native and non-native ligand poses by solving an optimization problem, in which native and non-native poses within each protein-ligand complex are compared to each other to minimize the prediction loss. Comparing poses of *different* complexes to each other would be meaningless since the 'native' and 'non-native' class labels are relevant *inside* each complex only. This method is very efficient for reconstructing a scoring function for pose prediction. Its design, however, limits the ability to predict the absolute binding affinities. Nonetheless, we can circumvent this problem using regression towards available affinity data. Therefore, I combined the binding constants, the Convex-PL energy, solvent, SASA, and flexibility descriptors in a regression model, which, as expected, lead to a considerable improvement of the CASF benchmark screening test results. As we wanted to preserve physics as much as possible, I designed the Convex-PL^R score to take the following form:

$$\text{Convex-PL}^R = \text{Convex-PL}^{scaled} + \text{correction term}, \tag{3.10}$$

and then trained the correction term using known binding affinities data. Here, Convex-PL^{scaled} is the Convex-PL energy scaled to correspond to the absolute scale of the binding affinities. This functional form is also used in the $\Delta_{\text{Vina}}\text{RF}_{20}$ scoring function [80]. I have tried solving both linear and non-linear kernel and ensemble regression problems and settled upon the linear ridge regression model:

$$\min_{\hat{\mathbf{w}}} \quad \sum_i ||\hat{\mathbf{w}}^T \hat{\mathbf{x}}_i - s_i||^2 + \alpha ||\hat{\mathbf{w}}||^2, \tag{3.11}$$

as it is easily regularizable, fast to train, and interpretable. Here, s_i is the binding constant of a complex i minus the Convex-PL^{scaled} score, $\hat{\mathbf{x}}_i = [S_{conf,A}, S_{conf,B}$, SASA descriptors, solvent structure vector $x^{sol}]$ is the corresponding feature vector, α is the regularization constant, and $\hat{\mathbf{w}}$ is the unknown vector of linear regression coefficients.

3.1.4 Model training and its limitations

I have trained Convex-PL^R on a subset of the PDBBind 2016 [33] general dataset. I have only used complexes that had K_d or IC50 data, and excluded the K_i values since K_i binding constants themselves are correlated with the size of molecules stronger than the other binding constants did, while our goal was to penalize such a correlation. As a consequence, Convex-PL predictions also correlated with K_i better than with the other binding constants. I have also excluded complexes satisfying a number of criteria listed in Table A.8. These include outlying complexes with very high binding constants or very low Convex-PL predictions, and complexes, for which the binding constant was known only approximately ($\sim$, $>$, and $<$ signs in the PDBBind index file). The total size of the training subset was equal to 8,972 complexes. Complexes used in the two CASF benchmarks were excluded from the dataset.

Initially, the model from Eq. 3.11 turned out to predict binding affinities very well but produced poor results on the screening tests. Indeed, the prediction of binding constants for the native complexes that do bind is a different task from a more general task of a binding free energy prediction for a protein-ligand pair that may not even bind in reality. Choosing model hyperparameters

with cross-validation on the affinity data only refines the affinity prediction model. Therefore, for the validation, I used the virtual screening test from the CASF-2013 benchmark. The goal of the training procedure was not to get the best correlation with the binding constants but to keep the balance between scoring and screening tasks. Such problems can also be solved by multi-task learning [81, 54, 91].

To reduce the number of free parameters, for each atom type, I summed up all the radial distribution functions representing solvent, merging them to a single feature per protein or ligand atom type. I have also applied additional regularization to the solvent radial distribution function descriptors multiplying them by coefficients $0 \leq \gamma \leq 1$. We chose the values for these coefficients using grid search, while the regularization parameter α was found by random search. Tables 3.1 and A.9 list the descriptors. The original Convex-PL potential uses a 10 Å cutoff for the pairwise interactions between its 42 ligand, and 23 protein atom types. In the reworked Convex-PL^R model, I use another version of the original potential, Convex-PL$^{5.2A}$, with a 5.2 Å cutoff, 40 ligand and 23 protein atom types listed in Table A.6. I trained it on the PDBBind 2016 general dataset enriched with 3,225 complexes from the Protein Data Bank, which contain ligands with halogen and sulfonamide groups. Truncation of the cutoff radius alone already decreases the virtual screening test bias towards bigger ligands by reducing the variance and the total number of pairwise interactions, as it can be seen in subsection 3.2 Results.

name	number
Convex-PL$^{5.2A,scaled}$	1
ligand flexibility $S_{conf,B}$	1
receptor flexibility $S_{conf,A}$	1
non-apolar and aromatic ligand SASA	1
non-apolar and aromatic protein SASA	1
sum of solvent radial distribution function bins for each protein and ligand atom type	63

Table 3.1: Descriptors used in the Convex-PL^R scoring function. Table A.9 lists the regression coefficients.

3.1.5 Integration with AutoDock Vina

To be able to use Convex-PL directly for sampling of the docking poses, we created VinaCPL – a version of AutoDock Vina [70] with Convex-PL embedded as a scoring function. Since Convex-PL is pairwise and distance-dependent, it could be naturally mapped to the AutoDock Vina pairwise interaction grids. In more detail, I updated Vina's atom types to the 42 ligand and 23 protein types and replaced the AutoDock Vina scoring function with Convex-PL. I additionally penalized intra-ligand clashes using the values of the Convex-PL protein-ligand potential mapped to the ligand-ligand atom types. Our modifications of AutoDock Vina also increased the *num_saved_min* internal parameter and skipped the RMSD-based clustering, so that more conformations could be generated for further re-scoring dependent on the provided *num_modes* command-line argument. However, I were unable to do the same with Convex-PL^R, as the direct computation of the SASA and grid-based solvent features in the sampling algorithm would be relatively inefficient, as they cannot be precomputed on the rigid pairwise grids. In our most recent docking protocols I apply a round of Convex-PL or Convex-PL^R rescoring to the poses sampled with VinaCPL to avoid the intra-ligand clash term in the final scores.

3.2 Results

With all these modifications, I obtained sufficient performance improvements on all the CASF tests except the ranking one, as it can be seen in Figures 4.5, 4.6. I have also assessed Convex-PL on a subset of DUD benchmark, and on the user submissions from the D3R challenges, the results are listed in Figure 4.8 and Table 4.2.

3.2.1 CASF benchmarks

Scoring and ranking

Convex-PL^R obtains scores with the top-ranked correlation to the experimental binding constants, which are plotted in Figure 3.2. Below I discuss several outliers present in the two benchmarks.

Scoring outliers

The binding site of the 3kwa complex includes a zinc atom, which is not parametrized in Convex-PL. Moreover, the 3kwa ligand itself is an alkyl chain, flexible but symmetric, and thus we are perhaps overestimating the contribution of its conformational entropy change upon binding. The binding site of the 4dew complex is located on an interface between the two monomers. This assembly was constructed by applying a crystallographic symmetry operator to the monomeric asymmetric subunit. This resulted in unnaturally small distances (up to 2.5 Å) between the heavy atoms of the ligand aromatic ring and the lysine amino group of the reconstructed monomer, and, in the underestimation of the binding constant. The 1hfs complex consists of a large molecule that binds at the interface of two protein monomers and forms multiple contacts with them. As a result, Convex-PL^R overestimates its binding affinity. The 4gid ligand is a peptide and one more bulky molecule, whose energy we overestimate. On the contrary, the 1zea complex, whose ligand is also a large peptide and an outlier in our scoring predictions, suffers from an excessive energy penalty driven by the solvent-accessible surface area descriptor.

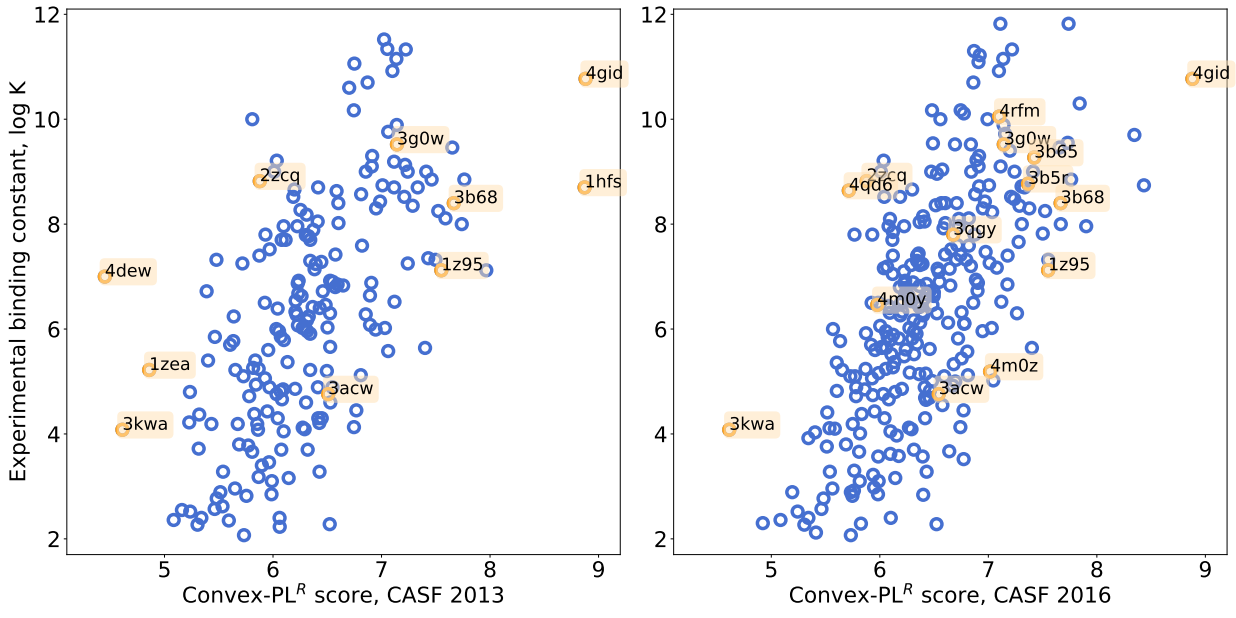


Figure 3.2: Convex-PL^R scores, obtained for the CASF benchmarks, versus experimental binding constants. PDB codes of the outliers and challenging complexes discussed in the text are highlighted in orange.

Both CASF-2013 and 2016 ranking test performance did not improve much upon supplementing Convex-PL with additional descriptors and even dropped in some cases. However, seven of eleven clusters of the CASF-2016 benchmark, in which Convex-PL^R made predictions with average Spearman correlation coefficients equal or less than 0.3, were on average wrongly predicted by the majority of other scoring functions assessed in the benchmark. Below I would like to discuss the reasons for the poor performance of Convex-PL^R on three clusters.

Ranking outliers

The 3g0w cluster that includes the 3b65, 3b5r, 3b68, and 1z95 complexes turned out to be one of the most complicated cases for Convex-PL. Here, each ligand contains halogen substituents, whose orientation-dependent interactions are incompletely described with the distance-dependent Convex-PL^R terms. The 3g0w ligand also comprises a nitrile group that could be underrepresented in the training set. Again, its π interactions with the protein ring [257] may not be fully described with the interactions deduced from pairwise distances. Overall, the binding affinity scores that we produced for this cluster are very close to each other, unlike the experimental binding constants, which spread to a broader range of values.

The 2zcq complex of the 2zcq cluster contains the magnesium ions interacting with the ligand. However, neither Convex-PL nor our additional descriptors are parametrized for the ligand-metal interactions, and, thus, we are overlooking the electrostatic interactions between the two magnesium ions and a phosphonosulfonate ligand headgroup [258]. Three other proteins of the cluster (4ea2, 2zcr, and 2zy1) are ranked correctly. Notably, all versions of Convex-PL and other top-performing scoring functions overscore the binding affinity of the 3acw complex, which I can hardly interpret.

An interesting case where both Convex-PL^R and the top-ranked scoring functions, including $\Delta_{\text{Vina}}\text{RF}_{20}$, show near-zero correlation is the 4rfm cluster consisting of the 4rfm, 4qd6, 3qgy, 4m0y, and 4m0z PDB structures corresponding to the interleukin-2 inducible T-cell kinase (ITK) inhibitors. Here, most of the scoring functions, including Convex-PL, disfavour the 4qd6 ligand, although its experimental binding affinity is high and should be ranked second. It seems that the structures of the complexes from the benchmark correspond to those marked as a biological assembly in the Protein Data Bank. However, the biological assembly deposited for 4qd6, along with the published analysis of its interactions with the ligand [259], does not include or describe the beta-sheet, which seems to be a part of domain swapping and is present in the original crystallographic structure. Conversely, the protein part that was involved in domain swapping as a beta-sheet in 4qd6 adopts another conformation in 4rfm, 3qgy, 4m0y, and 4m0z. In these complexes, it stops participating in domain swapping and forms a beta-sheet inside the main protein chain. As a result, interactions that contribute to the ligand-binding in 4rfm, 3qgy, and 4m0y are lost in 4qd6, because of the incomplete protein structure, leading to lower affinity predictions. The binding affinity score predicted with Convex-PL^R for the 4qd6 ligand in a complex with the two protein chains instead of one turned out to be closer to the experimental constant. The original 4m0z and 4m0y structures contain ligands that bind to the two sites of ITK [260], one of which was chosen per each protein for the benchmark creation. For some reason, Convex-PL and other scoring functions overestimate the affinity of the ligand binding to the allosteric site of the 4m0z structure. I have also noticed that the binding affinity provided in the benchmark for 4m0y corresponds to the allosteric pocket binding [260], while the ligand

chosen for the benchmark binds the ATP pocket. The correct choice of the 4m0y affinity also increases the ranking and scoring test performance of Convex-PL^R, as well as of several other scoring functions. This case particularly illustrates the vulnerability of binding energy prediction models to the interpretation of data in various steps of data collection and processing. These include uncertain information in the papers with multiple binding constants coming from different experiments, unreliable biological assembly information, incorrect interpretation of electron density of small molecules, etc.

Docking and screening tests

Convex-PL^R keeps and improves the high performance of Convex-PL in both CASF benchmarks. Improvement of the screening power was the central goal of modifications I applied to Convex-PL, and on these tests, we can clearly see the importance of additional descriptors. Convex-PL^R is top-ranked in all screening tests from the two benchmarks and outperforms other scoring functions in the enrichment prediction, as shown in Figures 4.5, 4.6. Overall, the inclusion of additional descriptors and a smaller value of pairwise interactions cutoff value finally allowed us to overcome the bias towards bigger ligands and tighter interfaces, as it is shown in Figure 3.3.

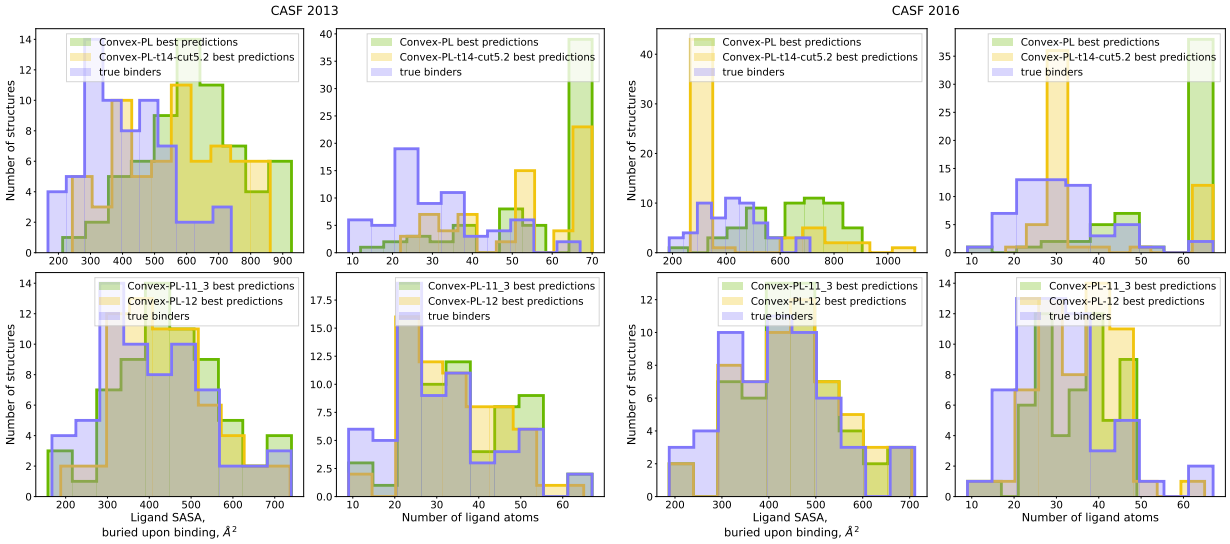


Figure 3.3: Histograms of average ligand buried SASA and number of atoms computed for the native "truly binding" ligands and decoy poses, top-ranked by different Convex-PL versions in the virtual screening tests from the CASF-2013 and CASF-2016 benchmarks.

3.2.2 D3R benchmarks

Figure 4.8 implies that Convex-PL^R shows similar results in pose prediction to those of Convex-PL. Both scoring functions demonstrate good performance in Grand Challenges 2 and 4, and quite low results in the pose prediction exercise of Grand Challenge 3. Their fail in the latter case may be related to the exposure of the wast binding pocket, and the fact that I do not take into account the explicit water molecules. None of the solvent-related features introduced in Colvex-PL^R aided in this case. Another reason may be related to a number of orientation-dependent interactions with halogen substituents and aromatic groups of the ligands used in the Challenge, that a distance-dependent potential may overlook. In the affinity prediction task, Convex-PL^R re-scoring yields

good results in the Grand Challenge 2, and, surprisingly, in the Grand Challenge 3. Notably, the correlation of binding constants from these challenges with buried ligand SASA values is stronger than in the case of the Grand Challenge 4, where Convex-PL^R fails.

3.2.3 DUD

To additionally measure the virtual screening performance of Convex-PL^R, I assessed it using the Directory of Useful Decoys (DUD) [165] dataset specifically designed for virtual screening benchmarking. As it was already mentioned in Section 1.3.4, for the evaluation I selected a subset of 31 targets from DUD that did not have co-factors.

Docking protocol and results

For each of the 31 DUD targets, I ran docking simulations using VinaCPL with *num_modes*= 300, and then re-scored the obtained docking poses with Convex-PL^R. The final score of each compound was obtained by averaging the scores of the top-10 Convex-PL^R predictions. The same procedure was then repeated for the original Convex-PL to obtain the scores that do not include intra-ligand clashes. We have already successfully applied earlier versions of this protocol in the D3R Grand Challenges 2 [207] and 4 [228], that are described below in Section III.
In this test, I assessed the Convex-PL versions together with a few state-of-the-art ones, namely, AutoDock Vina and Smina [177] with Vinardo chosen as the scoring function. AutoDock Vina, Smina, and VinaCPL were launched with the exhaustiveness parameter set to 10, other parameters except for the number of output conformations were left to their defaults. For AutoDock Vina and Smina, the number of output conformations was set to the maximum value of 25. I have tested several ways to define the docking binding boxes. I achieved the best virtual screening results for all three protocols with the binding box determined with the following procedure : *(i)* Target co-crystal ligand dimensions, namely the co-crystal box, were measured. *(ii)* Dimensions of the input ligand 3D conformation provided in the dataset, namely the ligand box, were measured. *(iii)* The input ligand box was aligned with the co-crystal box. *(iv)* Finally, for each dimension, the maximum size of the two boxes was chosen and multiplied by an arbitrarily chosen factor of 1.6. Notably, virtual screening results of both Vina and Vinardo achieved with such a box outperformed those reported previously [74].
Figures 3.4, 4.9 and Table 4.2 summarize virtual screening powers evaluated by measuring the ROC AUC, 5% enrichment factor, and BEDROC$_{\alpha=20}$ metrics. Here, Convex-PL^R improves the Convex-PL performance and shows the state-of-the-art results, slightly outperforming both Vina and Vinardo in early enrichment metrics and having fewer cases with ROC area values below 0.5.

3.3 Conclusions

This section presented Convex-PL^R – a reworked Convex-PL protein-ligand potential, derived from thermodynamical considerations. To circumvent a bias towards bigger protein-ligand interfaces, I supplemented our model with conformational entropic terms for the ligand and the binding pocket sidechain flexibility, and also those that account for the change in solvation upon binding. The latter was represented using analytical computation of the solvent-accessible surface area, together with a grid-based representation of the solvent. I have also integrated the original Convex-PL

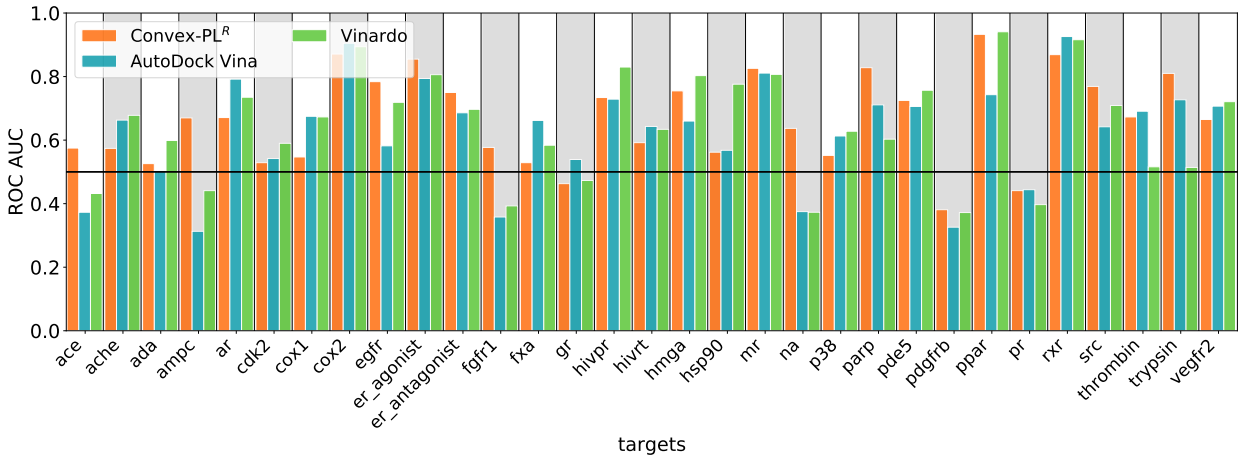


Figure 3.4: Per-target illustration of ROC AUC scores of virtual screening tests on the 31 targets from the DUD dataset for Convex-PL^R, AutoDock Vina, and Vinardo.

scoring function inside AutoDock Vina. Poses sampled with this method can then be re-scored with Convex-PL^R. I assessed Convex-PL^R on CASF 2013 and 2016, on the DUD virtual screening test, and the tests based on the D3R Challenges. It demonstrated the state-of-the-art results in all the tests except the pose prediction stage of D3R Grand Challenge 3 and the affinity prediction part of D3R Grand Challenge 4.

Chapter 4. KORP-PL

A common practice in protein-protein docking and protein structure quality assessment are the sidechain-free and orientation-dependent approaches [261, 262, 256, 263, 264, 265, 266, 267]. A number of orientation-dependent scoring functions also exist for protein-ligand docking [51, 45]. Being inspired by these methods, we combined our scoring functions development pipeline and the KORP [263] potential, which had been successfully used in protein and loop modeling. The success of KORP is rooted in the consideration of the full six-dimensional (6D) joint probability distribution function that only depends on the relative orientation between protein residues. For the protein-ligand interactions, we reduce the dependence of the pairwise potential to a 3D joint probability of observing an interacting ligand atom at a given relative position and orientation from a protein residue.

In collaboration with Pablo Chacón from Rocasolano Institute of Physical Chemistry, Spain, and Karina dos Santos Machado from Universidade Federal do Rio Grande, Brazil, we have developed a novel scoring function called KORP-PL. KORP-PL does not require sidechain atoms, and only three backbone atoms of the protein residue are needed. As a result, it is relatively fast, as each interaction requires only the computation of two spherical angles and a single distance. Despite its seeming simplicity, our approach yields state-of-the-art results.

My contribution in this collaboration was the participation in the theoretical formulation of the method, all programming to create KORP-PL on the basis of Convex-PL and KORP, and the majority of assessments. Validation on D3R, DUD, and DUD-E benchmarks was performed by me too. Assessments on the CASF benchmarks were done both by me and Karina Machado. Cross-validation to find the suitable parameters of the model's geometry was done by Karina Machado. These results were recently published in Bioinformatics [49].

4.1 Methods

4.1.1 Model of interactions

Our starting point was the 6D orientation-dependent knowledge-based potential KORP [263], which had been successfully used in protein and loop modeling. The main idea behind using the 6D statistics in KORP is that one can unambiguously define local coordinate frames for each of the protein residues. However, this model cannot be applied to small molecules owing to their higher chemical diversity. Therefore, we keep the reduced representation for the protein molecule, and use the all-atom representation for the ligand. A point in a 3D coordinate frame can be described with only 3 variables, unlike a relative orientation of one frame with respect to another, which requires 6 variables. This results in collecting the three-dimensional statistics for the involved interactions. As depicted in Figure 4.1, the relative position and orientation of a given interaction is specified by the ligand atom coordinates and by a 3D oriented frame (i.e. a local coordinate system) built from three backbone atoms C, C_α, and N of the protein residue. Therefore, only two spherical angles θ and φ and the distance r between the ligand atom and the center of the residue frame located at C_α are required. The interaction $\tilde{E}_{i,j}$ between a residue i and a ligand atom j is then derived from

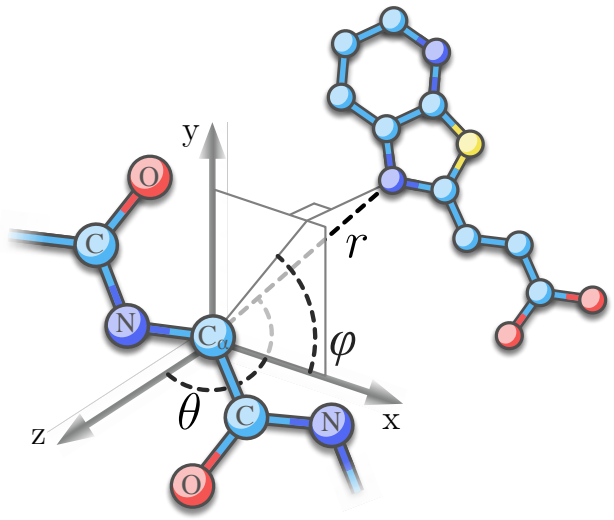


Figure 4.1: Schematic view of the relative orientation of a ligand molecule to a protein residue. The residue is represented with a 3D oriented frame built from three backbone atoms. The relative orientation of a ligand atom is described by two spherical angles, θ, the *polar angle* between the r and z vectors, and φ, the *azimuthal angle* between x and the projection of r into the xy plane.

the Boltzmann distribution,

$$\tilde{E}_{i,j}(\theta,\varphi,r) = -RT \ln \frac{P^{\text{obs}}_{aa,lig}(\theta,\varphi,r) + z}{P^{\text{ref}}(\theta,\varphi,r) + z}, \tag{4.1}$$

where RT is the Boltzmann factor, and $P^{\text{obs}}_{aa,lig}$ is the joint 3D probability of observing a protein amino acid i of a type aa and a ligand atom j of a type lig at a given distance r and orientation (θ,ψ) in a set of crystallographic structures. We should note that there is a full dependence between the three variables in $P^{\text{obs}}_{aa,lig}(\theta,\varphi,r)$. In order to counterbalance nonspecific residue-ligand interactions, we introduce the reference probability P^{ref} regardless of the type of interaction. It corresponds to the *reference state*, defined as an average distribution over the different amino acid and ligand types [247]. Also, we add the z constant to both the nominator and denominator of the above expression to prevent numerical instability for low-count statistics. Following the original KORP implementation, we zero-mean normalized individual contributions $\tilde{E}_{ij}(\theta,\varphi,r)$ at every distance to reduce distance-specific biases. More precisely, from each value of the interaction potential at (θ,φ,r) we subtracted an average taken at the same distance r over all θ and φ values,

$$E_{ij}(\theta,\varphi,r) = \tilde{E}_{ij}(\theta,\varphi,r) - \langle \tilde{E}_{ij}(\theta,\varphi,r) \rangle_{\theta,\varphi}. \tag{4.2}$$

The total protein-ligand interaction potential will be then the sum of all individual contributions E_k within a certain cutoff distance,

$$E = \sum_k E_k(\theta_k,\varphi_k,r_k). \tag{4.3}$$

The total number of protein residue types is equal to 20 and corresponds to the 20 standard amino acids. The set of 37 ligand atom types comprises 8 carbon types, 12 nitrogen types, 7 oxygen types, 4 sulfur types, 2 phosphorus types, and 4 types describing halogens (see Table A.7 for more details). Each ligand atom type is assigned using the Knodle library [243] in the same manner as

we did for the Convex-PL scoring function [55].

4.1.2 Training data

We derived KORP-PL using structures of protein-ligand complexes deposited in the PDBBind 2016 general dataset [268]. We excluded 373 structures intersecting with those from the CASF-2013 and CASF-2016 benchmarks. This resulted in 12,910 selected examples. Also, there were no intersections between PDBBind 2016 and examples from the D3R challenges that we use to compile our benchmark. We did not specifically preprocess the input structures. We did not remove homologous receptor structures since their bound ligands can be very diverse. In fact, previously we did not find any effect of excluding structures in the training set homologous to the ones in the test set on the prediction accuracy [55]. Nonetheless, we provide additional computational experiments excluding the test set structures from the training set at different levels of similarity. We collected the statistics using interactions within the range of radial distances r of (2 Å, 11 Å). This statistics were divided into 12 bins. The angular statistics were collected into 180 equiareal bins using a uniform angular sampling tessellation described elsewhere [269, 263].

4.1.3 Reweighing the potential for binding affinity predictions

Initial tests demonstrated rather poor performance of KORP-PL in affinity prediction exercises. This is mostly the result of the independence of E_{ij} terms from each other. Motivated by this observation, I devised a reweighing scheme to balance the contributions of each component E_{ij},

$$E^w = \sum_{i}^{N_{aa}} \sum_{j}^{N_{lig}} w_{ij} E_{ij}, \tag{4.4}$$

where N_{aa} is the number of amino acid types and N_{lig} is the number of ligand atom types. I computed the weighting factors by fitting experimental binding constants from the PDBBind 2016 general dataset. This was achieved by minimizing the squared error loss between the predicted energy and the logarithm of experimental binding constants using the L-BFGS-B algorithm [270] implemented in scipy [271].

For each protein-ligand complex, I first collected a matrix of E_{ij} contributions. To reduce the number of parameters we assumed that each weight w_{ij} can be decomposed to a product of two coefficients

$$w_{ij} = c_i r_j, \tag{4.5}$$

where $\dim(\mathbf{c}) = N_{aa}$ and $\dim(\mathbf{r}) = N_l$. Here, "r" stands for a row, and "c" stands for a column of the weights matrix of size $N_{aa} \times N_{lig}$. I then iteratively optimized the $\mathbf{c}$ and $\mathbf{r}$ vectors following Algorithm 1 and obtained the resulting weight matrix $\mathbf{w}$. This approach allowed us to reduce the dimensionality of the problem from $N_{aa}N_{lig}$ to $N_{aa} + N_{lig}$.

The obtained weights are listed in Table A.10. Interestingly, higher weights correspond to hydrophobic interactions, which will be mentioned below. Throughout the text, I will refer to the reweighed version of KORP-PL as to KORP-PL^w.

Algorithm 1: Optimization of the vectors **c** and **r** to find the reweighing coefficients.

Data: matrix of predicted energy contributions **E**, vector of logarithms of experimental binding constants **s**

Initialization: $\mathbf{c} = \mathbf{1}(N_{aa})$, $\mathbf{r} = \mathbf{1}(N_{lig})$, $t = 0$, optimization lower and upper boundary values $b_l = 0.8$ and $b_u = 10.0$

while *not converging* **do**

  if $t\%2 == 0$ **then**

$$\mathbf{c} = \text{argmin}||\mathbf{s} - \mathbf{c}\sum_{j=0}^{N_{lig}} r_j \mathbf{E}_j||^2 \tag{4.6}$$

  end

  else

$$\mathbf{r} = \text{argmin}||\mathbf{s} - \mathbf{r}\sum_{i=0}^{N_{aa}} c_i \mathbf{E}_i||^2 \tag{4.7}$$

  end

  t = t + 1

end

Output: $w_{i,j} = c_i r_j$

4.2 Results

4.2.1 CASF benchmarks

Figures 4.5 and 4.6 show the results obtained on the docking, scoring, ranking, and virtual screening tests from the CASF benchmarks. We can see that KORP-PL performs exceptionally well in the pose prediction exercise, despite being a coarse-grained scoring function. For the CASF-2016 benchmark, its success rate in finding a near-native pose within 2 Å RMSD as the best prediction is 85.6%.

Figures 4.5 and 4.6 (d-e) demonstrate the top-ranked performance of KORP-PL in both screening tests. These results are especially notable if considering the enrichment factor metric in CASF-2016, where all other tested scoring functions perform rather poorly. KORP-PL's affinity predictions, however, turned out to be worse than average. As a consequence, ranking power results are also worse or close to average when compared with the other scoring functions. To investigate the reasons leading to such rather poor performance, we plotted binding affinities predicted by KORP-PL versus the experimental binding constants. Figure 4.2 shows them colored according to the hydrophobic scale of the protein binding pockets suggested in the CASF benchmark [160]. We can see that KORP-PL often underestimates affinity values for complexes with hydrophobic pockets. We suppose that it happens due to the way we compute the *reference state* inherited from the original KORP 6D potential. Indeed, the 6D residue-residue interactions have a strong angular dependence, which is not the case for the protein-ligand setting. For example, the subtraction of the angular average in Eq. 4.2 will result in a near-zero potential for non-directional contributions. This is precisely the case for some of the hydrophobic interactions. It motivated us to introduce the reweighing scheme (see Eq. 4.4), which allowed us to partially compensate for this effect. Indeed, the KORP-PL^w potential performed considerably better than KORP-PL on the scoring

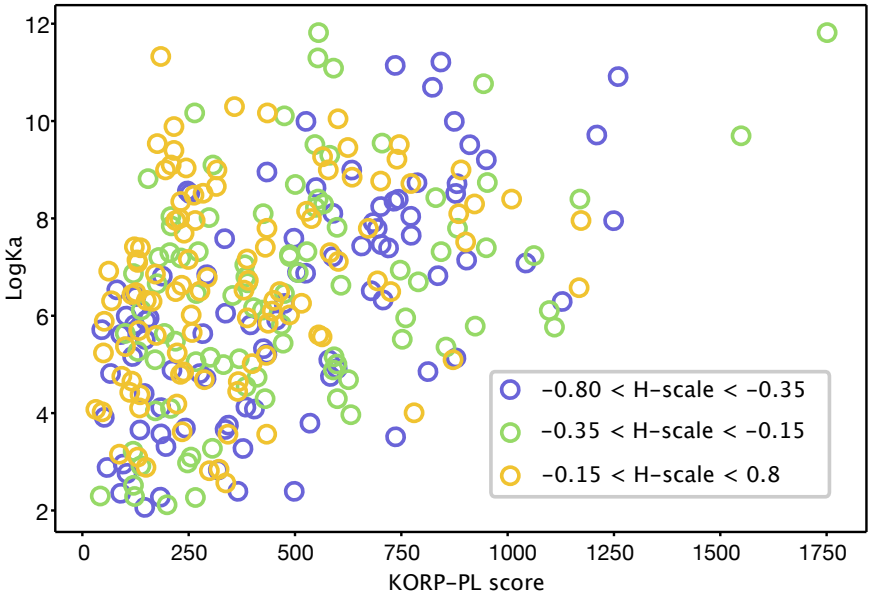


Figure 4.2: Scatter plot of KORP-PL scores versus the log K_a constants from CASF-2016 benchmark. Each point is colored according to the hydrophobicity of the protein binding pocket (H-scale, in logD units) as defined in [160]. The Pearson correlation coefficients between KORP-PL scores and log K_a constants, computed for three different H-scale groups, are: 0.63 for H-scales between -0.80 and -0.35, 0.45 for H-scales between -0.35 and -0.15, and 0.31 for H-scales between -0.15 and 0.8.

tests. However, its performance is still far from perfect and this is a subject for further investigation. We should also note that moderate performance of various scoring functions in affinity prediction tasks can be partially explained by the fact that experimental uncertainties of binding affinity data in current databases are often larger than one order of magnitude [272]. Such significant scatter is the result of different methodologies and accuracy of binding assays used in different research groups.

Figure A.21 contains further analysis of the correlation between the KORP-PL scores and a number of ligand properties computed for the CASF-2016 complexes. CASF benchmarks are derived from the PDBBind database and thus contain complexes similar to our training set. Thus, it is interesting and important to learn how much our results can *overfit* the input data. Therefore, we ran additional experiments and modified the training set by augmenting it with the intersection with the test set, and also removing a number of complexes based on the protein [273, 274] and ligand [129] shape similarity. After, we recomputed the CASF docking and screening tests to investigate the possible overfitting. These results are listed in Tables A.14-A.20 and discussed in Appendix A. Overall, removing the closest complexes (pocket TM-score > 0.8 and ligand shape Jaccard distance < 0.2) affects the metrics only marginally. Further elimination of about a thousand of more distant complexes (pocket TM-score > 0.5 and ligand shape Jaccard distance < 0.4) worsens the overall performance. Notably, high-quality docking predictions (Q1) are affected more than the low-quality ones (Q2-3). This indicates that for a successful high-resolution pose prediction, the training set must contain complexes with interactions that somewhat resemble those in the test set. Indeed, any statistical (Boltzmann in our case) approximation is limited if some features are not present or their distribution is unbalanced in the training set.

4.2.2 D3R benchmarks

Figure 4.8 demonstrates very good performance of KORP-PL in all pose prediction exercises derived from the D3R Challenges. KORP-PL also showed good results in the Grand Challenge 2 and Grand Challenge 4 affinity ranking tasks. However, we obtained near-zero correlations in affinity prediction of the cathepsin S complexes from the Grand Challenge 3.
D3R Grand Challenge 3 pose prediction test turned out to be an interesting case. In this exercise, the binding site is exposed to solvent and is surrounded by water molecules in the co-crystal structure as well as in some of the user submissions. We should specifically mention that we do not take explicit water molecules into account. KORP-PL showed excellent results in the pose prediction exercise compared to AutoDock Vina and both versions of Convex-PL. Although we cannot directly compare the pose prediction results with the full protocols evaluated in the challenge, only a few of them were successful, especially if no visual inspection and ligand-based methods were used [172]. This means that the selection of correct binding poses for the cathepsin S inhibitors could be a challenge for many scoring functions. For example, as can be seen in Figure 4.8, Convex-PL and Convex-PL^R failed in many cases to detect the correct binding mode, while AutoDock Vina and the simplistic ΔSAS were almost completely incapable of doing it. This could be caused by a combination of the following reasons. Firstly, we believe that by its design, KORP-PL is able to better catch directed interactions from target complexes, such as hydrogen and halogen bonding, and π-stacking [275]. Secondly, all the incorrect poses are located deeper in the binding pocket, forming more contacts than the native conformation. Most of the scoring functions tend to be biased towards the total number of protein-ligand contacts, which could lead to incorrect predictions for Convex-PL, Vina, and ΔSAS. As we have already discussed, KORP-PL underestimates some of the non-orientational hydrophobic interactions. However, in this particular case of D3R Grand Challenge 3, it helps to predict ligand positions that are not very buried in protein pockets.

4.2.3 DUD benchmark

To evaluate the performance of KORP-PL in large-scale virtual screening tasks, I have assessed it on the same 31 targets of the DUD benchmark, as I did for Convex-PL^R. For pose sampling, I took the results of VinaCPL and AutoDock Vina sampling that was discussed in subsection *Integration with AutoDock Vina* of Chapter 3. Then, I tested two rescoring protocols for both KORP-PL and KORP-PL^w. In the first one, I did re-scoring of the 300 VinaCPL poses per molecule followed by an averaging of the 10 top scores. In the second one, I re-scored the 25 AutoDock Vina poses. Figure 4.9 and Table 4.2 clearly show that both protocols with both scoring functions outperformed Convex-PL^R, as well as AutoDock Vina and Vinardo. The protocol based on VinaCPL turned out to perform considerably better than the one based on AutoDock Vina. It provided the median ROC AUC values greater than 0.8 and the median 5% enrichment factors that were greater than 7.

4.2.4 DUD-E benchmark

Motivated by the excellent performance of KORP-PL in the DUD benchmark, I assessed it on the 90 targets from the DUD-E benchmark. Again, complexes with co-factors were removed from the

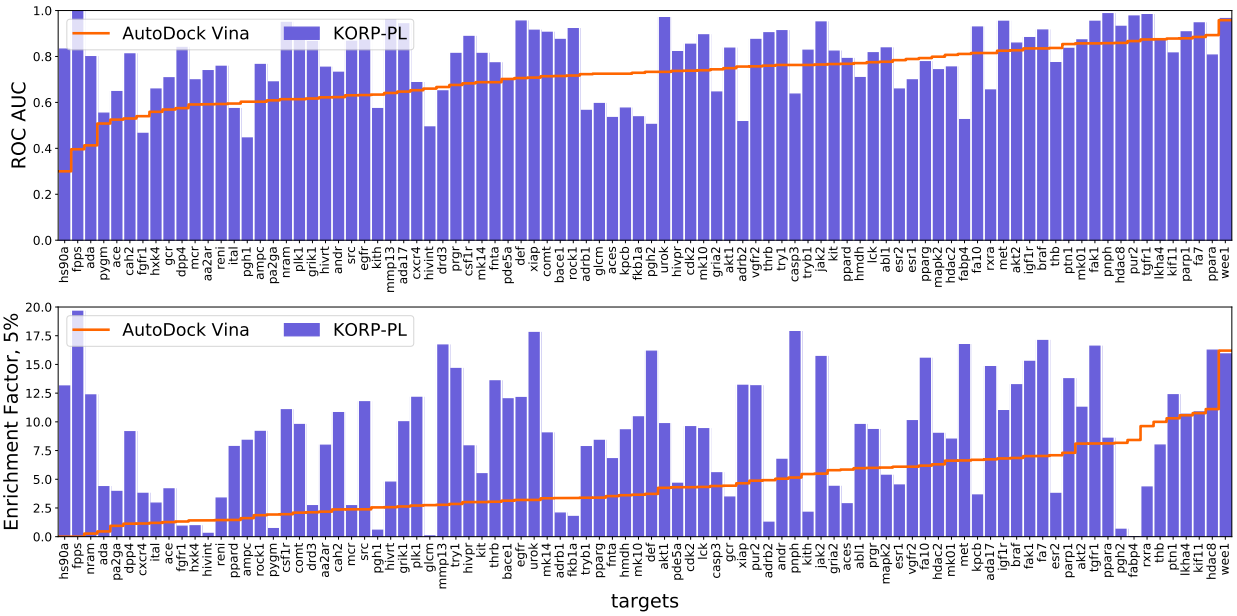


Figure 4.3: ROC AUC scores and 5% enrichment factors computed for the 90 targets from the DUD-E dataset with KORP-PL's rescoring of AutoDock Vina poses and AutoDock Vina.

evaluation. This time we used the same protocol with AutoDock Vina, as for the DUD benchmark, and a VinaCPL-based protocol. In the latter, we restricted the local optimization of the poses done on each Markov chain Monte Carlo (MCMC) step. In the original AutoDock Vina code, local optimization is done before the decision on rejection is taken. We, however, applied local optimization for only the accepted poses. Potentially, this results in lower accuracy of predictions, but increases the code efficiency that was crucial for us on such a big benchmark. If applied on the DUD benchmark, this protocol provided results, similar to those obtained with usual local optimization, as shown in Table 4.2.

As it can be seen in Table 4.1 and Figures 4.3, 4.4, KORP-PL and KORP-PL^w outperform AutoDock Vina in all the metrics, being almost twice better if considering the enrichment factors. Re-scoring of AutoDock Vina's poses works slightly better than VinaCPL rescoring, unlike it was in DUD. This makes KORP-PL comparable to some recent structure-based deep-learning models that demonstrate excellent virtual screening performance [85]. However, according to [92] that we have mentioned in the introduction, such scoring functions tend to achieve high performance on the DUD-E benchmark only if they have been originally trained on it, and thus probably learn hidden biases such as the decoy selection criteria that were used upon the benchmark construction.

4.3 Conclusions

This section described KORP-PL – a novel knowledge-based scoring function for protein-ligand interactions based on the backbone-only receptor and full-atom ligand representations. The receptor representation is adopted from the KORP scoring function, which was designed to model interactions in a protein molecule with a set of oriented coordinate frames unambiguously built on each protein residue. For the ligands, we have chosen a full-atom representation, as the local coordinate frames cannot be defined for chemically diverse small molecules. KORP-PL interaction potential is then derived using statistics of relative orientations and positions of ligand atoms in

Scoring function	ROC AUC		BEDROC $_{\alpha=20}$		EF5%	
	median	average	median	average	median	average
AutoDock Vina	0.725	0.712	0.231	0.260	3.701	4.461
KORP-PL / Vina	0.802	0.763	0.467	0.433	8.384	7.886
KORP-PL^w / Vina	0.802	0.759	0.404	0.424	7.287	7.661
KORP-PL / VinaCPL, local opt restricted	0.774	0.751	0.408	0.411	7.309	7.589
KORP-PL^w / VinaCPL, local opt restricted	0.782	0.748	0.373	0.396	6.566	7.178

Table 4.1: ROC AUC scores, 5% enrichment factors, and BEDROC [174] values computed for the 90 targets from the DUD-E dataset. Twelve targets with co-factors in the binding pocket were excluded from the 102 original targets. It is important to note here that our results for AutoDock Vina are slightly lower than those reported in [85], where the median and average ROC AUC, and median and average EF5% are equal to 0.740, 0.717, 4.228, and 4.485, respectively. This could be caused by the differences in the binding pocket detection or other docking protocol settings. Per-target evaluation results can be found in Table 8.8.

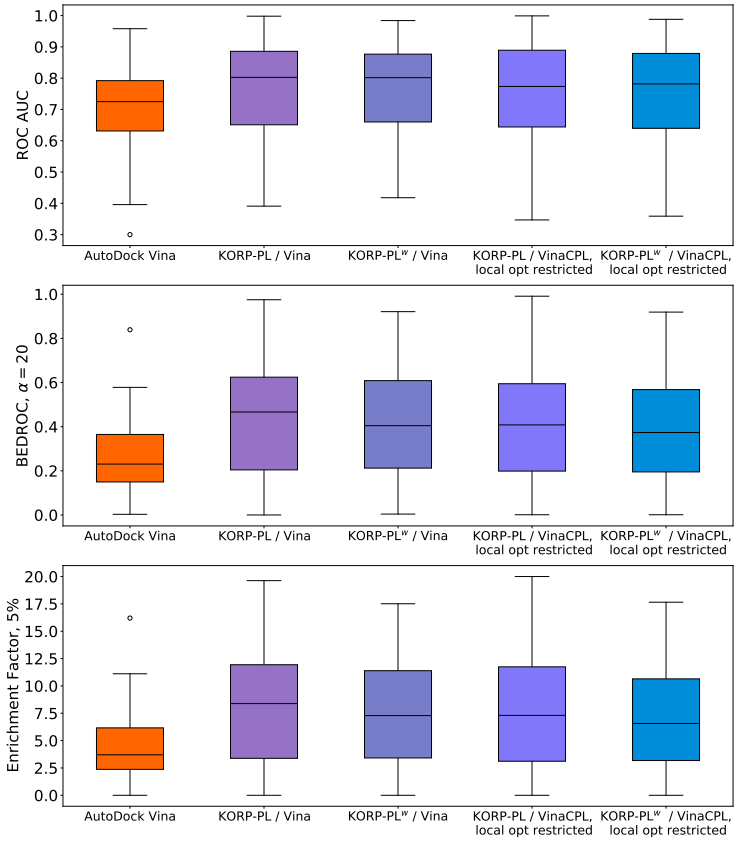


Figure 4.4: Box plots representing the ROC AUC, BEDROC, and the 5% enrichment factor metric values computed for the 90 targets from the DUD-E dataset with AutoDock Vina, KORP-PL re-scoring of VinaCPL poses, KORP-PL^w re-scoring of VinaCPL poses, KORP-PL re-scoring of AutoDock Vina poses, KORP-PL^w re-scoring of AutoDock Vina poses.

the local coordinate systems of protein residues.

We have demonstrated for the first time that a coarse-grained sidechain-free protein representation can be successfully used for very accurate predictions of ligand binding poses. Indeed, KORP-PL shows excellent pose prediction and screening results in CASF-2013 and CASF-2016 benchmarks, and even in pose prediction benchmarks compiled from the D3R Grand Challenges. KORP-PL also demonstrates outstanding results in the DUD-E virtual screening benchmark, where it considerably outperforms AutoDock Vina. Our affinity prediction performance is, however, lower than average, and much more work is required to advance developments in this direction. To investigate this

problem, we reweighed the contributions of residues and ligand atoms by minimizing the squared difference between the predicted and experimental affinities. This resulted in an increase of the weights of non-oriented hydrophobic residues and atoms. The reweighed scoring function, KORP-PL^w, demonstrates a better correlation with binding constants. Overall, this part of my work proposes a very efficient solution to circumvent the long-standing problem of sampling protein sidechain conformations in molecular docking. This paves the way for the development of a new generation of flexible docking approaches.

Convex-PL and KORP-PL assessment

Since the majority of the benchmarks were used for all the three developed scoring functions, this part of the manuscript lists the figures with their summarized evaluations. In more detail, it contains the results of evaluation of Convex-PL, Convex-PL^R, and the two versions of KORP-PL on the CASF, D3R-based, and DUD benchmarks.

CASF benchmark 2013

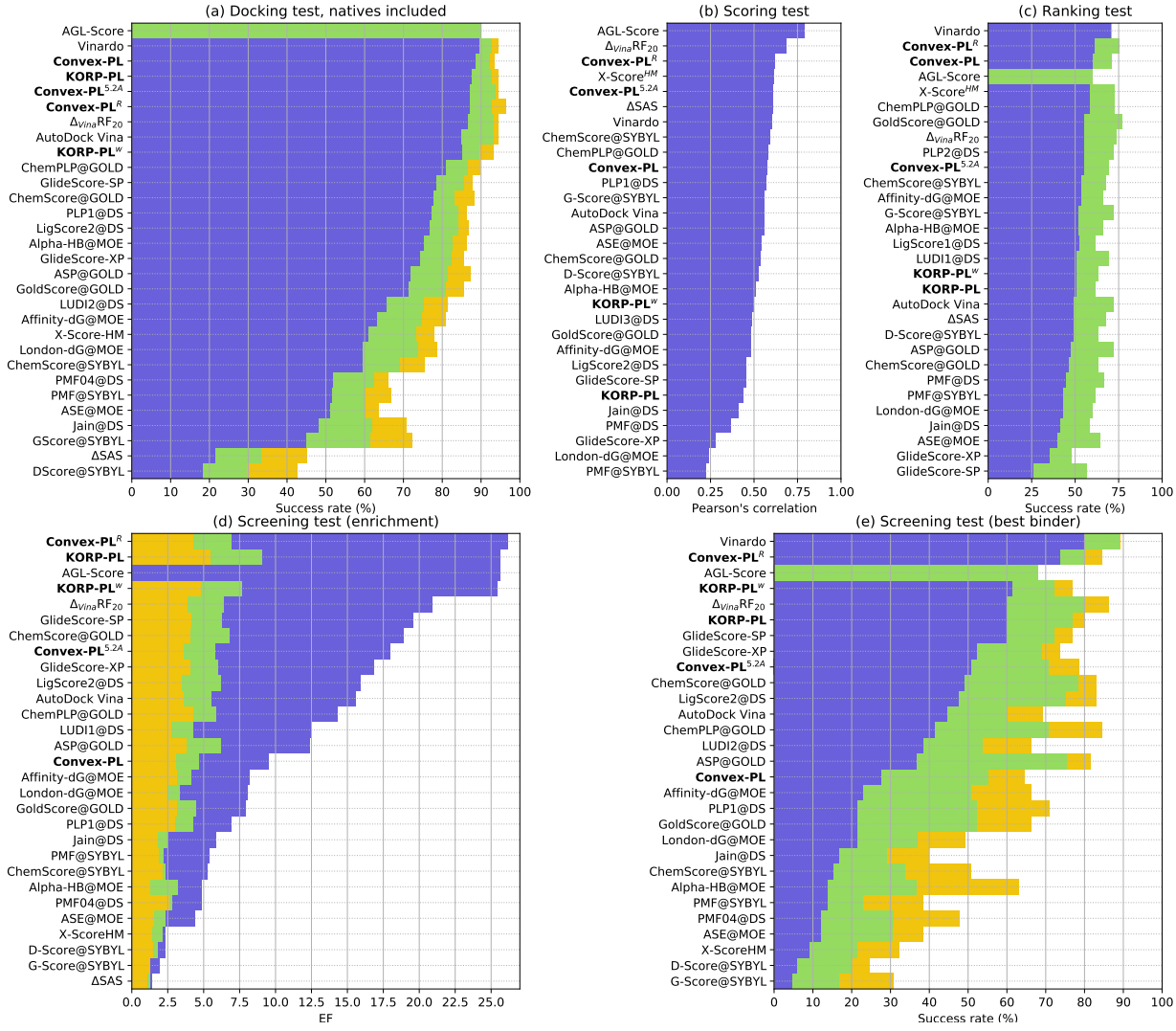


Figure 4.5: CASF-2013 benchmark results. (a) Success rate of finding a native or near-native pose within 2 Å RMSD in Top 1 (blue), Top 2 (green), and Top 3 (yellow) predictions. (b) Pearson's correlation between predicted scores and experimental log K_a constants. (c) Success rate of correct ranking of all the three ligands binding the target protein (blue), and ranking the best complex as the top one (green). (d) Enrichment factors computed considering 1% (blue), 5% (green), and 10% (yellow) of the top-ranked compounds. (e) Success rate of identifying the highest-affinity binder among the 1% (blue), 5% (green) or 10% (yellow) top-ranked ligands.

CASF benchmark 2016

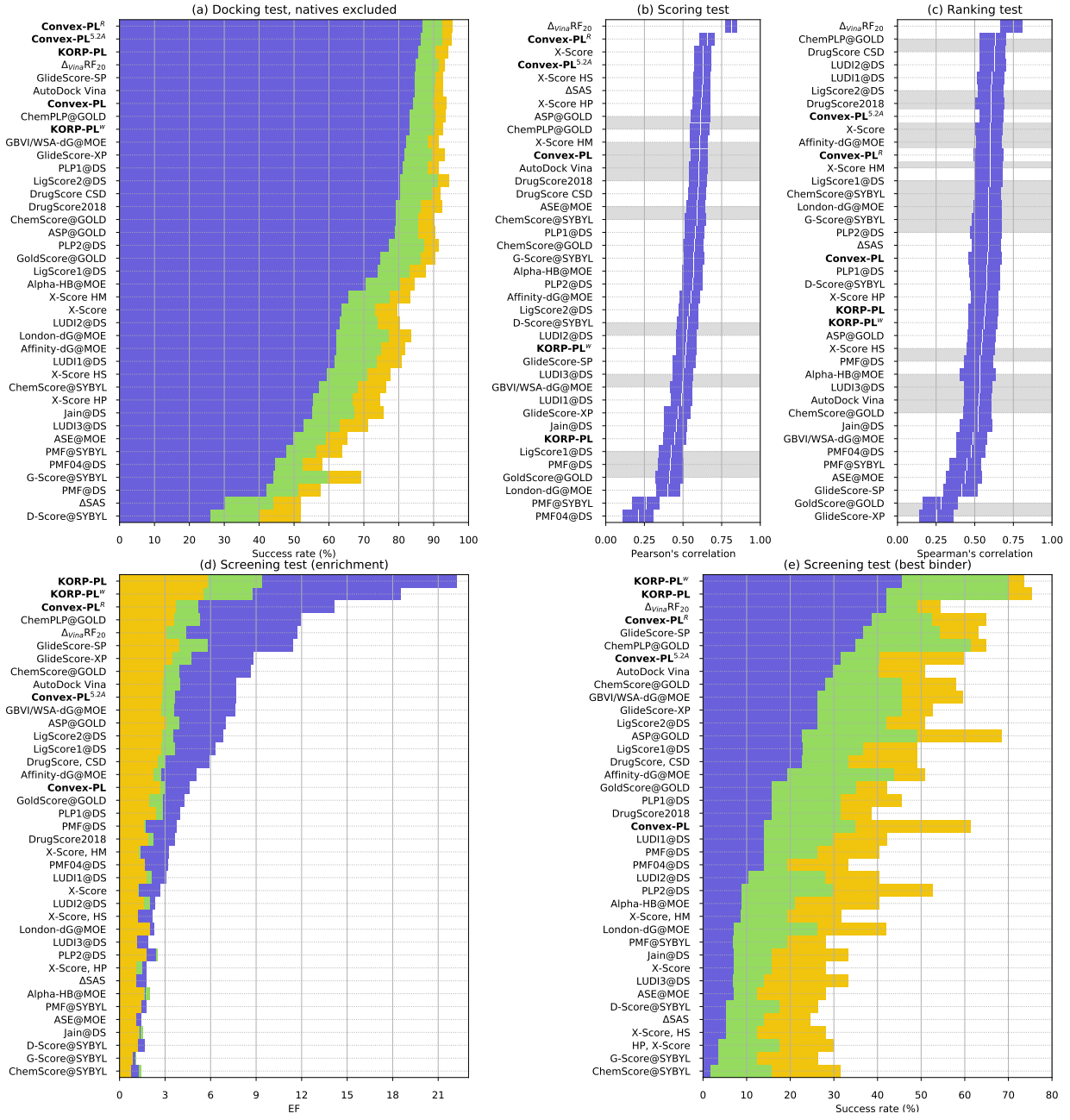
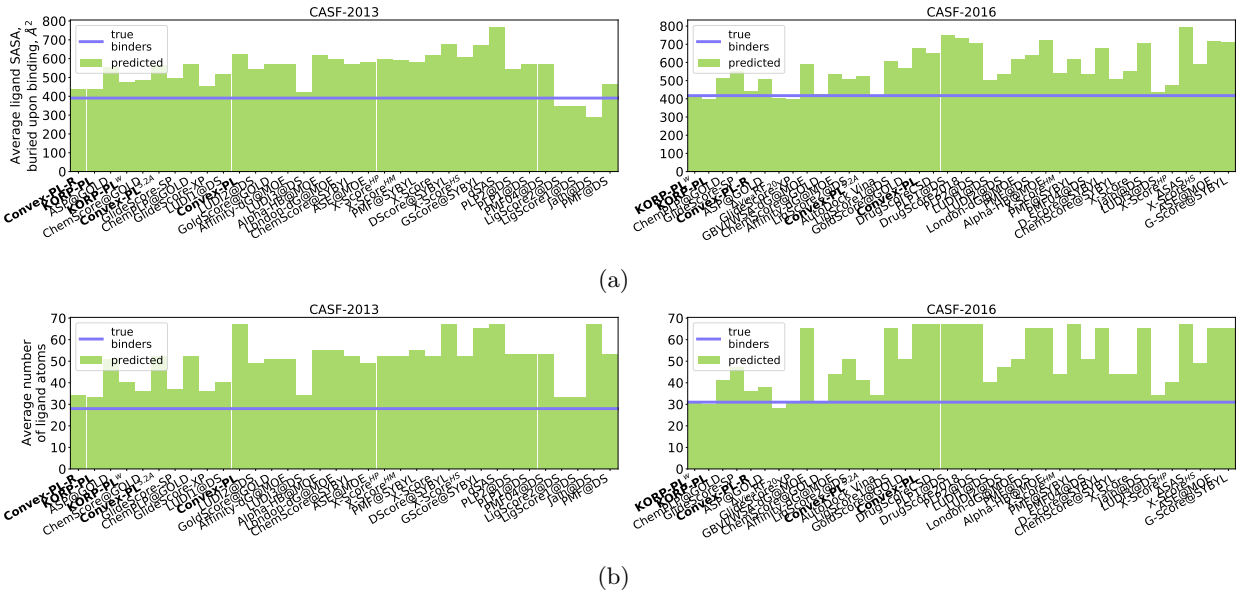


Figure 4.6: CASF-2016 benchmark results. (a) Success rate of finding a near-native pose within 2 Å RMSD in Top 1 (blue), Top 2 (green), and Top 3 (yellow) predictions. Native poses are excluded. (b) Pearson's correlation with confidential values between predicted scores and experimental log K_a. (c) Spearman's rank correlation with confidential values among the 57 clusters. (d) Enrichment factors computed considering 1% (blue), 5% (green), and 10% (yellow) of the top-ranked compounds. (e) Success rate of identifying the highest-affinity binder among the 1% (blue), 5% (green) or 10% (yellow) top-ranked ligands.

Ligand size biases in CASF screening tests



(a)

(b)

Figure 4.7: (a)-(b) The purple line represents the average values of buried SASAs and numbers of atoms computed for ligands that natively bind target proteins and should have been predicted as the most affine binders[1]. The green boxes correspond to decoys that were top-ranked by scoring functions assessed on the virtual screening test from the CASF benchmarks 2013 and 2016. The scoring functions are sorted by the ability to predict the highest affinity binder in the 5% of the top-ranked decoys. SASA values were computed with PyMOL's[96] *get_area()* function with *dot_solvent* set to 3.

[1] Or be among the most affine binders in several cases when the target protein was known to bind ligands with higher affinity but without co-crystal structure.

D3R benchmark

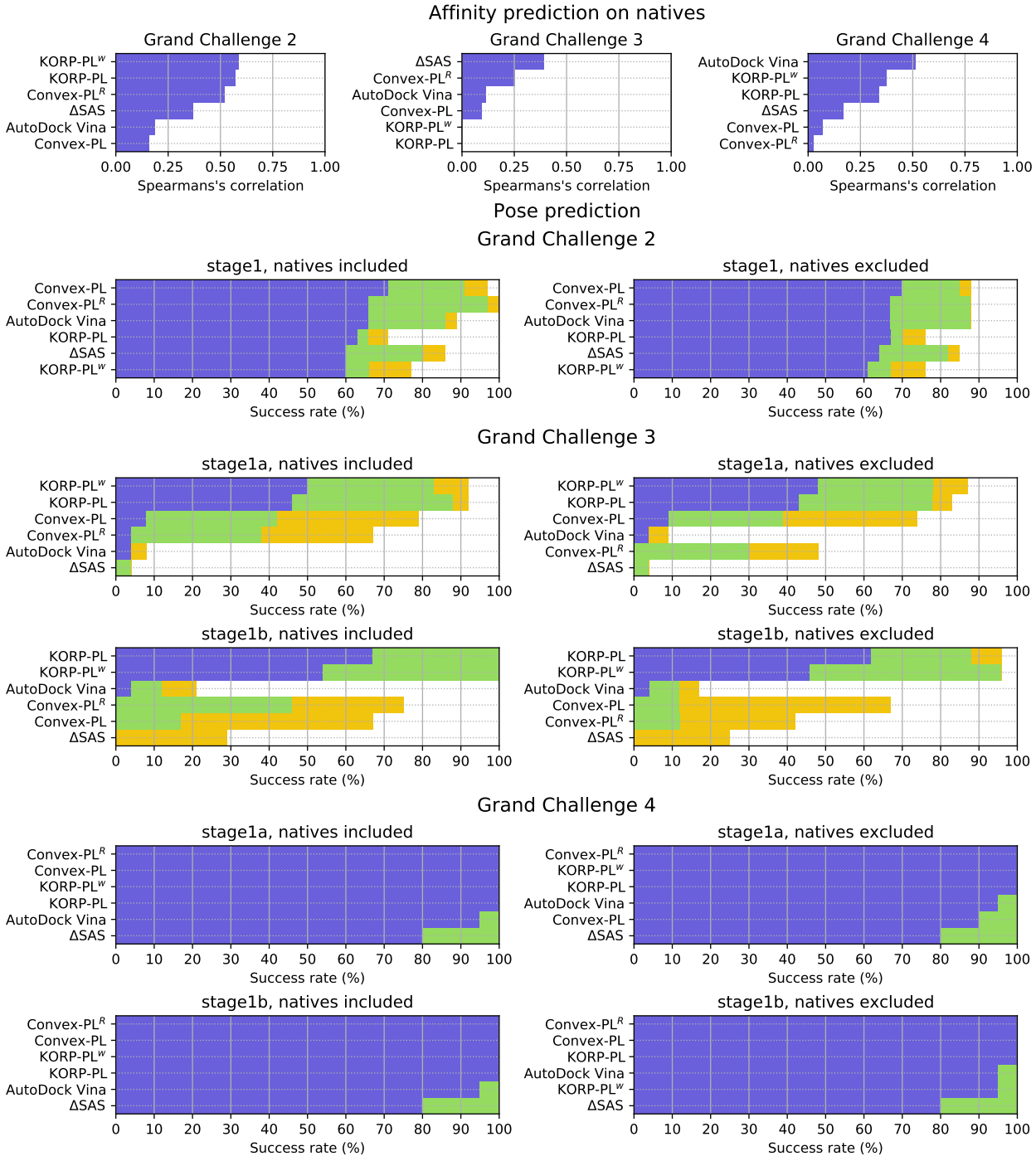


Figure 4.8: D3R pose prediction and scoring results. Success rates of finding a pose within 2 Å RMSD from the native conformation among the 1%, 5%, and 10% of top-ranked poses are shown in blue, green, and yellow, respectively. Scoring power is represented by the Spearman's correlation coefficient between the predicted and experimental binding constants. These success rates are computed with respect to the actual number of ligands, for which the poses with the desired RMSD values were present in the user submissions.

DUD benchmark

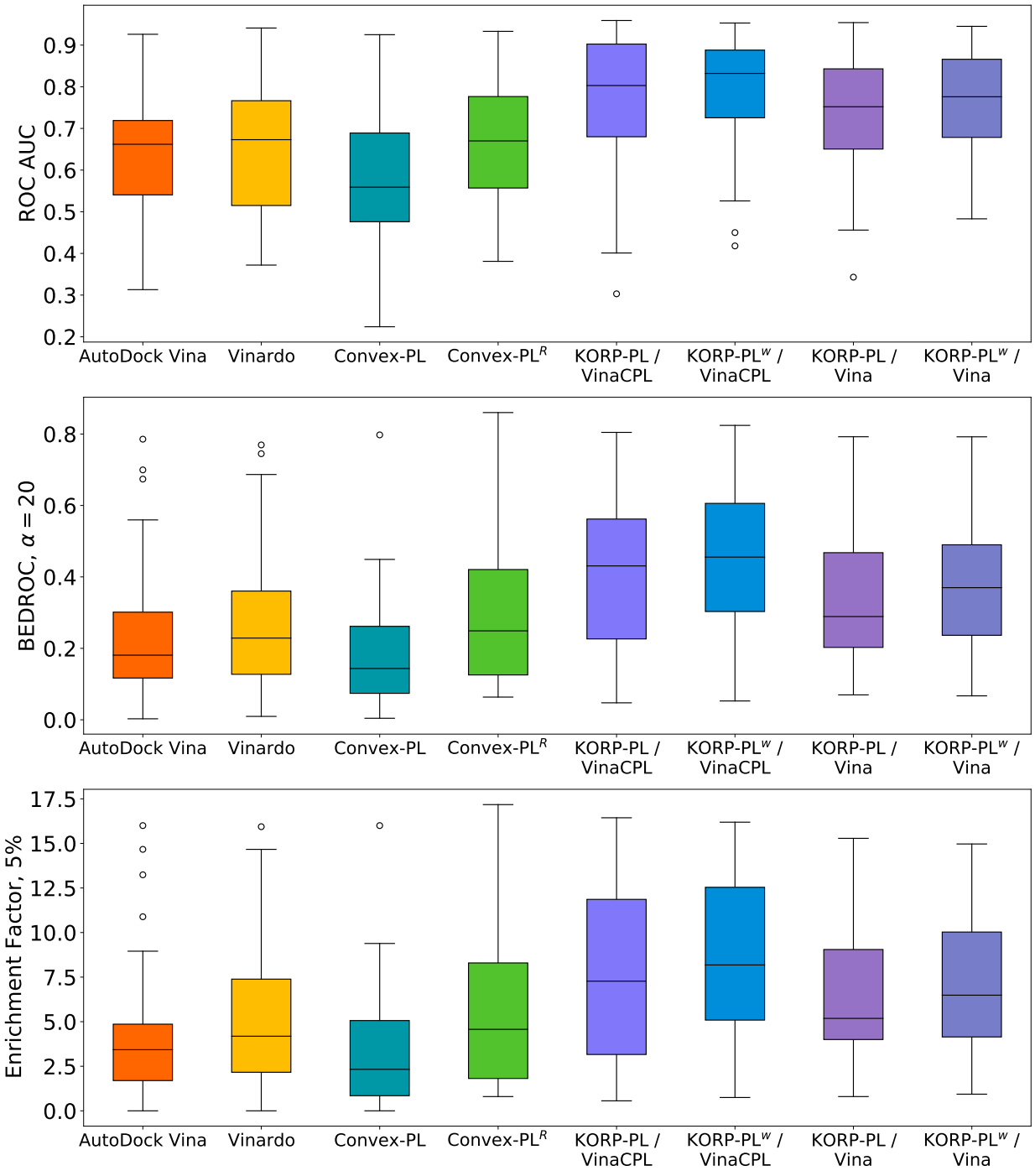


Figure 4.9: Box plots representing the ROC AUC, BEDROC, and the 5% enrichment factor metric values computed for the 31 targets from the DUD dataset with AutoDock Vina, Vinardo, Convex-PL, Convex-PL^R, KORP-PL re-scoring of VinaCPL poses, KORP-PL^w re-scoring of VinaCPL poses, KORP-PL re-scoring of AutoDock Vina poses, KORP-PL^w re-scoring of AutoDock Vina poses.

Scoring function	ROC AUC		BEDROC $_{\alpha=20}$		EF5%	
	median	average	median	average	median	average
Convex-PL^R	0.670	0.667	0.249	0.302	4.571	5.488
AutoDock Vina	0.662	0.626	0.181	0.246	3.429	4.452
Vinardo	0.673	0.646	0.229	0.276	4.186	5.199
Convex-PL	0.559	0.572	0.143	0.181	2.329	3.331
KORP-PL / VinaCPL	0.803	0.757	0.431	0.411	7.273	7.634
KORP-PL^w / VinaCPL	0.832	0.787	0.456	0.449	8.182	8.521
KORP-PL / Vina	0.752	0.727	0.289	0.342	5.185	6.301
KORP-PL^w / Vina	0.776	0.763	0.370	0.377	6.487	6.965
KORP-PL / VinaCPL, local opt restricted	0.812	0.746	0.435	0.383	7.347	6.849
KORP-PL^w / VinaCPL, local opt restricted	0.806	0.777	0.469	0.421	8.163	8.014

Table 4.2: ROC AUC scores, 5% enrichment factors, and BEDROC [174, 129] values computed for the 31 targets from the DUD dataset. Nine targets with co-factors in the binding pocket were excluded from the 40 original targets. Per-target evaluation results can be found in Tables A.28 and A.29.

Part III Validation on docking challenges

Chapter 5. D3R Grand Challenge 2

In 2016 we took part in the blind docking competition D3R Grand Challenge 2 [171]. The initial submissions were done both by me and by my scientific adviser Sergei Grudinin. The analysis of the challenge results, re-docking, and assessment of various docking protocols after the challenge was performed by me. These results were published in the Journal of Computer-Aided Molecular Design [207].

5.1 Challenge description

5.1.1 Challenge structure

D3R Grand Challenge 2 provided a blinded unpublished dataset containing the farnesoid X receptor (FXR) target, which was kindly contributed by Roche and curated by D3R. This dataset contained 36 crystal structures with resolution better than 2.6 Å supplied with binding affinity data (IC50s) for 102 compounds across five orders of magnitude, comprising four chemical series and 6 miscellaneous compounds. The organizers have also provided two subsets of 18 and 15 compounds for a separate evaluation of absolute binding affinity prediction, in which we have not participated.
The D3R Grand Challenge 2 (GC2) consisted of two stages. In the first stage, the goals were to predict crystallographic binding poses of 36 selected ligands spanning all chemical series, predict binding affinities for all 102 ligands, and to predict the relative binding affinities for two designated free energy subsets of 18 and 15 compounds. After the termination of the first stage, the organizers released the blinded co-crystal structures of the 36 FXR complexes selected for pose prediction. We should note that FXR_33 was excluded from the pose prediction evaluations because of the artifacts found in the co-crystal structure, and only 35 ligands were assessed. In the second stage, the goal was to predict binding affinities for all 102 ligands and the relative binding affinities for the two free energy subsets of 18 and 15 compounds using the released 36 co-crystal structures of protein-ligand complexes. After the end of the second stage, the experimental binding affinities (IC50) were made available.
We participated only in the pose prediction part of Stage 1 and assessed several docking protocols. Some of these were introduced earlier [276, 185], others were designed specifically for the FXR targets. Like many other participants, we obtained sufficiently small RMSD values for most of the benzimidazole-containing ligands. However, we did not succeed in docking of the majority of the other groups of compounds, being able to predict binding poses with the mean best RMSD equal to 2.96 Å. This was an average result, especially when compared with very accurate structure-based submissions utilizing known information on co-crystal ligands [203], or Free Energy Perturbation submissions [277]. After the end of the Challenge and subsequent release of the full crystallographic structures, we analyzed our protocols and conducted some new experiments, which improved the results of our initial submissions. The subsections below provide a detailed explanation of our submission and post-submission protocols.

5.1.2 Dataset for pose prediction

The 36 compounds selected for the exercise, of which 35 were evaluated, include benzimidazoles, isoxazoles, sulfonamides, spiranes and several other unclassified molecules. The organizers provided

them as a set of SMILES strings and 2D SDF files. The target receptor of the Grand Challenge 2 is the nuclear farnesoid X receptor (FXR). It is mainly expressed in the nuclei of liver cells, and is responsible for lipid concentration regulation. The RCSB database (also knows as Protein Data Bank) [20] contains a large variety of FXR structures with a different degree of homology to the target receptor, whose apo structure was also provided as an input. For example, for the homology threshold of 30%, the search over the RCSB database yields about 300 protein structures. Among those structures, the 3D conformation of the protein is mostly conserved, except for the bindig pocket region. The spacious binding pocket is formed by several alpha helices and loops, three of which are quite flexible, as it is shown in Figure 5.1. Despite the considerable amount of available FXR structures, we have not found any co-crystallized with spirane- or sulfonamide-containing ligands, making modeling of the corresponding targets from the exercise more challenging.

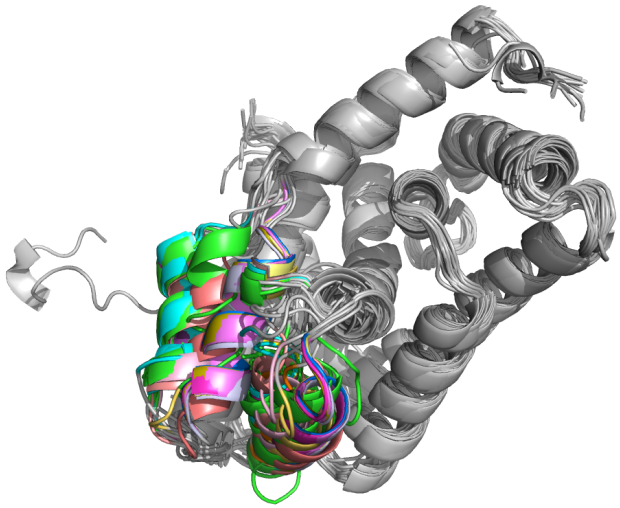


Figure 5.1: Proteins with a high sequence identity (higher than 90 %) to the target FXR apoprotein superposed to each other. Only the A chains are shown. The two helices with a high conformational freedom are highlighted.

5.2 Submission protocol

5.2.1 Structure preparation

For our submission, we generated 3D ligand structures starting from the corresponding SMILES strings using OpenBabel's *gen3d* command [278]. Then, we manually adjusted ligands with ambiguous symmetry using geometry of small molecules with similar functional groups from the RCSB database as an example. No additional structure optimization was performed at this step.

Regarding the receptors, we manually selected one model for each target from the RCSB database. We firstly pre-selected a subset of complexes using the homology threshold of 30 % and then found the best model for each ligand by ligand similarity identified based on the SMILES substrings of the functional groups. This procedure identified 23 template receptors for 36 targets. Among these, 17 had homology to the target sequence higher than 90 %. No additional structure optimization of the receptors was used.

5.2.2 Docking and rescoring

To generate putative binding poses, we used the AutoDock Vina software package with the default scoring function [70]. Using AutoDock Vina, we generated 1,000 docking solutions for the subsequent re-scoring. In the AutoDock Vina configuration files, the parameter *num_modes* was set to 1,000 and *exhaustiveness* to 100. We identified the receptor binding pocket based on the structures of homologous FXR proteins. All the rotatable bonds in ligands were kept flexible during the docking procedure, all the protein residues inside the binding pockets were kept rigid. We assigned the Gasteiger atomic partial charges and converted all receptors and ligands to the PDBQT format (the format is an extension of the PDB file format, which also allows representing a kinematic tree

of a molecule) using the AutoDockTools package [137]. For the Grand Challenge 2 submission, we did not use explicit hydrogens neither for the receptors nor for the ligands. This choice, however, is not methodologically correct when using AutoDock Vina with the default scoring function. Finally, we re-scored the obtained poses with the Convex-PL potential. No additional clustering of the final poses was made.

5.3 Protocols evaluated after the submission

Once the co-crystal FXR structures were released by the challenge organizers, we repeated the docking tests using the crystallographic structures of 35 (since FXR_33 target was skipped from the assessment) receptors with the corresponding ligands. To rigorously assess the performance of different docking strategies, in addition to the self-docking test, we designed seven other docking experiments. All these are summarized in Table 5.3 and listed below. In the additional docking experiments, we optimized several ligands and also mutated the receptors whose sequence was different from the target as it is described below.
In all these protocols we ran AutoDock Vina with the same parameters as in the submissions. A big difference, however, was in the application of the Convex-PL scoring function. This time we used VinaCPL, i.e. Convex-PL was integrated inside AutoDock Vina without the subsequent rescoring. We, again, did not use explicit hydrogens, however, in this case it was a correct decision, as Convex-PL does not take them into account. Finally, we selected the five top-scored structures after a simple clusterization procedure with a symmetry-adapted ligand RMSD threshold value of 1 Å.

Self-docking

In the self-docking experiment, we docked each co-crystal ligand to the corresponding receptor structure. If several chains were present in the released structures, we performed docking to all of them and then chose the pose with the highest score.

One ligand to one receptor

In the first experiment (exp. 1), we repeated the submission protocol but mutated all the receptors to the target sequence.

One ligand to multiple receptors

In the second experiment (exp. 2), we exhaustively docked all the ligands used in the submission protocols to each of the receptors used in the submission plus to all other receptors with homology higher than 90 % to the target receptor sequence. In this experiment, we mutated all the receptors to the target sequence using SCWRL4 [279]. This protocol may be seen as "one ligand to multiple receptors" docking. The receptor PDB codes used in this experiment are listed in Table 5.1.

Multiple ligands to multiple receptors

In the third experiment (exp. 3), we repeated exp. 2 for each of the targets with all the ligand models available as explained in more detail below.

1osh	1osv	1ot7	3bej	3dct	3dcu	3fli	3fxv	3gd2	3hc5	3hc6	3l1b
3okh	3oki	3olf	3omk	3omm	3oof	3ook	3p88	3p89	3rut	3ruu	3rvf
4oiv	4qe6	4qe8	4wvd	5iaw	5ick	1s0x	3dy6	3peq	3w5p	4r06	5eit

Table 5.1: PDB codes of the 36 receptors chosen from the RCSB database for the "multiple receptors" docking protocols. Proteins with low sequence identity are highlighted in grey.

One ligand to some receptors

In the forth experiment (exp. 4), we repeated exp. 2, where for each of the target we used only a few receptors whose bound ligand had the most similarity to the target ligand.

One ligand to multiple+self receptors

In the fifth experiment (exp. 5), we complemented exp. 2 with the co-crystal structures of the receptors.

One ligand to self receptor

In the sixth experiment (exp. 6), for each of the targets, we docked the ligand used in the submission to the co-crystal structure of the receptor. This protocol may be seen as "one ligand to self receptor" docking.

Self ligand to multiple receptors

In the seventh experiment (exp. 7), for each of the targets, we docked the co-crystal structure of the ligand to a number of receptors from exp. 2.

Ligand input optimization

In addition to the above-listed experiments, we also repeated exp.1, exp. 2, exp. 4, exp. 5, and exp. 6, where for each of the targets we complemented some difficult ligands from the submission with their optimized structures as explained below in more detail. These experiments are labeled "mod" in Table 5.3. The original experiments are respectively labeled "old".

5.3.1 Input structure preparation

Preparation of ligand structures

For all the additional experiments, we generated 3D ligand structures using OpenBabel *gen3d* command [278] complemented with fragments of similar small molecules in cases when OpenBabel was unable to provide reasonable structures. Regarding OpenBabel, its *gen3d* method generates ligands using a set of rules and templates, followed by a minimization in MMFF94 force-field [280]. Finally, it performs a conformational search followed once again by the ligand optimization.
For the additional computational experiments described above, we used the same ligand structures as in our submission. However, for "difficult" ligands, i.e. those whose docking to the co-crystal receptors resulted in RMSD values greater than 1.5 Å in all the top-5 poses without clusterization, we performed an additional structure optimization. More precisely, we optimized these ligands using force-field-based structure optimization with the MMFF94 force-field [280] implemented in the Avogadro package [121]. If the gradient-based optimization with MMFF94 did not converge,

PDB code (chain)	#	values for mutated residues in the binding pocket: mean molecular weight before	mean molecular weight after	mean pH at isoelectric point before	mean pH at isoelectric point after	mean index of hydrophobicity before	mean index of hydrophobicity after
1osv	5	129.34	106.51	6.34	5.81	18	25
1ot7 (AC)	4	118.12	110.87	5.23	5.76	26	21
1ot7 (BDE)	5	129.34	106.51	6.34	5.81	18	25
3bej	1	147.13	89.10	3.22	6.00	-31	41
3rvf	3	142.14	89.10	6.55	6.00	-17	41
3dct	2	126.11	89.10	4.45	6.00	-18	41
3dcu	2	126.11	89.10	4.45	6.00	-18	41
3gd2	2	126.11	89.10	4.45	6.00	-18	41
3hc5	2	126.11	89.10	4.45	6.00	-18	41
3hc6	2	126.11	89.10	4.45	6.00	-18	41
3p88	2	126.11	89.10	4.45	6.00	-18	41
3p89	2	126.11	89.10	4.45	6.00	-18	41
3rut	2	126.11	89.10	4.45	6.00	-18	41
3ruu	2	126.11	89.10	4.45	6.00	-18	41
4oiv	2	160.66	89.10	6.99	6.00	-23	41
4qe6	3	142.14	89.10	6.55	6.00	-17	41
4qe8 (AC)	2	160.66	89.10	6.99	6.00	-23	41
4qe8 (BD)	1	147.13	89.10	3.22	6.00	-31	41
5iaw	1	147.13	89.10	3.22	6.00	-31	41
1s0x	23	138.12	131.53	6.04	5.99	59	38
3dy6 (A)	23	133.76	139.46	5.86	5.89	49	50
3dy6 (B)	22	134.61	142.39	5.84	5.89	53	53
3peq	21	134.77	141.30	5.83	5.91	51	51
3w5p	42	138.04	138.45	5.91	6.10	51	43
4r06 (A)	22	142.89	140.26	5.71	5.89	52	60
4r06 (B)	17	134.68	127.38	5.79	5.64	19	41
5eit (A)	26	136.04	133.27	5.88	5.96	47	35
5eit (B)	25	136.20	132.71	5.90	6.06	50	38

Table 5.2: List of receptors with the mutated amino acids in the binding pocket used in the experiments. The binding pocket was defined as all the residues having at least one heavy atom within the distance of 4 Å from any atom of ligands that are available for the 36 receptor proteins and the co-crystal FXR protein targets. Proteins with a low sequence identity are highlighted in grey. Thirteen proteins had 100% sequence identity in the binding pocket and were excluded from the table.

we used the UFF force-field [281]. Although we used the MMFF94 force-field optimization both in OpenBabel (implicitly) and in Avogadro, the resulting conformations were not the same, perhaps due to the differences between the 3D structure generation algorithms, and also due to the fact that we used different input files, 2D SDF files for OpenBabel and SMILES strings for Avogadro. Finally, all the structures of ligands were converted to the PDBQT format using AutoDockTools with the default identification of rotatable bonds.

Preparation of receptor structures

To represent the structural heterogeneity of receptors, we selected all the structures from the RCSB database with the identity to the target sequence higher than 90 %. The identity was computed with the *BLASTP* program from the BLAST+ package [248, 249]. This resulted in a set of 30 receptors. For the submission, as mentioned above, we used 6 additional low-homologous receptors chosen by ligand similarity to the target ligands. These were also added to the receptors set, resulting in a pool of 36 structures. For each of the receptors, we mutated all the residues different from the apo protein sequence using the SCWRL4 tool [279] as it is described below. We did not model

residues that were missing in the crystallographic structures. To structurally align the receptors to the provided model, we use the *align* command from PyMOL [96] with the default options. Finally, all the receptor structures were converted to the PDBQT format using AutoDockTools. As in the submission, we treated all the receptors' bonds as rigid, i.e. no flexible residues were selected.

Mutations

Table 5.2 lists the summary of the total number and properties of the mutated residues in the binding pocket. It contains the mean molecular weights, pI and hydrophobicity indices of the mutated residues before and after the mutations. The reference values were taken from the Sigma-Aldrich webpage [282]. For all the highly homologous proteins listed in Table 5.2 only 1-5 residues were mutated in the binding site consisting of 39-60 residues. Three of these mutations occurred in all of these proteins and they considerably changed at least one property of the residues. These are the polar Ser342 and the hydrophilic Glu350 and Arg455 residues that were replaced by the hydrophobic alanine residues[1]. Ser342 is located close to majority of the available ligands, but its mutation preserves the residue's size and charge. Arg455 is located relatively close to several ligand aromatic groups, and thus its turn into a more hydrophobic alanine might be energetically favourable. Glu350 is situated rather close to several other ligands, but its hydrophilic glutamate side-chain is rotated outward from the binding pocket. Therefore, we did not expect these mutations to strongly influence the docking results. However, we found it better to apply them to the protein structures.
As for the low-identity proteins, although the mean values of the pI, size and hydrophobicity index remain similar, as it can be seen from Table 5.2, about a half of the mutations for each receptor lead to significant changes of some of the residue's properties.

Missing residues

As it was mentioned above, we did not model missing residues in the binding sites of the protein structures that we used for the submission. For example, in several proteins an entire helix (Gln271–Glu285, Arg268–Phe282, Gln257–Phe282 in 1osh, 3l1b, and 4oiv, respectively) or a small loop (Lys343-Gly347, Lys343-Pro345, Leu344-Ala346 in 3fxv, 4qe6, and 3bej, respectively) were missing. For these structures, inserting the helix model might have been a good solution, as its conformation is conserved in all the 30 highly homologous receptor models. This helix also appeared to have the same conformation in a number of reference co-crystal FXR complexes and it has no steric clashes with any of the correct ligand poses. The missing loop has a different conformation compared to the apo protein in most of the co-crystal receptor structures, and its modeling would have been way more challenging.

5.3.2 RMSD computation

We computed symmetry-adapted RMSD values with a modified *GetBestRMS()* function from the RDKit package [129]. Originally, this function aligns molecules taking into account the symmetry matchings by iterating over them to find the best alignment transformation with the lowest RMSD

[1]Residue numbers are given with respect to the 3rvf receptor and may differ slightly for other proteins. The corresponding apoprotein residues are Ala346, Ala354, Ala459.

value. In our modification, we preserved the iteration over all symmetry matchings and removed the alignment part to make the function only compute the smallest RMSD value among various symmetry matchings without superposing the molecules. This modified version of *GetBestRMS()* is now available in RDKit as *CalcRMS()*. We should note that our recalculated RMSD values of our submission are slightly larger compared to those reported in the D3R challenge results. For example, our computations resulted in the mean value for our submission of 2.96 Å versus 2.90 Å reported by the challenge organizers. This may happen due to the differences in RMSD computation and alignment procedures.

5.4 Results

5.4.1 Docking results

Table 5.3 lists the top-5 pose prediction results for all the docking experiments with Convex-PL as a scoring function. We obtained sufficiently good results for the self-docking experiment with the mean RMSD value to the solutions of 0.58 Å. This indicates that our method is always able to predict near-native ligand poses given as input co-crystal receptor conformations along with the correct local geometries of the ligands. The local geometry includes bond lengths and angles, because these are kept unchanged during the docking procedure.

To figure out why the mean RMSD value of our submission was about 5 times larger that the result of the self-docking experiment, we ran exp. 6 with docking ligands to the co-crystal receptors, which is summarized in the corresponding column of Table 5.3. With this protocol, we were unable to predict a binding pose within 3 Å RMSD in only a few cases. However, according to the exp. 5 column, the native receptor structure does not guarantee the highest docking score if there are other receptor structures present in the ensemble. It this experiment, the co-crystal receptor structures were complemented with the homologous FXR structures from the RCSB database. For example, for the ligands of FXR_4, FXR_10, FXR_11, and FXR_16 targets, our scoring function preferred non co-crystal receptors with considerably different ligand conformations.

Re-docking of the ligands to the same receptors that we used for the submission did not affect the mean RMSD, as it is seen from the column exp. 1 of Table 5.3. For some targets RMSD improved, for others it became worse. Considerable RMSD differences that can be seen for only several ligands could be the result of mutation of the residues located near the binding pocket (FXR_20, FXR_13, FXR_15). These differences remained even when we re-generated the ligand structures with a force-field-based optimization.

Columns exp. 2 and exp. 3 of Table 5.3 list the results of docking of each ligand (either single or multiple) to a set of homologous FXR structures found in the RCSB database. Overall, these experiments considerably improved the docking accuracy as compared to exp. 1. In case of "easy" ligands such as benzimidazoles, all the three strategies (exp. 1-3) performed well. However, for more complicated targets when the procedure of choosing a proper receptor is unclear, docking to multiple receptors allows to achieve better poses compared to the results of exp. 1, where we performed docking to a single receptor.

In exp. 3, which as we believed was the most rigorous blind experiment, we obtained a lower mean RMSD value of 2.33 Å than in both parts of exp. 2 (2.15 Å and 2.21 Å). FXR_11 target contributed the most to this RMSD difference, as the top-2 predictions for the "modified" and

top-4 predictions for the "old" ligands consist only of poses with RMSD values greater than 8 Å, which are not rejected by the subsequent clusterization procedure in exp. 3. The best pose was chosen properly for only 6 out of 13 "modified" ligands (FXR_2, FXR_3, FXR_8, FXR_17, FXR_22, FXR_34). Although the number of the ligands that we re-optimized is rather small to make rigorous conclusions, we would suggest to be careful with the "multiple ligand" docking protocols, as these may require better clusterization algorithms and a visual inspection of the obtained poses for each of the ligands.

Docking to the smaller sets of receptors, chosen for each ligand by ligand similarity, improved pose prediction for FXR_2 and the submission version of the FXR_11 ligand as it is shown in the exp. 4 column of Table 5.3. However, this strategy excluded some potentially good candidates from the docking set of receptors. For example, in the case of FXR_15 target in exp. 2, for four receptors we were able to obtain rather low-scored poses within RMSD < 3.5 Å and the top-scored conformation with RMSD of 4.87 Å, which preserved the sulphonamide group in the near-native position. Nevertheless, only one of these four receptors was in the set that we chose for the FXR_15 target in this experiment.

Below we would like to discuss several cases of the molecules that turned out to be the most challenging for our docking algorithms.

5.4.2 Importance of the receptor conformation

The most difficult target for our protocols appeared to be FXR_18. Its ligand contains the same functional groups (chlorine, amide, aromatic and cyclohexane rings, planar trigonal nitrogen in the center of the molecule) as some of the benzimidazole ligands from the complexes available at RCSB, namely, 3oki, 3omk, 3oof and several other structures of human FXRs obtained by Roche in 2011 [283]. Our scoring protocol correctly identifies all these proteins from a larger set of receptor candidates, however, the co-crystal ligand binding pose differs from our predictions by RMSD of more than 9 Å. This can be the result of the high conformational mobility of four of the binding pocket helices, as highlighted in Figure 5.2 (a) for the highest scored receptor model 3oki, which differ by 6.34 Å RMSD from the co-crystal structure.

More precisely, besides the two helices that are known to be flexible for the farnesoid X receptors, as we have mentioned in the dataset description, two other helices of the binding site of our top-scored receptor 3oki have different conformations compared to the co-crystal receptor structure. These are connected with a loop and located on the opposite side of the binding pocket, as it is shown at the left side of Figure 5.2 (a). This region seems to be relatively conserved in all the proteins that we used for the "multiple receptors" docking protocols. The correct receptor conformation superposed with these proteins can be seen in Figure 5.2 (b), where it is "buried" under the other proteins' conformations except for a short Trp473–Thr466 helix, which is located near one of the rings of the FXR_18 ligand. In the co-crystal structure, this cyclohexane ring has a different 3D configuration from all the other ligands from the targets we used for the docking experiments, as it is shown in Figure 5.2 (c). Thus, the correct docking poses of this part of the ligand result in steric clashes with the flexible helices of the non co-crystal receptor structures.

A similar situation is observed for another difficult FXR_4 target. Here, our docking algorithm prefers the above-mentioned receptors found in complex with benzimidazole ligands, in particular, the 3oki one. Its ligand has a benzole ring with a chlorine substitute and two cyclohexane rings.

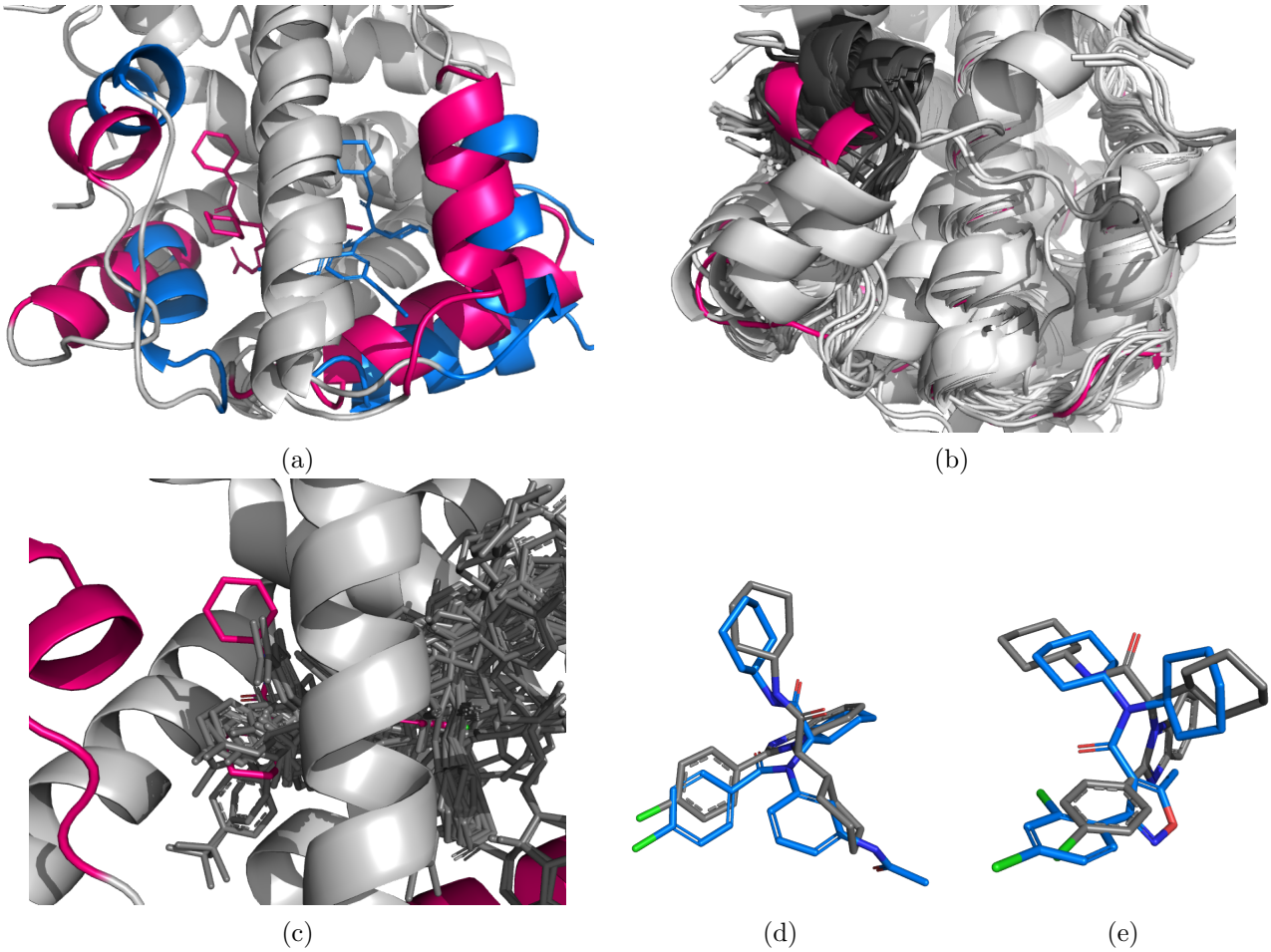


Figure 5.2: (a) Reference structure of the FXR apoprotein in complex with the FXR_18 ligand (red) and the result of docking to the 3oki protein (blue). Conformationally conserved chains are shown in grey. The two helices on the right have a high conformational freedom in all the homologous structures of FXR. (b) Co-crystal structure of apoprotein superposed with 36 proteins used in "multiple receptors" docking protocols. Here, the trp473–thr466 helix (the residue numbers may differ for different proteins) is shown in red for the reference structure and in dark grey for the other proteins. (c) Co-crystal ligand conformation and 30 ligands found in the highly homologous proteins (dark grey). (e) The best predicted ligand pose for FXR_18 (RMSD=9.34, blue) superposed with the 3oki ligand (chain A, grey). (d) The best predicted ligand pose for FXR_4 (RMSD=6.74, blue) superposed with the 3oki ligand (chain C, grey).

Figure 5.2 (d), (e) show the best predicted poses for FXR_4 and FXR_18 ligands superposed with the ligands of the 3oki structure. These poses, however, vary significantly from the correct solutions. Similarly to the case of FXR_18, the conformation of one of the 3oki helices differs from the co-crystal structure of FXR_4. Thus, the structure of the 3oki receptor would once again cause clashes with the co-crystal pose of the ligand.

We have previously mentioned the FXR_15 ligand, for which we were able to find only three receptors, docking to which allowed to obtain poses with RMSD values of about 3–3.5 Å (in exp. 2). For each of these receptors, the sulfonamide-containing part of the ligand is predicted correctly, while the flexible "tails" adopt other conformations to avoid steric clashes with the protein side chains. Overall, from the "self ligand to multiple receptors" experiments we can see that we would not be able to predict neither of FXR_4, FXR_15, and FXR_18 docking poses without the co-crystal protein structures even if we use the co-crystal geometries of the ligands.

Finally, we would like to mention the importance of the correct receptor sequence in the binding pocket. As it can be seen from a comparison of Table 5.3 with Table 5.4, the mean RMSD values

for the experiments with the mutated proteins are lower except for exp. 1. In this experiment, docking several ligands to low-homologous receptors became less successful after mutating a big portion of their residues.

5.4.3 Importance of the ligand geometry

In the results obtained for docking the ligand models to the co-crystal receptors (exp. 6), the spiro and sulfonamide-containing FXR_11 ligand docking poses have the lowest quality with RMSD of more than 10 Å. Figure 5.3 (a) shows the 2D structure of the FXR_11 ligand. As the self-docking results are close to the co-crystal poses, it seems that we have certain problems with the 3D structures of the FXR_11 ligand that we generated from its SMILES string. We should note that visually these structures look very similar to each other and to the co-crystal ligand.
Although the clusterization procedure helped to achieve a near-native pose for the ligand that we used in submission, the best scored poses of this ligand are 'flip-flopped' in comparison to the co-crystal pose resulting in high RMSD values of about 9 Å. Therefore, we re-generated its 3D structure once again in Avogadro, which lead to even worse docking pose of 10 Å RMSD and caused us to thoroughly examine the structure of this ligand. A closer look on the reference co-crystal ligand conformation reveals rather small differences in valence and trihedral angles that, however, seem to be crucial for the correct docking. In particular, we have found two regions where small changes in geometry lead to improvement of docking of the ligand generated in Avogadro to the reference receptor. More precisely, the C25 atom of the FXR_11 ligand has sp3 hybridization, although its geometry may be somehow distorted by the adjoined aromatic ring, while the value of the C26-C25-N13 angle in the reference structure shown in Figure 5.3 (b) is more than 120°. However, we were able to only generate structures with this angle's values varying from 111° to 115° depending on the force-field we used (MMFF94 or UFF), optimization precision, and the algorithm (OpenBabel or Avogadro). The second part of the molecule, whose geometry seems to be important, was the N1 nitrogen and atoms connected to it that are shown in Figures 5.3 (c)-(d). Here, the nitrogen atom has rather planar geometry, probably due to the vicinity of a sulfonamide group, while the structures that we generated in Avogadro were more tetrahedral, as one may expect from an sp3 nitrogen. Even if the co-crystal nitrogen geometry would be less planar, the chirality of N1 could lead to similar problems. What is interesting, although OpenBabel uses the MMFF94 force-field for the ligand optimization, it produced even more flat geometries than the co-crystal N1 had. More precisely, the normalized triple product of the bonds formed by N1 and its neighbors equals to -0.07 for the ligand generated in OpenBabel, while for the cases shown in Figure 5.3 these are equal to 0.24 and 0.52 for (c) and (d), correspondingly. Making changes in these two regions of the molecule lead to an improvement of the docking poses, as it can be seen in Figures 5.3 (e)-(f) and their relative scores compared to the results of docking to 3oki and 3oof receptors. The latter poses were highly scored but their conformations were very different from the reference.
Although this section emphasizes the importance of the initial ligand structure, we should not forget about the significance of the correct receptors' conformations. For example, the exp. 7 column of Table 5.3 lists docking results of the co-crystal ligand structures to multiple receptor models. Here, docking of the FXR_11 structure to 36 homologous receptors found in RCSB database did not yield any low-RMSD results. Even though the 3gd2 receptor has a very similar conformation to

the co-crystal structure of the FXR_11 receptor, one of its residues is too close to the co-crystal ligand structure, and thus we are only able to obtain a pose with RMSD value of 2.84 Å, where a half of the ligand conformation is predicted correctly, while the rest is distorted to avoid a clash. Docking with flexible side chains might be a solution in this situation, however the large size of the binding pocket entangles the choice of proper flexible residues in the blind experiment.

While we were able to detect two specific regions of the FXR_11 ligand that improve the docking results, we could not do the same for, for example, the FXR_2 ligand, although it is smaller than FXR_11. The FXR_2 ligand has only three rotatable bonds and aromatic rings that have rather classical, or "template", geometry. Moreover, all the structures that we generated with OpenBabel and Avogadro had some differences in valence angle values from the experimental structure, and the fact that one of our structures improved the docking results seems in some sense to be rather random.

Figure 5.4 shows a histogram representing differences of valence angle and bond length values between the 35 FXR structures that we generated using OpenBabel and the co-crystal structures. The majority of angles and bond lengths optimized by OpenBabel appear to be close to the co-crystal values. Indeed, as can be computed from the training data of our Knodle method [243], the mean standard deviation of the value of the valence angle for a given hybridization (sp2 or sp3, for example) is about 5°. The mean standard deviation of a bond length between atoms of the same type is about 5 pm. Figure 5.4 shows that only a few differences in angles and bond lengths between the modeled and co-crystal structures exceed these values. In particular, as it was already shown in Figure 5.3 (b), it happens for some angles formed by atoms connected to aromatic rings, while the differences for the aromatic rings themselves never exceed 5 pm. Considerable differences of terminal oxygens' local geometry may occur either due to ambiguous protonation states, or due to the inaccuracy in structural alignment. Other atoms with highly diverging geometries include several sulfurs, oxygens from ester and ether functional groups, and several atoms belonging to non-aromatic cycles.

5.5 Conclusions

Overall, the GC2 exercise provided us an excellent opportunity to assess the performance of multiple docking protocols on a set of ligands bound to the FXR receptor, whose multiple structures in different bound conformations can be found in the RCSB database. This challenge was the first time when we used VinaCPL. For the initial submission, we used the closest receptor template from RCSB as measured by the similarity of the bound ligand and obtained the mean RMSD value of 2.96 Å.

Later on, we ran a series of additional computational experiments, including self-docking, to evaluate the overall performance of different protocols. These included docking of ligands that we used in the submission, several ligands generated with other algorithms and co-crystal ligand structures to the receptors used for the submission, proteins with high sequence identity to the target apo protein, and experimentally obtained co-crystal apoprotein structures.

We have discovered two major hurdles in the successful predictions. A comparison of docking to the structures taken from RCSB and docking to the native receptors revealed the crucial role of protein flexibility. In the case of known correct protein structures, it was possible to obtain poses

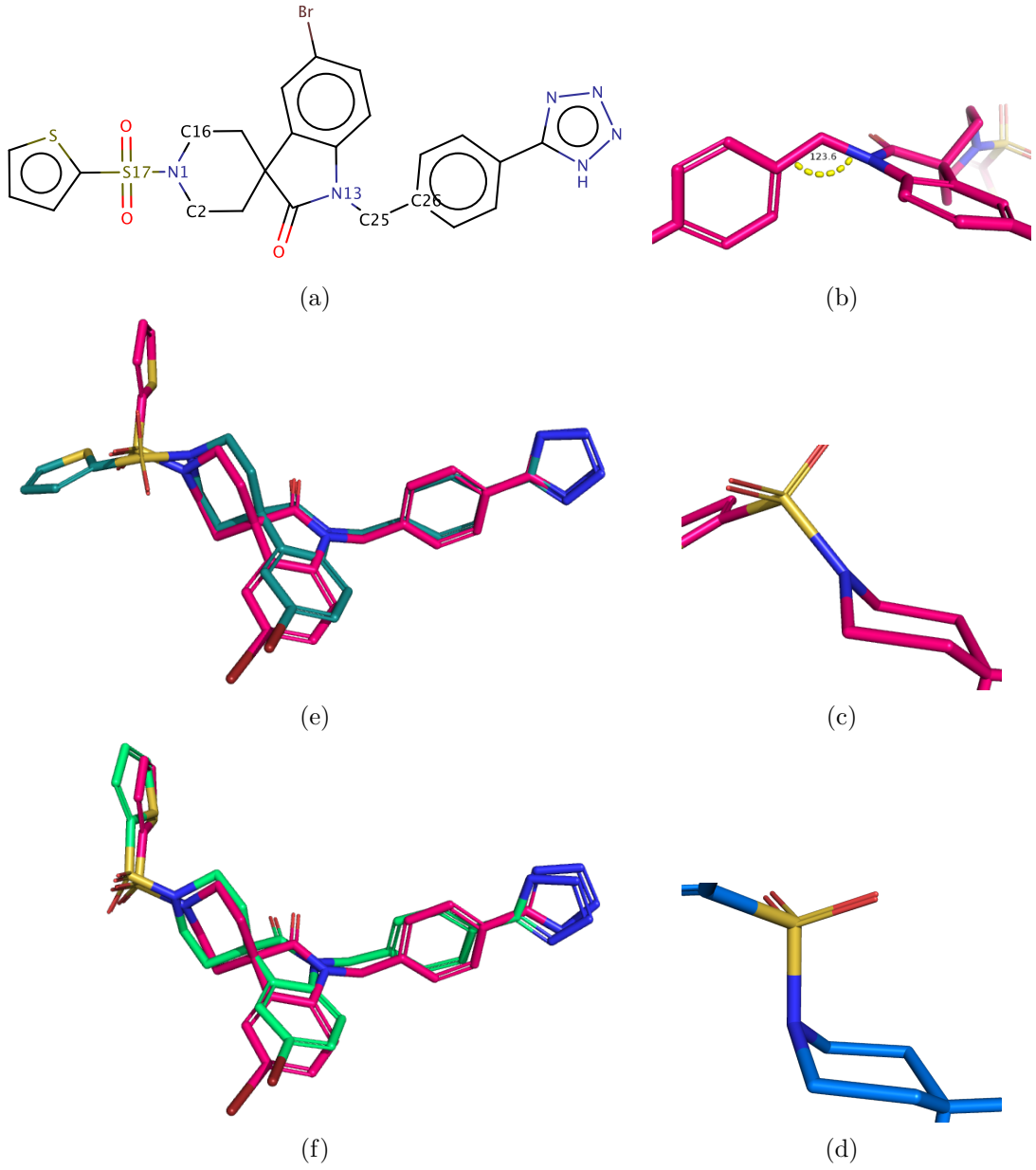


Figure 5.3: FXR_11 docking examples. Experimentally obtained reference structure is shown in red, the ligand that we generated in Avogadro is shown in blue. (a) 2D structure. (b) C26-C25-N13 valence angle of the co-crystal structure. (c)-(d) Planar geometry of the nitrogen atom. (e) The best-scored pose of docking of the structure with corrected C26-C25-N13 angle. (f) The best-scored pose of docking of the structure with corrected C26-C25-N13 angle and corrected N1 atom geometry.

with RMSD values close to 1 Å, which is more than twice better than our submission results. The wide binding pocket of the FXR apoprotein allows, on the one hand, docking of ligands of different sizes without any special procedures designed to enlarge or open the pocket. On the other hand, a scoring function may prefer selecting a smaller ligand in the wrong place of the binding pocket, especially if it is the binding site for a similar ligand, as we have seen in the example of the FXR_18 target. Applying a small number of binding site mutations for the highly homologous proteins turned out to be effective. Conversely, mutating of nearly a half of the binding site for the low homologous proteins worsens the docking results.

Another hurdle of our docking protocol was the ligand 3D structure generation and the resulting local geometries of ligands. As it was mentioned in the *Blind docking challenges* subsection of

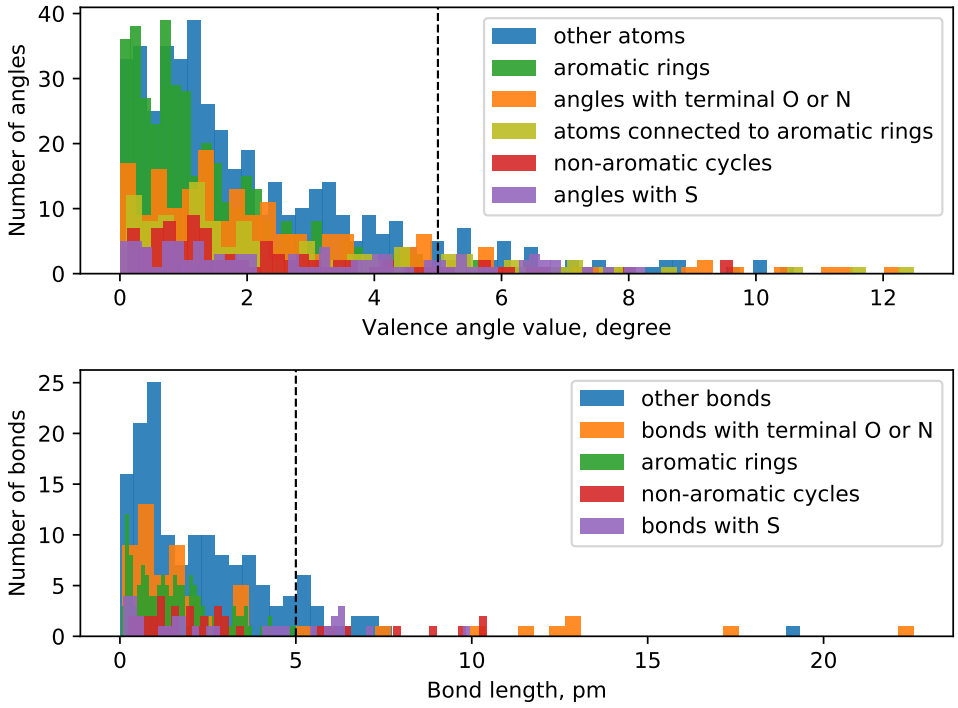


Figure 5.4: Valence angle and bond length differences between the ligands generated with OpenBabel and co-crystal structures. Dashed lines at 5° and 5 pm provide a visual indication of mean standard deviations of valence angles and bond length values, correspondingly, expected at the same hybridizations and bond orders.

Introduction, conformers generated by algorithms, which usually involve a force-field based minimization, may have slightly different bond length and valence angle values compared to the co-crystal structure. These small differences accumulate and it can be impossible to achieve a correct structure with dihedral angles sampling. We have shown in the example of the FXR_11 target that even small changes in the local geometry of the ligand cannot be bypassed by sampling in torsion coordinates and thus considerably affect the docking results. Our protocols evaluation demonstrates that docking experiments with the correct structures of ligands allow obtaining poses with better mean RMSD values than other approaches. However, the choice of a correct receptor still makes a major contribution to the quality of the final docking results. We should note that the importance of 3D ligand structures is not a widely discussed topic in docking literature, and in the current study we have tested only two 3D ligand structure building algorithms. For example, the subsequent Grand Challenge 4 has illustrated this problem in the case of macrocycle conformer generation. Conformer generation algorithms are often benchmarked by measuring the RMSD of the obtained conformers in an ensemble to the crystallographic ligand structures. However, in many non-rigid-body protocols, only one conformer of the ensemble is taken, and only its local geometry that is left unsampled during docking makes sense. Therefore, I suppose that a more thorough comparison of other methods on a wider dataset with an analysis of how well is the local geometry reproduced will definitely provide useful results.

As for the best blind docking strategy, our experiments have shown that the best results could be achieved for cross-docking of all ligands to all available highly homologous receptor structures. Depending on the chosen method of ligands generation, with this protocol, we obtained poses with RMSD varying from 2.15 to 2.33 Å versus the 2.96 Å from our initial submission without any visual

FXR id	self docking	submission	exp. 1		exp. 2		exp. 3	exp. 4		exp. 5		exp. 6		exp. 7
			mod	old	mod	old		mod	old	mod	old	mod	old	
1	0.347	5.467	5.898		3.050		3.050	3.050		0.549		0.549		4.076
2	0.304	5.166	3.088	3.083	7.931	7.453	7.453	3.088	3.083	0.447	7.453	0.447	5.158	6.448
3	0.466	5.196	6.928	6.905	3.694	3.678	3.679	3.985	3.990	1.982	3.678	1.982	5.032	1.281
4	1.346	3.499	6.119		6.741		6.741	6.741		6.741		1.372		6.638
5	0.157	0.773	0.678		0.678		0.678	0.678		0.673		0.673		0.620
6	0.313	0.474	0.412		0.425		0.425	0.425		0.391		0.423		0.456
7	0.865	1.212	1.176		0.883		0.883	0.883		0.748		0.748		0.991
8	0.789	0.668	0.976	0.706	0.951	0.723	0.723	0.951	0.723	0.951	0.723	0.883	1.631	0.603
9	0.462	0.455	0.482		0.495		0.495	0.495		0.530		0.645		0.396
10	0.561	4.558	4.616		3.199		3.199	3.648		3.199		1.158		2.227
11	1.027	6.355	7.164	7.010	3.268	5.223	8.929	10.007	2.840	3.268	5.223	10.014	1.047	6.905
12	1.177	4.811	6.311		3.187		3.187	3.385		1.308		1.308		2.568
13	0.337	5.686	1.179		0.679		0.679	0.679		0.392		0.392		0.502
14	0.503	0.729	0.588		0.553		0.553	0.553		0.617		0.635		0.434
15	0.265	4.146	9.677		4.875		4.875	9.088		1.277		1.277		4.300
16	0.333	7.622	9.341	9.909	2.313	3.751	3.751	2.313	3.751	2.085	3.751	2.085	1.620	1.064
17	0.709	9.562	9.528	6.820	2.224	2.158	2.158	2.224	2.158	2.224	1.931	1.667	1.931	1.381
18	0.405	6.411	7.504		9.207		9.207	9.207		0.653		0.653		8.805
19	0.359	0.626	0.704		0.695		0.695	0.695		0.653		0.653		0.623
20	0.627	4.006	0.848		0.824		0.824	0.824		0.969		0.749		0.700
21	0.478	0.527	0.717		0.680		0.680	0.680		0.660		0.786		0.518
22	0.259	1.876	2.332	1.922	2.248	1.956	1.956	2.248	1.956	0.783	1.746	0.783	2.016	2.075
23	0.325	7.285	7.350	7.281	2.745	3.726	3.600	2.745	3.557	2.745	1.488	2.806	1.488	2.542
24	0.428	0.805	0.683		0.629		0.629	0.629		0.627		0.584		0.424
25	0.390	0.670	0.533		0.556		0.556	0.556		0.556		0.486		0.416
26	0.447	0.451	1.039		0.509		0.509	0.509		0.634		0.634		0.849
27	0.434	0.929	1.371		1.395		1.395	1.395		1.389		1.389		0.412
28	0.608	0.774	1.004		0.503		0.503	0.503		0.439		0.439		0.386
29	0.528	0.812	1.027		0.704		0.704	0.704		0.660		0.660		0.853
30	0.575	1.358	2.164	1.948	1.696	1.056	1.630	1.696	1.056	1.530	1.056	1.530	1.815	1.608
31	0.626	1.702	0.992	1.658	1.003	1.068	1.068	1.003	1.068	0.998	1.068	0.998	0.937	0.893
32	0.456	2.019	1.968	2.015	2.046	2.032	2.082	2.046	2.032	1.995	2.018	1.983	1.994	2.131
34	1.552	4.906	4.568	4.551	2.886	2.390	2.390	2.866	2.751	1.729	1.422	1.305	1.422	2.270
35	0.774	0.883	0.895		0.826		0.826	0.826		0.544		0.559		0.841
36	0.774	1.159	1.352		0.952		0.952	0.952		0.952		0.974		1.029
mean	0.572	2.959	3.177	3.103	2.150	2.213	2.333	2.351	2.173	1.311	1.621	1.264	1.253	1.950

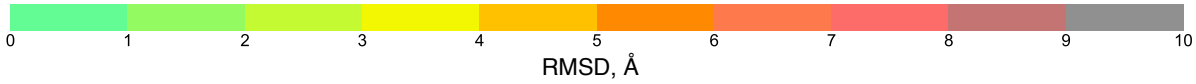


Table 5.3: Top-5 docking pose prediction results for the D3R Grand Challenge 2. The 'mod' column represents ligands that were created in Avogadro and minimized with either UFF or MMFF94 force-field. The 'old' one represents ligands that we used for the submission. All the RMSD values are computed with RDKit. The following experiments are listed in the table: self docking – docking of the co-crystal ligands to the co-crystal receptors; submission – results of the submission, where one ligand was docked to a single receptor for each of the targets; exp. 1 – "one ligand to one receptor"; exp. 2 – "one ligand to multiple receptors"; exp. 3 – "multiple ligands to multiple receptors"; exp. 4 – "one ligand to some receptors"; exp. 5 – "one ligand to multiple+self receptors"; exp. 6 – "one ligand to self receptor"; exp. 7 – "self ligand to multiple receptors".

inspection of the input structures and the obtained results.

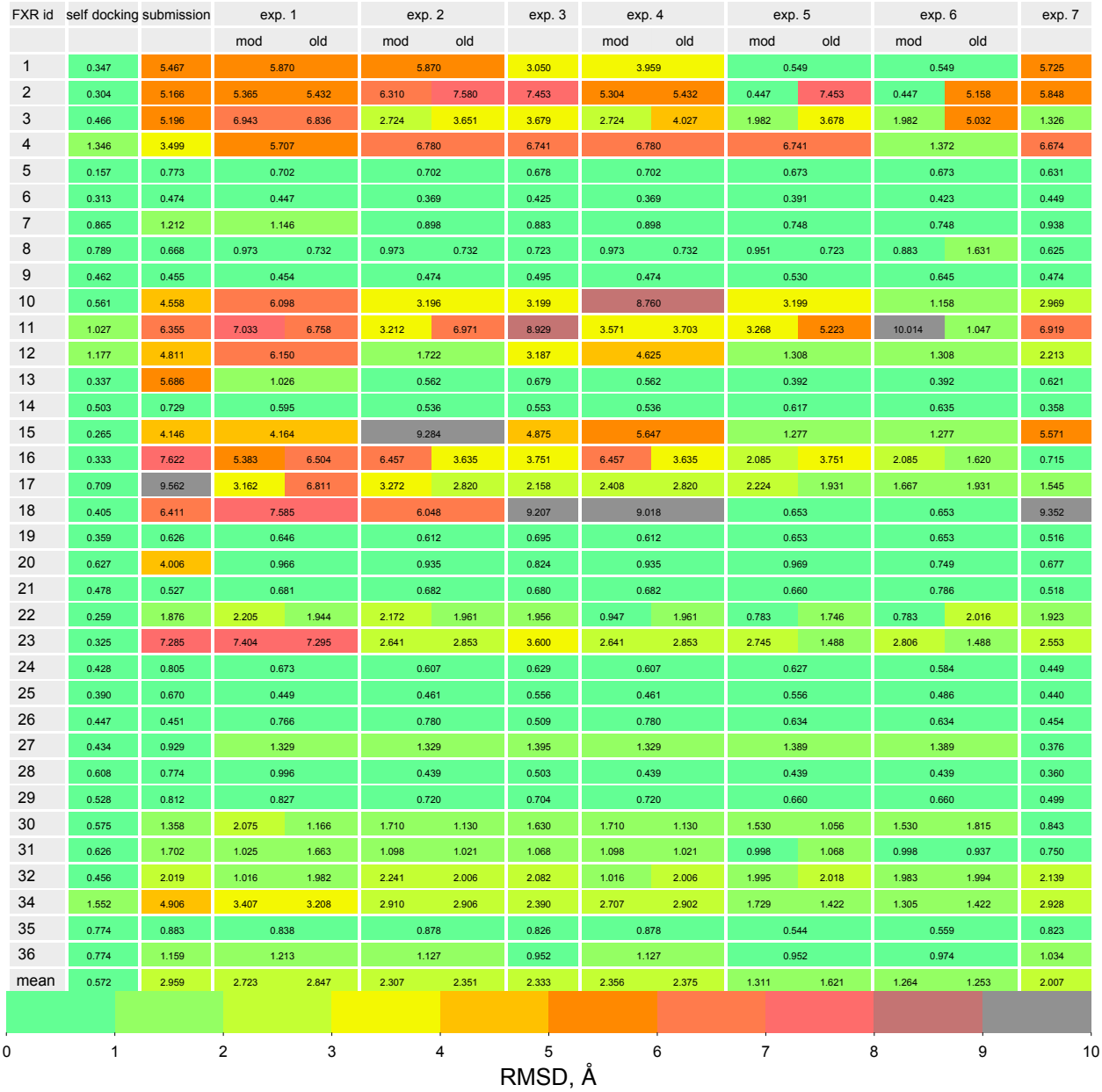


Table 5.4: Top-5 docking pose prediction results for proteins without mutations. The 'mod' column represents ligands that were created in Avogadro and minimized with either UFF or MMFF94 force-field. The 'old' one represents ligands that we used for the submission. All the RMSD values are computed with RDKit. The following experiments are listed in the table: self docking – docking of the co-crystal ligands to the co-crystal receptors; submission – results of the submission, where one ligand was docked to a single receptor for each of the targets; exp. 1 – "one ligand to one receptor"; exp. 2 – "one ligand to multiple receptors"; exp. 3 – "multiple ligands to multiple receptors"; exp. 4 – "one ligand to some receptors"; exp. 5 – "one ligand to multiple+self receptors"; exp. 6 – "one ligand to self receptor"; exp. 7 – "self ligand to multiple receptors".

Chapter 6. CAPRI round 41

Our team is regularly participating in another blind challenge called Critical Assessment of PRediction of Interactions (CAPRI). It is dedicated to the prediction of protein-protein, and more recently also protein-peptide, interactions, which is usually done by specifically designed protein-protein docking algorithms and scoring functions. For most of the CAPRI cases, our protein-ligand docking algorithms and Convex-PL as a scoring function are not the best choices. Nonetheless, we were able to apply Convex-PL in one of the CAPRI rounds, in which the target ligands were small saccharide molecules. This blind challenge consisted of two stages: pose prediction, and scoring, where the participants were asked to score all the complexes submitted during the pose prediction stage. We have not published a detailed report and analysis of participation in this round of CAPRI. Besides CAPRI 41, I have also participated in CAPRI 44 to assess the possibilities of scoring peptide conformations with Convex-PL. These results, however, are not presented in this thesis.

6.1 Challenge description

The target receptors of the CAPRI round 41 were two arabino-oligosaccharide binding proteins, AbnE, and a catalytic mutant (E201A) of AbnB. The organizers suggested to predict the binding mode of four arabinosaccharides consisting of 3, 4, 5, and 6 mono-saccharides (1,5-alpha-L-arabinohexaose (A6), 5-alpha-L-arabinopentaose (A5), 1,5-alpha-L-arabinotetraose (A4), and 1,5-alpha-L-arabinotriose (A3)). Participants were asked to predict binding of all the four compounds with AbnE, and the A5 compound with AbnB. This blind challenge consisted of two stages: pose prediction, and scoring, where the participants were asked to score all the complexes, submitted during the pose prediction stage.

6.2 Input structures preparation

6.2.1 Protein structure preparation

For the pose prediction stage, we modeled the target protein structures from the given sequences using MODELLER [284], I-TASSER [285], and IntFOLD3 [286] servers. We have also found several highly homologous structures in the Protein Data Bank. Notably, one of them, 3d5z, matched the second target of the exercise exactly, and was bound to a ligand of interest.
Moreover, to approach the protein flexibility problem, we generated an additional set of target protein conformations by applying *non-linear Normal Mode Analysis* implemented in the NOLB package [287]. It allows to find a collective motion that could lead to a conformational shift of a protein to the conformation of its known homologue, as a combination of oscillations along several modes. We superposed the structures of modeled receptors with a number of homologous structures obtained from the PDB that are listed in Table 6.1 to model the possible oscillations that could transform one protein structure to another one. We than saved 'snapshots' of the combinations of these oscillations along the trajectory, and used these conformations for docking. Some proteins required small mutations and missing residues recovery that I did with SCWRL4 [279].

4c1t 3d5z 3k00 3k01 3k02 **3d5z** 3d61 5ci5 5f7v

Table 6.1: PDB codes of the proteins that we used to model the receptor in CAPRI 41.

6.2.2 Ligand structure preparation

No SMILES strings and no stereoisometry was provided by the organizers. I created the SMILES strings from the arabinosaccharide ligands found in RCSB, mostly from the 3d5z one. I have generated the ligands using OpenBabel's [278] *gen3d* command. Conversion to the PDBQT format for AutoDock Vina's input was done with AutoDockTools.

6.3 Docking and scoring

I ran docking simulations in VinaCPL with all available receptor conformations. In addition to the Convex-PL score, our collaborators from the Institute of Biology Paris-Seine provided *conservation scores* (JET2 score) [288], ranking how often each residue is present in the homologous proteins. Such scores are designed to identify critical residues for a protein structure or functions. Since Convex-PL is a linear combination of pairwise interactions between each protein and ligand atom within a cutoff radius, we separately computed the interactions with each residue. Then, we computed a weighted sum of these interactions, taking the JET2 scores as weights of each residue. Finally, we re-scored the obtained poses with a linear combination of Convex-PL and Convex-PL-JET2 scores.

6.4 Results

Using the Convex-PL-JET2 scores (P24 predictor in the challenge results), we were able to obtain poses that were labeled by the CAPRI organizers as medium-quality in 4 of 5 cases, and high-quality poses in 1 of 5 cases. This is a good result in comparison with other predictors, as the number of medium-quality poses was rather low in 2 of 5 cases, and none of the participants acquired high-quality poses for any of the targets besides the one, in which we also acquired such predictions. The majority of the problems we encountered were driven by the saccharide nature of the ligands. For example, as it can be seen in Figure 6.1 , the two terminals of the A5 ligand differ by the presence of either the "-C-OH" group, or the "-OH" group, which contribute similar interaction potentials to the resulting score. Consequently, Convex-PL often equally scored poses with opposite orientations and considerably different RMSD values. It also seems that we did not

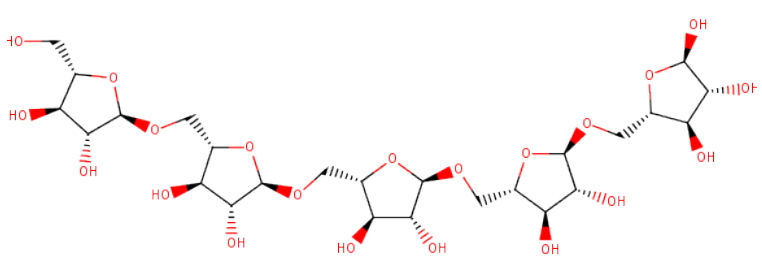


Figure 6.1: 1,5-alpha-L-arabinopentaose (A5) ligand

manage to correctly model the stereoisometry and the cycle geometry of the ligand input structures in some cases. Since the cycle geometries cannot be sampled on the fly in a docking program, this

could worsen the pose prediction. The problem with equivalent scoring of two opposite orientations of a ligand can be also related to the wrong stereochemistry of an input structure. None of the orientations could be close to the energy minimum owing to the problems in local geometry that might cause the loss of interactions important for binding. Moreover, besisedes the cycle geometries, the conformational space of the glycosidic linkages is, in reality, much more restricted than it could be modeled with AutoDock Vina, and especially with VinaCPL that has an extremly simplistic intramolecular energy representation.

Overall, docking of carbohydrates seems to be a challenging task that requires considerably gentler structure preparation and intramolecular energy computation, compared to what we had. After the end of this CAPRI round I found out that there exist a number of algorithms specifically for carbohydrate modeling [289, 290], and chemical libraries of glycans [291]. For example, there exist the so-called Carbohydrate Intrinsic (CHI) energy functions that represent the distributions of torsional angles of the glycosidic ligands from the Protein Data Bank [292]. Such energy functions can be integrated into the docking tools [292, 293]. Some participants of this round of CAPRI used glycan-specific protocols, for example, in Rosetta [294].

Chapter 7. D3R Grand Challenge 3

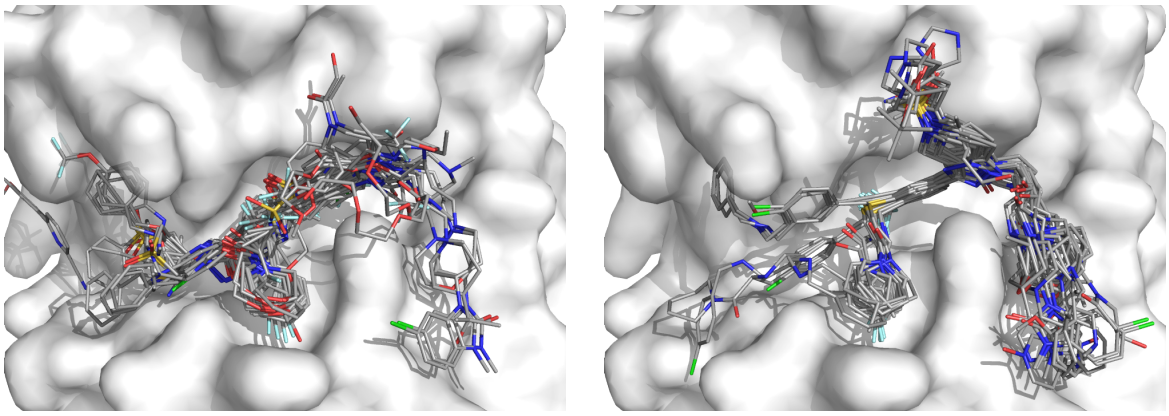
We have also participated in the D3R Grand Challenge 3 [172] sub-challenge with the Cathepsine S (CatS) receptor as a target protein. CatS inhibition is a target for regulating immune hyper-responsiveness [295]. The set of ligands contained quite flexible molecules, most of which simultaneously had several diverse chemically active groups. The results of this work are not published.

7.1 Challenge description

At this time, based on the lessons learned from the D3R GC2 results, the pose prediction Stage 1 of the Challenge was divided into Stage 1a, and Stage 1b, in which the crystallographic structures of receptors were revealed. We participated in all stages of the CatS sub-challenge. Structure files were provided by Janssen. Simultaneously to the CatS sub-challenge, several more affinity prediction subchallenges with different targets were held.

7.1.1 Challenge data

For Stage 1 and Stage 2, a set of 24 and 136 ligands was suggested, respectively, for the pose and affinity prediction. All of them were non-covalently binding. As always, input data was provided as a set of SMILES strings. Challenge organizers have also provided an apo structure of the receptor.



(a) (b)

Figure 7.2: Superposition of ligands binding to the cathepsine protein. All complexes were aligned. Co-crystal pose of the CatS receptor in complex with the CatS-14 ligand is shown for reference. (a) Ligands binding to the cathepsine receptors found in the Protein Data Bank. (b) Co-crystal poses of the 24 target ligands from the pose prediction stage.

7.2 Stage 1

I have generated the input ligands with two algorithms, OpenBabel's *gen3d* [278], and RDKit's *EmbedMolecule* [129] functions. As for the receptor, I collected 44 structures from the Protein Data Bank based on homology, and supplemented them with the two apo structures provided by the organizers. Since the binding pocket of CatS is rather flexible, I performed docking to all these 46 structures. PDBQT-formatted files for both ligands and receptors were created with AutoDockTools[137].

I used AutoDock Vina [70] with integrated Convex-PL as a scoring function to do the sampling. In this earlier version of VinaCPL, local optimization before the acceptance was restricted. For docking, I chose a cubic binding box with a 24 Å side. In Stage 1a we also used our visual inspection to detect the best poses. For Stage 1b we decided not to inspect the poses visually, and tried two sampling algorithms, one of which was our in-house development based on Rapidly-exploring Random Trees (RRT).
The binding site of CatS is quite spacious, as it can be seen in Figure 7.2, and we obtained quite a lot of rather different poses for each target ligand. Unfortunately, we were unable to obtain low-RMSD poses for a considerable number of target ligands neither for the unknown, nor for the known correct co-crystal structure of the protein. The best result we achieved was 3.9 Å average RMSD for the closest poses that was obtained in Stage 1a with additional visual inspection. Other submissions had even worse RMSD values spanning up to 8.9 Å in Stage 1b. The 8.9 Å-submission, however, was still better than the median of all Stage 1b submissions' average RMSD, that perfectly illustrates the difficulty of the target. Interestingly, I have noticed that poses generated with RDKit had higher scores than those obtained with OpenBabel.

7.3 Stage 2

The goal of the scoring stage was to predict relative binding affinities of 136 ligands in complex with the CatS protein. Native poses of 19 ligands were already known from the answers of the previous stage. I did structure preparation in a similar manner to Stage 1.
We revised the results of our rather unsuccessful docking experiments that were done for the pose prediction stage and tried to introduce several improvements of the protocol. First of all, we created an enhanced training set with more halogen and sulfonamide groups based taken from the Protein Data Bank, and re-traind Convex-PL on this dataset. This training set was later used to train the Convex-PL$^{5.2A}$ from Chapter 3. In this Challenge I also found out that the intra-ligand clash term of the Convex-PL score used in our AutoDock Vina modification leads to less accurate scoring, although it is still required for feasible ligand poses sampling. Therefore since the Grand Challenge 3, we started re-scoring all the poses with the usual Convex-PL version after docking. After the pose prediction stage answers were revealed, it turned out that all the target ligands were well aligned, as it is shown in Figure 7.1b. Therefore we introduced a simplistic ligand-based filtering protocol that preferred only those poses, in which the positions of the key functional groups were spatially close to those known from the first stage answers. With this approach we obtained affinity predictions with a 0.39 Spearman correlation coefficient, that were ranked 14 of 75 submissions.

7.4 Conclusions

Although our pose prediction results were relatively unsuccessful, D3R Grand Challenge 3 demonstrated a number of weak spots of our protocol. First of all, it showed me the importance of a better analysis of the binding pocket. Prediction of explicit water molecule positions could probably improve our docking results. Later on, it turned out that the implicit SASA and grid solvent model of Convex-PL^{R} was not able to correctly predict near-native poses neither. Another problem was related to the multiple aromatic and halogen-containing groups of the Challenge's ligands that require a better representation of oriented interactions in a scoring function. While both versions

of Convex-PL were unable to properly score them, partially orientation-dependent KORP-PL potential was able to successfully detect near-native poses in the benchmarking experiments that we did after the end of the Challenge.

Overall, an analysis of the publications describing user submissions to the Challenge from subsection 1.3.5 implies that the most successful structures were template-based. Classical fully structure-based docking pipelines, like the one we used for Stage 1a, did not provide good RMSDs in the majority of the cases, except the Iorga's group submission [219], if considering the fully automatic protocols without human intervention. Our own predictions in the scoring exercise of Stage 2 were perhaps relatively successful because of the application of ligand-based filtering.

Chapter 8. D3R Grand Challenge 4

In 2018 we participated in the D3R Grand Challenge 4 sub-challenge targeting the beta secretase 1 (BACE) receptor [173]. Both the submission and the analysis of the docking results were performed by me. These results were published in the Journal of Computer-Aided Molecular Design [228].

8.1 Challenge description

8.1.1 Challenge structure

Similar to the Grand Challenge 3, Grand Challenge 4 consisted of the pose prediction Stage 1a and Stage 1b with unknown and known co-crystal receptor structures, and of a scoring Stage 2. The goal of Stage 1 was to predict the correct binding poses of the ligands. Later on, Stage 2 targeted affinity or free binding energy estimation for a larger set of ligands. It was also possible to participate in the affinity prediction in both substages of Stage 1. However, we only took part in pose prediction parts of Stage 1 substages, and in Stage 2.
This challenge provided interesting examples of macrocycle docking. Macrocycles are often described as large non-peptidic cyclic molecules. Modeling of cyclic molecules generally poses multiple computational challenges related to the preservation of molecular topology upon sampling of cycle conformations. One of the approaches consists in splitting the macrocycle, sampling its split parts, followed by solving the loop closure problem. There are efficient sampling methods specifically developed for cyclic peptides [296]. However, to the best of our knowledge, there are no free [297] methods for macrocycle sampling in torsion coordinates, which are essential for computationally efficient docking protocols.

8.1.2 Challenge data

BACE is a transmembrane aspartic-acid protease that is responsible for the cleavage of the amyloid precursor protein. This leads to amyloid-β peptide formation [298]. Beta amyloid is the main component of amyloid plaques found in brains of Alzheimer's disease patients, therefore activity regulation of beta-secretase is one of the promising Alzheimer's treatment strategies [299].
BACE substrate is normally a polypeptide in the extended β strand conformation. Potential BACE inhibitors are designed to mimic this property, which can be achieved with macrocyclization [300]. BACE binding pocket contains several sub-sites, which are partially or totally occupied by the inhibitor [301, 302]. One of the types of aspartic protease inhibitors are hydroxyethylamine-containig compounds, binding with hydrogen bonds to the aspartate residues.
This challenge focused on 158 hydroxyethylamine inhibitors provided by Novartis. Twenty of them were used in the pose prediction of Stage 1. These were one acyclic and 19 macrocyclic compounds. Later on, 154 inhibitors were used in the affinity (IC50) prediction of Stage 2. Most of them were cyclic with the cycle length varying between 14 and 17 atoms, with diverse substituents and cycle structures. In this chapter we will refer to these compounds as to BACE_[ID], with ID ranging between 1 and 158.

8.2 Methods used for docking and scoring

Below I will briefly describe the computational approaches used for docking and scoring throughout the challenge. I was adapting the algorithms used for ligand conformer generation and some of the scoring function parameters between the stages based on the analysis of the previous results. Therefore, our structure preparation procedures and submission protocols will be described and analyzed in the Submission protocols and discussion subsection and in the discussion of results.
I applied the following docking pipeline in all the stages. Binding pocket was centered on the co-crystal ligand geometrical center. Box sizes were set to (22, 22, 25) Å with respect to the orientation of the original structure. All ligand conformations were cross-docked to all the chosen receptors with VinaCPL. I generated 400 poses for each ligand conformation for the subsequent re-scoring. In the AutoDock Vina configuration files, the parameter *num_modes* was set to 400 and *exhaustiveness* to 10. Local optimization was applied to only those poses that were accepted after a MCMC step. PDBQT-formatted structures were generated in the AutoDockTools package [137], where I kept all rotatable bonds in the ligands to be flexible. Explicit hydrogens were removed from the molecules. In our parametrization, ligand protonation states are defined by the atom types, which are assigned according to the ligand 3D geometry. These were generated from the provided SMILES strings using RDKit functions, as it is explained in more detail below. Receptor atom types corresponded to those at neutral pH. Receptors were considered to be rigid.
Then, I re-scored the obtained poses with the Convex-PL potential, and the earlier versions of the Convex-PL^R that are referred in this section as *enhanced Convex-PL*. The captions of evaluation tables list the description of the Convex-PL parameters I used during the computational experiments. I used the same grid solvent representation and ligand flexibility estimation, as in Convex-PL^R. The number of SASA terms, however, was slightly different. One of our submissions also included terms that approximated the conformational entropy of the receptor sidechains, computed in a different way than in Convex-PL^R. I estimated the entropy using a volume accessible to each of the sidechains normalized by its solvent-accessible surface area. Then I computed a set of 20 descriptors, one per each of the amino acid types, using the following equation,

$$\text{receptor flexibility}_a = \log \prod_{i}^{\#\ residues_a} v_i \frac{s_{i,unbound}}{s_{i,single}},$$

where the product is taken over all amino acids of the same type located at the interface with the ligand. Here, a is a type of amino acid, v_i is a precomputed constant volume of a sphere that is obtained by the rotation of the sidechain of type a around its C_β carbon, $s_{i,unbound}$ is the solvent-accessible surface area of the residue i computed for the receptor molecule in the unbound state, and $s_{i,single}$ is the total surface area of the same residue, if it is extracted from the receptor.
Finally, the best poses were clustered with the 0.5 Å threshold using the best-scored structures as seeds for the new clusters. The resulting scores in Stage 2 were averaged over the top 10 predictions for each compound.
I computed symmetry-adapted RMSD values with the *CalcRMS()* function from the RDKit package [129]. Unlike the Grand Challenge 2, the RMSD values I obtained corresponded to those reported in the official evaluation results. Receptor alignment was done with the PyMOL 1.8.6 [96] *align* function.

8.3 Stage 1a

For the first stage, we intended to use a simple and robust protocol with a minimal amount of user intervention, and also without using ligand-based approaches. Therefore I chose cross-docking of flexible ligands with multiple conformations of rigid cycles, to several receptor structures.

8.3.1 Structure preparation

Starting from the provided SMILES strings, I generated 1,000 3D conformations for each macrocyclic ligand using RDKit's [129] *EmbedMolecule()* function [119] with default parameters. I then clustered these conformations with respect to the pairwise locations of the cycle atoms using hierarchical clustering from *scipy.cluster.hierarchy* with a threshold of 0.2 Å. One conformation from each cluster was then selected for docking. For the acyclic BACE_20 I generated one conformation using RDKit's *EmbedMolecule()* function.
The Protein Data Bank contains more than 300 highly homologous structures of the BACE receptor, whose binding site seems to be rather conserved. Out of these 300 receptors, I selected 38 fully homologous structures for the acyclic BACE_20 docking. Nine of them were crystallized together with cyclic ligands and thus I chose them for the BACE_1-19 docking. Table 8.1 lists the PDB codes of selected structures. Apart from removing solvent molecules I did not do any other modifications of the selected structures.

2f3e	2f3f	3dv1	3dv5	3k5c	4dpf	4dpi	4gmi	4k8s	2fdp	2g94	2hm1	2iqg
2p4j	2qk5	2qmd	2qmf	2qmg	2qp8	2zjn	3cib	3cic	3dm6	3duy	3i25	3ixj
3ixk	3k5d	3k5f	3k5g	3kyr	3l58	3l5e	3lnk	3veu	4gid	4k9h	5dqc	

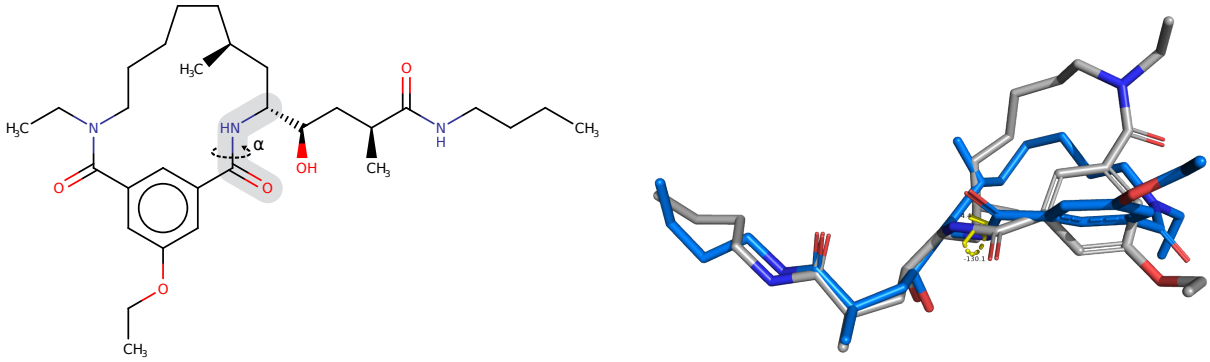
Table 8.1: PDB codes of protein structures selected for Stage 1a docking. Structures highlighted in gray were used for docking of the acyclic BACE_20 ligand only.

8.3.2 Evaluation results

It turned out that all cyclic ligand conformations generated by RDKit had an incorrectly sampled dihedral angle between the atoms of an amide group leading to a cis conformation instead of the native trans one. This angle is denoted as α in Figure 8.1, and is a part of all the cycle-containing ligands of Stage 1. This resulted in completely wrong geometry of the whole neighborhood of the amide group, which could not be fixed by docking due to the macrocycle rigidity. An example of an incorrectly predicted cycle conformation is shown in Figure 8.1, where the inclination of the cycle plane is different from the native geometry. In many cases, this also lead to flipped and shifted ligand docking poses, which produced high RMSD values. I have noticed this amide bond sampling problem at the very end of the Stage 1a timeframe, and submitted two predictions where the flipped and shifted poses were rejected based on the cycle similarity with the co-crystallized ligands. One more submission also used visual inspection. Overall, improper cycle conformations lead to lower than average and average in case of the manual or automatic pose rejection results listed in Table 8.2. Using the automatic pipeline without rejection of unrealistic poses, we obtained satisfactory low RMSDs for only a few ligands, one of which was the acyclic BACE_20.

id	scoring function	rejection of unrealistic conformations	visual inspection	mean RMSD, Å		
				average	closest	top-1
biw3a	enhanced Convex-PL	✓	✓	1.99	1.40	1.82
jit54	enhanced Convex-PL	✓	-	2.78	1.72	2.64
bsrv5	enhanced Convex-PL	✓	-	2.88	1.77	2.64
buck5	enhanced Convex-PL	-	-	3.90	2.52	3.99
maej5	enhanced Convex-PL	-	-	3.92	2.57	3.99
s4fu0	original Convex-PL	-	-	5.45	3.77	5.47

Table 8.2: Stage 1a evaluation results. Here I applied different versions of the enhanced Convex-PL function. The *jit54* and *buck5* submissions included the type-specific interactions with displaced solvent and Convex-PL score computed with a 5.2 Å distance cutoff. The *bsrv5* and *maej5* submissions included the solvent-accessible surface areas and the Convex-PL score computed with a 5.2 Å distance cutoff. In the *biw3a* submission, I chose the highest-ranked poses scored with the three versions of Convex-PL used in all the other Stage 1a submissions, and rejected some poses based on visual inspection.



(a) (b)

Figure 8.1: BACE_1 ligand. (a) Incorrectly sampled torsion angle of the amide group present in most of the 158 compounds is highlighted in light gray. On average, the dihedral angle α's value differs by more than 100° from the ones found in crystallographic structures. (b) The native ligand conformation is shown in blue, our top-scored pose is shown in gray. It can be seen that the wrong α value leads to the incorrect conformation of the cycle.

8.4 Stage 1b

8.4.1 Structure preparation

For Stage 1b, crystallographic structures of all the receptors were revealed by the challenge organizers, and we used them to repeat the docking calculations. I removed the water molecules, and no other additional modifications were applied to the receptor structures.

Learning from the Stage 1a experience, I changed the way to sample ligand cycles. Initially I only tried to sample more conformations (up to 10,000). However, it turned out that in all of them RDKit produced the wrong α value of the dihedral angle despite different combinations of parameters in the *EmbedMolecule()* function. I then tried to minimize all conformers using a force field with a constraint on the wrongly predicted dihedral angle. The constraint applied with the UFF force field implemented in RDKit did not affect the final results. Also, constrained minimization using the MMFF94 [303] force field resulted in very distorted structures. Although at this stage it could have been possible to simply use another tool for conformer generation, not all of them are free, and I also felt being somewhat challenged to make RDKit generate better conformations. Finally, I decided to try the *coordMap* option of the *EmbedMolecule()* function, which rejects conformations

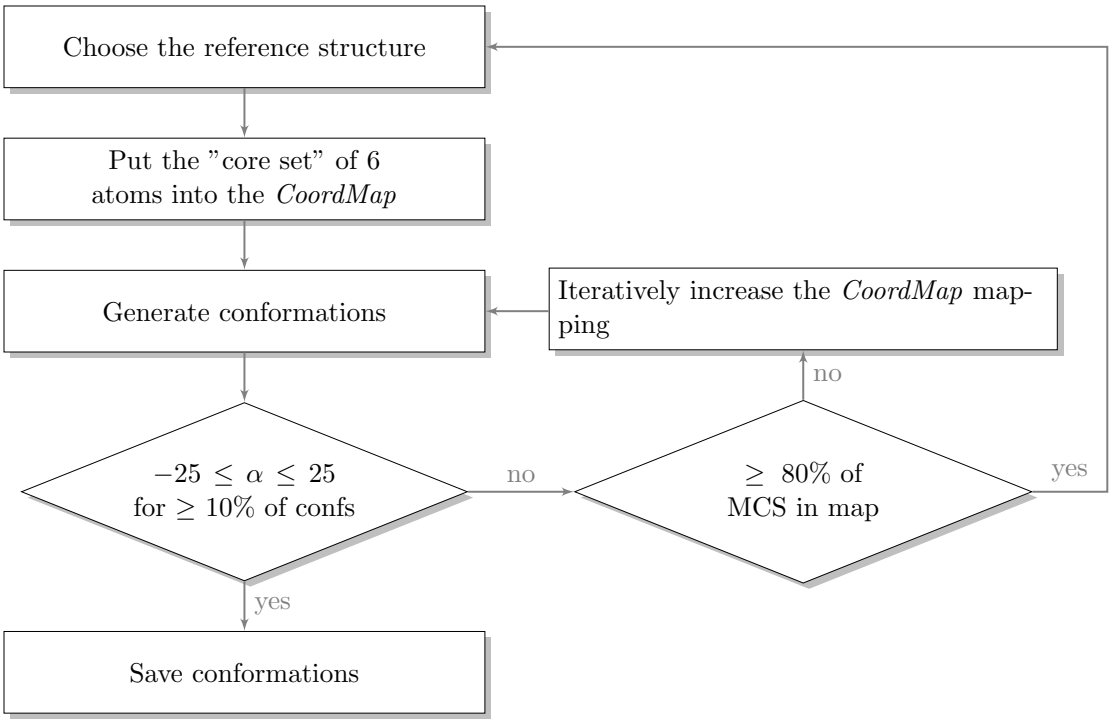
where the distances between specified atoms' positions are different from those passed through the *coordMap* argument, up to a certain threshold. When using only the 4 dihedral angle atoms, conformational sampling results did not change and the angle was still wrongly sampled. I have tried to tweak internal threshold of this map-based reduction in the RDKit source code, but it did not improve the results. Therefore, I increased the size of the map, pushing ourselves to a more ligand-based setup. Figure 8.2a schematically represents an algorithm for the map generation used for cyclic ligands.

I started with computing the maximum common substructures (MCS0) between the cycles (including non-rotatable cycle substituents) of each target ligand and the cycles of the 9 ligands co-crystallized with proteins listed in Table 8.1. I also computed the maximum common substructures between the entire ligands (MCS). For each target ligand, I chose a reference ligand based on the MCS0 size. Then, I selected 4 atoms corresponding to the wrongly predicted amide group, and two carbon atoms bound to them, including one from the hydroxyethylamine group. These are shown in yellow in Figure 8.2b and will be referenced as a "core set". The mapping of these 6 atom indices in the target ligand structure to the coordinates from the reference ligand structure were provided as a *coordMap* argument to the conformer generating function. I then computed α value of the generated conformers. If more than 10% α values were lying between $-25°$ and $25°$, I saved the conformers and proceeded to the next target ligand. If not, I iteratively increased the map based on a set of rules illustrated in Figure 8.2b until 10% of structures would have the correct amide bond conformation. If more than 80% of the MCS was included into the map without providing good conformers, I moved to the next reference structure. If three reference structures were not sufficient, I aligned them to each other based on the coordinates of the atoms of the "core set", and used the union of the MCSs of both reference molecules to create a new mapping. After at least 10% of good conformations was achieved, I stopped the algorithm and saved the molecules. If $\geq 70\%$ of conformations were generated with α values inside the $[-25°, 25°]$ threshold interval, I squeezed this interval to $[-10°, 10°]$ and rejected outlying conformations.

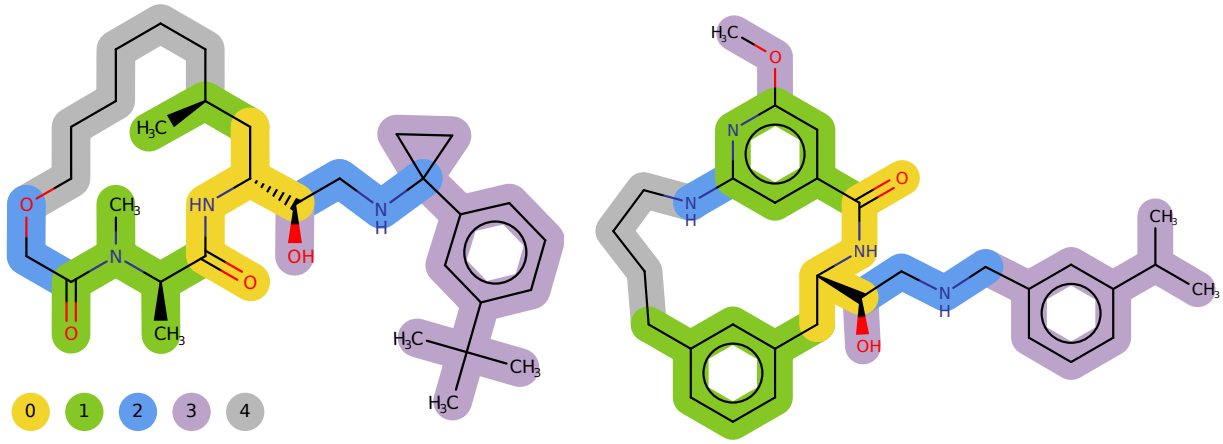
Overall, even though I did not manage to find out what exactly led to the cycle sampling problems, this approach finally allowed us to create structures with correct α angle for all macrocyclic targets.

8.4.2 Evaluation results

This approach lead to low-RMSD results, summarized in Table 8.3. The mean RMSD of the closest pose of all our submissions was less than 1 Å. Figure 8.3 shows several examples of the poses I obtained in Stage 1b. The enhanced versions of Convex-PL on average predict binding poses more accurately compared to the original version. For example, the top-1 ranked pose of the BACE_12 ligand in the *dhueb* submission was considerably shifted and rotated with respect to the native pose, which resulted in the 10.53 Å RMSD. In the *nyrou* submission I obtained 0.80 Å RMSD. However, the biggest contribution to this performance improvement was driven not by the additional descriptors, but by the change of the interaction cutoff distance to 5.2 Å, which is smaller than the default value of 10 Å. This smaller cutoff value was used to train the enhanced versions of Convex-PL in the *nyrou* and *vfkn2* submissions.



(a) Algorithm 1. Schematic description of an algorithm for conformer generation in RDKit driven by distance constraints. It was implemented in python3 using RDKit. Please see main text for more details.



BACE_4 BACE_12

(b) Examples of ligand mapping priority. Each color represents a different priority, which are ranked from 0 to 4. On each iteration of the algorithm an atom (or a group of atoms in case of rings) was added to the map with the following priorities. (1) Atoms with minimal topological distance from the "core set", amide groups of the cycle, aromatic substituents topologically close to the "core set". (2) Carbons and nitrogen of the hydroxyethylamine group, non-carbon atoms of the cycle. (3) Atoms of the "tails", oxygen of the hydroxyethylamine group. (4) Rest of the macrocycle atoms topologically far from the "core set", hydroxyl and carboxyl substituents of the macrocycle. I tried to use as few of these atoms as possible since they adopt the most diverse conformations as compared between the cycles, and I would not like to occasionally freeze them.

id	scoring function	mean RMSD, Å		
		average	closest	top-1
nyrou	enhanced Convex-PL	0.98	0.84	0.89
vfkn2	enhanced Convex-PL	0.99	0.84	0.89
mjevm	enhanced Convex-PL	1.14	0.79	1.00
dhueb	original Convex-PL	1.56	0.90	1.60

Table 8.3: Stage 1b evaluation results. Enhanced version of Convex-PL used in the *nyrou* submission was trained on the interactions with the volume displaced solvent and the original Convex-PL score computed with a 5.2 Å cutoff. The *vfkn2* submission included solvent-accessible surface area descriptors and the Convex-PL score computed with a 5.2 Å cutoff. Scoring function used in the *mjevm* submission included included interactions with the volume of displaced solvent and the original Convex-PL score computed with a 4.8 Å cutoff.

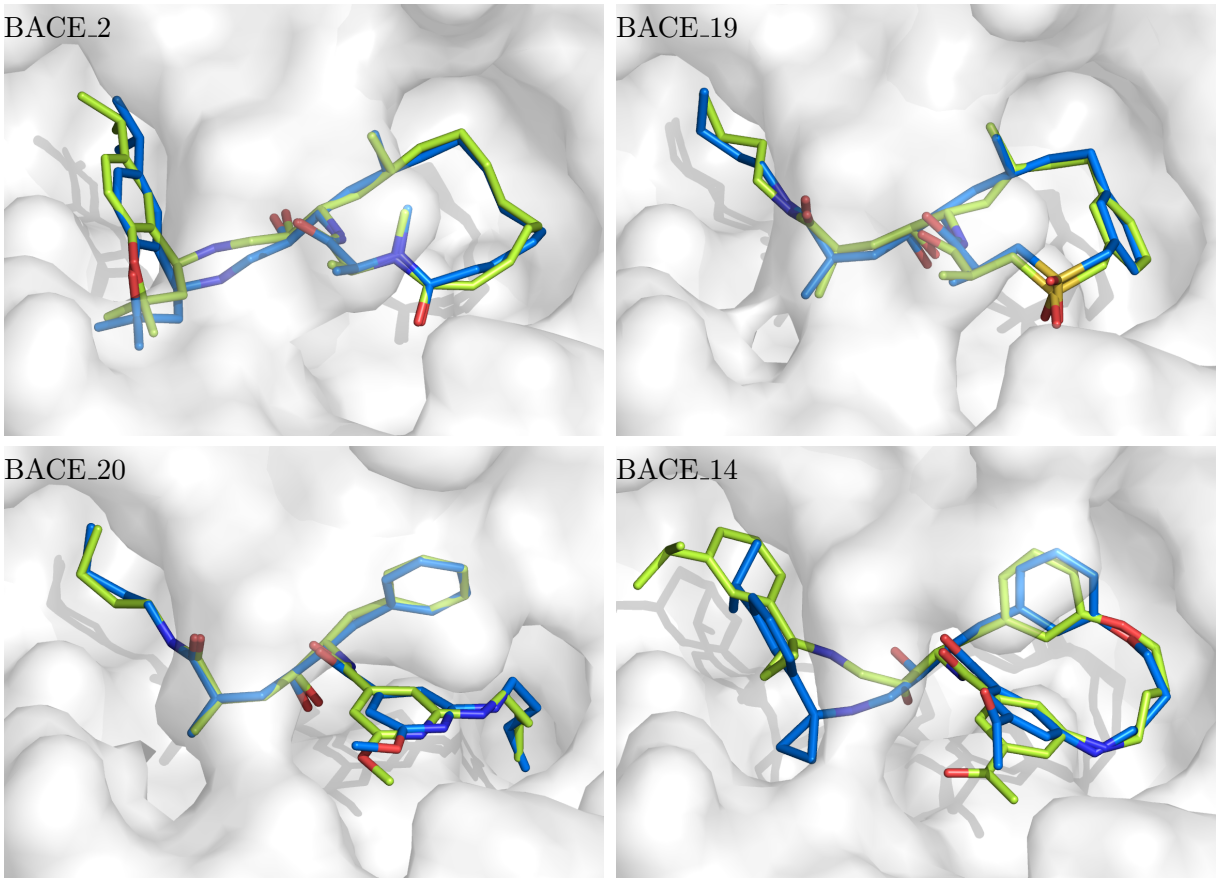


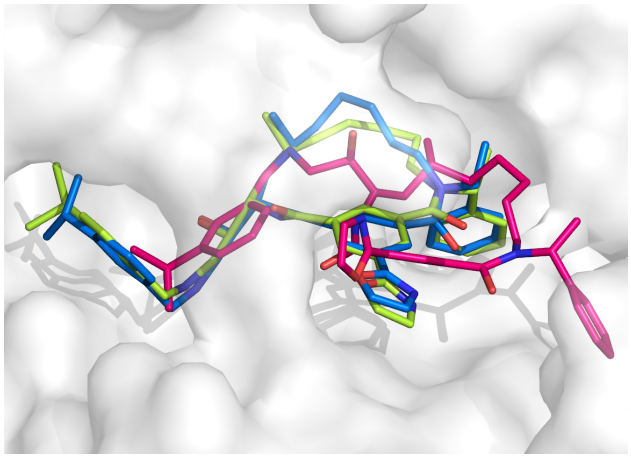
Figure 8.3: Examples of the closest poses from our Stage 1b *nyrou* submission. Crystallographic structures are shown in blue, our predictions are shown in green. Bond orders are not shown.

scoring function	mean RMSD, Å		
	average	closest	top1
enhanced Convex-PL	1.54	0.89	1.22

Table 8.4: Stage 1a redocking results. The scoring function includes the solvent grid features, atomic SASA values, and the original Convex-PL$^{5.2\ A}$ score.

The low contribution of additional descriptors can be explained by the fact that all of them are related to the interactions that a molecule could have with displaced solvent. The BACE binding pocket is not very open to solvent, and the fraction of ligand surface that could be exposed to solvent does not change much even between the poses with 10 Å RMSD difference. Therefore, the sums of additional descriptors' contributions were very close to each other for the majority of ligand poses.

8.5 Stage 1a redocking



(a) BACE_7 ligand poses. Crystallographic structure is shown in blue, our Stage 1a prediction from the *buck5* submission is shown in red, the redocking pose is shown in green. Bond orders are not shown.

To check how did the macrocycle conformer quality influenced the results of Stage 1a, we repeated the ensemble docking of the BACE_1-19 ligand structures prepared for Stage 1b to the set of 9 receptors used in Stage 1a.
As it could be expected, better ligand structures considerably improved the pose prediction. Without manual inspection or pose filtering, we obtained the subangstrom mean RMSD value for the closest pose shown in Table 8.4. Figure 8.4a illustrates the redocking pose of the BACE_7, which is superimposed with the one I submitted for Stage 1a. Here it can be clearly seen how did the bad initial conformation from our submission (red) lead to a considerable shift of the ligand inside the pocket (blue).

8.6 Stage 2

Stage 2 was dedicated to the scoring exercises. The goal was to correctly predict the relative binding affinities of the set of 154 molecules binding the BACE receptor. The 20 crystallographic structures of complexes from Stage 1 were already revealed at this stage.

8.6.1 Structure preparation

Since the amount of computations required for docking of all the 154 compounds was considerably higher compared to Stage 1, and more protein structures became available for docking, I first selected a set of target structures for each compound. The BACE_1 – BACE_20 ligands were docked into the co-crystal receptors. For the rest of the cyclic ligands I first extracted the fragments containing the macrocycle only, and the macrocycle with some substituents, such as aromatic rings. I then computed the maximum common substructures of these fragments with the ligands with known co-crystal structures, and selected the receptors with maximum MCS size resulting in 4 - 12 receptors per each compound. Receptors for the acyclic BACE_145 and BACE_146 ligands were chosen based on the overall MCS size.
To create the cyclic ligand structures, I followed the algorithm applied in Stage 1b with several

modifications. The pool of reference ligands now included the 20 co-crystal structures from Stage 1. In some cases, I visually inspected the results and supervised the process of macrocycle structure generation.

8.6.2 Evaluation results

I ran out of time and have not finished docking of all the conformations of macrocyclic molecules. I have submitted two sets of predictions containing about 60% and 80% of all docked conformations to see how the result will change depending on these numbers. This resulted in Kendall τ of 0.12 for the first subset's best prediction, and 0.14 for the second, which are listed in Table 8.5. I can see that regardless the cutoff value, the ligand flexibility descriptor, which estimates the conformational entropy change upon binding, improved the results in all the enhanced submissions. The scoring function used in the submission with the highest Kendall τ, *xx4i5*, was trained on both solvent-related and entropy-related descriptors. Unlike the Stage 1 pose prediction exercise, where solvent-related descriptors almost did not contribute to the comparison of the poses, here they do influence the results, since binding poses of different ligands are now compared to each other.
I have also evaluated the ability of our enhanced scoring function to predict binding affinities based on the docking poses generated by other predicting teams. To do so, I firstly rescored all the available submissions of structure-based predictor teams with the scoring function used in the *xx4i5* submission. Secondly, I also applied local optimization to the ligand positions in the binding sites using VinaCPL. I then recomputed the affinity scores. Figure 8.5 shows the rescoring results. One can see that our approach does not improve the predictions of the best submitters (those with Kendall $\tau > 0.15$). Local optimization improves the results from 0.09 to 0.11 τ averaged over all the predictions without and with local optimization, respectively. Our own submissions got also slightly improved after the re-scoring.
Interestingly, we obtain relatively good affinity predictions with Kendall τ equal to 0.24 when using the docking poses submitted by the second-best structure-based affinity predictor *urt76*. This result gets worse if the local optimization is applied prior to computing the affinities. This fact indicates that we probably failed to predict correct binding poses for some of the compounds from the Stage2 dataset. In the meantime, the evaluation of Convex-PL from the previous Part shown in Figure 4.8, which was done on co-crystal poses, produced similar correlation to the one computed in our submission. Although Convex-PL^R is closer to the scoring function used in *xx4i5* submission than Convex-PL is, its evaluation on the same co-crystal dataset resulted in a worse correlation.

8.7 Conclusions

This docking exercise provided us a unique opportunity to model macrocyclic ligands that bind to protein targets. The conformational space of macrocycles is quite restricted by their topology, and sampling of the cycle conformations in torsion coordinates is a challenging problem. We have started with a fully structure-based and automated docking procedure with pre-generated rigid macrocycles. However, at the end of Stage 1a, I analyzed the docking results and discovered that the majority of our macrocycle conformations had unrealistic geometry. Therefore, I supplemented the docking protocol with constraints based on the structure of similar ligands. Finally, we converged to a stable pipeline that resulted in sufficiently low subangstrom RMSDs of binding poses.

id	scoring function	% initial conformations docked	Kendall's τ	Spearman's ρ
xx4i5	enhanced Convex-PL	80%	0.14	0.21
dzyxt	enhanced Convex-PL	80%	0.13	0.19
u7r6y	enhanced Convex-PL	80%	0.12	0.19
kzsv5	enhanced Convex-PL	60%	0.12	0.18
i88wa	original Convex-PL	80%	0.12	0.18
q6mvt	enhanced Convex-PL	60%	0.11	0.16

Table 8.5: Stage 2 affinity prediction results. Submissions *dzyxt* and *kzsv5* were scored only with two descriptors, the Convex-PL score computed with a 10 Å cutoff and the ligand flexibility. The *u7r6y* submission was scored using the ligand flexibility and the Convex-PL score computed with a 5.2 Å cutoff. The *xx4i5* and *q6mvt* submissions correspond to the scoring function trained on interactions with the volume of the displaced solvent, SASA values, ligand flexibility, flexibility of the interacting receptor residues, and the Convex-PL score computed with a 5.2 Å cutoff.

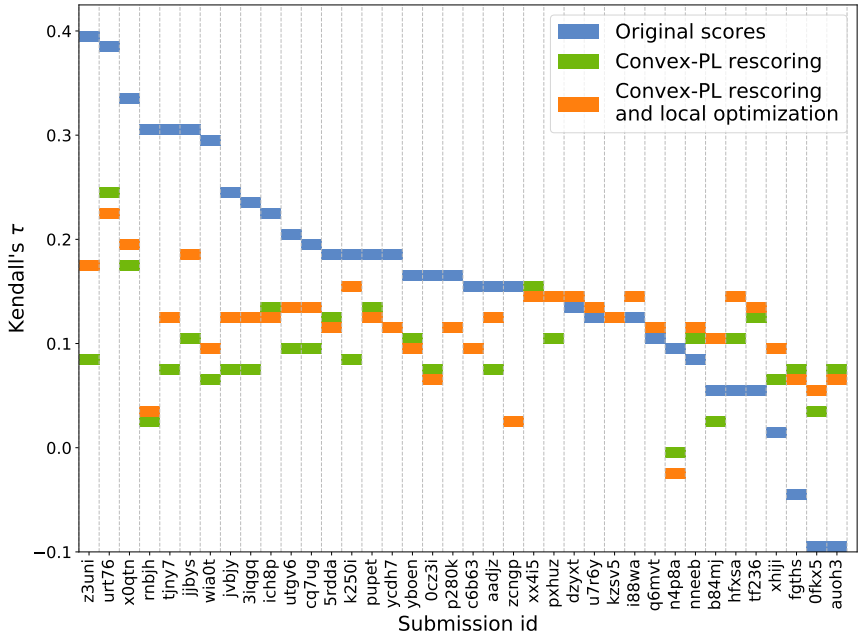


Figure 8.5: Re-scoring of the available structure-based submissions computed with the scoring function that was used in the *xx4i5* submission. All scores were rounded up to the second digit, as in the evaluation results chart. Submissions *dxji8* and *pngkk* were excluded from the comparison due to the incorrect receptor structures. Submissions *6jyjp* and *ufr7g* were excluded from the comparison because the provided ligand chemical structures did not correspond to the original structures.

Our results in Stage 1b were ranked 4th out of 70 if considering the average closest-pose RMSD. During the restricted challenge timeframe, I have not tried other algorithms of fast ligand conformer generation besides the one implemented in RDKit. Yet, I believe that the problems we encountered with the amide bond conformation undersampling in cycles deserve further research and investigation. Notably, a recent update of RDKit's 3D conformers generation [120] improves their macrocycle sampling algorithms with additional heuristics. It would also be worth to mention here BRICARD [304], a free tool designed for macrocycle conformational sampling that we were unaware of during the Challenge.

In this challenge, I have tested Convex-PL in combination with various descriptors, which was a part of Convex-PL^R development. However, of all the modifications, the cutoff distance change

from 10 Å to 5 Å lead to the most considerable improvement of the pose prediction accuracy. For the affinities prediction, though, additional descriptors worked better.

In the affinity predictions, I also relied on the values suggested by our scoring function. The resulting correlations turned out to be average compared to the other structure-based methods. One of the reasons for such performance is that, as I believe, we did not manage to obtain good binding poses for all the 154 ligands in Stage 2. For example, if we applied our scoring function to the pose predictions of some of the best submitters, we could considerably improve our own result. After rescoring of other predictors' submissions, I also noticed that local gradient-based pose optimization on average led to better binding affinity predictions.

Part IV Conclusions

Summary of the thesis

The goal of this thesis was to study the possible formulations of optimization problems for the prediction of protein-ligand interactions, implement them as scoring functions, and assess on the benchmarks and docking challenges.

We have developed Convex-PL – a novel knowledge-based pairwise distance-dependent scoring function for protein-ligand interactions, which is deduced by solving a convex optimization problem. Convex-PL is validated on CASF and D3R-compiled benchmarks and is integrated into a molecular docking program. However, Convex-PL can be biased to bigger ligand molecules and tighter protein-ligand interfaces, which is crucial for virtual screening applications. We have also created a version of AutoDock Vina, VinaCPL, with Convex-PL integrated as a scoring function. Later on, we have developed Convex-PL^R – a machine learning-based scoring function incorporating additional solvent and entropic terms. Convex-PL^R demonstrates better affinity prediction and virtual screening performance if compared to Convex-PL. It also seems to be less biased towards bigger protein-ligand interfaces in the CASF benchmark virtual screening test, if compared to Convex-PL. Being inspired by the protein potential KORP, we have collaboratively developed the first coarse-grained orientation-dependent knowledge-based scoring function for protein-ligand interactions called KORP-PL. It was validated on CASF, D3R-compiled, DUD, and DUD-E benchmarks, and has proved excellent pose prediction and virtual screening abilities. However, more work should be done to improve its relative affinity prediction performance.

We have participated in the pose prediction stage of the D3R Grand Challenge 2 docking challenge. Our initial submission had a relatively average quality of RMSD prediction compared to the other challenge participants. After the challenge was over, we re-ran multiple docking protocols and considerably improved our results. We have participated in round 41 of the CAPRI challenge that was dedicated to a non-peptide ligand docking. We were able to obtain good results when using a consensus score based on Convex-PL and residue conservation scores. We have participated in the cathepsin S sub-challenge of the D3R Grand Challenge 3. Pose prediction in this exercise turned out to be very challenging for Convex-PL. However, we later found out that KORP-PL is able to successfully identify near-native poses. Interestingly, Convex-PL demonstrated better than average results in affinity prediction. We have participated in the macrocycle ligand docking sub-challenge of the D3R Grand Challenge 4. In its Stage 1b, we were able to obtain sub-angstrom pose predictions ranked 4th out of 70 (if considering the closest submitted pose), with a fully automatic structure-based approach with ligand-based geometrical constraints for initial cycle conformers generation. However, our results in the affinity prediction exercise were rather average.

Technical summary

Convex-PL and KORP-PL are written in C++, and together with the Knodle library consist of about 60,000 lines of source code (pre-loaded scoring vectors and external libraries are not counted). The binaries are available at https://team.inria.fr/nano-d/software. Supporting Python scripts written for tests and docking challenges are about 100,000 lines of source code.

Publications

Publications containing the direct contributions of this thesis

Thesis results were reported in 4 journal publications, 1 more publication is in preparation. Additional work related to the thesis content is distributed in 5 journal publications.

1. Kadukova M., Grudinin S.. Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. *J. Comp. Aid. Mol. Des.* – 2017

2. Kadukova M., Grudinin S.. Docking of small molecules to farnesoid X receptors using AutoDock Vina with the Convex-PL potential: lessons learned from D3R Grand Challenge 2. *J. Comp. Aid. Mol. Des.* – 2018

3. Kadukova M., Chupin V., Grudinin S.. Docking rigid macrocycles using Convex-PL, AutoDock Vina, and RDKit in the D3R Grand Challenge 4. *J. Comp. Aid. Mol. Des.* – 2020

4. Kadukova M., dos Santos Machado K., Chacón P., Grudinin S.. KORP-PL: a coarse-grained knowledge-based scoring function for protein-ligand interactions. *Bioinformatics* – 2020

5. Kadukova M., Chupin V., Grudinin S.. Convex-PL^R – Revisiting affinity predictions and virtual screening using physics-informed machine learning. *In preparation*

Publications related to the thesis

1. Grudinin S., Kadukova M., Eisenbarth A., et. al.. Predicting binding poses and affinities for protein - ligand complexes in the 2015 D3R Grand Challenge using a physical model with a statistical parameter estimation.. *J. Comp. Aid. Mol. Des.* – 2016

   This paper describes our collaborative participation in the D3R Grand Challenge 2015.

2. Kadukova M., Grudinin S.. Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates. *J. Chem. Inf. Model.* – 2016

   This paper presents Knodle – our tool for ligand perception from 3D coordinates, types assignment, and file conversion. I have developed it as a part of my Master thesis. Knodle in currently used in all our scoring functions.

3. Lensink M., et. al.. Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins* – 2019

   Besides CAPRI round 41, I participated in several more CAPRI rounds to understand whether Convex-PL can be used for peptide scoring. It turned out that Convex-PL is not the best option for big peptides, however, my contributions resulted in the co-authorship of this overview paper.

4. Igashov I., Olechnovic K., Kadukova M., Venclovas Č., Grudinin S.. VoroCNN: Deep convolutional neural network built on 3D Voronoi tessellation of protein structures. *Bioinformatics* – 2021

I participated in the initial formulation of the 3D Voronoi tesselation-based method for quality assessment of the protein structures. This paper describes a graph convolutional network, with a graph built on the protein Voronoi tesselation, in which each cell represents an interatomic contact.

5. Varaksa T., Bukhdruker S., Grabovec I., et. al.. Hydroxylation of Antitubercular Drug Candidate, SQ109, by Mycobacterial Cytochrome P450. *Int. J. Mol. Sci.* – 2020

   This paper studies the mechanisms of the antitibercular drug candidate SQ109 functioning. I conducted docking experiments to show the possibility of hydrogen bonds formation between a hydroxylized version of SQ109 and its target receptor.

Appendix A

A.1 D3R Benchmark construction

Table A.1 summarizes the data available from the D3R Grand Challenges 2, 3, and 4 that we were using to construct the benchmark.

Table A.1: D3R Benchmark composition

Challenge # (target)	Stage	Total number of examples	Known binding constants	Number of submissions	Number of ligand targets with RMSD less than		
					1 Å	2 Å	3 Å
2 (farnesoid X receptor)	1	35	35 (IC50)	52	24	33	34
3 (cathepsin S)	1a	24	19 (IC50)	52	12	23	24
	1b			47	14	24	24
4 (beta-secretase 1)	1a	20	16 (IC50)	76	20	20	20
	1b			71	20	20	20

Table A.2 lists the user submissions from the D3R challenges that were incomplete or contained errors.

Table A.2: Comments on some of the D3R Challenge submissions.

Challenge	Stage	Submission ID/File	Comments
2	1	5rqrx/3BEJ-FXR_15-1.pdb 5rqrx/3BEJ-FXR_3-5.pdb 5rqrx/3BEJ-FXR_5-3.pdb 5rqrx/3BEJ-FXR_16-2.pdb cfn8u/3bej-FXR_1-2.pdb cfn8u/3bej-FXR_17-5.pdb cfn8u/3bej-FXR_23-4.pdb cfn8u/3bej-FXR_4-4.pdb cfn8u/3bej-FXR_5-5.pdb	lines with HETATM entry describing an atom with atom name=Y, res name=YT3, element symbol=Y3+ were removed for correct file parsing in AutoDock Vina
		04hag	FXR_5 pose predictions were not submitted
		41tia	FXR_35, FXR_36 pose predictions were not submitted
		piwli	FXR_35 pose predictions were not submitted
		m00nf	FXR_36 pose predictions were not submitted
		00ulb	FXR_9 and FXR_11 pose predictions were not submitted
		touhi	FXR_1, FXR_2, FXR_3, FXR_4, FXR_5, FXR_10, FXR_11, FXR_12, FXR_15, FXR_16, FXR_17, FXR_18, FXR_23, FXR_30, FXR_31, FXR_32, FXR_34 pose predictions were not submitted
		6tnqb, h67ea	The whole submission was excluded, because its .mol files contained more than 600 hydrogen atoms with the same coordinate.
3	1a	8bafj, pskfu	CatS_12 and CatS_18 pose predictions were not submitted
		bnexw, ccxri, cmv4q, pskfu, r87j3	CatS_23 pose predictions were not submitted.
		va7rj	CatS_8 pose predictions were not submitted.
		csz33	We modified all pdb files by filling the element symbol fields so that AutoDock Tools could convert the pdbs file to pdbqt.
		45h2u	Atoms from different residues are mixed and sorted by the element symbol, therefore KORP-PL probably wrongly creates the frames. In multiple files we modified the pdb files replacing 'A' in the element symbol with 'C' so that AutoDock Tools could convert the pdb files to pdbqt.
		csz33/3IEJ-CatS_14-1.mol	Was excluded because of misplaced nitrogen and carbon (atom ids 6 and 8).

		fcneq/3KWN-CatS_13-1.mol fcneq/3KWN-CatS_20-1.mol	Were converted from mol to mol with Knodle, since the default file contained errors, to which most of the mol parsers were sensitive.
	1b	3prcf, ciam7, kazpn, r5tfm, tjtp6	Strings were truncated to the z coordinate since the fields positions were not following the PDB standard. Element symbol fields were filled so that AutoDock Tools would be able to convert the pdb files to pdbqt.
		fojgi/gabj-CatS_14-1.mol	Was excluded because of misplaced nitrogen and carbon (atom ids 6 and 8).
		a7mwg/TUHD-CatS_18-2.mol a7mwg/TUHD-CatS_18-5.mol 8siib/JJOD-CatS_23-1.mol bjws5/JJOD-CatS_23-1.mol bjws5/JJOD-CatS_23-1.mol phfin/MEKM-CatS_13-1.mol phfin/MEKM-CatS_13-2.mol phfin/MEKM-CatS_13-3.mol phfin/MEKM-CatS_13-4.mol phfin/MEKM-CatS_13-5.mol	Were excluded because of missing bonds.
4	1a	st72s	BACE_17 and BACE_18 pose predictions were not submitted.
		w020k	BACE_18 pose predictions were not submitted.
		54duu, m0yfh, zb4ps	Non-standard entries (A, OA, HD, SA) of the element symbol column were replaced with element symbols so that AutoDock Tools could parse them.
		qk28f/5YGX-BACE_8-1.mol qk28f/5YGX-BACE_8-2.mol qk28f/5YGX-BACE_8-3.mol qk28f/5YGX-BACE_8-4.mol qk28f/5YGX-BACE_8-5.mol	Were excluded because of the bad quality, many bonds between atoms are absent.
	1b	rckaz	BACE_15 and BACE_16 pose predictions were not submitted.
		ozin4/BA08-BACE_8-1.mol ozin4/BA08-BACE_8-2.mol ozin4/BA08-BACE_8-3.mol ozin4/BA08-BACE_8-4.mol ozin4/BA08-BACE_8-5.mol phzxv/BA08-BACE_8-1.mol phzxv/BA08-BACE_8-2.mol phzxv/BA08-BACE_8-3.mol phzxv/BA08-BACE_8-4.mol phzxv/BA08-BACE_8-5.mol	Were excluded because of the bad quality, many bonds between atoms are absent.
		ozin4/BA20-BACE_20-1.mol ozin4/BA20-BACE_20-2.mol ozin4/BA20-BACE_20-3.mol ozin4/BA20-BACE_20-4.mol ozin4/BA20-BACE_20-5.mol	Were excluded because these submissions contained BACE_19 and not BACE_20.

A.2 DUD and DUD-E co-factors

Tables A.3 and A.4 list the targets excluded from the evaluation because of co-factors.

Table A.3: A list of 9 DUD targets excluded from the evaluation because of co-factors.

DUD targets excluded because of co-factors
alr2, comt, gart, gpb, pnp, sahh, tk, dhfr, inha

Table A.4: A list of 12 DUD-E targets excluded from the evaluation because of co-factors.

DUD-E targets excluded because of co-factors
aldr, aofb, cp2c9, cp3a4, dhi1, dyr, inha, mp2k1, nos1, pyrd, sahh, tysy

A.3 Atom types and some coefficients

Table A.5 contains the conversion of Knodle atom types to the ligand atom types used in Convex-PL and their chemical interpretation.

Table A.5: Correspondence between the sets of atom types used in this work.

Knodle original type	41 types used in Convex PL	description [a]
H.ac	-	acidic H (bonded to O.3ac, N.im, N.sam or N.ohac)
H.onh	-	amide NH
H.n	-	bonded to other nitrogens
H.o	-	bonded to other oxygens
H.0	-	all other hydrogens
C.ar6p	C.ar6x	sp2 carbon with a pos charged resonance structure in a protonated 6-membered heteroaromatic ring
C.ar6x	C.ar6x	sp2 carbon in a 6-membered heteroaromatic ring
C.ar6	C.ar6	sp2 carbon in a benzene ring
C.arp	C.ar	sp2 carbon with a positive charged resonance structure in other protonated heteroaromatic rings
C.arx	C.ar	sp2 carbon in other heteroaromatics
C.ar	C.ar	sp2 carbon in other aromatics
C.2r3o	C.sp2	carbonyl carbon in cyclopropanone or cyclopropenone
C.2r3x	C.sp2	sp2 carbon in heterocyclic 3-membered rings
C.2r3	C.sp2	sp2 carbon in 3-membered rings
C.3r3x	C.sp3	sp3 carbon in heterocyclic 3-membered rings
C.3r3	C.sp3	sp3 carbon in 3-membered rings
C.1n	C.sp1	sp carbon in cyano groups
C.1p	C.sp1	sp carbon with one heavy atom bonded
C.1s	C.sp1	sp carbon with two heavy atoms bonded
C.co2h	C.co2	sp2 carbon in explicitly protonated COOH groups
C.co2	C.co2	sp2 carbon in COO- groups (also set if protonation state is unknown)
C.es	C.sp2	carbonyl carbon in ester groups or anhydrides
C.hal	C.sp3	carbonyl carbon in acidhalogenides
C.am	C.sp2	carbonyl carbon in amides
C.o	C.sp2	other carbonyl carbon
C.s	C.sp2	thionyl carbon
C.gu	C.guh	sp2 carbon in unprotonated guanidino groups
C.guh	C.guh	sp2 carbon in protonated guanidino groups (also set if protonation state is unknown)
C.mi	C.guh	sp2 carbon in unprotonated amidino groups
C.mih	C.guh	sp2 carbon in protonated amidino groups (also set if protonation state is unknown)
C.n	C.sp2	sp2 carbon in imines
C.2p	C.sp2	other sp2 carbon with one heavy atom bonded
C.2s	C.sp2	other sp2 carbon with two heavy atoms bonded
C.2t	C.sp2	other sp2 carbon with 3 heavy atoms bonded
C.et	C.sp3	sp3 carbon in ethers
C.ohp	C.sp3	sp3 carbon in primary alcoholes
C.ohs	C.sp3	sp3 carbon in secondary alcoholes
C.oht	C.sp3	sp3 carbon in tertiary alcoholes
C.3n	C.sp3	other sp3 carbon bonded to nitrogen

C.3p	C.sp3	other sp3 carbon with one heavy atom bonded
C.3s	C.sp3	other sp3 carbon with two heavy atoms bonded
C.3t	C.sp3	other sp3 carbon with 3 heavy atoms bonded
C.3q	C.sp3	other sp3 carbon with 4 heavy atoms bonded
N.ar6p	N.arp	positive charged nitrogen in 6-membered aromatics (e.g. pyridinium or NAD+)
N.ar6	N.arA	sp2 nitrogen in 6-membered aromatics
N.arp	N.arp	sp2 nitrogen in protonated aromatics (e.g both nitrogens in protonated imidazole
N.ar2	N.arA	sp2 nitrogen in aromatics with two bonded atoms (corresponding to sybyl type N.2)
N.ar3	N.arA	sp2 nitrogen in aromatics with 3 heavy atoms (corresponding to sybyl type N.pl3)
N.ar3h	N.arA	sp2 nitrogen in aromatics with 2 heavy atoms and one hydrogen (corresponding to sybyl type N.pl3)
N.r3	N.3s	sp3 in aziridine or azirene rings
N.az	N.1	middle nitrogen in azides
N.1	N.1	other sp nitrogen
N.o2	N.o2	in nitro groups
N.ohac	N.oh	in hydroxamic acids
N.oh	N.oh	in hydroxylamines
N.ims	N.amA	imide nitrogen with two heavy atoms bonded
N.imt	N.amA	imide nitrogen with 3 heavy atoms bonded
N.amp	N.amA	carbon- or thionamide with one heavy atom bonded
N.ams	N.amA	carbon- or thionamide with two heavy atoms bonded
N.amt	N.amA	carbon- or thionamide with 3 heavy atoms bonded
N.samp	N.amA	sulfonamide with one heavy atom bonded
N.sams	N.amA	sulfonamide with two heavy atoms bonded
N.samt	N.amA	sulfonamide with 3 heavy atoms bonded
N.sim	N.amA	sulfonimide
N.gu1	N.guh	NH in unprotonated guanidino group (only if explicitly protonated)
N.gu2	N.guh	NH2 in unprotonated guanidino group (only if explicitly protonated)
N.guh	N.guh	nitrogen in protonated guanidino group (also set if protonation state is unknown)
N.mi1	N.guh	NH in unprotonated amidino group (only if explicitly protonated)
N.mi2	N.guh	NH2 in unprotonated amidino group (only if explicitly protonated)
N.mih	N.guh	nitrogen in protonated amidino group (also set if protonation state is unknown)
N.aap	N.amA	primary aromatic amine (hybridization can't be determined exactly)
N.aas2	N.amA	sp2 hybridized secondary aromatic amine
N.aas3	N.3s	sp3 hybridized secondary aromatic amine
N.aat2	N.amA	sp2 hybridized tertiary aromatic amine
N.aat3	N.3t	sp3 hybridized tertiary aromatic amine
N.2n	N.2s	sp2 nitrogen bonded to another nitrogen
N.2p	N.2p	other sp2 nitrogen with one heavy atom
N.2s	N.2s	other sp2 nitrogen with two heavy atoms
N.2t	N.amA	other sp2 nitrogen with three heavy atoms
N.3n	N.3n	sp3 nitrogen bonded to another nitrogen
N.3p	N.3p	sp3 nitrogen with one heavy atom bonded
N.3s	N.3s	sp3 nitrogen with two heavy atoms bonded
N.3t	N.3t	sp3 nitrogen with 3 heavy atoms bonded
N.4q	N.4	sp3 nitrogen with 4 bonded heavy atoms
N.4h	N.4	sp3 nitrogen with 4 bonded atoms (at least 1 hydrogen)
O.ar	O.ar	aromatic oxygen
O.r3	O.3et	in oxiran ring
O.n	O.n	oxygen in nitro groups
O.2n	O.n	oxygen in nitroso groups
O.noh	O.3ac	sp3 oxygen in hydroxylamine or hydroxamic acid
O.2co2	O.carb	sp2 oxygen in COOH (sp2 bonded to C.co2h)
O.2es	O.carb	sp2 oxygen in esters or anhydrids
O.2hal	O.carb	sp2 oxygen in acidhalogenides
O.am	O.carb	in carbonamides
O.co2	O.co2	in COO- or CSO-
O.2po	O.carb	sp2 oxygen in P=O (non deprotonated groups)
O.2so	O.carb	sp2 oxygen in S=O (non deprotonated groups)
O.2p	O.co2	sp2 oxygen in OPO3H- or PO3H- or POO-
O.2s	O.co2	sp2 oxygen in OSO3- or SO3- or POO- or deprotonated sulfonamides

O.3po	O.3et	sp3 oxygen with 2 heavy atoms bonded to at least one phosphor
O.3so	O.3et	sp3 oxygen with 2 heavy atoms bonded to at least one sulfur
O.carb	O.carb	in other carbonyl groups
O.o	O.3et	in peroxo groups
O.3ac	O.3ac	OH in COOH, CSOH, POOHOH, POOH or SOOOH
O.ph	O.3oh	phenolic hydroxyl group
O.3oh	O.3oh	hydroxyl group
O.3es	O.3et	sp3 oxygen in esters or anhydrids
O.3eta	O.3et	aromatic ether
O.3et	O.3et	aliphatic ether
S.ar	S.3	aromatic sulfur
S.r3	S.3	in thiiran ring
S.thi	S.2	thionyl group
S.o	S.o	in SO
S.o2h	S.o	in protonated sulfonamide or other SO2
S.o3h	S.o	in SO3
S.o4h	S.o	in OSO3
S.o2	S.o	in SO2 or deprotonated sulfonamides (or unknown protonation state)
S.o3	S.o	in SO3- (or unknown protonation state)
S.o4	S.o	in OSO3- (or unknown protonation state)
S.osn	S.o	in OSO2S- and etc (or unknown protonation state)
S.2p	S.o	sp2 oxygen in OPO2SH- or PO2SH- or POS-
S.2	S.2	in CSO- COS- or other sp2
S.sh	S.3	in SH groups
S.s	S.3	in S-S bonds
S.3	S.3	other sp3 sulfur
P.r3	P.3	in phosphiran rings
P.o	P.o	in PO
P.o2h	P.o	in not deprotonated PO2 groups
P.o3h	P.o	in not deprotonated PO3 groups
P.o4h	P.o	in not deprotonated PO4 groups
P.o2	P.o	in deprotonated PO2 groups (or unknown protonation state)
P.o3	P.o	in deprotonated PO3 groups (or unknown protonation state)
P.o4	P.o	in deprotonated PO4 groups (or unknown protonation state)
P.3	P.3	other sp3
F.0	F.0	bonded fluor
F.i	F.i	fluor ion
Cl.0	Cl.0	bonded chlorine
Cl.i	Cl.i	chlorine ion
Br.0	Br.0	bonded bromine
Br.i	Br.i	bromine ion
I.0	I.0	bonded iod
I.i	-	iod ion
O.h2o	-	water oxygen
Li	-	Li
Na	-	Na
Mg	-	Mg
Al	-	Al
Si	-	Si
K	-	K
Ca	-	Ca
Cr.th	-	Cr.th
Cr.oh	-	Cr.oh
Mn	-	Mn
Fe	-	Fe
Co	-	Co
Cu	-	Cu
Zn	-	Zn
Se	-	Se
Mo	-	Mo
Sn	-	Sn

Ni	-	Ni
Hg	-	Hg
B	-	B
As	-	As

[a] 160 types were adopted from the fconv extended types set, 4 types were added in Knodle.

Table A.6 lists the ligand atom types used in Convex-PL-^{R} and their chemical interpretation.

Table A.6: 40 ligand atom types used in Convex-PL^R

Type name	Description
C.ar6x	sp2 carbon in a 6-membered heteroaromatic ring
C.ar6	sp2 carbon in a benzene ring
C.ar	sp2 carbon in other aromatics
C.sp1	sp carbon
C.sp3	sp3 carbon
C.co2	sp2 carbon in COOH and COO- groups
C.guh	sp2 carbon in amidino and guanidino groups
C.sp2	other sp2 carbon
N.arp	charged nitrogen in protonated aromatics
N.arA	other nitrogen in aromatic rings
N.1	sp nitrogen
N.o2	nitrogen in nitro groups
N.oh	nitrogen in hydroxamic acids and in hydroxylamines
N.amA	other sp2 nitrogen with three heavy atoms; nitrogen in sulfonamide, sulfonimide, carbon- or thionamide, aromatic amine, imide
N.guh	nitrogen in guanidino or amidino group
N.2p	other sp2 nitrogen with one heavy atom
N.2s	other sp2 nitrogen bonded to the two heavy atoms
N.3p	other sp3 nitrogen with one heavy atom bonded
N.3s	other sp3 nitrogen with 2 heavy atoms bonded
N.3t	other sp3 nitrogen with 3 heavy atoms bonded
N.4	other sp3 nitrogen with 4 bonded atoms
O.3et	sp3 oxygen in ethers, esters, anhydrids, oxiran rings; sp3 oxygen with 2 heavy atoms bonded to at least one phosphor or sulfur
O.n	oxygen in nitro groups
O.3ac	sp3 oxygen in hydroxylamine or hydroxamic acid, COOH, CSOH, POOHOH, POOH or SOOOH
O.carb	sp2 oxygen in esters, anhydrids, carbonamides, acidhalogenides, COOH, POOHOH, POOH, SOOOH, and in other carbonyl groups
O.co2	sp2 oxygen in OSO3-, SO3-, POO-, deprotonated sulfonamides, COO-, CSO-, OPO3H-, PO3H-, POO-, peroxo groups
O.3oh	hydroxyl group
O.ar	aromatic oxygen
S.r	aromatic sulfur; sulfur in a thiiran ring
S.o	sulfur in SO, SO3, CSO-, COS-, OSO3, OPO2SH-, PO2SH-, POS-, thionyl group, SO3-, OSO3-
S.o2	sulfur in SO2 and sulfonamides
S.3	other sulfur
P.o	phosphorus in groups with oxygen
P.3	phosphorus in phosphiran rings; other sp3 phosphorus
F.0	fluorine
Cl.0	chlorine
Br.0	bromine

I.0	iodine
F.i	fluorine ion
Hal.i	chlorine ion; bromine ion; iodine ion

Table A.7 lists the ligand atom types used in KORP-PL and their chemical interpretation.

Table A.7: 37 ligand atom types used in KORP-PL.

Type name	Description
C.ar6x	sp2 carbon in a 6-membered heteroaromatic ring
C.ar6	sp2 carbon in a benzene ring
C.ar	sp2 carbon in other aromatics
C.sp1	sp carbon
C.sp3	sp3 carbon
C.co2	sp2 carbon in COOH and COO- groups
C.guh	sp2 carbon in amidino and guanidino groups
C.sp2	other sp2 carbon
N.arp	charged nitrogen in protonated aromatics
N.arA	other nitrogen in aromatic rings
N.1	sp nitrogen
N.o2	nitrogen in nitro groups
N.oh	nitrogen in hydroxamic acids and in hydroxylamines
N.amA	other sp2 nitrogen with three heavy atoms; nitrogen in sulfonamide, sulfonimide, carbon- or thionamide, aromatic amine, imide
N.guh	nitrogen in guanidino or amidino group
N.p	other sp2 or sp3 nitrogen with one heavy atom bonded
N.2s	other sp2 nitrogen bonded to the two heavy atoms
N.3s	other sp3 nitrogen with 2 heavy atoms bonded
N.3t	other sp3 nitrogen with 3 heavy atoms bonded
N.4	other sp3 nitrogen with 4 bonded atoms
O.3et	sp3 oxygen in ethers, esters, anhydrids, oxiran rings; sp3 oxygen with 2 heavy atoms bonded to at least one phosphor or sulfur
O.n	oxygen in nitro groups
O.3ac	sp3 oxygen in hydroxylamine or hydroxamic acid, COOH, CSOH, POOHOH, POOH or SOOOH
O.carb	sp2 oxygen in esters, anhydrids, carbonamides, acidhalogenides, COOH, POOHOH, POOH, SOOOH, and in other carbonyl groups
O.co2	sp2 oxygen in OSO3-, SO3-, POO-, deprotonated sulfonamides, COO-, CSO-, OPO3H-, PO3H-, POO-, peroxo groups
O.3oh	hydroxyl group
O.ar	aromatic oxygen
S.r	aromatic sulfur; sulfur in a thiiran ring
S.o	sulfur in SO, SO3, CSO-, COS-, OSO3, OPO2SH-, PO2SH-, POS-, thionyl group, SO3-, OSO3-
S.o2	sulfur in SO2 and sulfonamides
S.3	other sulfur
P.o	phosphorus in groups with oxygen
P.3	phosphorus in phosphiran rings; other sp3 phosphorus
F.0	fluorine
Cl.0	chlorine
Br.0	bromine
I.0	iodine

A.3.1 Selection criteria for the complexes from the PDBBind 2016 general set used in the linear regression model training

Table A.8: Number of PDB codes included and excluded from the linear regression training set.

Condition	Number of complexes satisfying the condition, excluding already filtered complexes	Number of complexes satisfying the condition, total
PDBBind 2016 general set	13308	13308
CASF 2013 and 2016	-373	373
K_i binding constants	-3687	3927
$\approx$ known binding constants	-69	83
$>, <, \geq, \leq$ binding constants	-160	203
"incorrect" label in the index file	-9	15
$-\log K \geq 13$	-5	9
Convex-PL$^{score} \leq 0$	-24	31
other	-7	7
Resulting number of complexes	8972	

A.3.2 Linear regression coefficients of Convex-PL^R

Table A.9: Feature weights. SASA descriptors for Convex-PL^R are computed as atomic SASA of protein and ligand atoms located within a 7 Å cutoff from each other. All features are scaled to the $[0, 1]$ interval.

Feature	Weights Convex-PL^R	Feature	Weights Convex-PL^R
ligand flexibility	-1.659	SASA, receptor atom type C.3n	0.000
"solvent" grid rdf, receptor atom type C.3n	0.069	SASA, receptor atom type C.3p	0.000
"solvent" grid rdf, receptor atom type C.3p	0.166	SASA, receptor atom type C.3s	0.000
"solvent" grid rdf, receptor atom type C.3s	0.088	SASA, receptor atom type C.3t	0.000
"solvent" grid rdf, receptor atom type C.3t	0.134	SASA, receptor atom type C.am	-0.501
"solvent" grid rdf, receptor atom type C.am	0.071	SASA, receptor atom type C.ar6	-0.501
"solvent" grid rdf, receptor atom type C.ar6	0.060	SASA, receptor atom type C.arp	-0.501
"solvent" grid rdf, receptor atom type C.arp	0.045	SASA, receptor atom type C.arx	-0.501
"solvent" grid rdf, receptor atom type C.arx	-0.028	SASA, receptor atom type C.co2	-0.501
"solvent" grid rdf, receptor atom type C.co2	0.032	SASA, receptor atom type C.guh	-0.501
"solvent" grid rdf, receptor atom type C.guh	0.002	SASA, receptor atom type N.3p	-0.501
"solvent" grid rdf, receptor atom type N.3p	0.008	SASA, receptor atom type N.ams	-0.501
"solvent" grid rdf, receptor atom type N.ams	0.065	SASA, receptor atom type N.amt	-0.501
"solvent" grid rdf, receptor atom type N.amt	0.021	SASA, receptor atom type N.ar2	-0.501
"solvent" grid rdf, receptor atom type N.ar2	-0.024	SASA, receptor atom type N.arp	-0.501
"solvent" grid rdf, receptor atom type N.arp	0.052	SASA, receptor atom type N.guh	-0.501
"solvent" grid rdf, receptor atom type N.guh	0.002	SASA, receptor atom type O.3ac	-0.501
"solvent" grid rdf, receptor atom type O.3ac	0.007	SASA, receptor atom type O.am	-0.501
"solvent" grid rdf, receptor atom type O.am	0.075	SASA, receptor atom type O.carb	-0.501
"solvent" grid rdf, receptor atom type O.carb	0.052	SASA, receptor atom type O.co2	-0.501
"solvent" grid rdf, receptor atom type O.co2	0.003	SASA, receptor atom type S.3	0.000
"solvent" grid rdf, receptor atom type S.3	0.051	SASA, receptor atom type S.sh	0.000
"solvent" grid rdf, receptor atom type S.sh	0.070	SASA, receptor atom type Se	0.000
"solvent" grid rdf, receptor atom type Se	0.000	SASA, ligand atom type C.ar6x	-2.612
"solvent" grid rdf, ligand atom type C.ar6x	0.087	SASA, ligand atom type C.ar6	-2.612
"solvent" grid rdf, ligand atom type C.ar6	0.170	SASA, ligand atom type C.ar	-2.612
"solvent" grid rdf, ligand atom type C.ar	0.081	SASA, ligand atom type C.sp1	0.000
"solvent" grid rdf, ligand atom type C.sp1	0.052	SASA, ligand atom type C.sp2	0.000
"solvent" grid rdf, ligand atom type C.sp2	-0.001	SASA, ligand atom type C.sp3	0.000
"solvent" grid rdf, ligand atom type C.sp3	-0.025	SASA, ligand atom type C.co2	-2.612
"solvent" grid rdf, ligand atom type C.co2	-0.119	SASA, ligand atom type C.guh	-2.612
"solvent" grid rdf, ligand atom type C.guh	-0.007	SASA, ligand atom type N.arp	-2.612
"solvent" grid rdf, ligand atom type N.arp	0.005	SASA, ligand atom type N.arA	-2.612
"solvent" grid rdf, ligand atom type N.arA	0.085	SASA, ligand atom type N.3s	-2.612
"solvent" grid rdf, ligand atom type N.3s	0.079	SASA, ligand atom type N.1	0.000
"solvent" grid rdf, ligand atom type N.1	0.063	SASA, ligand atom type N.o2	-2.612
"solvent" grid rdf, ligand atom type N.o2	0.016	SASA, ligand atom type N.oh	-2.612
"solvent" grid rdf, ligand atom type N.oh	0.065	SASA, ligand atom type N.amA	-2.612

"solvent" grid rdf, ligand atom type N.amA	-0.095	SASA, ligand atom type N.guh	-2.612
"solvent" grid rdf, ligand atom type N.guh	-0.004	SASA, ligand atom type N.2p	-2.612
"solvent" grid rdf, ligand atom type N.2p	0.026	SASA, ligand atom type N.2s	-2.612
"solvent" grid rdf, ligand atom type N.2s	0.016	SASA, ligand atom type N.3p	-2.612
"solvent" grid rdf, ligand atom type N.3p	0.008	SASA, ligand atom type N.3t	-2.612
"solvent" grid rdf, ligand atom type N.3t	0.032	SASA, ligand atom type N.4	-2.612
"solvent" grid rdf, ligand atom type N.4	-0.056	SASA, ligand atom type O.3et	-2.612
"solvent" grid rdf, ligand atom type O.3et	-0.033	SASA, ligand atom type O.n	-2.612
"solvent" grid rdf, ligand atom type O.n	0.016	SASA, ligand atom type O.3ac	-2.612
"solvent" grid rdf, ligand atom type O.3ac	0.068	SASA, ligand atom type O.carb	-2.612
"solvent" grid rdf, ligand atom type O.carb	-0.001	SASA, ligand atom type O.co2	-2.612
"solvent" grid rdf, ligand atom type O.co2	-0.138	SASA, ligand atom type O.3oh	-2.612
"solvent" grid rdf, ligand atom type O.3oh	-0.153	SASA, ligand atom type O.ar	-2.612
"solvent" grid rdf, ligand atom type O.ar	0.031	SASA, ligand atom type S.3	0.000
"solvent" grid rdf, ligand atom type S.3	0.021	SASA, ligand atom type S.r	-2.612
"solvent" grid rdf, ligand atom type S.r	0.044	SASA, ligand atom type S.o	-2.612
"solvent" grid rdf, ligand atom type S.o	0.003	SASA, ligand atom type S.o2	-2.612
"solvent" grid rdf, ligand atom type S.o2	0.201	SASA, ligand atom type P.3	0.000
"solvent" grid rdf, ligand atom type P.3	0.000	SASA, ligand atom type P.o	-2.612
"solvent" grid rdf, ligand atom type P.o	-0.107	SASA, ligand atom type F.0	-2.612
"solvent" grid rdf, ligand atom type F.0	0.086	SASA, ligand atom type Cl.0	-2.612
"solvent" grid rdf, ligand atom type Cl.0	0.055	SASA, ligand atom type Br.0	-2.612
"solvent" grid rdf, ligand atom type Br.0	0.003	SASA, ligand atom type I.0	-2.612
"solvent" grid rdf, ligand atom type I.0	0.025	SASA, ligand atom type F.i	-2.612
"solvent" grid rdf, ligand atom type F.i	-0.002	SASA, ligand atom type Cl.i	-2.612
"solvent" grid rdf, ligand atom type Cl.i	-0.000	side-chain flexibility	0.000

A.3.3 Reweighing coefficients of KORP-PL

Table A.10 lists the weights $w_{ij} = c_i r_j$ obtained after the reweighing procedure designed for better affinity predictions.

Table A.10: Weights of residue-ligand atom interactions used in KORP-PL^w.

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
C.ar6x	1.017	1.017	1.017	1.017	1.017	1.017	1.555	1.017	1.440	1.017	1.692	1.017	3.330	1.630	1.017	1.017	1.017	1.017	2.718	1.017
C.ar6	1.178	1.178	1.178	1.178	1.178	1.178	1.800	1.178	1.667	1.178	1.959	1.178	3.856	1.887	1.178	1.178	1.178	1.178	3.148	1.178
C.ar	1.188	1.188	1.188	1.188	1.188	1.188	1.816	1.188	1.681	1.188	1.976	1.188	3.889	1.903	1.188	1.188	1.188	1.188	3.175	1.188
C.sp1	0.687	0.687	0.687	0.687	0.687	0.687	1.050	0.687	0.972	0.687	1.142	0.687	2.248	1.100	0.687	0.687	0.687	0.687	1.835	0.687
C.sp2	0.913	0.913	0.913	0.913	0.913	0.913	1.396	0.913	1.293	0.913	1.520	0.913	2.991	1.464	0.913	0.913	0.913	0.913	2.442	0.913
C.sp3	1.617	1.617	1.617	1.617	1.617	1.617	2.472	1.617	2.289	1.617	2.690	1.617	5.295	2.591	1.617	1.617	1.617	1.617	4.322	1.617
C.co2	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
C.guh	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.arp	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.arA	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.3s	0.994	0.994	0.994	0.994	0.994	0.994	1.519	0.994	1.407	0.994	1.653	0.994	3.254	1.592	0.994	0.994	0.994	0.994	2.656	0.994
N.1	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.o2	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.oh	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.amA	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.guh	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.p	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.2s	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
N.3t	0.660	0.660	0.660	0.660	0.660	0.660	1.008	0.660	0.934	0.660	1.097	0.660	2.160	1.057	0.660	0.660	0.660	0.660	1.763	0.660
N.4	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
O.3et	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
O.n	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
O.3ac	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
O.carb	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
O.co2	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
O.3oh	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
O.ar	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
S.3	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
S.r	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
S.o	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
S.o2	1.452	1.452	1.452	1.452	1.452	1.452	2.220	1.452	2.056	1.452	2.416	1.452	4.756	2.328	1.452	1.452	1.452	1.452	3.882	1.452
P.3	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
P.o	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
F.0	1.001	1.001	1.001	1.001	1.001	1.001	1.530	1.001	1.416	1.001	1.664	1.001	3.276	1.603	1.001	1.001	1.001	1.001	2.674	1.001
Cl.0	1.213	1.213	1.213	1.213	1.213	1.213	1.854	1.213	1.717	1.213	2.017	1.213	3.972	1.944	1.213	1.213	1.213	1.213	3.242	1.213

Br.0	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608
I.0	0.608	0.608	0.608	0.608	0.608	0.608	0.930	0.608	0.861	0.608	1.012	0.608	1.992	0.975	0.608	0.608	0.608	0.608	1.626	0.608

A.4 CASF benchmarks

A.4.1 Evaluation of Convex-PL, Convex-PL$^{5.2A}$, Convex-PL^{R}, KORP-PL, and KORP-PL^{w} on the CASF benchmarks

Tables A.11 – A.13 list the results of Convex-PL, Convex-PL$^{5.2A}$, Convex-PL^{R}, KORP-PL, and KORP-PL^{w} in the CASF-2013 and CASF-2016 benchmarks. All results except Vinardo, AutoDock Vina, $\Delta_{Vina}\mathrm{RF}_{20}$ in CASF-2013, and our scoring functions in both benchmarks were taken from the supplementary information of CASF benchmark papers [94, 160]. Vinardo and $\Delta_{Vina}\mathrm{RF}_{20}$ results in the CASF-2013 were taken from the corresponding papers [74, 80]. AutoDock Vina results in the CASF-2013 were also taken from [80].

Docking test

Table A.11: Success rates of finding native and near-native docking poses with RMSD smaller than 1 Å , 2 Å , and 3 Å among 1, 2, and 3 top-ranked structures scored with KORP-PL, KORP-PL^{w}, and Convex-PL in the CASF-2013 and CASF-2016 docking tests.

Convex-PL

	CASF-2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Top 1	81.03	88.72	92.31	71.28	86.15	90.77	85.61	89.82	93.33	75.09	84.21	90.18
Top 2	86.67	92.31	94.87	78.46	89.23	93.85	90.88	94.74	95.44	82.46	89.82	92.98
Top 3	89.74	93.33	95.90	83.08	91.28	94.87	93.33	96.84	97.54	85.26	93.68	95.79

Convex-PL$^{5.2A}$

	CASF 2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Top 1	80.51	87.18	90.77	72.31	84.10	88.72	83.86	89.12	93.33	75.09	86.32	91.93
Top 2	87.18	93.85	96.41	78.46	90.77	94.87	90.88	96.14	97.19	83.86	92.63	95.44
Top 3	89.23	94.36	97.95	81.54	91.79	96.92	92.28	97.89	98.95	85.61	95.09	97.89

Convex-PL^{R}

	CASF 2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Top 1	81.03	88.72	90.77	71.28	84.10	87.18	84.56	90.53	94.04	75.44	86.67	91.23
Top 2	88.21	93.85	96.41	78.46	90.77	94.87	91.23	96.49	97.19	83.86	92.63	96.14
Top 3	90.77	94.87	97.44	82.05	91.79	96.41	92.63	98.25	98.95	85.61	95.44	98.25

KORP-PL

	CASF-2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Top 1	81.03	87.69	92.31	68.21	84.10	91.28	81.05	89.12	92.98	68.77	85.61	91.93
Top 2	86.15	92.82	96.41	75.38	88.72	94.87	87.37	93.33	95.79	77.89	90.53	95.44
Top 3	89.23	94.36	97.95	81.54	91.28	96.92	91.58	96.84	97.54	82.11	94.04	97.19

KORP-PL^w

	CASF-2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Top 1	77.44	85.13	91.28	67.18	81.54	90.26	78.25	84.91	89.47	67.37	83.16	89.12
Top 2	82.05	89.74	94.36	75.90	86.67	93.33	85.61	92.28	95.09	78.25	90.18	94.04
Top 3	86.67	93.33	96.92	81.54	90.77	96.41	88.07	94.39	98.25	81.05	92.63	97.89

Scoring and ranking tests

Table A.12: Scoring and ranking test results. r_p, r_s, and τ correspond to the Pearson's and Spearman's correlation coefficients and Kendall's τ, correspondingly. In CASF-2013, the ranking performance is measured as a success rate of (high-level) correct ranking of all the three ligands binding the target protein, and (low-level) ranking the best complex as the top one.

Scoring function	CASF-2013			CASF-2016		
	Scoring	Ranking		Scoring	Ranking	
	r_p	"high"	"low"	r_p	r_s	τ
Convex-PL	0.574	60.0	70.8	0.607	0.584	0.505
Convex-PL$^{5.2A}$	0.608	55.4	69.2	0.631	0.607	0.519
Convex-PL^R	0.613	50.8	70.8	0.661	0.604	0.516
KORP-PL	0.439	50.8	61.5	0.447	0.57	0.498
KORP-PL^w	0.5	50.8	63.1	0.521	0.561	0.498

Screening tests

Table A.13: CASF-2013 and CASF-2016 screening test results. Enrichment, best binder, and best target correspond to enrichment factors, a success rate of finding the strongest binding ligand for each protein target, and a success rate to find the strongest binding target for each ligand.

Scoring function	CASF-2013						CASF-2016								
	Enrichment			Best binder			Enrichment			Best binder			Best target		
	Top, %			Top, %			Top, %			Top, %			Top, %		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Convex-PL	9.5	4.7	3.1	27.7	55.4	64.6	4.6	3.1	2.7	14.0	35.1	61.4	19.6	30.5	40.0
Convex-PL$^{5.2A}$	18.0	5.8	3.6	50.8	70.8	78.5	7.7	3.6	2.8	31.6	40.4	59.6	17.5	29.1	40.7
Convex-PL^R	25.8	7.9	4.6	70.8	83.1	89.2	14.1	5.2	3.7	38.6	52.6	64.9	17.9	28.4	40.0
KORP-PL	25.6	9.1	5.5	60.0	76.9	80.0	22.2	9.4	5.8	42.1	70.2	75.4	15.1	27.0	37.9
KORP-PL^w	25.4	7.7	4.9	61.5	72.3	76.9	18.5	8.8	5.5	45.6	70.2	73.7	13.3	26.0	37.2

A.5 Overfitting experiments for KORP-PL in the CASF tests

In the tables below we will refer to the four subsets, listed in Table A.14. These are the *reference training set*, the *overfitted* set, the *closest removed* subset, and the *distant removed* subset.

Re-training KORP-PL

Tables A.15 - A.17 list the *difference* between the metrics computed for KORP-PL (see Tables S3 and S5) and the scores computed for the KORP-PL versions trained on different subsets of the training set: $\Delta\text{Score} = \text{Score}_{\text{KORP-PL}^{subset}} - \text{Score}_{\text{KORP-PL}}$.

We started the experiments by training on the overfitted subset. As expected, it confirmed the common practice of the necessity of excluding all test proteins of CASF-2013 / 2016 benchmark

Table A.14: Subsets of PDBBind 2016 chosen to analyze possible overfitting. Shape similarity between protein binding pockets is represented with TM-score computed using *Kpax* [273, 274], shape similarity between ligands is represented with so-called Jaccard (Tanimoto) distance computed using RDKit's *ShapeTanimotoDist* [129].

subset name	subset description
reference training set	PDBBind 2016 with excluded 373 complexes from CASF-2013 and 2016
$\text{subset}^{\text{overfitted}}$	PDBBind 2016 including the 373 complexes from CASF-2013 and 2016
$\text{subset}^{\text{closest removed}}$	reference training set with excluded 174 complexes similar to the ones from CASF-2013/2016 (TM-score > 0.8 and T < 0.2)
$\text{subset}^{\text{distant removed}}$	reference training set with excluded 1033 complexes similar to the ones from CASF-2013/2016 (TM-score > 0.5 and T < 0.4)

from the PDBBind training set. As can be seen, not removing the intersection with the test cases produces significantly overestimated metrics.

We then excluded from the training set the 174 most similar complexes with pocket TM-score greater than 0.8 and ligand similarity distance less than 0.2 to the test set. Without entering in a complex discussion about ligand similarity, removing similar cases resulted in a slight change of performance in the docking test of CASF-2016 and a slight drop in its screening test results. More precisely, in the docking test, 81.1% success rate of finding a native or 1A near-native pose in the top-ranked pose dropped by 0.4% – however, several other metrics with less accurate thresholds became better. For example, the 93.3% success rate of finding a native or 2A near-native pose in the first 2 top-ranked poses rose by 0.4% to 93.7%. In CASF-2013, all metrics became slightly worse. In the screening test, the 1% enrichment factor in CASF-2016 dropped down from 22.2 to 21.5, and increased from 25.6 to 26.1 in CASF-2013. Overall, excluding similar structures from the training set, we see only marginal changes in the overall performance. These results are in agreement with our previous results with Convex-PL[55].

Further exclusion of 1033 complexes with pocket TM-score greater than 0.5 and the ligand similarity distance less than 0.4 resulted in a more notable performance change. Note that in this case, we removed the majority of the complexes with fold resemblance of binding pockets of the test set complexes. In this scenario, we are sure that we remove all distant/related cases but it is likely that we also jeopardize key statistical information around the test cases. In this stringent scenario, the docking test success rate dropped depending on the required quality of the near-native pose. Indeed, the success rate of finding top-1 pose within 1A RMSD from the native one (Top1Q1) in CASF-2016 dropped by 5.6%, while the top1q2 success rate dropped by 3.2%. Similarly, Top3Q1, Top3Q2, Top3Q3 drop by 1.8%, 0.4%, and 0%, respectively. Success rates in CASF-2013 decreased more distinctly, with the biggest accuracy drop of 10.8% for Top1Q1 and the smallest drop of 2.6 for Top2Q2. As for the screening tests, the 1% enrichment factor decreased considerably from 22.2 to 16.4. At the same time, the 1% enrichment factor in CASF-2013 dropped only from 25.6 to 23.2. Notably, the 16.4 enrichment factor in CASF-2016 is still higher in comparison with other scoring functions evaluated on this benchmark. Such a performance shift in the CASF-2016 screening test indicates that KORP-PL has initially learned some specific interactions. Our results also indicate that for a successful high-resolution pose prediction (Q1, and to a lesser extent Q2), the training set must contain complexes with interactions that somewhat resemble those in the test set. Indeed, any Boltzmann approximation is limited when a feature is not present (or its distribution is underestimated) in the statistics from the training set.

Table A.15: Difference between the CASF docking test success rates computed between the scores produced by the KORP-PL modifications trained on different training set subsets and the reference version of KORP-PL.

KORP-PL trained on subset$^{\text{overfitted}}$

	CASF 2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3
Top 1	7.69	5.64	3.07	3.07	4.11	1.54	15.09	7.72	5.27	9.83	6.32	3.51
Top 2	6.16	2.05	1.03	5.65	2.05	1.03	11.23	5.97	3.51	7.02	4.91	2.45
Top 3	3.08	1.54	0.51	2.05	1.03	0.52	7.37	2.81	2.11	6.31	3.85	1.76

KORP-PL, trained on subset$^{\text{closest removed}}$

	CASF 2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3
Top 1	-2.57	-2.05	-1.54	-3.08	-1.54	-1.54	-0.35	0.35	0.0	-0.35	0.0	0.00
Top 2	-1.02	-1.03	-1.03	-0.51	-0.51	-1.54	0.0	0.35	0.35	0.0	0.70	0.35
Top 3	-0.51	-0.51	-1.03	-1.03	-0.51	-1.02	0.35	0.0	0.0	0.70	0.0	0.00

KORP-PL, trained on subset$^{\text{distant removed}}$

	CASF 2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3
Top 1	-10.77	-3.59	-3.08	-10.26	-5.13	-4.10	-5.61	-3.16	-2.45	-5.26	-4.91	-3.51
Top 2	-5.12	-2.56	-3.59	-2.05	-2.05	-4.10	-1.05	-0.35	-0.35	-1.75	-0.71	-0.70
Top 3	-6.15	-2.57	-3.08	-5.64	-2.56	-3.07	-1.76	-0.35	0.0	-2.81	-0.36	-0.35

Table A.16: Difference between the CASF scoring and ranking test results rates computed between the scores produced by the KORP-PL modifications trained on different training set subsets and the reference version of KORP-PL.

KORP-PL training subset	CASF 2013			CASF-2016	
	Δ Scoring	Δ Ranking		Δ Scoring	Δ Ranking
	r_p	"high"	"low"	r_p	r_s
subset$^{\text{overfitted}}$	0.010	3.0	3.1	0.017	-0.017
subset$^{\text{closest removed}}$	0.009	1.5	-1.5	0.007	-0.010
subset$^{\text{distant removed}}$	-0.009	-6.2	-6.1	-0.021	-0.037

Table A.17: Difference between the CASF screening test results computed between the scores produced by the KORP-PL modifications trained on different training set subsets and the reference version of KORP-PL.

KORP-PL training subset	CASF 2013						CASF-2016					
	Δ Enrichment			Δ Best binder			Δ Enrichment			Δ Best binder		
	Top, %			Top, %			Top, %			Top, %		
	1	5	10	1	5	10	1	5	10	1	5	10
subset$^{\text{overfitted}}$	4.1	0.5	0.3	10.8	3.1	4.6	3.6	1.1	0.6	8.8	7.0	8.8
subset$^{\text{closest removed}}$	0.5	-0.1	-0.2	1.5	0.0	0.0	-0.7	-0.0	0.0	-1.7	-1.8	1.8
subset$^{\text{distant removed}}$	-2.4	-1.8	-1.1	-7.7	-13.8	-7.7	-5.8	-1.0	-0.4	-10.5	-14.1	-7.0

Re-training the weights of KORP-PL^w

Tables A.18 - A.20 list the *difference* between the metrics computed for KORP-PL^w (see Tables S3 and S5) and the scores computed for the KORP-PL^w versions trained on different subsets of the training set: $\Delta\text{Score} = \text{Score}_{\text{KORP-PL}^{w,subset}} - \text{Score}_{\text{KORP-PL}^w}$.

Table A.18: Difference between the CASF docking test success rates computed between the scores produced by KORP-PL^w modifications with weights trained on the reference training set subsets and the reference version of KORP-PL^w.

KORP-PL^w, weights trained on subset$^{\text{closest removed}}$

	CASF 2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3
Top 0	-0.0	-0.0	0.51	-0.0	0.51	-0.0	-0.36	-0.35	0.0	-0.35	-0.70	-0.35
Top 1	0.0	0.52	0.51	0.51	1.02	0.52	0.35	-0.35	-0.0	-0.0	-0.36	-0.00
Top 2	-0.52	0.0	0.0	-1.03	-0.0	0.0	0.35	-0.0	-0.0	0.35	0.0	0.00

KORP-PL^w, weights trained on subset$^{\text{distant removed}}$

	CASF 2013						CASF-2016					
	With natives			Without natives			With natives			Without natives		
	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3	ΔQ1	ΔQ2	ΔQ3
Top 0	-0.0	-0.51	0.0	-0.0	-1.03	-1.03	0.70	0.35	0.0	0.70	-0.35	-0.70
Top 1	0.51	0.0	-0.51	-0.0	-0.0	-1.02	0.71	-0.35	-0.35	1.05	-0.36	-0.36
Top 2	0.51	0.0	0.0	-1.54	-1.03	-0.51	0.35	-0.0	-0.36	0.35	0.0	-0.35

Table A.19: Difference between the CASF scoring and ranking test results rates computed between the scores produced by KORP-PL^w modifications with weights trained on the reference training set subsets and the reference version of KORP-PL^w.

KORP-PL^w weights training subset	CASF 2013			CASF-2016	
	Δ Scoring	Δ Ranking		Δ Scoring	Δ Ranking
	r_p	"high"	"low"	r_p	r_s
subset$^{\text{closest removed}}$	-0.002	1.5	1.5	-0.002	0.004
subset$^{\text{distant removed}}$	-0.003	-1.6	0.0	-0.005	-0.005

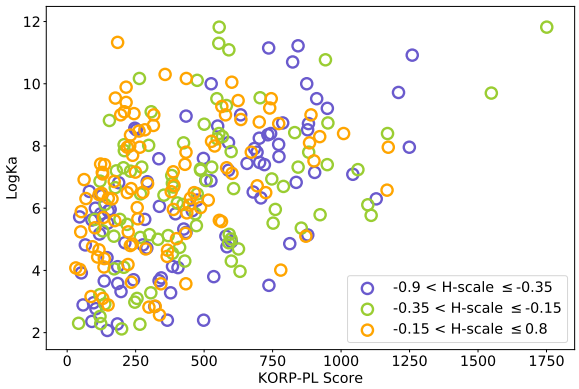
Table A.20: Difference between the CASF screening test results computed between the scores produced by KORP-PL^w modifications with weights trained on the reference training set subsets and the reference version of KORP-PL^w.

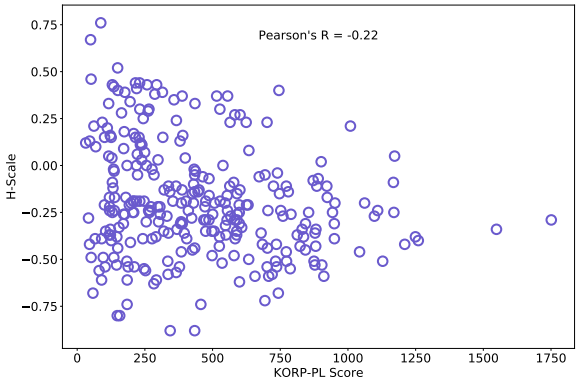
KORP-PL^w weights training subset	CASF 2013						CASF-2016					
	Δ Enrichment			Δ Best binder			Δ Enrichment			Δ Best binder		
	Top, %			Top, %			Top, %			Top, %		
	1	5	10	1	5	10	1	5	10	1	5	10
subset$^{\text{closest removed}}$	1.0	0.4	0.1	1.6	1.5	0.0	1.1	0.2	0.0	3.5	1.7	0.0
subset$^{\text{distant removed}}$	2.1	0.5	-0.1	1.6	1.5	0.0	1.6	0.0	-0.1	3.5	1.7	1.7

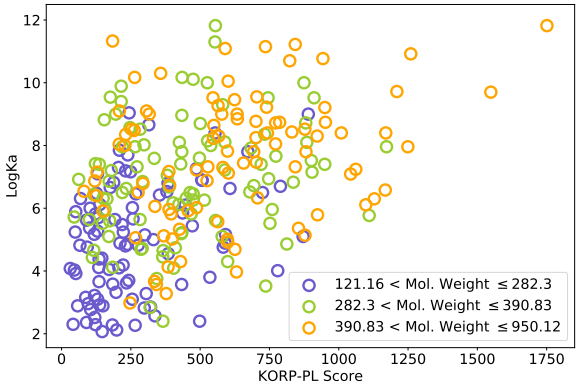
A.5.1 Analysis of the dependencies between KORP-PL scores and the properties of protein-ligand complexes

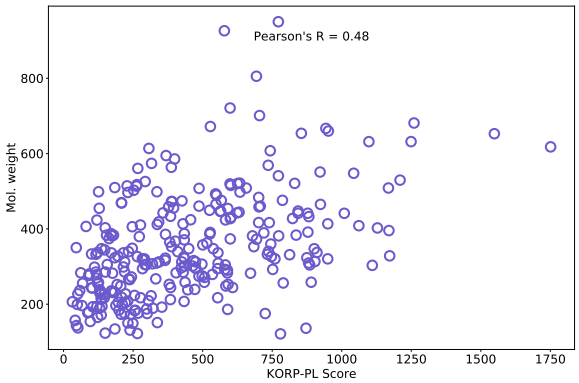
We have computed a range of ligand properties – ligand molecular weight, ligand number of atoms, ligand number of rotational bonds, ligand SASA buried upon binding, ligand SASA, ligand surface area, ligand volume, and colored the KORP-PL scores vs binding constant plots with respect to the values of these properties. Table A.21 indicates that, as the majority of knowledge-based scoring functions, KORP-PL score positively correlates with the ligand molecule size.

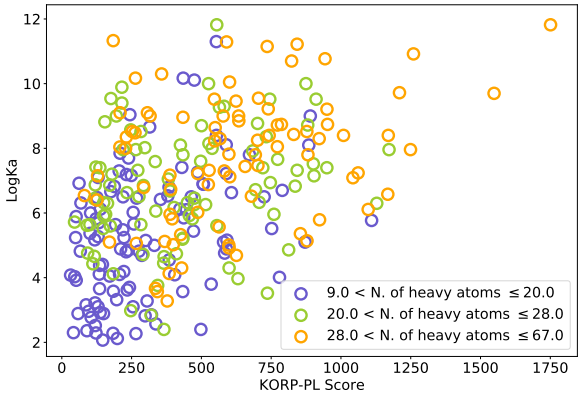
Table A.21: Left: KORP-PL scores and experimental values of the binding constants colored by the complex and ligand properties: hydrophobicity of the binding pocket, ligand molecular weight, ligand number of atoms, ligand number of rotational bonds, ligand SASA buried upon binding, ligand SASA, ligand surface area, ligand volume. Right: KORP-PL scores plotted versus protein-ligand complex properties.

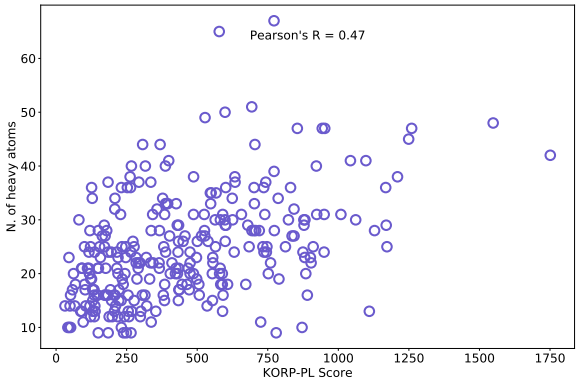


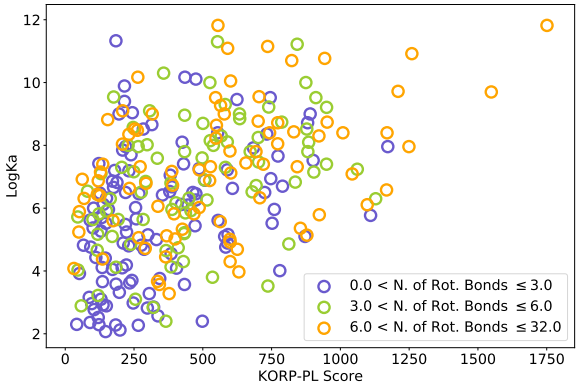


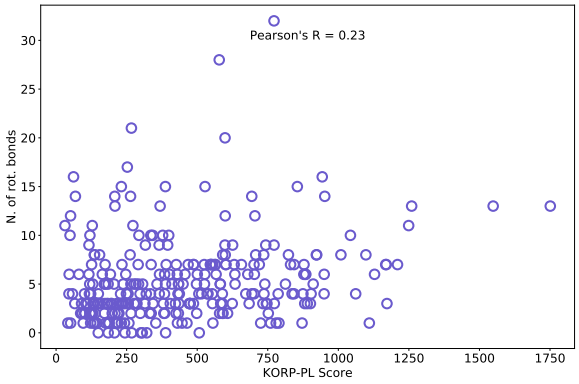


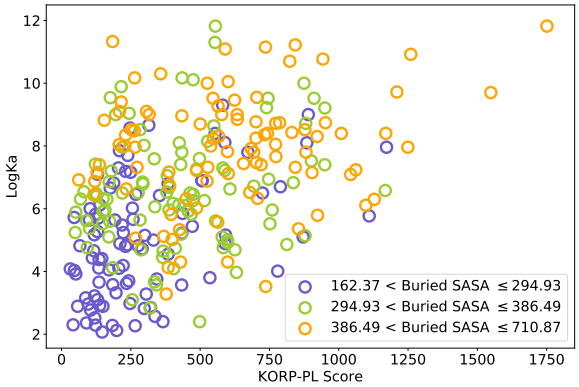


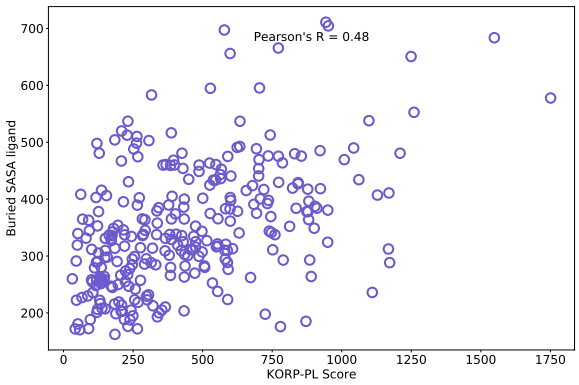


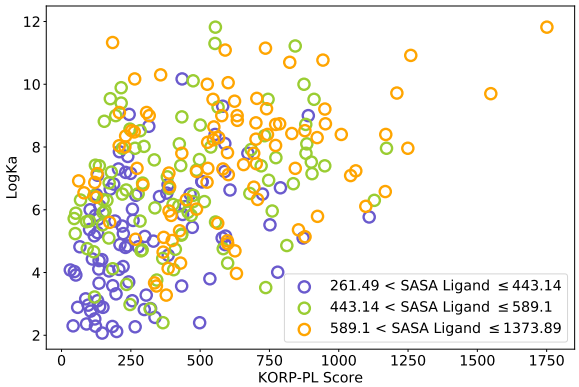


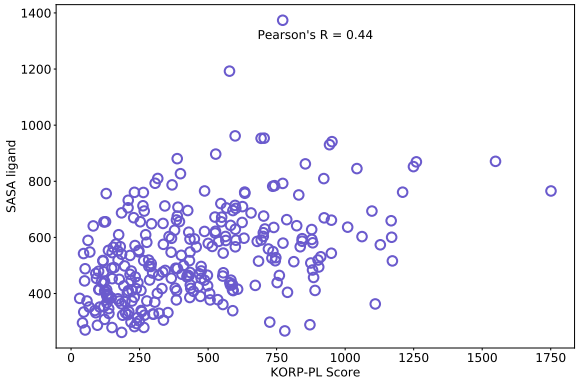


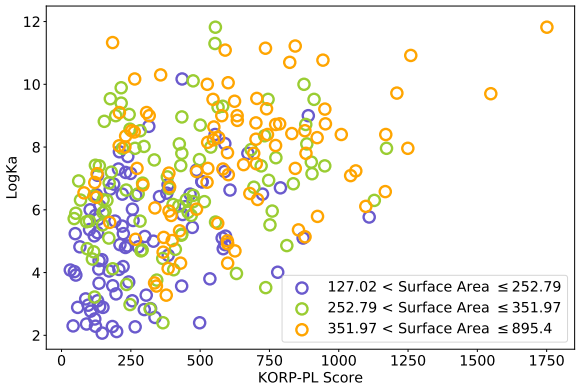


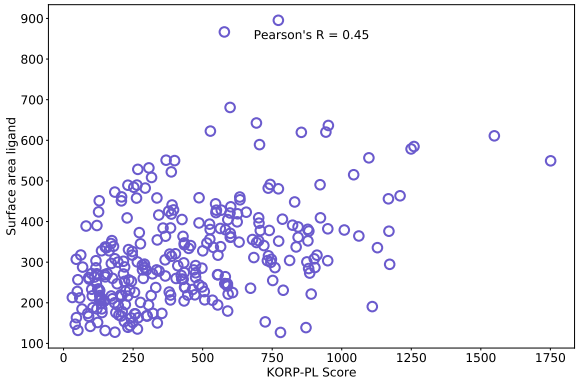


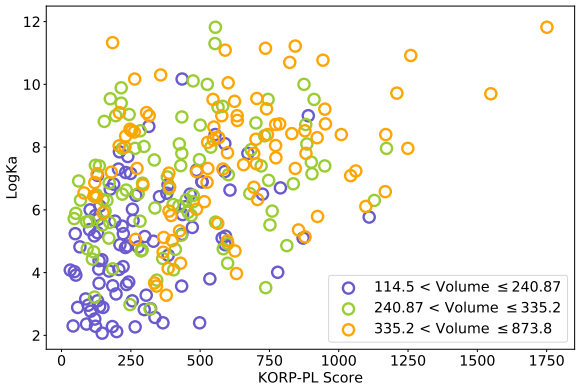


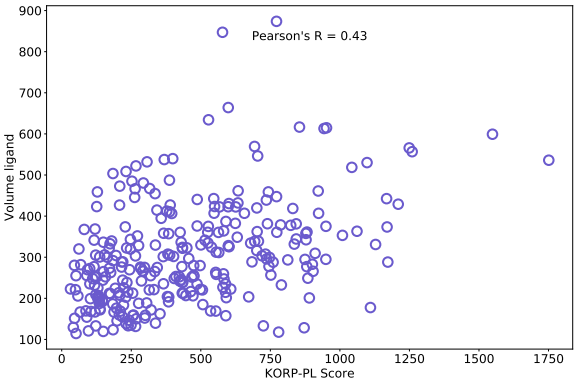












A.6 D3R Benchmark results

Tables A.23, A.24, A.25, A.26, A.27, A.22 list the results of Convex-PL, Convex-PL^R, KORP-PL, KORP-PL^w, AutoDock Vina, and ΔSAS in the benchmark obtained from the D3R Challenges structures, user submissions, and binding constants.

Table A.22: Correlation between the predicted binding affinity of the native poses from the D3R Grand Challenge 2, Grand Challenge 3, and Grand Challenge 4, and the experimentally obtained binding constants. r_p, r_s correspond to the Pearson's and Spearman's correlation coefficients, τ denotes the Kendall's τ coefficient.

Scoring test, native structures

Scoring function	Correlation coefficients								
	GC2			GC3			GC4		
	r_p	r_s	τ	r_p	r_s	τ	r_p	r_s	τ
Convex-PL	0.238	0.159	0.123	0.148	0.092	0.065	0.165	0.071	0.067
Convex-PL^R	0.525	0.522	0.334	0.226	0.251	0.207	0.157	0.026	0.033
KORP-PL	0.56	0.57	0.44	0.023	-0.77	-0.04	0.25	0.34	0.15
KORP-PL^w	0.564	0.587	0.415	0.014	-0.008	-0.006	0.343	0.376	0.233
AutoDock Vina	0.249	0.187	0.129	0.195	0.112	0.075	0.407	0.515	0.333
ΔSAS	0.49	0.37	0.24	0.44	0.39	0.33	0.53	0.17	0.12

Table A.23: Success rates of finding native and near-native docking poses with RMSD smaller than 1 Å , 2 Å , and 3 Å in the 1%, 5%, and 10% of the structures taken from the D3R Grand Challenge 2 user submissions.

Pose prediction test, D3R Grand Challenge 2

Scoring function	With natives									Without natives								
	Q1			Q2			Q3			Q1			Q2			Q3		
	top, %			top, %			top, %			top, %			top, %			top, %		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Convex-PL	0.51	0.8	0.94	0.71	0.91	0.97	0.83	0.94	0.97	0.62	0.83	0.83	0.7	0.85	0.88	0.79	0.88	0.88
Convex-PL^R	0.51	0.89	0.97	0.66	0.97	1.0	0.74	1.0	1.0	0.71	0.79	0.92	0.67	0.88	0.88	0.74	0.88	0.88
KORP-PL	0.46	0.63	0.66	0.63	0.66	0.71	0.63	0.74	0.86	0.67	0.92	0.96	0.67	0.7	0.76	0.65	0.76	0.88
KORP-PL^w	0.49	0.66	0.71	0.6	0.66	0.77	0.66	0.77	0.86	0.67	0.92	0.92	0.61	0.67	0.76	0.65	0.79	0.88
AutoDock Vina	0.51	0.69	0.8	0.66	0.86	0.89	0.71	0.91	0.94	0.71	0.88	0.88	0.67	0.88	0.88	0.71	0.91	0.91
ΔSAS	0.29	0.57	0.71	0.6	0.8	0.86	0.63	0.86	0.91	0.42	0.75	0.88	0.64	0.82	0.85	0.65	0.88	0.91

Table A.24: Success rates of finding native and near-native docking poses with RMSD smaller than 1 Å , 2 Å , and 3 Å in the 1%, 5%, and 10% of the structures taken from the D3R Grand Challenge 3 Stage 1a user submissions.

Pose prediction test, D3R Grand Challenge 3

Scoring function	With natives									Without natives								
	Q1			Q2			Q3			Q1			Q2			Q3		
	top, %			top, %			top, %			top, %			top, %			top, %		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Convex-PL	0.04	0.17	0.54	0.08	0.42	0.79	0.25	0.71	0.92	0.08	0.17	0.5	0.09	0.39	0.74	0.25	0.71	0.88
Convex-PL^R	0.04	0.25	0.5	0.04	0.38	0.67	0.12	0.58	0.88	0.0	0.08	0.17	0.0	0.3	0.48	0.08	0.58	0.79
KORP-PL	0.25	0.62	0.83	0.46	0.88	0.92	0.67	0.96	1	0.25	0.67	0.75	0.43	0.78	0.83	0.62	0.88	0.92
KORP-PL^w	0.29	0.54	0.79	0.5	0.83	0.92	0.62	0.96	1	0.42	0.58	0.75	0.48	0.78	0.87	0.58	0.88	0.96
AutoDock Vina	0	0	0	0.04	0.04	0.08	0.04	0.12	0.12	0	0	0	0.04	0.04	0.09	0.04	0.12	0.12
ΔSAS	0	0	0	0	0.04	0.04	0.08	0.12	0.25	0	0	0	0	0.04	0.04	0.08	0.12	0.25

Table A.25: Success rates of finding native and near-native docking poses with RMSD smaller than 1 Å , 2 Å , and 3 Å in the 1%, 5%, and 10% of the structures taken from the D3R Grand Challenge 3 Stage 1b user submissions.

Pose prediction test, D3R Grand Challenge 3

Scoring function	With natives									Without natives								
	Q1			Q2			Q3			Q1			Q2			Q3		
	top, %			top, %			top, %			top, %			top, %			top, %		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Convex-PL	0.0	0.08	0.62	0.0	0.17	0.67	0.04	0.29	0.75	0.0	0.07	0.86	0.0	0.12	0.67	0.04	0.25	0.75
Convex-PL^R	0.04	0.42	0.67	0.0	0.46	0.75	0.0	0.67	0.79	0.0	0.0	0.14	0.0	0.12	0.42	0.0	0.46	0.58
KORP-PL	0.38	0.88	0.92	0.67	1	1	0.79	1	1	0.5	0.64	0.86	0.62	0.88	0.96	0.75	0.96	1
KORP-PL^w	0.42	0.75	0.88	0.54	1	1	0.67	1	1	0.5	0.71	0.79	0.46	0.96	0.96	0.62	0.96	1
AutoDock Vina	0	0.08	0.12	0.04	0.12	0.21	0.04	0.12	0.33	0	0.07	0.07	0.04	0.12	0.17	0.04	0.12	0.29
ΔSAS	0	0	0.12	0	0	0.29	0	0.12	0.33	0	0	0.14	0	0	0.25	0	0.12	0.33

Table A.26: Success rates of finding native and near-native docking poses with RMSD smaller than 1 Å , 2 Å , and 3 Å in the 1%, 5%, and 10% of the structures taken from the D3R Grand Challenge 4 Stage 1a user submissions.

Pose prediction test, D3R Grand Challenge 4

Scoring function	With natives									Without natives								
	Q1			Q2			Q3			Q1			Q2			Q3		
	top, %			top, %			top, %			top, %			top, %			top, %		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Convex-PL	0.8	1.0	1.0	1	1	1	1	1	1	0.7	0.95	0.95	0.9	1.0	1.0	0.95	1.0	1.0
Convex-PL^R	0.95	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.85	0.95	1.0	1.0	1.0	1.0	1.0	1.0
KORP-PL	0.95	1	1	1	1	1	1	1	1	0.65	0.9	1	1	1	1	0.95	1	1
KORP-PL^w	0.95	1	1	1	1	1	1	1	1	0.65	0.9	1	1	1	1	1	1	1
AutoDock Vina	0.5	0.95	1	0.95	1	1	0.95	1	1	0.5	0.95	1	0.95	1	1	0.95	1	1
ΔSAS	0.4	0.75	0.9	0.8	1	1	0.9	1	1	0.4	0.75	0.9	0.8	1	1	0.9	1	1

Table A.27: Success rates of finding native and near-native docking poses with RMSD smaller than 1 Å , 2 Å , and 3 Å in the 1%, 5%, and 10% of the structures taken from the D3R Grand Challenge 4 Stage 1b user submissions.

Pose prediction test, D3R Grand Challenge 4

Scoring function	With natives									Without natives								
	Q1			Q2			Q3			Q1			Q2			Q3		
	top, %			top, %			top, %			top, %			top, %			top, %		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Convex-PL	0.65	0.95	1.0	1	1	1	1	1	1	0.65	0.95	0.95	1	1	1	1	1	1
Convex-PL^R	0.85	0.9	0.95	1.0	1.0	1.0	1.0	1.0	1.0	0.7	0.75	0.9	1.0	1.0	1.0	1.0	1.0	1.0
KORP-PL	0.95	1	1	1	1	1	1	1	1	0.6	0.95	1	1	1	1	1	1	1
KORP-PL^w	1	1	1	1	1	1	1	1	1	0.65	1	1	0.95	1	1	1	1	1
AutoDock Vina	0.55	1	1	0.95	1	1	1	1	1	0.55	1	1	0.95	1	1	1	1	1
ΔSAS	0.15	0.7	0.8	0.8	1	1	0.8	1	1	0.15	0.7	0.8	0.8	1	1	0.8	1	1

A.7 DUD

Table A.28: ROC AUC, EF5%, and BEDROC$_{\alpha=20}$ values computed for the results of virtual screening done by AutoDock Vina, Vinardo, Convex-PL^R, and Convex-PL for the 31 targets from the DUD datset.

	ROC AUC				EF5%				BEDROC$_{\alpha=20}$			
Target	AutoDock Vina	Vinardo	Convex-PL^R	Convex-PL	AutoDock Vina	Vinardo	Convex-PL^R	Convex-PL	AutoDock Vina	Vinardo	Convex-PL^R	Convex-PL
ace	0.373	0.432	0.575	0.451	1.633	2.041	7.755	4.082	0.116	0.136	0.378	0.202
ache	0.663	0.678	0.574	0.568	3.738	4.299	1.869	2.617	0.236	0.258	0.131	0.143
ada	0.503	0.599	0.526	0.224	0.0	2.564	1.026	0.0	0.013	0.128	0.091	0.004
ampc	0.313	0.441	0.670	0.461	0.952	0.952	0.952	0.0	0.029	0.059	0.064	0.016
ar	0.792	0.735	0.671	0.655	10.886	7.595	4.304	5.316	0.560	0.445	0.230	0.275
cdk2	0.542	0.590	0.529	0.494	2.50	4.722	3.333	4.167	0.176	0.261	0.215	0.237
cox1	0.675	0.673	0.547	0.456	7.20	8.0	0.80	0.80	0.389	0.362	0.077	0.072
cox2	0.905	0.894	0.871	0.542	13.239	12.911	10.235	0.376	0.70	0.687	0.567	0.037
egfr	0.582	0.719	0.784	0.785	2.484	3.789	6.863	7.116	0.132	0.231	0.385	0.396
er_agonist	0.794	0.806	0.855	0.597	8.955	9.254	8.955	0.895	0.459	0.493	0.478	0.082
er_antagonist	0.686	0.697	0.750	0.720	4.615	7.179	8.205	5.128	0.276	0.352	0.408	0.271
fgfr1	0.358	0.393	0.577	0.532	0.50	0.840	3.167	1.667	0.030	0.057	0.164	0.080
fxa	0.662	0.584	0.529	0.687	1.781	2.50	1.507	2.329	0.118	0.152	0.090	0.151
gr	0.539	0.473	0.463	0.516	2.308	1.795	2.051	1.795	0.141	0.111	0.129	0.109
hivpr	0.729	0.830	0.734	0.559	3.871	8.065	8.387	3.548	0.260	0.397	0.433	0.165
hivrt	0.643	0.634	0.592	0.559	4.651	4.186	4.651	2.326	0.257	0.223	0.237	0.141
hmga	0.660	0.803	0.755	0.463	3.429	4.0	4.571	0.571	0.181	0.229	0.309	0.036
hsp90	0.568	0.776	0.562	0.269	0.0	4.324	2.162	0.0	0.040	0.20	0.153	0.019
mr	0.811	0.807	0.826	0.772	14.667	14.667	13.333	5.333	0.674	0.647	0.598	0.253
na	0.375	0.373	0.637	0.493	0.0	0.0	1.224	0.0	0.003	0.009	0.122	0.007
p38	0.613	0.628	0.552	0.601	2.335	1.234	1.762	1.630	0.179	0.126	0.120	0.117
parp	0.711	0.603	0.828	0.691	4.0	2.286	8.571	2.286	0.218	0.113	0.456	0.120
pde5	0.706	0.757	0.725	0.489	5.0	5.0	8.409	2.727	0.322	0.359	0.503	0.188
pdgfrb	0.326	0.372	0.381	0.360	1.294	1.302	1.647	1.059	0.082	0.080	0.089	0.076
ppar	0.743	0.941	0.933	0.925	4.471	15.938	17.177	16.0	0.222	0.769	0.860	0.798
pr	0.444	0.397	0.441	0.297	1.482	0.0	1.482	0.741	0.063	0.024	0.076	0.039
rxr	0.926	0.916	0.869	0.847	16.0	14.0	12.0	6.0	0.786	0.745	0.597	0.288
src	0.642	0.709	0.769	0.612	1.761	5.223	7.296	3.774	0.119	0.250	0.408	0.194
thrombin	0.691	0.516	0.673	0.707	4.722	2.50	3.889	5.0	0.281	0.134	0.249	0.307
trypsin	0.727	0.514	0.810	0.790	2.041	4.082	5.714	9.388	0.157	0.177	0.344	0.449
vegfr2	0.707	0.721	0.665	0.606	7.50	5.909	6.818	6.591	0.407	0.352	0.392	0.326
mean	0.626	0.646	0.667	0.572	4.452	5.199	5.488	3.331	0.246	0.276	0.302	0.181
median	0.662	0.673	0.670	0.559	3.429	4.186	4.571	2.329	0.181	0.229	0.249	0.143

Table A.29: ROC AUC, EF5%, and BEDROC$_{\alpha=20}$ values computed for the results of virtual screening done by KORP-PL and KORP-PL^w rescoring of VinaCPL and AutoDock Vina poses for the 31 targets from the DUD datset.

	ROC AUC						EF5%						BEDROC$_{\alpha=20}$					
Target	KORP-PL / VinaCPL	KORP-PL^w / VinaCPL	KORP-PL / Vina	KORP-PL^w / Vina	KORP-PL / VinaCPL, local opt restricted	KORP-PL^w / VinaCPL, local opt restricted	KORP-PL / VinaCPL	KORP-PL^w / VinaCPL	KORP-PL / Vina	KORP-PL^w / Vina	KORP-PL / VinaCPL, local opt restricted	KORP-PL^w / VinaCPL, local opt restricted	KORP-PL / VinaCPL	KORP-PL^w / VinaCPL	KORP-PL / Vina	KORP-PL^w / Vina	KORP-PL / VinaCPL, local opt restricted	KORP-PL^w / VinaCPL, local opt restricted
ace	0.779	0.741	0.658	0.628	0.821	0.756	7.347	7.347	4.490	3.674	7.347	8.163	0.431	0.437	0.235	0.237	0.435	0.436
ache	0.303	0.418	0.343	0.483	0.302	0.418	0.561	0.748	1.121	0.935	1.495	0.935	0.048	0.053	0.073	0.067	0.066	0.056
ada	0.692	0.670	0.658	0.663	0.676	0.659	3.077	4.103	5.128	4.615	3.590	4.615	0.245	0.220	0.328	0.311	0.259	0.215
ampc	0.917	0.907	0.842	0.857	0.859	0.872	16.191	16.191	4.762	6.667	8.571	16.191	0.805	0.824	0.258	0.347	0.553	0.723
ar	0.739	0.850	0.701	0.806	0.725	0.849	2.025	5.823	1.013	2.025	1.772	6.329	0.202	0.328	0.148	0.193	0.209	0.338
cdk2	0.803	0.826	0.752	0.769	0.793	0.805	5.278	5.833	4.444	4.722	3.889	6.111	0.302	0.348	0.281	0.296	0.261	0.315
cox1	0.474	0.571	0.456	0.544	0.50	0.606	1.60	2.40	0.80	1.60	1.60	3.20	0.101	0.160	0.070	0.111	0.159	0.209
cox2	0.473	0.526	0.516	0.539	0.468	0.504	1.315	1.080	1.315	1.315	0.939	0.986	0.085	0.063	0.092	0.088	0.099	0.069
egfr	0.956	0.949	0.954	0.945	0.952	0.943	15.790	14.40	15.284	13.60	14.863	13.558	0.802	0.743	0.781	0.701	0.772	0.687
er_agonist	0.710	0.891	0.635	0.856	0.713	0.892	4.179	8.955	2.090	6.269	2.388	8.955	0.232	0.456	0.145	0.350	0.240	0.474
er_antagonist	0.906	0.832	0.864	0.845	0.886	0.803	10.769	10.769	10.256	10.769	8.205	10.769	0.529	0.551	0.536	0.558	0.512	0.531
fgfr1	0.866	0.885	0.676	0.682	0.907	0.930	11.50	12.833	4.0	5.833	13.333	14.0	0.607	0.671	0.205	0.297	0.682	0.710
fxa	0.959	0.953	0.689	0.679	0.909	0.896	16.438	14.931	9.589	8.904	14.658	10.959	0.726	0.704	0.483	0.455	0.596	0.568
gr	0.642	0.736	0.690	0.776	0.655	0.752	2.308	4.359	4.103	5.897	3.077	5.385	0.169	0.279	0.198	0.326	0.199	0.311
hivpr	0.872	0.872	0.896	0.897	0.833	0.854	5.161	6.452	8.710	10.645	4.839	5.806	0.384	0.425	0.469	0.524	0.339	0.391
hivrt	0.668	0.715	0.643	0.678	0.666	0.706	4.186	6.046	3.256	3.256	4.651	6.046	0.287	0.335	0.212	0.215	0.30	0.327
hmga	0.590	0.776	0.492	0.715	0.554	0.745	8.0	8.0	4.0	5.143	7.429	7.429	0.443	0.427	0.201	0.236	0.444	0.406
hsp90	0.767	0.825	0.799	0.827	0.726	0.812	3.243	4.324	6.487	6.487	4.324	2.703	0.221	0.275	0.423	0.446	0.151	0.218
mr	0.694	0.902	0.760	0.921	0.667	0.891	1.333	10.667	5.333	10.667	0.0	12.0	0.144	0.515	0.224	0.508	0.147	0.507
na	0.899	0.834	0.781	0.734	0.899	0.826	13.061	13.061	10.204	8.980	12.245	12.653	0.603	0.635	0.473	0.462	0.590	0.627
p38	0.401	0.450	0.715	0.762	0.425	0.485	1.234	1.894	1.189	2.467	0.661	1.674	0.079	0.123	0.128	0.210	0.074	0.118
parp	0.923	0.878	0.770	0.740	0.812	0.765	10.286	9.143	9.714	8.571	9.714	9.714	0.533	0.516	0.467	0.422	0.490	0.493
pde5	0.842	0.806	0.828	0.803	0.839	0.806	9.318	8.182	8.409	7.045	8.636	8.182	0.572	0.498	0.538	0.468	0.576	0.508
pdgfrb	0.506	0.549	0.593	0.630	0.533	0.571	5.059	5.882	4.941	6.471	4.824	6.471	0.277	0.332	0.289	0.370	0.30	0.369
ppar	0.946	0.920	0.908	0.890	0.898	0.854	15.529	14.353	13.882	12.235	12.706	8.0	0.780	0.721	0.701	0.614	0.587	0.469
pr	0.709	0.648	0.611	0.592	0.720	0.663	6.667	2.222	5.185	2.222	6.667	3.704	0.332	0.178	0.273	0.166	0.354	0.202
rxr	0.912	0.948	0.799	0.890	0.916	0.945	9.0	15.0	6.0	11.0	8.0	13.0	0.543	0.659	0.315	0.482	0.496	0.606
src	0.872	0.880	0.880	0.886	0.838	0.850	14.465	14.465	15.094	14.969	12.453	11.447	0.740	0.762	0.792	0.792	0.575	0.594
thrombin	0.854	0.815	0.844	0.808	0.849	0.806	12.222	12.222	8.333	9.444	10.833	10.278	0.552	0.577	0.408	0.427	0.468	0.498
trypsin	0.941	0.941	0.927	0.927	0.939	0.930	12.245	12.245	9.388	10.612	10.204	9.388	0.50	0.539	0.426	0.498	0.466	0.509
vegfr2	0.855	0.885	0.853	0.875	0.855	0.879	7.273	10.227	6.818	8.864	8.409	9.773	0.463	0.558	0.417	0.502	0.487	0.559
mean	0.757	0.787	0.727	0.763	0.746	0.777	7.634	8.521	6.301	6.965	6.849	8.014	0.411	0.449	0.342	0.377	0.383	0.421
median	0.803	0.832	0.752	0.776	0.812	0.806	7.273	8.182	5.185	6.487	7.347	8.163	0.431	0.456	0.289	0.370	0.435	0.469

A.8 DUD-E

Table A.30: ROC AUC values, BEDROC values for $\alpha = 20$, and 5% enrichment factors computed on the DUD-E benchmark for KORP-PL, KORP-PL^w, and AutoDock Vina. Average results are computed including and excluding the 12 targets containing co-factors that are not parametrized in KORP-PL.

		ROC AUC					BEDROC $_{\alpha=20}$					EF 5%				
Target	Has cofactor	AutoDock Vina	KORP-PL / Vina	KORP-PL^w / Vina	KORP-PL / VinaCPL, local opt restricted	KORP-PL^w / VinaCPL, local opt restricted	AutoDock Vina	KORP-PL / Vina	KORP-PL^w / Vina	KORP-PL / VinaCPL, local opt restricted	KORP-PL^w / VinaCPL, local opt restricted	AutoDock Vina	KORP-PL / Vina	KORP-PL^w / Vina	KORP-PL / VinaCPL, local opt restricted	KORP-PL^w / VinaCPL, local opt restricted
aa2ar		0.592	0.738	0.728	0.745	0.732	0.120	0.453	0.437	0.410	0.376	2.180	7.962	7.725	7.227	6.682
abl1		0.777	0.837	0.804	0.836	0.80	0.337	0.530	0.499	0.459	0.420	5.966	9.763	9.559	8.136	7.254
ace		0.525	0.647	0.646	0.836	0.827	0.091	0.276	0.282	0.536	0.539	1.270	4.159	4.309	8.568	8.618
aces		0.725	0.534	0.612	0.483	0.558	0.311	0.152	0.278	0.092	0.195	5.843	2.861	4.940	1.476	3.343
ada		0.413	0.799	0.685	0.743	0.624	0.044	0.309	0.191	0.225	0.173	0.458	4.351	2.595	3.282	2.672
ada17		0.647	0.942	0.923	0.911	0.870	0.380	0.80	0.712	0.750	0.611	6.736	14.807	12.972	14.015	11.032
adrb1		0.723	0.565	0.565	0.498	0.508	0.199	0.110	0.127	0.052	0.094	3.362	2.052	2.227	0.873	1.572
adrb2		0.756	0.516	0.499	0.439	0.437	0.271	0.069	0.108	0.044	0.085	4.922	1.253	1.790	0.626	1.298
akt1		0.749	0.836	0.820	0.890	0.887	0.226	0.502	0.473	0.535	0.530	4.255	9.835	8.842	10.118	9.787
akt2		0.827	0.858	0.822	0.887	0.879	0.421	0.584	0.578	0.623	0.627	8.105	11.263	10.632	12.0	11.684
aldr	yes	0.708	0.499	0.433	0.428	0.359	0.256	0.047	0.034	0.034	0.006	4.455	0.545	0.455	0.545	0.0
ampc		0.603	0.765	0.718	0.752	0.686	0.085	0.470	0.414	0.401	0.348	1.613	8.387	7.419	7.419	6.452
andr		0.623	0.731	0.790	0.723	0.790	0.319	0.369	0.376	0.403	0.371	5.048	6.730	6.272	6.730	5.966
aofb	yes	0.575	0.519	0.418	0.518	0.398	0.183	0.085	0.076	0.062	0.044	3.214	1.429	1.190	0.833	0.714
bace1		0.715	0.874	0.873	0.840	0.843	0.173	0.641	0.635	0.516	0.547	3.134	12.0	11.835	10.103	9.814
braf		0.835	0.915	0.916	0.899	0.896	0.383	0.667	0.659	0.628	0.627	6.853	13.227	13.147	11.793	11.873
cah2		0.530	0.811	0.822	0.742	0.781	0.136	0.593	0.594	0.461	0.466	2.371	10.802	10.802	8.335	8.192
casp3		0.763	0.636	0.652	0.687	0.684	0.250	0.322	0.281	0.247	0.193	4.413	5.559	4.871	4.069	3.152
cdk2		0.738	0.853	0.859	0.843	0.854	0.265	0.505	0.551	0.440	0.518	4.311	9.574	9.950	8.371	9.699
comt		0.714	0.905	0.873	0.963	0.90	0.094	0.505	0.444	0.772	0.513	2.093	9.767	8.837	16.512	10.698
cp2c9	yes	0.601	0.568	0.586	0.598	0.612	0.088	0.109	0.189	0.120	0.198	1.421	1.421	3.825	2.186	3.388
cp3a4	yes	0.576	0.595	0.637	0.616	0.664	0.152	0.109	0.130	0.155	0.192	2.340	1.504	2.061	2.340	3.287
csf1r		0.683	0.887	0.892	0.901	0.903	0.128	0.585	0.607	0.505	0.509	1.958	11.049	11.399	10.070	9.510
cxcr4		0.653	0.686	0.671	0.639	0.637	0.091	0.232	0.213	0.228	0.208	1.148	3.770	3.607	3.934	3.770
def		0.706	0.954	0.956	0.954	0.950	0.199	0.867	0.867	0.894	0.857	3.727	16.149	16.398	17.516	16.646
dhi1	yes	0.766	0.431	0.565	0.439	0.586	0.218	0.039	0.117	0.020	0.153	3.776	0.501	1.927	0.116	2.852
dpp4		0.575	0.839	0.809	0.853	0.824	0.074	0.479	0.390	0.406	0.365	1.131	9.138	6.766	7.173	6.339
drd3		0.667	0.650	0.663	0.632	0.664	0.144	0.156	0.170	0.121	0.172	2.126	2.697	3.063	1.943	2.971
dyr	yes	0.677	0.881	0.801	0.775	0.629	0.207	0.565	0.412	0.220	0.090	3.675	10.106	7.350	3.640	1.378
egfr		0.632	0.871	0.848	0.890	0.870	0.162	0.651	0.609	0.685	0.641	3.197	12.115	11.563	13.678	12.548
esr1		0.789	0.698	0.730	0.758	0.752	0.363	0.250	0.234	0.343	0.276	6.093	4.498	4.083	6.029	4.785
esr2		0.784	0.658	0.714	0.671	0.693	0.384	0.204	0.228	0.196	0.195	7.092	3.765	4.235	3.059	3.059
fa10		0.815	0.928	0.915	0.925	0.913	0.384	0.846	0.782	0.793	0.719	6.187	15.530	14.318	14.369	12.854
fa7		0.885	0.946	0.934	0.917	0.901	0.402	0.890	0.877	0.836	0.780	7.027	17.081	16.865	15.784	14.486
fabp4		0.811	0.525	0.554	0.432	0.468	0.485	0.029	0.054	0.002	0.009	8.421	0.0	0.351	0.0	0.0
fak1		0.857	0.953	0.957	0.957	0.959	0.425	0.784	0.801	0.826	0.835	7.018	15.263	15.614	16.491	17.018
fgfr1		0.540	0.465	0.465	0.454	0.453	0.40	0.283	0.305	0.224	0.252	1.322	0.909	0.909	0.579	0.826
fkb1a		0.729	0.537	0.705	0.402	0.70	0.233	0.127	0.293	0.016	0.159	3.370	1.758	4.542	0.073	1.978
fnta		0.688	0.772	0.718	0.803	0.743	0.214	0.398	0.298	0.459	0.303	3.534	6.785	5.059	7.778	5.118
fpps		0.396	0.998	0.983	0.999	0.988	0.004	0.975	0.820	0.991	0.887	0.0	19.624	16.808	20.0	17.653
gcr		0.569	0.707	0.805	0.725	0.809	0.274	0.222	0.350	0.271	0.384	4.440	3.446	5.684	4.725	6.679
glcm		0.725	0.595	0.731	0.430	0.70	0.261	0.022	0.212	0.002	0.176	2.748	0.064	2.236	0.0	1.597
gria2		0.744	0.645	0.659	0.669	0.655	0.315	0.224	0.234	0.339	0.344	5.791	4.377	4.377	5.926	5.926
grik1		0.617	0.865	0.780	0.838	0.729	0.149	0.523	0.346	0.451	0.217	2.632	10.0	6.711	8.816	3.289
hdac2		0.807	0.754	0.686	0.708	0.593	0.341	0.503	0.420	0.367	0.145	6.303	8.992	7.311	6.723	2.269
hdac8		0.859	0.931	0.923	0.925	0.885	0.578	0.852	0.864	0.810	0.733	11.111	16.239	16.667	15.641	14.017
hivint		0.660	0.493	0.512	0.441	0.451	0.091	0.024	0.047	0.012	0.019	1.422	0.284	0.474	0.095	0.095
hivpr		0.737	0.821	0.876	0.840	0.898	0.198	0.481	0.561	0.485	0.593	3.011	7.90	9.176	8.043	9.692
hivrt		0.622	0.753	0.701	0.763	0.708	0.160	0.285	0.298	0.286	0.295	2.570	4.745	5.074	4.745	4.712
hmdh		0.771	0.708	0.802	0.637	0.774	0.228	0.534	0.553	0.377	0.428	3.612	9.298	9.632	6.221	7.023
hs90a		0.30	0.832	0.821	0.790	0.777	0.003	0.664	0.693	0.523	0.464	0.0	13.120	13.440	10.40	9.920
hxk4		0.559	0.658	0.633	0.681	0.60	0.092	0.077	0.104	0.125	0.038	1.417	0.945	1.575	1.890	0.315
igf1r		0.835	0.882	0.880	0.938	0.937	0.363	0.578	0.577	0.742	0.759	6.814	10.973	10.885	15.398	15.310
inha	yes	0.706	0.391	0.446	0.347	0.456	0.318	0.0	0.004	0.001	0.006	5.915	0.0	0.0	0.0	0.0
ital		0.595	0.573	0.619	0.505	0.558	0.055	0.173	0.207	0.154	0.233	1.202	2.918	3.348	2.575	4.034
jak2		0.765	0.950	0.946	0.929	0.929	0.324	0.786	0.754	0.701	0.693	5.490	15.686	14.510	14.510	14.248
kif11		0.878	0.815	0.846	0.844	0.892	0.566	0.594	0.639	0.599	0.668	10.761	10.863	11.371	11.066	11.980

kit		0.767	0.823	0.820	0.838	0.841	0.198	0.324	0.355	0.289	0.312	3.016	5.476	6.032	5.397	5.556
kith		0.634	0.573	0.536	0.512	0.459	0.336	0.156	0.201	0.124	0.163	5.455	2.121	2.879	1.667	2.879
kpcb		0.725	0.575	0.586	0.590	0.603	0.365	0.205	0.241	0.196	0.232	6.694	3.629	4.839	3.790	4.274
lck		0.775	0.816	0.816	0.828	0.824	0.258	0.512	0.493	0.452	0.444	4.334	9.40	8.873	8.433	8.082
lkha4		0.875	0.874	0.844	0.891	0.795	0.503	0.561	0.487	0.522	0.305	10.574	10.656	8.852	9.836	5.738
mapk2		0.799	0.742	0.697	0.718	0.684	0.336	0.313	0.319	0.253	0.281	6.019	5.340	5.146	3.786	4.951
mcr		0.591	0.698	0.813	0.688	0.782	0.185	0.189	0.313	0.261	0.272	2.383	2.694	5.596	4.456	4.456
met		0.825	0.953	0.954	0.957	0.961	0.365	0.836	0.841	0.836	0.855	6.639	16.721	16.803	16.557	16.803
mk01		0.857	0.872	0.890	0.890	0.904	0.374	0.503	0.621	0.527	0.683	6.619	8.489	11.223	10.791	12.518
mk10		0.740	0.894	0.880	0.867	0.857	0.20	0.546	0.504	0.503	0.482	3.656	10.430	9.247	10.0	8.817
mk14		0.688	0.813	0.796	0.822	0.815	0.194	0.477	0.484	0.508	0.520	3.344	9.027	8.984	9.727	9.683
mmp13		0.641	0.958	0.942	0.948	0.934	0.157	0.860	0.807	0.826	0.80	2.775	16.686	15.491	15.80	15.337
mp2k1	yes	0.635	0.712	0.676	0.704	0.672	0.064	0.212	0.146	0.207	0.198	0.992	3.554	2.397	4.132	4.050
nos1	yes	0.637	0.762	0.579	0.697	0.601	0.116	0.163	0.081	0.226	0.131	1.880	2.650	1.197	3.846	2.137
nram		0.614	0.948	0.945	0.945	0.947	0.031	0.670	0.716	0.674	0.730	0.270	12.342	12.973	12.162	13.153
pa2ga		0.609	0.689	0.675	0.659	0.648	0.077	0.198	0.125	0.311	0.259	0.945	3.937	1.890	6.929	5.039
parp1		0.879	0.907	0.877	0.884	0.843	0.397	0.701	0.545	0.645	0.514	7.305	13.747	10.377	12.129	9.730
pde5a		0.702	0.70	0.713	0.713	0.706	0.235	0.277	0.284	0.311	0.294	4.306	4.646	4.731	5.297	4.901
pgh1		0.603	0.445	0.457	0.447	0.460	0.151	0.040	0.050	0.030	0.049	2.550	0.558	0.637	0.398	0.717
pgh2		0.733	0.504	0.523	0.495	0.521	0.419	0.043	0.055	0.036	0.048	8.173	0.640	0.866	0.527	0.678
plk1		0.614	0.895	0.862	0.877	0.848	0.153	0.627	0.604	0.561	0.556	2.710	12.129	12.0	12.129	11.484
pnph		0.858	0.986	0.984	0.986	0.982	0.339	0.932	0.921	0.936	0.919	5.150	17.854	17.511	18.112	17.511
ppara		0.893	0.806	0.734	0.772	0.712	0.432	0.463	0.350	0.404	0.287	8.125	8.566	6.287	7.390	5.368
ppard		0.768	0.791	0.788	0.882	0.879	0.115	0.387	0.330	0.601	0.524	1.458	7.847	6.319	11.597	10.486
pparg		0.793	0.777	0.758	0.762	0.736	0.220	0.463	0.426	0.453	0.395	3.402	8.382	7.497	8.216	7.109
prgr		0.676	0.813	0.862	0.820	0.854	0.331	0.523	0.516	0.497	0.457	5.991	9.324	9.234	8.829	8.108
ptn1		0.854	0.835	0.834	0.806	0.788	0.574	0.70	0.701	0.527	0.556	10.311	12.356	12.711	9.156	9.867
pur2		0.867	0.976	0.960	0.959	0.944	0.412	0.920	0.892	0.879	0.828	4.876	13.134	12.935	12.836	11.642
pygm		0.508	0.553	0.479	0.549	0.534	0.10	0.054	0.056	0.060	0.068	1.930	0.702	1.053	0.877	1.228
pyrd	yes	0.823	0.637	0.449	0.623	0.444	0.406	0.081	0.065	0.092	0.070	7.612	1.194	1.194	1.493	1.045
reni		0.593	0.757	0.830	0.706	0.827	0.111	0.239	0.318	0.114	0.261	1.447	3.359	3.979	1.034	3.256
rock1		0.717	0.921	0.920	0.890	0.889	0.117	0.521	0.532	0.452	0.457	1.872	9.163	9.064	7.192	7.882
rxra		0.815	0.654	0.688	0.732	0.737	0.517	0.237	0.191	0.262	0.216	9.630	4.321	3.333	4.938	3.704
sahh	yes	0.767	0.435	0.474	0.450	0.441	0.364	0.020	0.016	0.014	0.001	5.263	0.105	0.105	0.105	0.0
src		0.631	0.866	0.859	0.841	0.839	0.136	0.616	0.577	0.449	0.452	2.383	11.745	10.638	8.40	8.424
tgfr1		0.874	0.982	0.983	0.982	0.985	0.439	0.857	0.873	0.823	0.856	8.114	16.584	17.153	15.587	16.441
thb		0.837	0.773	0.780	0.750	0.778	0.544	0.431	0.397	0.373	0.377	10.0	7.976	7.262	6.548	7.143
thrb		0.760	0.903	0.888	0.869	0.847	0.207	0.744	0.663	0.690	0.572	3.043	13.566	12.288	12.520	10.407
try1		0.763	0.912	0.905	0.877	0.858	0.181	0.793	0.760	0.730	0.635	2.850	14.644	14.195	13.536	11.504
tryb1		0.763	0.827	0.821	0.740	0.732	0.214	0.445	0.349	0.365	0.220	3.392	7.836	6.667	6.316	3.860
tysy	yes	0.820	0.728	0.632	0.631	0.509	0.427	0.248	0.117	0.143	0.039	6.945	4.051	1.608	2.058	0.386
urok		0.733	0.969	0.921	0.981	0.962	0.198	0.916	0.748	0.940	0.804	3.203	17.778	13.922	18.954	14.706
vgfr2		0.757	0.874	0.872	0.866	0.855	0.319	0.540	0.573	0.505	0.532	6.097	10.097	10.774	9.226	9.903
wee1		0.958	0.965	0.973	0.927	0.943	0.839	0.769	0.853	0.581	0.723	16.204	15.912	16.788	12.409	14.745
xiap		0.708	0.914	0.916	0.928	0.937	0.262	0.727	0.738	0.734	0.741	4.651	13.178	14.264	14.419	14.419
Average, 102 targets		0.712	0.763	0.759	0.751	0.748	0.260	0.433	0.424	0.411	0.396	4.461	7.886	7.661	7.589	7.178
Median, 102 targets		0.725	0.802	0.802	0.774	0.782	0.231	0.467	0.404	0.408	0.373	3.701	8.384	7.287	7.309	6.566
Average, 90 targets		0.715	0.785	0.786	0.775	0.777	0.264	0.472	0.465	0.451	0.436	4.528	8.637	8.423	8.365	7.922
Median, 90 targets		0.731	0.815	0.818	0.825	0.812	0.234	0.502	0.459	0.453	0.436	3.692	9.082	8.840	8.276	7.568

Abbreviations

CASF Comparative Assessment of Scoring Functions

D3R Drug Design Data Resource

DUD Directory of Useful Decoys

DUD-E Directory of Useful Decoys - Enhanced

ITC Isothermal titration calorimetry

MCMC Markov chain Monte Carlo

NMR Nuclear magnetic resonance

PDB Protein Data Bank

RMSD Root mean squared deviation

SASA Solvent-accessible surface area

SAXS Small angle X-ray scattering

SPR Surface plasmon resonance

SVM Support Vector Machines

Bibliography

[1] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.*, 72(3):1047–1069, 1997.

[2] M. K. Gilson and H.-X. Zhou. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.*, 36:21–42, 2007.

[3] A. Tamura and P. L. Privalov. The entropy cost of protein association. *J. Mol. Biol.*, 273(5):1048–1060, 1997.

[4] C. E. A. Chang, W. Chen, and M. K. Gilson. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U. S. A.*, 104(5):1534–1539, 2007.

[5] C. H. Reynolds and M. K. Holloway. Thermodynamics of ligand binding and efficiency. *ACS Med. Chem. Lett.*, 2(6):433–437, 2011.

[6] M. K. Gilson. An Introduction to Protein-Ligand Binding for BindingDB Users, 2010. `http://bindingdb.org/bind/BindingDB-Intro2a.pdf`.

[7] V. Kairys, L. Baranauskiene, M. Kazlauskiene, et al. Binding affinity in drug design: experimental and computational techniques. *Expert opinion on drug discovery*, 14(8):755–768, 2019.

[8] G. U. Nienhaus. *Protein-ligand interactions: methods and applications*, volume 305. Springer, 2005.

[9] X. Du, Y. Li, Y. L. Xia, et al. Insights into protein–ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.*, 17(2):1–34, 2016.

[10] D. M. Jameson and G. Mocz. Fluorescence polarization/anisotropy approaches to study protein-ligand interactions. In G. U. Nienhaus, editor, *Protein-Ligand Interactions*, pages 301–322. Springer, 2005.

[11] P. A. Van Der Merwe. Surface plasmon resonance. In S. E. Harding and B. Z. Chowdhry, editors, *Protein-ligand interactions: hydrodynamics and calorimetry*, volume 1, pages 137–170. Oxford Univ Press Oxford, UK, 2001.

[12] M. M. Pierce, C. S. Raman, and B. T. Nall. Isothermal titration calorimetry of protein-protein interactions. *Methods A Companion to Methods Enzymol.*, 19(2):213–221, 1999.

[13] C. Yung-Chi and W. H. Prusoff. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.*, 22(23):3099–3108, 1973.

[14] R. A. Engh and R. Huber. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A*, 47(4):392–400, 1991.

[15] Bernhard Rupp. *Biomolecular Crystallography*. Garland Science, 2010.

[16] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. 280:5705–5736, 2013.

[17] I. Pelczer and S. Szalma. Multidimensional NMR and data processing. *Chem Rev*, 91(7):1507–1524, 1991.

[18] K. M. Yip, N. Fischer, E. Paknia, et al. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161, 2020.

[19] M. Wu and G. C. Lander. How low can we go? structure determination of small biological complexes using single-particle cryo-em. *Current Opinion in Structural Biology*, 64:9–16, 2020.

[20] P. W. Rose, A. Prlić, A. Altunkaya, et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.*, 45(D1):D271–D281, 2017.

[21] O. S. Smart, V. Horský, S. Gore, et al. Validation of ligands in macromolecular structures determined by X-ray crystallography. *Acta Crystallogr. Sect. D Struct. Biol.*, 74:228–236, 2018.

[22] C. X. Weichenberger, E. Pozharski, and B. Rupp. Visualizing ligand molecules in twilight electron density. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 69(2):195–200, 2013.

[23] A. Cereto-Massagué, M. J. Ojeda, R. P. Joosten, et al. The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites. *J. Cheminform.*, 5(7):1–9, 2013.

[24] A. Meyder, E. Nittinger, G. Lange, et al. Estimating Electron Density Support for Individual Atoms and Molecular Fragments in X-ray Structures. *J. Chem. Inf. Model.*, 57(10):2437–2447, 2017.

[25] R. P. Joosten and G. Vriend. PDB improvement starts with data deposition. *Science*, 317(5835):195–196, 2007.

[26] S. Boutet, L. Lomb, G. J. Williams, et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science*, 337(6092):362–364, 2012.

[27] D. P. Frueh, A. C. Goodrich, S. H. Mishra, and S. R. Nichols. NMR methods for structural studies of large monomeric and multimeric proteins. *Current opinion in structural biology*, 23(5):734–739, 2013.

[28] S. Gražulis, A. Daškevič, A. Merkys, et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic acids research*, 40(D1):D420–D427, 2012.

[29] A. Gaulton, A. Hersey, M. Nowotka, et al. The ChEMBL database in 2017. *Nucleic Acids Res.*, 45(D1):D945–D954, 11 2016.

[30] M. Davies, M. Nowotka, G. Papadatos, et al. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.*, 43(W1):W612–W620, 2015.

[31] M. K. Gilson, T. Liu, M. Baitaluk, et al. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, 44(D1):D1045–D1053, 2016.

[32] Y. Wang, J. Xiao, T. O. Suzek, et al. PubChem's BioAssay database. *Nucleic Acids Res.*, 40(D1):D400–D412, 2012.

[33] Z. Liu, M. Su, L. Han, et al. Forging the basis for developing protein–ligand interaction scoring functions. *Acc. Chem. Res.*, 50(2):302–309, 2017.

[34] M. L. Benson, R. D. Smith, N. A. Khazanov, et al. Binding MOAD, a high-quality protein–ligand database. *Nucleic Acids Res.*, 36(suppl_1):D674–D678, 2007.

[35] L. Wang, B. Berne, and R. A. Friesner. On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. *Proc. Natl. Acad. Sci. U.S.A*, 109(6):1937–1942, 2012.

[36] N. Homeyer and H. Gohlke. FEW: A workflow tool for free energy calculations of ligand binding. *J. Comput. Chem.*, 34(11):965–973, 2013.

[37] L. Wang, Y. Wu, Y. Deng, et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.*, 137(7):2695–2703, 2015. PMID: 25625324.

[38] J. Liu and R. Wang. Classification of current scoring functions. *J. Chem. Inf. Model.*, 55(3):475–482, 2015. PMID: 25647463.

[39] I. Muegge and Y. C. Martin. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.*, 42(5):791–804, 1999.

[40] H. F. Velec, H. Gohlke, and G. Klebe. DrugScoreCSD-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.*, 48(20):6296–6303, 2005.

[41] W. Mooij and M. L. Verdonk. General and targeted statistical potentials for protein–ligand interactions. *Proteins: Struct., Funct., Bioinf.*, 61(2):272–287, 2005.

[42] S.-Y. Huang and X. Zou. Mean-Force Scoring Functions for Protein–Ligand Binding. *Annu. Rep.-Comp. Chem.*, 6:280–296, 2010.

[43] H. Zhou and J. Skolnick. GOAP: a Generalized Orientation-Dependent, All-Atom Statistical Potential for Protein Structure Prediction. *Biophys. J.*, 101(8):2043–2052, 2011.

[44] S.-Y. Huang and X. Zou. Scoring and Lessons Learned with the CSAR Benchmark Using an Improved Iterative Knowledge-Based Scoring Function. *J. Chem. Inf. Model.*, 51(9):2097–2106, 2011.

[45] G. Neudert and G. Klebe. DSX: a knowledge-based scoring function for the assessment of protein–ligand complexes. *J. Chem. Inf. Model.*, 51(10):2731–2745, 2011.

[46] P. Popov and S. Grudinin. Knowledge of Native Protein-Protein Interfaces Is Sufficient To Construct Predictive Models for the Selection of Binding Candidates. *J. Chem. Inf. Model.*, 55(10):2242–55, Oct 2015.

[47] Z. Zheng and K. M. Merz. Development of the Knowledge-Based and Empirical Combined Scoring Algorithm (KECSA) to Score Protein–Ligand Interactions. *J. Chem. Inf. Model.*, 53(5):1073–1083, 2013. PMID: 23560465.

[48] P. Chen, Y. Ke, Y. Lu, et al. DLIGAND2: an improved knowledge-based energy function for protein–ligand interactions using the distance-scaled, finite, ideal-gas reference state. *J. Cheminformatics*, 11(1):52, 2019.

[49] M. Kadukova, K. d. S. Machado, P. Chacón, and S. Grudinin. KORP-PL: a coarse-grained knowledge-based scoring function for protein-ligand interactions. *Bioinformatics*, 2020.

[50] H. Gohlke, M. Hendlich, and G. Klebe. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, 295(2):337–356, 2000.

[51] M. Kulharia, R. S. Goody, and R. M. Jackson. Information theory-based scoring function for the structure-based prediction of protein-ligand binding affinity. *J. Chem. Inf. Model.*, 48(10):1990–1998, 2008.

[52] S.-Y. Huang and X. Zou. Inclusion of Solvation and Entropy in the Knowledge–Based Scoring Function for Protein–Ligand Interactions. *J. Chem. Inf. Model.*, 50(2):262–273, 2010. PMID: 20088605.

[53] T. Debroise, E. I. Shakhnovich, and N. Cheron. A hybrid knowledge-based and empirical scoring function for protein–ligand interaction: SMoG2016. *J. Chem. Inf. Model*, 57(3):584–593, 2017.

[54] Z. Yan and J. Wang. Incorporating specificity into optimization: evaluation of spa using csar 2014 and casf 2013 benchmarks. *J. Comput.-Aided Mol. Des.*, 30(3):219–227, 2016.

[55] M. Kadukova and S. Grudinin. Convex-pl: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. *J. Comput. Aided. Mol. Des.*, 31(10):943–958, 2017.

[56] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, et al. Charmm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.

[57] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[58] T. J. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz. Dock 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.*, 15(5):411–428, 2001.

[59] D. A. Case, T. E. Cheatham, T. Darden, et al. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.*, 26(16):1668–1688, 2005.

[60] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

[61] S. Yin, L. Biedermannova, J. Vondrasek, and N. V. Dokholyan. MedusaScore: An accurate force field-based scoring function for virtual drug screening. *J. Chem. Inf. Model.*, 48(8):1656–1662, 2008.

[62] H.-J. Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.*, 8(3):243–256, 1994.

[63] A. N. Jain. Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J. Comput.-Aided Mol. Des.*, 10(5):427–440, 1996.

[64] M. D. Eldridge, C. W. Murray, T. R. Auton, et al. Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.*, 11(5):425–445, 1997.

[65] G. M. Morris, D. S. Goodsell, R. S. Halliday, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, 19(14):1639–1662, 1998.

[66] R. Wang, L. Lai, and S. Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.*, 16(1):11–26, 2002.

[67] R. A. Friesner, J. L. Banks, R. B. Murphy, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.*, 47(7):1739–1749, 2004. PMID: 15027865.

[68] R. A. Friesner, R. B. Murphy, M. P. Repasky, et al. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein- ligand complexes. *J. Med. Chem.*, 49(21):6177–6196, 2006.

[69] O. Korb, T. Stutzle, and T. E. Exner. Empirical Scoring Functions for Advanced Protein–Ligand Docking With Plants. *J. Chem. Inf. Model.*, 49(1):84–96, 2009.

[70] O. Trott and A. J. Olson. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.*, 31(2):455–461, 2010.

[71] S.-H. Wang, Y.-T. Wu, S.-C. Kuo, and J. Yu. Hotlig: A molecular surface-directed approach to scoring protein–ligand interactions. *J. Chem. Inf. Model.*, 53(8):181–2195, 2013.

[72] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, et al. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.*, 10(4):1–7, 2014.

[73] H. Hogues, T. Sulea, and E. O. Purisima. Evaluation of the wilma-sie virtual screening method in community structure–activity resource 2013 and 2014 blind challenges. *J. Chem. Inf. Model.*, 56(6):955–964, 2015.

[74] R. Quiroga and M. A. Villarreal. Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PloS One*, 11(5):e0155183, 2016.

[75] S. Piotto, L. Di Biasi, R. Fino, et al. Yada: a novel tool for molecular docking calculations. *J. Comput.-Aided Mol. Des.*, 30(9):753–759, 2016.

[76] P. J. Ballester and J. B. Mitchell. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics*, 26(9):1169–1175, 2010.

[77] S. L. Kinnings, N. Liu, P. J. Tonge, et al. A Mach. Learn.-Based Method to Improve Docking Scoring Functions and its Application to Drug Repurposing. *J. Chem. Inf. Model.*, 51(2):408–419, 2011. PMID: 21291174.

[78] D. Zilian and C. A. Sotriffer. SFCscore RF: a random forest-based scoring function for improved affinity prediction of protein–ligand complexes. *J. Chem. Inf. Model*, 53(8):1923–1933, 2013.

[79] G.-B. Li, L.-L. Yang, W.-J. Wang, et al. ID-Score: A New Empirical Scoring Function Based on a Comprehensive Set of Descriptors Related to Protein–Ligand Interactions. *J. Chem. Inf. Model.*, 53(3):592–600, 2013. PMID: 23394072.

[80] C. Wang and Y. Zhang. Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest. *J. Comput. Chem.*, 38(3):169–177, 2017.

[81] H. M. Ashtawy and N. R. Mahapatra. Task-Specific Scoring Functions for Predicting Ligand Binding Poses and Affinity and for Screening Enrichment. *J. Chem. Inf. Model*, 58(1):119–133, 2017.

[82] J. Lu, X. Hou, C. Wang, and Y. Zhang. Incorporating explicit water molecules and ligand conformation stability in machine-learning scoring functions. *J. Chem. Inf. Model*, 59(11):4540–4549, 2019.

[83] D. D. Nguyen and G.-W. Wei. Agl-score: Algebraic graph learning score for protein–ligand binding scoring, ranking, docking, and screening. *J. Chem. Inf. Model*, 59(7):3291–3304, 2019.

[84] H. Li, K. H. Sze, G. Lu, and P. J. Ballester. Machine-learning scoring functions for structure-based virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 7(March):1–21, 2020.

[85] M. Ragoza, J. Hochuli, E. Idrobo, et al. Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model*, 57(4):942–957, 2017.

[86] H. Jiang, M. Fan, J. Wang, et al. Guiding Conventional Protein–Ligand Docking Software with Convolutional Neural Networks. *J. Chem. Inf. Model.*, 2020.

[87] C. Shen, J. Ding, Z. Wang, et al. From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(1):e1429, 2020.

[88] I. Wallach, M. Dzamba, and A. Heifets. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855*, 2015.

[89] J. Jiménez, M. Skalic, G. Martinez-Rosell, and G. De Fabritiis. K-deep: Protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *J. Chem. Inf. Model*, 58(2):287–296, 2018.

[90] D. S. Karlov, S. Sosnin, M. V. Fedorov, and P. Popov. graphdelta: Mpnn scoring function for the affinity prediction of protein–ligand complexes. *ACS Omega*, 2020.

[91] P. Francoeur, T. Masuda, J. Sunseri, et al. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design. *J. Chem. Inf. Model*, 60(9):4200–4215, 2020.

[92] L. Chen, A. Cruz, S. Ramsey, et al. Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. *PloS One*, 14(8):e0220113, 2019.

[93] J. Scantlebury, N. Brown, F. Von Delft, and C. M. Deane. Dataset augmentation allows deep learning-based virtual screening to better generalize to unseen target classes, and highlight important binding interactions. *J. Chem. Inf. Model*, 60(8):3722–3730, 2020.

[94] Y. Li, M. Su, Z. Liu, et al. Assessing protein–ligand interaction scoring functions with the casf-2013 benchmark. *Nat. Protoc.*, 13(4):666, 2018.

[95] M. Su, Q. Yang, Y. Du, et al. Comparative assessment of scoring functions: the casf-2016 update. *J. Chem. Inf. Model*, 59(2):895–913, 2018.

[96] Schrödinger, LLC. The pymol molecular graphics system, version 1.3. 2011.

[97] E. Krissinel and K. Henrick. Detection of protein assemblies in crystals. In *International Symposium on Computational Life Science*, pages 163–174. Springer, 2005.

[98] C. W. Murray and M. L. Verdonk. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput. Aided. Mol. Des.*, 16(10):741–753, 2002.

[99] J. Wang, W. Wang, S. Huo, et al. Solvation model based on weighted solvent accessible surface area. *J. Phys. Chem. B*, 105(21):5055–5067, 2002.

[100] A. Ben-Naim. Statistical Potentials Extracted from Protein Structures: Are These Meaningful Potentials? *J. Chem. Phys.*, 107(9):3698–3706, 1997.

[101] T. Lazaridis and M. Karplus. Effective energy function for proteins in solution. *Proteins Struct. Funct. Genet.*, 52(2):176–192, 2003.

[102] P. Labute. The generalized born/volume integral implicit solvent model: estimation of the free energy of hydration using london dispersion instead of atomic surface area. *J. Comput. Chem.*, 29(10):1693–1698, 2008.

[103] I. Reulecke, G. Lange, J. Albrecht, et al. Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the hyde scoring function. *ChemMedChem: Chemistry Enabling Drug Discovery*, 3(6):885–897, 2008.

[104] N. Schneider, G. Lange, S. Hindle, et al. A consistent description of hydrogen bond and dehydration energies in protein–ligand complexes: methods behind the hyde scoring function. *J. Comput. Aided. Mol. Des.*, 27(1):15–29, 2013.

[105] Y. Cao and L. Li. Improved protein–ligand binding affinity prediction by using a curvature-dependent surface-area model. *Bioinformatics*, 30(12):1674–1680, 2014.

[106] S. Genheden, T. Luchko, S. Gusarov, et al. An MM/3D-RISM approach for ligand binding affinities. *J. Phys. Chem. B*, 114(25):8505–8516, 2010.

[107] J. M. Swanson, R. H. Henchman, and J. A. McCammon. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.*, 86(1 I):67–74, 2004.

[108] X. Zou, Y. Sun, and I. D. Kuntz. Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *J. Am. Chem. Soc.*, 121(35):8033–8043, 1999.

[109] M. S. Lee and M. A. Olson. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys. J.*, 90(3):864–877, 2006.

[110] S. L. N. Y. NY. Schrödinger release 2019-1: Watermap. `https://www.schrodinger.com/watermap`.

[111] D. Cappel, W. Sherman, and T. Beuming. Calculating water thermodynamics in the binding site of proteins–applications of watermap to drug discovery. *Curr. Top. Med. Chem*, 17(23):2586–2598, 2017.

[112] M. D. L. U. Pinner. Flap/waterflap. `http://www.moldiscovery.com/software/flap/`.

[113] A. Sridhar, G. A. Ross, and P. C. Biggin. Waterdock 2.0: Water placement prediction for holo-structures with a pymol plugin. *PloS One*, 12(2):e0172743, 2017.

[114] M. F. Lensink, G. Brysbaert, N. Nadzirin, et al. Blind prediction of interfacial water positions in capri. *Proteins: Struct., Funct., Bioinf.*, 82(4):620–632, 2014.

[115] Y. Li, Y. D. Gao, M. K. Holloway, and R. Wang. Prediction of the favorable hydration sites in a protein binding pocket and its application to scoring function formulation. *J. Chem. Inf. Model*, 2020.

[116] V. Le Guilloux, P. Schmidtke, and P. Tuffery. Fpocket: an open source platform for ligand pocket detection. *BMC bioinf.*, 10(1):1–11, 2009.

[117] E. Lionta, G. Spyrou, D. Vassilatis, and Z. Cournia. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr. Top. Med. Chem.*, 14(16):1923–1938, 2014.

[118] S. Grudinin, E. Laine, and A. Hoffmann. Predicting protein functional motions: an old recipe with a new twist. *Biophys. J.*, 2020.

[119] S. Riniker and G. A. Landrum. Better informed distance geometry: using what we know to improve conformation generation. *J. Chem. Inf. Model*, 55(12):2562–2574, 2015.

[120] S. Wang, J. Witek, G. A. Landrum, and S. Riniker. Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *J. Chem. Inf. Model*, 60(4):2044–2058, 2020.

[121] M. D. Hanwell, D. E. Curtis, D. C. Lonie, et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.*, 4(1):17, 2012.

[122] P. C. Hawkins, A. G. Skillman, G. L. Warren, et al. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.*, 50(4):572–584, 2010.

[123] J. Gasteiger, C. Rudolph, and J. Sadowski. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.*, 3(6 PART C):537–547, 1990.

[124] M. A. Miteva, F. Guyon, and P. Tufféry. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.*, 38(SUPPL. 2):622–627, 2010.

[125] K. S. Watts, P. Dalal, R. B. Murphy, et al. ConfGen: A conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.*, 50(4):534–546, 2010.

[126] N. Yoshikawa and G. R. Hutchison. Fast, efficient fragment-based coordinate generation for Open Babel. *J. Cheminform.*, 11(1):1–9, 2019.

[127] N. O. Friedrich, F. Flachsenberg, A. Meyder, et al. Conformator: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.*, 59(2):731–742, 2019.

[128] J. M. Blaney and J. S. Dixon. Distance geometry in molecular modeling. *Reviews in computational chemistry*, pages 299–335, 1994.

[129] G. Landrum. Rdkit: Open-source cheminformatics. http://www.rdkit.org.

[130] N. S. Pagadala, K. Syed, and J. Tuszynski. Software for molecular docking: a review. *Biophys. Rev.*, 9(2):91–102, 2017.

[131] R. L. DesJarlais, R. P. Sheridan, G. L. Seibel, et al. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.*, 31(4):722–729, 1988.

[132] D. Kozakov, R. Brenke, S. R. Comeau, and S. Vajda. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins Struct. Funct. Genet.*, 65(2):392–406, 2006.

[133] D. Padhorny, D. R. Hall, H. Mirzaei, et al. Protein–ligand docking using FFT based sampling: D3R case study. *J. Comput. Aided. Mol. Des.*, 32(1):225–230, 2018.

[134] F. Ding, S. Yin, and N. V. Dokholyan. Rapid flexible docking using a stochastic rotamer library of ligands. *J. Chem. Inf. Model.*, 50(9):1623–1632, 2010.

[135] R. Abagyan and M. Totrov. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.*, 235(3):983–1002, 1994.

[136] G. Jones, P. Willett, R. C. Glen, et al. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267(3):727–748, 1997.

[137] G. M. Morris, R. Huey, W. Lindstrom, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30(16):2785–2791, 2009.

[138] W.-H. Shin and C. Seok. Galaxydock: protein–ligand docking with flexible protein side-chains. *J. Chem. Inf. Model*, 52(12):3225–3232, 2012.

[139] L. Hoffer and D. Horvath. S4MPLE - Sampler for multiple protein-ligand entities: Simultaneous docking of several entities. *J. Chem. Inf. Model.*, 53(1):88–102, 2013.

[140] O. Korb, T. Stützle, and T. E. Exner. An ant colony optimization approach to flexible protein–ligand docking. *Swarm Intel.*, 1(2):115–134, 2007.

[141] R. Abagyan, M. Totrov, and D. Kuznetsov. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.*, 15(5):488–506, 1994.

[142] I. W. Davis and D. Baker. RosettaLigand Docking with Full Ligand and Receptor Flexibility. *J. Mol. Biol.*, 385(2):381–392, 2009.

[143] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, 261(3):470–489, 1996.

[144] A. N. Jain. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.*, 46(4):499–511, 2003.

[145] V. Zoete, A. Grosdidier, and O. Michielin. Docking, virtual high throughput screening and in silico fragment-based drug design. *J. Cell. Mol. Med.*, 13(2):238–248, 2009.

[146] A. Scarpino, G. G. Ferenczy, and G. M. Keserü. Comparative Evaluation of Covalent Docking Tools. *J. Chem. Inf. Model.*, 58(7):1441–1458, 2018.

[147] P. Cozzini, G. E. Kellogg, F. Spyrakis, et al. Target flexibility: An emerging consideration in drug discovery and design. *J. Med. Chem.*, 51(20):6237–6255, 2008.

[148] M. Totrov and R. Abagyan. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol*, 18(2):178–184, 2008.

[149] R. M. Knegtel, I. D. Kuntz, and C. M. Oshiro. Molecular docking to ensembles of protein structures. *J. Mol. Biol.*, 266(2):424–440, 1997.

[150] G. Bottegoni, I. Kufareva, M. Totrov, and R. Abagyan. Four-dimensional docking: A fast and accurate account of discrete receptor flexibility in ligand docking. *J. Med. Chem.*, 52(2):397–406, 2009.

[151] M. Ignatov, C. Liu, A. Alekseenko, et al. Monte Carlo on the manifold and MD refinement for binding pose prediction of protein–ligand complexes: 2017 D3R Grand Challenge. *J. Comput. Aided. Mol. Des.*, 33(1):119–127, 2019.

[152] M. I. Zavodszky and L. A. Kuhn. Side-chain flexibility in protein-ligand binding: The minimal rotation hypothesis. *Protein Sci.*, 14(4):1104–1114, 2005.

[153] M. P. Beck and B. W. Lin. Some heuristics for the consensus ranking problem. *Comput. Oper. Res.*, 10(1):1–7, 1983.

[154] X. Ren, Y. S. Shi, Y. Zhang, et al. Novel Consensus Docking Strategy to Improve Ligand Pose Prediction. *J. Chem. Inf. Model.*, 58(8):1662–1668, 2018.

[155] K. Palacio-Rodríguez, I. Lans, C. N. Cavasotto, and P. Cossio. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci. Rep.*, 9(1):1–14, 2019.

[156] R. Wang, Y. Lu, and S. Wang. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.*, 46(12):2287–2303, 2003.

[157] G. L. Warren, C. W. Andrews, A.-M. Capelli, et al. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, 49(20):5912–5931, 2006. PMID: 17004707.

[158] T. Cheng, X. Li, Y. Li, et al. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.*, 49(4):1079–1093, 2009.

[159] Y. Li, L. Han, Z. Liu, and R. Wang. Comparative assessment of scoring functions on an updated benchmark: 2. evaluation methods and general results. *J. Chem. Inf. Model.*, 54(6):1717–1736, 2014.

[160] M. Su, Q. Yang, Y. Du, et al. Comparative assessment of scoring functions: the CASF-2016 update. *J. Chem. Inf. Model*, 59(2):895–913, 2018.

[161] M. J. Hartshorn, M. L. Verdonk, G. Chessari, et al. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.*, 50(4):726–741, 2007.

[162] R. D. Smith, J. Dunbar, J. B., P. M. Ung, et al. CSAR Benchmark Exercise of 2010: Combined Evaluation Across All Submitted Scoring Functions. *J. Chem. Inf. Model.*, 51:2115–2131, 2011.

[163] R. D. Smith, K. L. Damm-Ganamet, J. B. Dunbar, et al. CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. *J. Chem. Inf. Model.*, 56(6):1022–1031, 2016.

[164] H. A. Carlson, R. D. Smith, K. L. Damm-Ganamet, et al. CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J. Chem. Inf. Model.*, May 2016.

[165] N. Huang, B. K. Shoichet, and J. J. Irwin. Benchmarking sets for molecular docking. *J. Med. Chem.*, 49(23):6789–6801, 2006.

[166] S. G. Rohrer and K. Baumann. Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model.*, 49(2):169–184, 2009.

[167] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.*, 55(14):6582–6594, 2012.

[168] M. R. Bauer, T. M. Ibrahim, S. M. Vogel, and F. M. Boeckler. Evaluation and optimization of virtual screening workflows with DEKOIS 2.0–a public library of challenging docking benchmark sets. *J. Chem. Inf. Model*, 53(6):1447–1462, 2013.

[169] V. K. Tran-Nguyen, C. Jacquemard, and D. Rognan. LIT-PCBA: An unbiased data set for machine learning and virtual screening. *J. Chem. Inf. Model.*, 60(9):4263–4273, 2020.

[170] J. Sieg, F. Flachsenberg, and M. Rarey. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. *J. Chem. Inf. Model.*, 59(3):947–961, 2019.

[171] Z. Gaieb, S. Liu, S. Gathiaka, et al. D3R grand challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided. Mol. Des.*, 32(1):1–20, 2018.

[172] Z. Gaieb, C. D. Parks, M. Chiu, et al. D3R Grand Challenge 3: blind prediction of protein–ligand poses and affinity rankings. *J. Comput. Aided. Mol. Des.*, 33(1):1–18, 2019.

[173] C. D. Parks, Z. Gaieb, M. Chiu, et al. D3R Grand Challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided. Mol. Des.*, Jan 2020.

[174] J.-F. Truchon and C. I. Bayly. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.*, 47(2):488–508, 2007.

[175] K. L. Damm-Ganamet, R. D. Smith, J. B. Dunbar Jr, et al. CSAR Benchmark Exercise 2011–2012: Evaluation of Results from Docking and Relative Ranking of Blinded Congeneric Series. *J. Chem. Inf. Model.*, 53(8):1853–1870, 2013.

[176] S. Gathiaka, S. Liu, M. Chiu, et al. D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. *J. Comput. Aided. Mol. Des.*, 30(9):651–668, 2016.

[177] D. R. Koes, M. P. Baumgartner, and C. J. Camacho. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model*, 53(8):1893–1904, 2013.

[178] C. Yan, S. Z. Grinter, B. R. Merideth, et al. Iterative knowledge-based scoring functions derived from rigid and flexible decoy structures: evaluation with the 2013 and 2014 csar benchmarks. *J. Chem. Inf. Model.*, 56(6):1013–1021, 2015.

[179] V. Y. Martiny, F. Martz, E. Selwa, and B. I. Iorga. Blind pose prediction, scoring, and affinity ranking of the csar 2014 dataset. *J. Chem. Inf. Model.*, 56(6):996–1003, 2015.

[180] P. Prathipati and K. Mizuguchi. Integration of ligand and structure based approaches for csar-2014. *J. Chem. Inf. Model.*, 56(6):974–987, 2015.

[181] W.-H. Shin, G. R. Lee, and C. Seok. Evaluation of galaxydock based on the community structure–activity resource 2013 and 2014 benchmark studies. *J. Chem. Inf. Model.*, 56(6):988–995, 2015.

[182] J. Sunseri, M. Ragoza, J. Collins, and D. R. Koes. A d3r prospective evaluation of machine learning for protein-ligand scoring. *J. Comput.-Aided Mol. Des.*, 30(9):761–771, 2016.

[183] P. Prathipati, C. Nagao, S. Ahmad, and K. Mizuguchi. Improved pose and affinity predictions using different protocols tailored on the basis of data availability. *J. Comput.-Aided Mol. Des.*, 30(9):817–828, 2016.

[184] X. Zhu, W.-H. Shin, H. Kim, and D. Kihara. Combined approach of patch-surfer and pl-patchsurfer for protein–ligand binding prediction in csar 2013 and 2014. *J. Chem. Inf. Model.*, 56(6):1088–1099, 2015.

[185] S. Grudinin, M. Kadukova, A. Eisenbarth, et al. Predicting binding poses and affinities for protein - ligand complexes in the 2015 D3R Grand Challenge using a physical model with a statistical parameter estimation. *J. Comput. Aided. Mol. Des.*, 30(9):791–804, 2016.

[186] S.-Y. Huang, M. Li, J. Wang, and Y. Pan. Hybriddock: a hybrid protein–ligand docking protocol integrating protein-and ligand-based approaches. *J. Chem. Inf. Model.*, 56(6):1078–1087, 2015.

[187] A. Kumar and K. Y. Zhang. Application of shape similarity in pose selection and virtual screening in CSARdock2014 exercise. *J. Chem. Inf. Model.*, 56(6):965–973, 2015.

[188] A. Kumar and K. Y. Zhang. Prospective evaluation of shape similarity based pose prediction method in D3R Grand Challenge 2015. *J. Comput.-Aided Mol. Des.*, 30(9):685–693, 2016.

[189] I. Slynko, F. Da Silva, G. Bret, and D. Rognan. Docking pose selection by interaction pattern graph similarity: application to the d3r grand challenge 2015. *J. Comput.-Aided Mol. Des.*, 30(9):669–683, 2016.

[190] V. Salmaso, M. Sturlese, A. Cuzzolin, and S. Moro. Dockbench as docking selector tool: the lesson learned from d3r grand challenge 2015. *J. Comput.-Aided Mol. Des.*, 30(9):773–789, 2016.

[191] D. Fourches, E. Muratov, F. Ding, et al. Predicting binding affinity of CSAR ligands using both structure-based and ligand-based approaches. *J. Chem. Inf. Model.*, 53(8):1915–1922, 2013.

[192] R. Politi, M. Convertino, K. Popov, et al. Docking and scoring with target-specific pose classifier succeeds in native-like pose identification but not binding affinity prediction in the csar 2014 benchmark exercise. *J. Chem. Inf. Model.*, 56(6):1032–1041, 2016.

[193] D. Fourches, R. Politi, and A. Tropsha. Target-specific native/decoy pose classifier improves the accuracy of ligand ranking in the csar 2013 benchmark. *J. Chem. Inf. Model.*, 55(1):63–71, 2014.

[194] D. Santos-Martins. Interaction with specific hsp90 residues as a scoring function: validation in the d3r grand challenge 2015. *J. Comput.-Aided Mol. Des.*, 30(9):731–742, 2016.

[195] N. Deng, W. F. Flynn, J. Xia, et al. Large scale free energy calculations for blind predictions of protein–ligand binding: the d3r grand challenge 2015. *J. Comput.-Aided Mol. Des.*, 30(9):743–751, 2016.

[196] M. M. Ignjatović, O. Caldararu, G. Dong, et al. Binding-affinity predictions of hsp90 in the d3r grand challenge 2015 with docking, mm/gbsa, qm/mm, and free-energy simulations. *J. Comput.-Aided Mol. Des.*, 30(9):707–730, 2016.

[197] S. Ruiz-Carmona and X. Barril. Docking-undocking combination applied to the d3r grand challenge 2015. *J. Comput.-Aided Mol. Des.*, 30(9):805–815, 2016.

[198] A. S. Mey, J. Juárez-Jiménez, A. Hennessy, and J. Michel. Blinded predictions of binding modes and energies of hsp90-α ligands for the 2015 d3r grand challenge. *Bioorganic & Medicinal Chemistry*, 24(20):4890–4899, 2016.

[199] M. P. Baumgartner and C. J. Camacho. Choosing the optimal rigid receptor for docking and scoring in the csar 2013/2014 experiment. *J. Chem. Inf. Model.*, 56(6):1004–1012, 2015.

[200] E. Selwa, V. Y. Martiny, and B. I. Iorga. Molecular docking performance evaluated on the d3r grand challenge 2015 drug-like ligand datasets. *J. Comput.-Aided Mol. Des.*, 30(9):829–839, 2016.

[201] Z. Ye, M. P. Baumgartner, B. M. Wingert, and C. J. Camacho. Optimal strategies for virtual screening of induced-fit and flexible target in the 2015 d3r grand challenge. *J. Comput.-Aided Mol. Des.*, 30(9):695–706, 2016.

[202] P. Nedumpully-Govindan, D. B. Jemec, and F. Ding. Csar benchmark of flexible medusadock in affinity prediction and nativelike binding pose selection. *J. Chem. Inf. Model.*, 56(6):1042–1052, 2015.

[203] P. C.-H. Lam, R. Abagyan, and M. Totrov. Ligand-biased ensemble receptor docking (LigBEnD): a hybrid ligand/receptor structure-based approach. *J. Comput. Aided. Mol. Des.*, 32(1):187–198, 2018.

[204] R. Duan, X. Xu, and X. Zou. Lessons learned from participating in D3R 2016 Grand Challenge 2: compounds targeting the farnesoid X receptor. *J. Comput. Aided. Mol. Des.*, 32(1):103–111, 2018.

[205] P. da Silva Figueiredo Celestino Gomes, F. Da Silva, G. Bret, and D. Rognan. Ranking docking poses by graph matching of protein–ligand interactions: lessons learned from the D3R Grand Challenge 2. *J. Comput. Aided. Mol. Des.*, 32(1):75–87, 2018.

[206] E. Selwa, E. Elisée, A. Zavala, and B. I. Iorga. Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations. *J. Comput. Aided. Mol. Des.*, 32(1):273–286, 2018.

[207] M. Kadukova and S. Grudinin. Docking of small molecules to farnesoid X receptors using AutoDock Vina with the Convex-PL potential: lessons learned from D3R Grand Challenge 2. *J. Comput. Aided. Mol. Des.*, 32(1):151–162, 2018.

[208] V. Salmaso, M. Sturlese, A. Cuzzolin, and S. Moro. Combining self- and cross-docking as benchmark tools: the performance of DockBench in the D3R Grand Challenge 2. *J. Comput. Aided. Mol. Des.*, 32(1):251–264, 2018.

[209] M. Réau, F. Langenfeld, J.-F. Zagury, and M. Montes. Predicting the affinity of Farnesoid X Receptor ligands through a hierarchical ranking protocol: a D3R Grand Challenge 2 case study. *J. Comput. Aided. Mol. Des.*, 32(1):231–238, 2018.

[210] X. Fradera, A. Verras, Y. Hu, et al. Performance of multiple docking and refinement methods in the pose prediction D3R prospective Grand Challenge 2016. *J. Comput. Aided. Mol. Des.*, 32(1):113–127, 2018.

[211] Y.-D. Gao, Y. Hu, A. Crespo, et al. Workflows and performances in the ranking prediction of 2016 D3R Grand Challenge 2: lessons learned from a collaborative effort. *J. Comput. Aided. Mol. Des.*, 32(1):129–142, 2018.

[212] M. P. Baumgartner and D. A. Evans. Lessons learned in induced fit docking and metadynamics in the Drug Design Data Resource Grand Challenge 2. *J. Comput. Aided. Mol. Des.*, 32(1):45–58, 2018.

[213] S. Bhakat, E. Åberg, and P. Söderhjelm. Prediction of binding poses to FXR using multi-targeted docking combined with molecular dynamics and enhanced sampling. *J. Comput. Aided. Mol. Des.*, 32(1):59–73, 2018.

[214] E. P. Raman, W. Yu, S. K. Lakkaraju, and A. D. MacKerell Jr. Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J. Chem. Inf. Model*, 53(12):3384–3398, 2013.

[215] A. Kumar and K. Y. Zhang. Shape similarity guided pose prediction: lessons from D3R Grand Challenge 3. *J. Comput. Aided. Mol. Des.*, 33(1):47–59, 2019.

[216] P. I. Koukos, L. C. Xue, and A. M. Bonvin. Protein–ligand pose and affinity prediction: Lessons from D3R Grand Challenge 3. *J. Comput. Aided. Mol. Des.*, 33(1):83–91, 2019.

[217] D. D. Nguyen, Z. Cang, K. Wu, et al. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. *J. Comput. Aided. Mol. Des.*, 33(1):71–82, 2019.

[218] P. C.-H. Lam, R. Abagyan, and M. Totrov. Hybrid receptor structure/ligand-based docking and activity prediction in ICM: development and evaluation in D3R Grand Challenge 3. *J. Comput. Aided. Mol. Des.*, 33(1):35–46, 2019.

[219] L. Chaput, E. Selwa, E. Elisee, and B. I. Iorga. Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations. *J. Comput. Aided. Mol. Des.*, 33(1):93–103, 2019.

[220] X. He, V. H. Man, B. Ji, et al. Calculate protein–ligand binding affinities with the extended linear interaction energy method: application on the Cathepsin S set in the D3R Grand Challenge 3. *J. Comput. Aided. Mol. Des.*, 33(1):105–117, 2019.

[221] B. Xie and D. D. Minh. Alchemical Grid Dock (AlGDock) calculations in the D3R Grand Challenge 3. *J. Comput. Aided. Mol. Des.*, 33(1):61–69, 2019.

[222] P. C.-H. Lam, R. Abagyan, and M. Totrov. Macrocycle modeling in ICM: benchmarking and evaluation in D3R Grand Challenge 4. *J. Comput. Aided. Mol. Des.*, 33(12):1057–1069, 2019.

[223] S. Kotelnikov, A. Alekseenko, C. Liu, et al. Sampling and refinement protocols for template-based macrocycle docking: 2018 D3R Grand Challenge 4. *J. Comput. Aided. Mol. Des.*, 34(2):179–189, 2020.

[224] A. Kumar and K. Y. Zhang. Improving ligand 3D shape similarity-based pose prediction with a continuum solvent model. *J. Comput. Aided. Mol. Des.*, 33(12):1045–1055, 2019.

[225] Y. Yang, J. Lu, C. Yang, and Y. Zhang. Exploring fragment-based target-specific ranking protocol with machine learning on cathepsin S. *J. Comput. Aided. Mol. Des.*, 33(12):1095–1105, 2019.

[226] L. El Khoury, D. Santos-Martins, S. Sasmal, et al. Comparison of affinity ranking using AutoDock-GPU and MM-GBSA scores for BACE-1 inhibitors in the D3R Grand Challenge 4. *J. Comput. Aided. Mol. Des.*, 33(12):1011–1020, 2019.

[227] D. Santos-Martins, J. Eberhardt, G. Bianco, et al. D3r grand challenge 4: prospective pose prediction of bace1 ligands with autodock-gpu. *J. Comput. Aided. Mol. Des.*, 33(12):1071–1081, 2019.

[228] M. Kadukova, V. Chupin, and S. Grudinin. Docking rigid macrocycles using Convex-PL, AutoDock Vina, and RDKit in the D3R Grand Challenge 4. *J. Comput. Aided. Mol. Des.*, 34(2):191–200, 2020.

[229] J. Zou, C. Tian, and C. Simmerling. Blinded prediction of protein–ligand binding affinity using Amber thermodynamic integration for the 2018 D3R grand challenge 4. *J. Comput. Aided. Mol. Des.*, 33(12):1021–1029, 2019.

[230] E. Elisée, V. Gapsys, N. Mele, et al. Performance evaluation of molecular docking and free energy calculations protocols using the D3R Grand Challenge 4 dataset. *J. Comput. Aided. Mol. Des.*, 33(12):1031–1043, 2019.

[231] A. Basciu, P. I. Koukos, G. Malloci, et al. Coupling enhanced sampling of the apo-receptor with template-based ligand conformers selection: performance in pose prediction in the D3R Grand Challenge 4. *J. Comput. Aided. Mol. Des.*, 34(2):149–162, 2020.

[232] B. Wang and H.-L. Ng. Deep neural network affinity model for BACE inhibitors in D3R Grand Challenge 4. *J. Comput. Aided. Mol. Des.*, 34(2):201–217, 2020.

[233] O. Yakovenko and S. J. Jones. Modern drug design: the implication of using artificial neuronal networks and multiple molecular dynamic simulations. *J. Comput. Aided. Mol. Des.*, 32(1):299–311, 2018.

[234] J. Sunseri, J. E. King, P. G. Francoeur, and D. R. Koes. Convolutional neural network scoring and minimization in the D3R 2017 community challenge. *J. Comput. Aided. Mol. Des.*, 33(1):19–34, 2019.

[235] D. D. Nguyen, K. Gao, M. Wang, and G.-W. Wei. MathDL: mathematical deep learning for D3R Grand Challenge 4. *J. Comput. Aided. Mol. Des.*, 34(2):131–147, 2020.

[236] E. Neveu, D. W. Ritchie, P. Popov, and S. Grudinin. PEPSI-Dock: a detailed data-driven protein-protein interaction potential accelerated by polar Fourier correlation. *Bioinformatics*, 32(17):i693–i701, Sep 2016.

[237] S. Grudinin, P. Popov, E. Neveu, and G. Cheremovskiy. Predicting Binding Poses and Affinities in the CSAR 2013-2014 Docking Exercises Using the Knowledge-Based Convex-PL Potential. *J. Chem. Inf. Model.*, 56(6):1053–1062, Nov 2016.

[238] V. Vapnik. *The nature of statistical learning theory*. Springer, 2000.

[239] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, 2004.

[240] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, volume 2, pages 1137–1145, 1995.

[241] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ Press, 2004.

[242] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods*. MIT press Cambridge, MA, 1998.

[243] M. Kadukova and S. Grudinin. Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates. *J. Chem. Inf. Model.*, 56(8):1410–1419, 2016.

[244] R. Wang, X. Fang, Y. Lu, et al. The PDBbind Database: Methodologies And Updates. *J. Med. Chem.*, 48(12):4111–9, June 2005.

[245] M.-y. Shen and A. Sali. Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, 15(11):2507–2524, 2006.

[246] H. Zhou and Y. Zhou. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.*, 11(11):2714–2726, 2002.

[247] R. Samudrala and J. Moult. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J. Mol. Biol.*, 275(5):895–916, 1998.

[248] C. Camacho, G. Coulouris, V. Avagyan, et al. Blast+: architecture and applications. *BMC Bioinf.*, 10(1):421, 2009.

[249] G. M. Boratyn, A. A. Schäffer, R. Agarwala, et al. Domain enhanced lookup time accelerated blast. *Biology direct*, 7(1):12, 2012.

[250] Convex-PL^R – Revisiting affinity predictions and virtual screening using physics-informed machine learning. Submitted.

[251] F. Giordanetto, S. Cotesta, C. Catana, et al. Novel scoring functions comprising QXP, SASA, and protein side-chain entropy terms. *J. Chem. Inf. Comput. Sci.*, 44(3):882–893, 2004.

[252] M. J. Sternberg and J. S. Chickos. Protien side-chain conformational entropy derived from fusion data-comparison with other empirical scales. *Protein Eng. Des. Sel.*, 7(2):149–155, 1994.

[253] S. D. Pickett and M. J. Sternberg. Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.*, 231(3):825–839, 1993.

[254] K. V. Klenin, F. Tristram, T. Strunk, and W. Wenzel. Derivatives of molecular surface area and volume: Simple and exact analytical formulas. *J. Comput. Chem.*, 32(12):2647–2653, 2011.

[255] K. Klenin, F. Tristram, T. Strunk, and W. Wenzel. Achieving numerical stability in analytical computation of the molecular surface and volume. *From Computational Biophysics to Systems Biology (CBSB11)–Celebrating Harold Scheraga's 90th Birthday*, 8:75, 2012.

[256] M. Karasikov, G. Pagès, and S. Grudinin. Smooth orientation-dependent scoring function for coarse-grained protein quality assessment. *Bioinformatics*, 35(16):2801–2808, 2019.

[257] A. A. Nirschl, Y. Zou, S. R. Krystek Jr, et al. N-aryl-oxazolidin-2-imine muscle selective androgen receptor modulators enhance potency through pharmacophore reorientation. *J. Med. Chem.*, 52(9):2794–2798, 2009.

[258] C.-I. Liu, G. Y. Liu, Y. Song, et al. A cholesterol biosynthesis inhibitor blocks staphylococcus aureus virulence. *Science*, 319(5868):1391–1394, 2008.

[259] G. Trani, J. J. Barker, S. M. Bromidge, et al. Design, synthesis and structure–activity relationships of a novel class of sulfonylpyridine inhibitors of interleukin-2 inducible t-cell kinase (itk). *Bioorg. Med. Chem. Lett*, 24(24):5818–5823, 2014.

[260] S. Han, R. M. Czerwinski, N. L. Caspers, et al. Selectively targeting an inactive conformation of interleukin-2-inducible t-cell kinase by allosteric inhibitors. *Biochem. J*, 460(2):211–222, 2014.

[261] J. Zhang and Y. Zhang. A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. *PloS One*, 5(10):e15386, 2010.

[262] W. Elhefnawy, L. Chen, Y. Han, and Y. Li. ICOSA: A distance-dependent, orientation-specific coarse-grained contact potential for protein structure modeling. *J. Mol. Biol.*, 427(15):2562–2576, 2015.

[263] J. R. Lopez-Blanco and P. Chacon. KORP: knowledge-based 6D potential for fast protein and loop modeling. *Bioinformatics*, 2019.

[264] A. Liwo, P. Arłukowicz, C. Czaplewski, et al. A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape: application to the UNRES force field. *Proc. Natl. Acad. Sci. U.S.A.*, 99(4):1937–1942, 2002.

[265] A. W. Senior, R. Evans, J. Jumper, et al. Protein structure prediction using multiple deep neural networks in the 13th critical assessment of protein structure prediction (CASP13). *Proteins: Struct. Func. Bioinfo.*, 87(12):1141–1148, 2019.

[266] W. Zheng, Y. Li, C. Zhang, et al. Deep-learning contact-map guided protein structure prediction in CASP13. *Proteins: Struct. Func. Bioinfo.*, 87(12):1149–1164, 2019.

[267] A. Kryshtafovych, T. Schwede, M. Topf, et al. Critical assessment of methods of protein structure prediction (CASP)– Round XIII. *Proteins: Struct. Func. Bioinfo.*, 87(12):1011–1020, 2019.

[268] R. Wang, X. Fang, Y. Lu, et al. The PDBbind database: methodologies and updates. *J. Med. Chem*, 48(12):4111–4119, 2005.

[269] B. Beckers and P. Beckers. A general rule for disk and hemisphere partition into equal-area cells. *Computational Geometry*, 45(7):275–283, 2012.

[270] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw (TOMS)*, 23(4):550–560, 1997.

[271] P. Virtanen, R. Gommers, T. E. Oliphant, et al. SciPy 1.0–fundamental algorithms for scientific computing in Python. *Nat. Methods*, 17(3):261–272, 2020.

[272] H. Wätzig, I. Oltmann-Norden, F. Steinicke, et al. Data quality in drug discovery: the role of analytical performance in ligand binding assays. *J. Comput.-Aided Mol. Des.*, 29(9):847–865, 2015.

[273] Y. Zhang and J. Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Struct. Func. Bioinfo.*, 57(4):702–710, 2004.

[274] D. W. Ritchie, A. W. Ghoorah, L. Mavridis, and V. Venkatraman. Fast protein structure alignment using Gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics*, 28(24):3274–3281, 2012.

[275] S. Salentin, S. Schreiber, V. J. Haupt, et al. PLIP: fully automated protein–ligand interaction profiler. *Nucleic acids research*, 43(W1):W443–W447, 2015.

[276] S. Grudinin, P. Popov, E. Neveu, and G. Cheremovskiy. Predicting binding poses and affinities in the csar 2013–2014 docking exercises using the knowledge-based convex-pl potential. *J. Chem. Inf. Model.*, 56(6):1053–1062, 2015.

[277] C. Athanasiou, S. Vasilakaki, D. Dellis, and Z. Cournia. Using physics-based pose predictions and free energy perturbation calculations to predict binding poses and relative binding affinities for FXR ligands in the D3R Grand Challenge 2. *J. Comput. Aided. Mol. Des.*, 32(1):21–44, 2018.

[278] N. M. OLBoyle, M. Banck, C. A. James, et al. Open Babel: an Open Chemical Toolbox. *J. Cheminform.*, 3:33, 2011.

[279] G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack. Improved prediction of protein side-chain conformations with scwrl4. *Proteins: Struct., Funct., Bioinf.*, 77(4):778–795, 2009.

[280] T. A. Halgren. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of Mmff94. *J. Comput. Chem.*, 17(5-6):490–519, 1996.

[281] A. K. Rappé, C. J. Casewit, K. Colwell, et al. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society*, 114(25):10024–10035, 1992.

[282] Sigma-Aldrich Amino Acids Reference Chart. `http://www.sigmaaldrich.com/life-science/metabolomics/learning-center/amino-acid-reference-chart.html`.

[283] H. G. Richter, G. M. Benson, K. H. Bleicher, et al. Optimization of a novel class of benzimidazole-based farnesoid X receptor (FXR) agonists to improve physicochemical and ADME properties. *Bioorganic Med. Chem. Lett.*, 21(4):1134–1140, 2011.

[284] N. Eswar, B. Webb, M. A. Marti-Renom, et al. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinformatics*, 15(1):5–6, 2006.

[285] Y. Zhang. I-tasser server for protein 3d structure prediction. *BMC Bioinf.*, 9(1):40, 2008.

[286] M. T. Buenavista, D. B. Roche, and L. J. McGuffin. Improvement of 3d protein models using multiple templates guided by single-template model quality assessment. *Bioinformatics*, 28(14):1851–1857, 2012.

[287] A. Hoffmann and S. Grudinin. Nolb: Nonlinear rigid block normal-mode analysis method. *J. Chem. Theory Comput.*, 13(5):2123–2134, 2017.

[288] E. Laine and A. Carbone. Local geometry and evolutionary conservation of protein surfaces reveal the multiple recognition patches in protein-protein interactions. *PLoS computational biology*, 11(12), 2015.

[289] R. J. Woods. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem. Rev.*, 118(17):8005–8024, 2018.

[290] I. Y. Chernyshov and P. V. Toukach. REStLESS: automated translation of glycan sequences from residue-based notation to SMILES and atomic coordinates. *Bioinformatics*, 34(15):2679–2681, 2018.

[291] P. V. Toukach and K. S. Egorova. New Features of Carbohydrate Structure Database Notation (CSDB Linear), As Compared to Other Carbohydrate Notations. *J. Chem. Inf. Model.*, 60(3):1276–1289, 2019.

[292] A. K. Nivedha, S. Makeneni, L. Foley, et al. Importance of Ligand Conformational Energies in Carbohydrate Docking : Sorting the Wheat from the Chaff. *J. Comput. Chem.*, 1(706), 2013.

[293] A. K. Nivedha, D. F. Thieker, S. Makeneni, et al. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.*, 12(2):892–901, 2016.

[294] S. S. Roy Burman, M. L. Nance, J. R. Jeliazkov, et al. Novel sampling strategies and a coarse-grained score function for docking homomers, flexible heteromers, and oligosaccharides using Rosetta in CAPRI rounds 37–45. *Proteins Struct. Funct. Bioinforma.*, 88(8):973–985, 2020.

[295] M. K. Ameriks, S. D. Bembenek, M. T. Burdett, et al. Diazinones as P2 replacements for pyrazole-based cathepsin S inhibitors. *Bioorg. Med. Chem. Lett*, 20(14):4060–4064, 2010.

[296] M. Jusot, D. Stratmann, M. Vaisset, et al. Exhaustive exploration of the conformational landscape of small cyclic peptides using a robotics approach. *J. Chem. Inf. Model*, 58(11):2355–2368, 2018.

[297] D. Sindhikara, S. A. Spronk, T. Day, et al. Improving accuracy, diversity, and speed with prime macrocycle conformational sampling. *J. Chem. Inf. Model*, 57(8):1881–1894, 2017.

[298] R. Vassar, D. M. Kovacs, R. Yan, and P. C. Wong. The β-secretase enzyme bace in health and alzheimer's disease: regulation, cell biology, function, and therapeutic potential. *Journal of Neuroscience*, 29(41):12787–12794, 2009.

[299] F. Prati, G. Bottegoni, M. L. Bolognesi, and A. Cavalli. BACE-1 Inhibitors: From Recent Single-Target Molecules to Multitarget Compounds for Alzheimer's Disease: Miniperspective. *J. Med. Chem.*, 61(3):619–637, 2017.

[300] S. Hanessian, G. Yang, J.-M. Rondeau, et al. Structure-based design and synthesis of macroheterocyclic peptidomimetic inhibitors of the aspartic protease β-site amyloid precursor protein cleaving enzyme (bace). *J. Med. Chem.*, 49(15):4544–4567, 2006.

[301] J. B. Jordan, D. A. Whittington, M. D. Bartberger, et al. Fragment-linking approach using 19f nmr spectroscopy to obtain highly potent and selective inhibitors of β-secretase. *J. Med. Chem.*, 59(8):3732–3749, 2016.

[302] S. Butini, S. Brogi, E. Novellino, et al. The structural evolution of β-secretase inhibitors: a focus on the development of small-molecule inhibitors. *Curr. Top. Med. Chem*, 13(15):1787–1807, 2013.

[303] P. Tosco, N. Stiefl, and G. Landrum. Bringing the mmff force field to the rdkit: implementation and validation. *J. Cheminformatics*, 6(1):37, 2014.

[304] E. A. Coutsias, K. W. Lexa, M. J. Wester, et al. Exhaustive conformational sampling of complex fused ring macrocycles using inverse kinematics. *J. Chem. Theory Comput.*, 12(9):4674–4687, 2016.

Drug discovery is a very expensive process consisting of multiple phases. Computer simulations provide an opportunity to scale and speed up its early stages by performing the initial screening of drug candidates and modeling their interactions with the target receptors. Such modeling is generally performed with molecular docking methods. 3D models of the drug candidates are superposed with 3D receptor models by specific algorithms that are able to estimate the binding free energy change and detect the best configuration of the molecular complex. The binding free energy can be approximately evaluated with the so-called scoring functions. This thesis presents the development and assessment of several protein-ligand scoring functions with the associated docking and screening protocols. Research carried out in the scope of this thesis resulted in the development of three novel scoring functions, namely Convex-PL, Convex-PL-R, and KORP-PL. Convex-PL is a knowledge-based pairwise distance-dependent scoring function for protein-ligand interactions, which is deduced by solving a quadratic optimization problem. Our motivation for its development was to prove that a knowledge-based scoring function can be derived by solving a classification convex optimization problem and also to demonstrate that the non-native ligand poses for the classification can be obtained with rigid constant-RMSD transformations of the native ones. Convex-PL is validated on several benchmarks and integrated into AutoDock Vina. Although a great variety of scoring functions have been designed throughout the years of the field development, a number of challenges in scoring functions creation remain unsolved. Many of the all-purpose scoring functions demonstrate worse performance in the virtual screening tests, compared to the precision with which they are able to predict co-crystal binding poses. After an analysis of the publicly available results of the virtual screening test of the CASF benchmarks, we have discovered that many of the scoring functions are biased towards favouring bigger protein-ligand interfaces. One of the reasons for such preference is insufficient consideration of the interactions with the solvent and also mistreating the entropic contributions. To address this problem, we have developed Convex-PL-R – a machine learning-based scoring function that incorporates additional solvent and entropic terms. It demonstrates better affinity prediction and virtual screening performance if compared to Convex-PL. Finally, we have developed KORP-PL – the first coarse-grained orientation-dependent knowledge-based scoring function. KORP-PL proves the concept that a scoring function with a coarse-grained representation of the receptor is suitable for protein-ligand interactions prediction. We believe that its sidechain-free nature will pave the way for novel molecular docking algorithms that will be able to overcome the receptor flexibility problem. Despite being a coarse-grained scoring function, KORP-PL is very successful in pose prediction tests and also on the virtual screening benchmarks. From a practical point of view, the thesis author has participated in several docking challenges that resulted in publications containing various docking protocols evaluation. This thesis comprises the description of participation in the D3R Grand Challenges 2, 3, and 4, and in the CAPRI round 41.

La découverte de médicaments est un processus très coûteux composé de plusieurs phases. Les simulations informatiques offrent la possibilité de passer à l'échelle et d'accélérer les premières phases, en effectuant le criblage initial des candidats médicaments et en modélisant leurs interactions avec les récepteurs cibles. Une telle modélisation est généralement réalisée avec des méthodes d'amarrage moléculaire (« docking »). Les modèles 3D des candidats médicaments sont superposés aux modèles de récepteurs 3D par des algorithmes spécifiques capables d'estimer le changement d'énergie libre de liaison et de détecter la meilleure configuration du complexe moléculaire. L'énergie libre de liaison peut être évaluée approximativement avec des fonctions « de score ». Cette thèse présente le développement et l'évaluation de plusieurs fonctions de score protéine-ligand, avec les protocoles d'amarrage et de criblage associés.

Les recherches menées dans le cadre de cette thèse ont abouti au développement de trois nouvelles fonctions de score, à savoir Convex-PL, Convex-PL-R et KORP-PL. Convex-PL est une fonction de score dédiée aux interactions protéine-ligand, qui dépend des distances entre paires d'atomes et se base sur la connaissance (« knowledge-based »). Nous l'avons dérivée de la résolution d'un problème d'optimisation quadratique. Notre motivation était de prouver qu'une fonction de score basée sur la connaissance peut être dérivée en résolvant un problème d'optimisation convexe de classification, et également de démontrer que les « poses » non natives du ligand utiles pour la classification peuvent être obtenues avec des transformations rigides à déviation (RMSD) constante de la pose native. Convex-PL a été validée sur plusieurs benchmarks et intégrée à AutoDock Vina.

Bien qu'une grande variété de fonctions de score ait été conçue au cours des années de développement du domaine, un certain nombre de défis dans la création de ces fonctions restent à relever. De nombreuses fonctions de score polyvalentes démontrent des performances moins bonnes dans les tests de criblage virtuels, relativement t à la précision avec laquelle elles sont capables de prédire les poses de liaison co-cristallisées. Après une analyse des résultats du test de criblage virtuel des benchmarks CASF, nous avons découvert que de nombreuses fonctions de score favorisent les interfaces les plus grandes. L'une des raisons de cette préférence est une prise en compte insuffisante des interactions avec le solvant et également une mauvaise utilisation des contributions entropiques. Pour résoudre ce problème, nous avons développé Convex-PL-R - une fonction de score basée sur l'apprentissage automatique (« machine learning ») qui incorpore des termes de solvatation et entropiques supplémentaires. Elle fournit de meilleures prédictions d'affinité et de meilleures performances de criblage virtuel par rapport à Convex-PL.

Enfin, nous avons développé KORP-PL - la première fonction de score guidée par les données et dépendante de l'orientation des particules en résolution gros-grain. KORP-PL prouve le concept selon lequel une fonction de score basée sur une représentation gros-grain du récepteur convient pour la prédiction des interactions protéine-ligand. Nous pensons que sa nature sans chaînes latérales ouvrira la voie à de nouveaux algorithmes d'amarrage moléculaire capables de surmonter le problème de flexibilité des récepteurs. Bien qu'il s'agisse d'une fonction de score à gros grains, KORP-PL est très efficace dans les tests de prédiction de pose et également dans les criblages virtuels.

D'un point de vue pratique, l'auteur de la thèse a participé à plusieurs défis d'amarrage qui ont abouti à des publications présentant l'évaluation de divers protocoles d'amarrage. Cette thèse comprend la description de la participation aux Grands Défis D3R 2, 3 et 4 et au round 41 de CAPRI.