Artificial Intelligence in Medicine 63 (2015) 135–152

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Survey Paper

Machine learning in computational docking

Mohamed A. Khamis[a,*], Walid Gomaa[a,b], Walaa F. Ahmed[a]

[a] *Cyber-Physical Systems Lab, Egypt-Japan University of Science and Technology (E-JUST), P.O. Box 179, New Borg El-Arab City, 21934 Alexandria, Egypt*
[b] *Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt*

ARTICLE INFO

Article history:
Received 27 July 2014
Received in revised form 8 January 2015
Accepted 9 February 2015

Keywords:
Machine learning
Random forest
Support vector machine
Drug discovery
Computational docking
Scoring function
Virtual screening
Complex binding affinity
Ligands ranking accuracy
Force field interaction
Pharmacophore fingerprint

ABSTRACT

Objective: The objective of this paper is to highlight the state-of-the-art machine learning (ML) techniques in computational docking. The use of smart computational methods in the life cycle of drug design is relatively a recent development that has gained much popularity and interest over the last few years. Central to this methodology is the notion of computational docking which is the process of predicting the best pose (orientation + conformation) of a small molecule (drug candidate) when bound to a target larger receptor molecule (protein) in order to form a stable complex molecule. In computational docking, a large number of binding poses are evaluated and ranked using a scoring function. The scoring function is a mathematical predictive model that produces a score that represents the binding free energy, and hence the stability, of the resulting complex molecule. Generally, such a function should produce a set of plausible ligands ranked according to their binding stability along with their binding poses. In more practical terms, an effective scoring function should produce promising drug candidates which can then be synthesized and physically screened using high throughput screening process. Therefore, the key to computer-aided drug design is the design of an efficient highly accurate scoring function (using ML techniques).
Methods: The methods presented in this paper are specifically based on ML techniques. Despite many traditional techniques have been proposed, the performance was generally poor. Only in the last few years started the application of the ML technology in the design of scoring functions; and the results have been very promising.
Material: The ML-based techniques are based on various molecular features extracted from the abundance of protein–ligand information in the public molecular databases, e.g., protein data bank bind (PDBbind).
Results: In this paper, we present this paradigm shift elaborating on the main constituent elements of the ML approach to molecular docking along with the state-of-the-art research in this area. For instance, the best random forest (RF)-based scoring function [35] on PDBbind v2007 achieves a Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.803 while the best conventional scoring function achieves 0.644 [34]. The best RF-based ranking power [6] ranks the ligands correctly based on their experimentally determined binding affinities with accuracy 62.5% and identifies the top binding ligand with accuracy 78.1%.
Conclusions: We conclude with open questions and potential future research directions that can be pursued in smart computational docking; using molecular features of different nature (geometrical, energy terms, pharmacophore), advanced ML techniques (e.g., deep learning), combining more than one ML models.

1. Introduction

Traditional drug design paradigms – such as random screening [1] and chance discovery [2] are essentially trial and error techniques. Thus they are very *time consuming* (∼10–15 years), Hospital very expensive ($300 M), with extremely low yield (e.g., as mentioned in Ref. [3], over the last 50 years, only 500,000 compounds have been tested for anti-cancer, only 25 are in wide use today). On the other hand, *computer-aided drug design* (CADD) [4] is target specific, structure-based, automatic, fast, and very low cost with high success rate.

The core process of CADD is *computational docking*. Computational docking is the process of predicting the best orientation

* Corresponding author. Tel.: +20 3 309 4075; mobile: +20 100 638 2428.
E-mail addresses: mohamed.khamis@ejust.edu.eg (M.A. Khamis), walid.gomaa@ejust.edu.eg (W. Gomaa), wella_dba@yahoo.com (W.F. Ahmed).

http://dx.doi.org/10.1016/j.artmed.2015.02.002
0933-3657/© 2015 Elsevier B.V. All rights reserved.

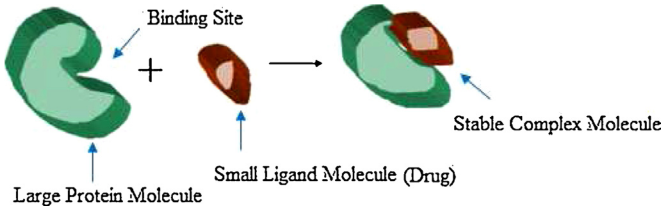


Fig. 1. Docking of a ligand with a target protein: fitting puzzle pieces.

and conformation of a small molecule (drug ligand) when bound to a target large receptor molecule (protein) in order to form a stable complex molecule. This amounts to predicting the *binding free energy* (negative value in kcal/mol unit), and hence the stability of the complex molecule resulting from the docking process. The predicted binding free energy is usually calculated using a molecular mechanics (MM) force field. A predicted *binding affinity* inhibition constant (IC_{50}), K_i, or K_d a positive value in nanomolar (nM) unit is then derived from the predicted binding free energy. This latter value is verified by comparing with the experimentally measured binding affinity. The aim of the docking process is the activation/suppression of the target protein for/from performing some *functionality*.

The DNA includes the *encoding* (program) of the protein formation. Every protein molecule consists of a chain of *amino-acids*. There are about 20 amino-acids (e.g., leucine, alanine, serine, etc.). For instance, a compound of 1 Carbon atom, 1 Nitrogen atom, and 1 Oxygen atom may form a specific *amino-acid*. The *active site* (binding site) of the protein is the *pocket* in which the atoms of the small ligand molecule (key) binds to the nearby amino-acids of the large protein molecule (lock). Fig. 1 illustrates the docking process of a ligand into the active site of a target protein.[1]

In molecular docking, a large number of binding poses are evaluated using a *scoring function* [6]. A *scoring function* is a mathematical predictive model that produces a score that represents the *binding free energy* of a binding pose. The result of the docking process is a set of ligands ranked according to their predicted binding scores. In addition, novel ligands can be designed from scratch (known by *de novo* design) based on the 3D structure of the *binding pocket* of the target protein and then assessed using a *scoring function*. Top-ranked ligands which are the most promising *drug candidates* are then synthesized and physically screened using the high-throughput screening (HTS) process.

Some computational methods, e.g., *virtual screening* are employed to complement HTS by minimizing the number of ligands to be *physically screened*. Virtual screening is a computer-based technique used to identify promising compounds (e.g., ligands) when binding to a target molecule (e.g., protein) of known structure [7]. There exist two types of virtual screening: *protein-based* and *ligand-based* [6]. On one hand, the most popular approach to predict the binding affinity in virtual screening is structure-based (protein-based). In this type, ***physicochemicalinteractions*** between a ligand and a protein are deduced from the **3D structures** of both molecules. On the other hand, in ligand-based approaches, only ligands that are ***biochemicallysimilar*** to the ones known to bind to the target protein are screened. Ligand-based screening does not directly take information about the target protein into account. Thus, it may not identify *novel chemicals* as hits. Therefore, this method is used when the 3D structure of the target protein is not available. Nevertheless, the 3D structures of protein–ligand complexes are now available on a wide range in molecular databases. Thus, interest in the *protein-based approach* has increased. The protein-based approach is more accurate than the ligand-based approach due to the inclusion of shape and volume information that are extracted from the protein's 3D structure during the screening process.

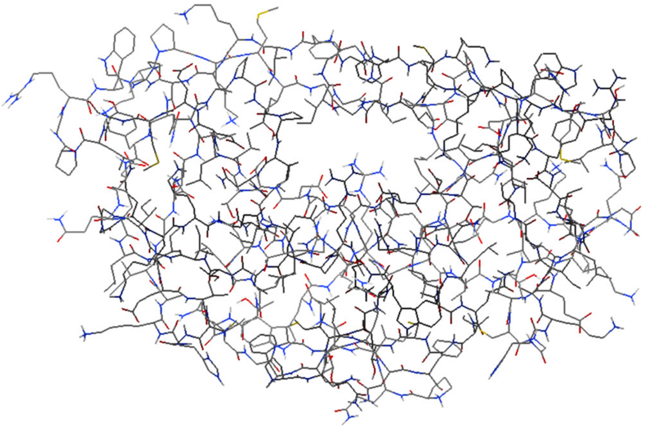


Fig. 2. Protein tertiary structure: HIV-1 protease (hsg1.pdb).

There are two inputs to any *docking algorithm*: (1) the *tertiary structure* (i.e., 3D structure; atom types and coordinates) of the target protein which is determined by biophysical or prediction techniques and (2) a database of potential ligands (small drug molecules).

Many research works have been done recently on target proteins and potential ligands (drugs) bound at the protein active site. For instance, in Ref. [8], the authors study the adaptability of binding residues and flap region of the anti-HIV drug (TMC-114) resistance in the human immunodeficiency virus-1 (HIV-1) protease mutants. In Ref. [9], the authors present structural basis for the resilience of Darunavir (TMC-114) resistance major flap mutations of HIV-1 protease. In Ref. [10], the author presents the role of the extended A-loop region in auto-activation of the mutant prime target in gastrointestinal stromal tumor therapy from a molecular dynamics (MD) simulation insight. In Ref. [11], the authors present an in silico investigation of molecular mechanism of laminopathy caused by a point mutation (R482W) in lamin A/C protein. In Ref. [12], the authors present an in silico analysis of the relationship between mutation of serine residue at 315th position in Mycobacterium tuberculosis (M.tb) catalase-peroxidase enzyme and Isoniazid susceptibility. In Ref. [13], the authors present the use of long term MD simulation in predicting cancer associated single nucleotide polymorphisms. In Ref. [14], the authors present a drug resistance mechanism of the bacterial pyrazinamidase enzyme in M.tb.

Fig. 2 shows an example of a target protein which is the HIV-1 protease. The figure is created using the auto dock tools (ADT) [15], where the corresponding data file hsg1.pdb is retrieved from the protein data bank (PDB) database [16]. Fig. 3 shows the clinically

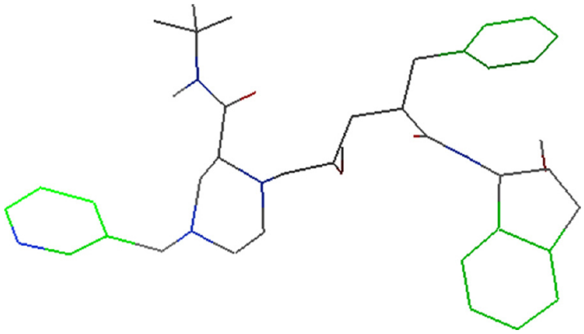


Fig. 3. Ligand structure: Indinavir (ind.pdb).

[1] The original version of Fig. 1 is available at [5].

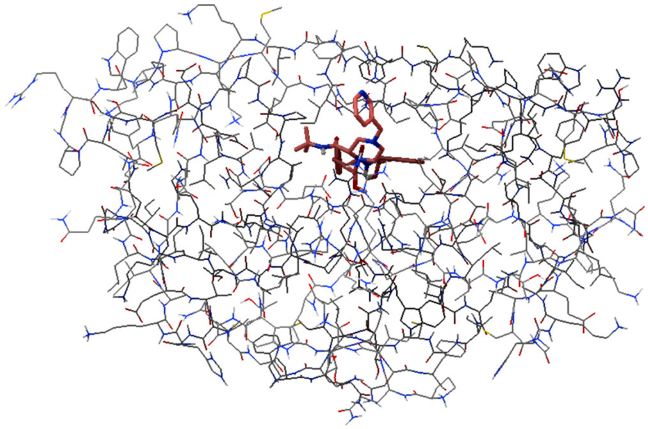


Fig. 4. Complex molecule chemical and physical characteristics, e.g., specific pose of the docked ligand Indinavir when fit into the binding pocket of the receptor protein HIV-1 protease. The ligand is displayed by the *sticks and balls* view (for bonds and atoms respectively) in order to be easily visible inside the pocket of the protein.

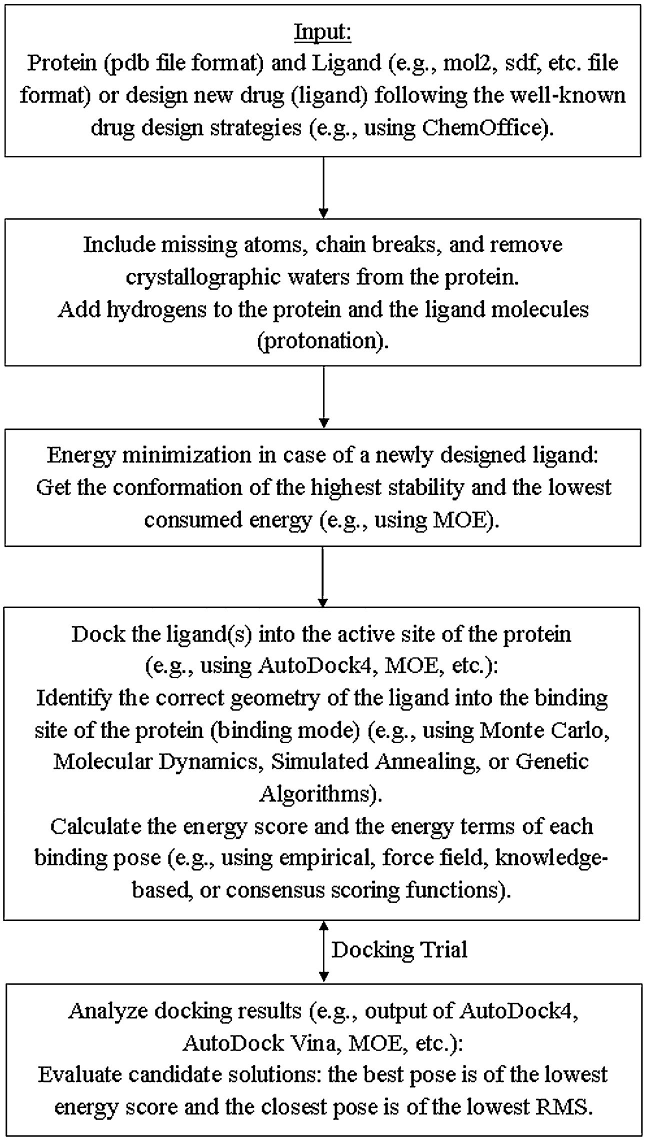
approved HIV-1 protease inhibitor Indinavir. The figure is created using the ADT [15], where the corresponding data file ind.pdb is retrieved from the PDB database [16].

Two outputs are generated by the *docking algorithm*: (1) a description of a *novel* ligand (the drug molecule) (Fig. 3) and (2) a description of (near) optimal stable complex (i.e., with optimal binding mode). Such description of an (near) optimal mode consists of two components: (i) the *relative orientation* of the ligand with respect to the receptor protein; this is essentially a geometric description with respect to some frame of reference and (ii) the *conformation* of the ligand and receptor protein when bound together; this is essentially a description of conformational changes, that are, changes in the shapes of the underlying molecules in order to fit/bind together. The chemical and physical characteristics of the resulting complex molecule is presented in Fig. 4.

There are three elements of *computational docking*. The first is the *representation* of the chemical molecules, i.e., protein and ligands. The second main element is searching the *orientation and conformation space* for optimal poses. The third and final element is *ranking* the potential solutions (according to the binding free energy). Fig. 5 depicts the major steps of any *docking algorithm*.

In this paper, we present the state-of-the-art research in CADD, particularly the application of the machine learning (ML) technology in computational docking. Nevertheless, the work surveyed in this paper spans many research areas that cover the following disciplines: (1) molecular biology where we clarify concepts such as binding affinity which characterizes the stability of the candidate drug, (2) MD where a great deal of understanding of the docking process and the stability of the resulting complex molecule depend on the dynamics and the different energy terms among the docked molecules, e.g., Van der Waals (VdW), electrostatic, etc., (3) pharmacology through the study of the pharmacophore features of the underlying molecules as well as the design of new ligands (novel drugs), (4) quantum physics where important issues in the design of new drugs have to do with quantum effects, and (5) high performance computing (HPC) where efficient parallel techniques need to be designed and applied in order to cope with the huge computational power involved in the docking process.

The remainder of this paper is organized as follows. In Section 2, we give the necessary chemical, physical, and molecular biology background. Section 3 presents the notion of a scoring function and its typical functionalities. These include scoring power, ranking power, docking power and screening power. We specifically focus on the use of ML techniques in the design of efficient highly accurate scoring function and give a survey of the state-of-the-art research in that direction. In Section 4, we present the currently popular molecular software along with their ability to perform computational docking. Section 5 presents a literature review of the popular molecular databases and the information provided by each. In Section 6, we explain the different kinds of molecular features that can be extracted from molecular databases and by using molecular software. These include geometric features, physical force field energy terms, and pharmacophore features. The use of combinations of these features can affect the prediction ability of the ML techniques used in computational docking. Finally, Section 7 concludes the paper and proposes key points for research extensions.



Input:
Protein (pdb file format) and Ligand (e.g., mol2, sdf, etc. file format) or design new drug (ligand) following the well-known drug design strategies (e.g., using ChemOffice).

Include missing atoms, chain breaks, and remove crystallographic waters from the protein.
Add hydrogens to the protein and the ligand molecules (protonation).

Energy minimization in case of a newly designed ligand:
Get the conformation of the highest stability and the lowest consumed energy (e.g., using MOE).

Dock the ligand(s) into the active site of the protein (e.g., using AutoDock4, MOE, etc.):
Identify the correct geometry of the ligand into the binding site of the protein (binding mode) (e.g., using Monte Carlo, Molecular Dynamics, Simulated Annealing, or Genetic Algorithms).
Calculate the energy score and the energy terms of each binding pose (e.g., using empirical, force field, knowledge-based, or consensus scoring functions).

Analyze docking results (e.g., output of AutoDock4, AutoDock Vina, MOE, etc.):
Evaluate candidate solutions: the best pose is of the lowest energy score and the closest pose is of the lowest RMS.

Fig. 5. The major steps of computational docking.

2. Background: chemical, physical and biological

2.1. Introduction

In this section, we introduce the chemical, physical, and biological background that is necessary to understand the docking process and the consequent automation of drug design. As mentioned above, the first element in computational docking is the representation of the underlying molecules. Generally, the biological molecule can be represented in a variety of ways; in other words using different types of features based on a *feature selection process*. *Geometric-based features* can be used to represent a molecule, for example, the coordinates and orientations of different types of atoms with respect to some fixed frame of reference. Another example of feature could be fixing a Carbon atom, then counting the number of Nitrogen atoms positioned in the ball centered at the Carbon atom with particular radius. Such features are essentially static and are used for geometric-matching. Their determination is fast and robust. Nevertheless, they cannot capture the *dynamical changes* in the protein–ligand pair. Thus, in contrast to the geometric features, *atomic representation* of the exposed residues (amino-acids) can be adopted. For instance, in a specific binding pose the number of Hydrogen bonds between the atoms of the ligand and the nearby amino-acids of the protein indicates a better binding mode. Such representation can be acquired by MD simulation, e.g., using molecular operating environment (MOE) [17] and the ranking of the bound protein–ligand complex is based on the potential energy functions. This representation captures the *dynamical features* of the *physico-chemical* complex characteristics. However, such a representation needs huge computational resources. Consequently, HPC is currently being actively applied in the MD simulation. For instance, Peng et al. [18] accelerate MD simulation on a many-core computing platform. In addition, Richardson [19] implements parallel algorithms for solving the time-dependent Schrödinger equation, Eq. (1), on the CM-2 supercomputer.

2.2. Quantum chemistry

Quantum chemistry methods provide a *rigorous* description of molecular systems [20]. Such methods are used to solve the quantum Schrödinger equation, Eq. (1); and are generally applicable, for example, in energetics, kinetics, magnetic spectra, vibrational spectra, optical spectra, chemical insights, etc.

$$i\hbar\frac{\delta}{\delta t}\psi(x,t)=\hat{\mathrm{H}}\psi(x,t), \tag{1}$$

where $\hat{\mathrm{H}}$ is the self-adjoint Hamiltonian operator for the system under study and $\psi(x, t)$ is the system's wave function.

As mentioned in Ref. [21], almost all existing docking scoring functions, e.g. those based on physical force fields, involve the fitting of data from experiments and calculations based on quantum mechanics. For instance, the energy terms of the force field scoring function involve some interaction parameters whose values should be theoretically available from quantum mechanical calculations [22]. However, *quantum mechanics* is very time consuming with respect to computational simulation and calculations. Thus, in practice these parameters are derived empirically by adjusting their values to properly predict the geometry of well known compounds. Then, these parameters are used to calculate the structures of new compounds [22]. The accuracy of these parameters is crucial to MM calculations.

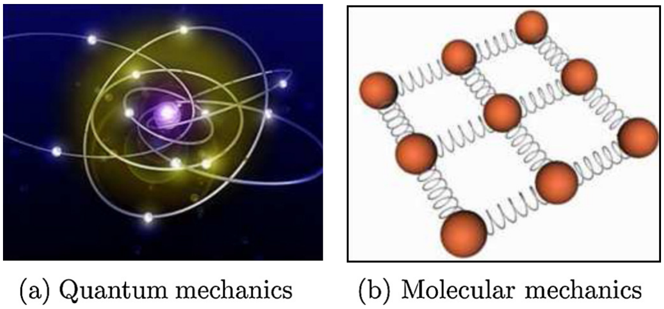

(a) Quantum mechanics (b) Molecular mechanics

Fig. 6. Quantum mechanics versus MM [20].

2.3. Molecular mechanics

In order to model large molecular systems, it is demanding to reduce the complexity of the system. MM is a *non-quantum* classical mechanical technique for handling *large* molecular systems. Thus, MM methods are thousands times faster than quantum chemistry methods [20]. Quantum mechanics considers the atoms as collections of electrons and nuclei, whereas MM considers the atoms as soft or hard spheres, where covalent chemical bonds are treated as springs [20], see Fig. 6. In addition, MM methods use classical potential energy equations.

The potential energy equations used to calculate the energies in MM methods and the parameters/constants used in the equations are known as a *Force Field* [20]. There are many force fields designed for different purposes [20], e.g., assisted model building with energy refinement (AMBER) [23], chemistry at Harvard molecular mechanics (CHARMM) [24], etc. In any such model the binding free energy is calculated using an analytical equation; its form differs according to the docking software program that implements the model: AMBER [23], AutoDock4 [25], MOE [17], etc. For instance, as mentioned in Ref. [22], the total steric energy (kcal/mol) of a molecule can be expressed as a sum of the energies of the bonded and non-bonded interactions:

$$\begin{aligned}
&E_{\text{steric-energy}} = E_{\text{bonded-covalent}} + E_{\text{non-bonded}},\\
&E_{\text{bonded-covalent}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{str\_bend}} + E_{\text{oop}} + E_{\text{tor}},\\
&E_{\text{non-bonded}} = E_{VdW} + E_{\text{qq}}.
\end{aligned} \tag{2}$$

As the protein–ligand interaction is a kind of organic interaction the main kinds of bonds are either covalent or weaker (rather than ionic). Eq. (2) depicts the different components of both the covalent bond and the weaker more physical bond between any pair of atoms involved in the protein–ligand complex obtained after the docking process. These components can vary according to the

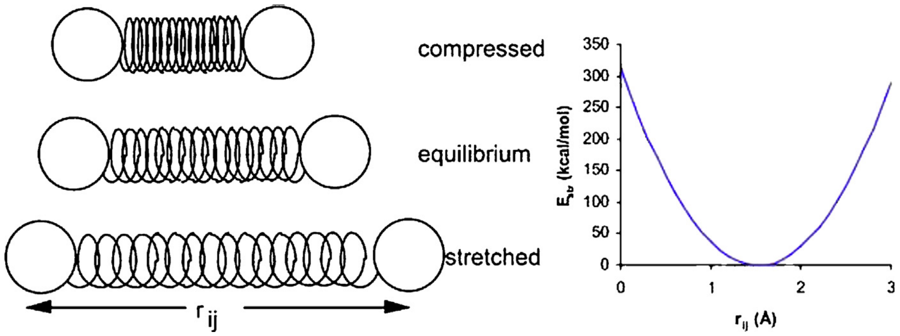


Fig. 7. Modeling the molecular system by a spring – stretching/compressing [22].

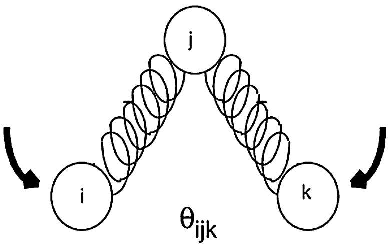


Fig. 8. Modeling the molecular system by a spring – bending [22].

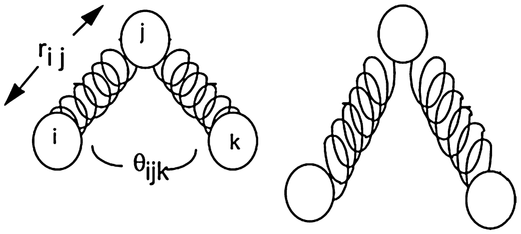


Fig. 9. Modeling the molecular system by a spring – stretching-bending [22].

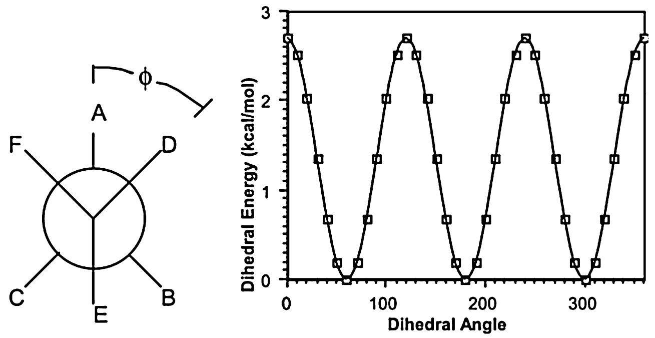


Fig. 11. Modeling the molecular system by a spring – torsion [22].

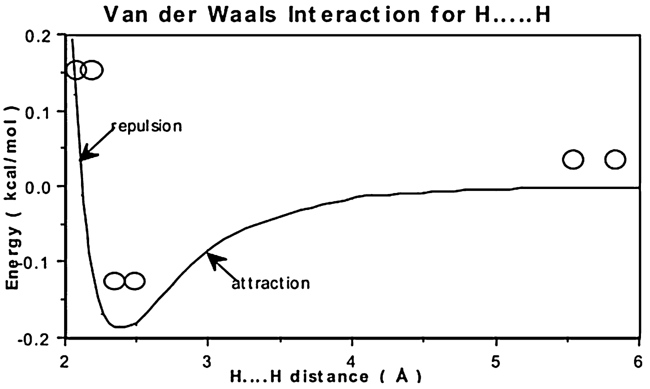


Fig. 12. VdW interactions between two hydrogen atoms in a molecule [22].

underlying MM model and/or the docking software that computes the free energy. This variety of models of the underlying MM comes from the fact that there is no single best set of force field parameters due to the diversity in the types of compounds [22]. For instance, as mentioned in Ref. [22], the MM2 force field works best on hydrocarbons since most of the known compounds used to derive the force field parameters were hydrocarbons. In the following, we will explain briefly the different energy components. E_{str} represents the energy required to stretch or compress a covalent bond between two atoms [22] (for example, under stress force); see Fig. 7. E_{bend} is the energy required to bend a bond from its equilibrium angle θ_0 [22]; see Fig. 8. $E_{\text{str\_bend}}$ is the stretch-bend interaction energy that takes into account the observation that when a bond is bent, the two associated bond lengths increase; see Fig. 9. E_{oop} is the out of planning bending energy required to deform a planar group of atoms (a group of atoms arranged in a plane) from its equilibrium angle ω_0; see Fig. 10. E_{tor} is the torsion energy needed to rotate about bonds [22] (can be a kind of squeezing force); see Fig. 11. It can be noticed that all such energy terms are concerned with the basic structure of the underlying molecule.

E_{VdW} is the steric exclusion and long-range attraction VdW energy (quantum mechanics origins) [26]; it represents the electro-dynamic forces: physical interaction derived from chemical features.

As mentioned in Ref. [22], the VdW interactions are extremely important in determining the 3D structure of many biomolecules especially proteins. Fig. 12 shows the VdW energy as a function of the distance between two hydrogen atoms [22]. When two atoms are far from each other attraction arises, and when they are close enough repulsion occurs [22]. Nevertheless, the repulsions (not the attractions) are often the most important in determining the 3D structure of the molecule [22]. Particularly, a measure of the size of an atom is its VdW radius [22]. The lowest point on the curve in Fig. 12 represents the point that gives the lowest most favorable energy of interaction between the two atoms; the distance at such point is the sum of the VdW radii of the two atoms [22].

One of the famous equations that describes this force between two generic atoms i and j (e.g., Carbon and Hydrogen) is the following:

$$E_{VdW,ij} = \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \qquad (3)$$

where A_{ij} and B_{ij} are empirical parameters that control the depth and position (interatomic distance) of the potential energy for a given pair i, j of non-bonded interacting atoms [20], R_{ij} represents

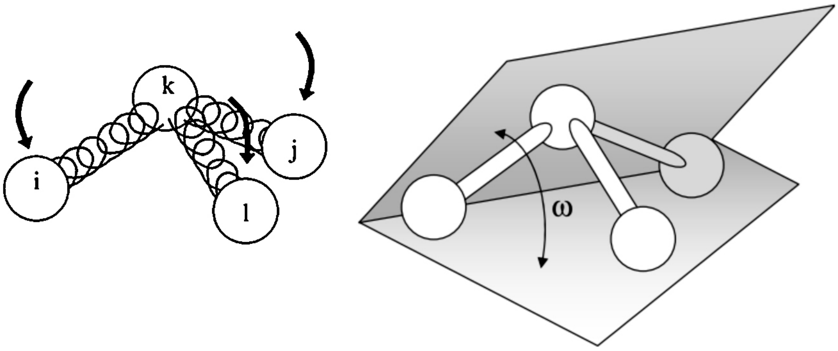


Fig. 10. Modeling the molecular system by a spring – out-of-planning bending [22].

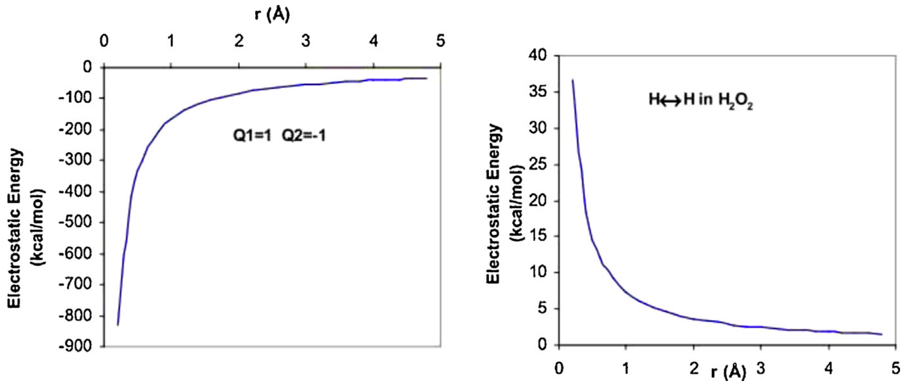


Fig. 13. (a) Coulomb attraction of a positive and a negative charge. (b) Coulomb repulsion of the two hydrogens in H_2O_2 [22].

the distance between the atoms i and j, i.e., interatomic distance. Finally, E_{qq} (in Eq. (2)) is the Coulomb potential function for electrostatic interactions of charges [22]:

$$E_{qq,ij} = \frac{cQ_iQ_j}{4\pi\epsilon r_{ij}} \tag{4}$$

The Q_i and Q_j are the partial atomic charges for atoms i and j separated by a distance r_{ij} and ϵ is the relative dielectric constant [22]. The Coulomb attraction of a positive and a negative charge and a Coulomb repulsion of the two hydrogens in H_2O_2 are shown in Fig. 13.

Note that the VdW and the Coulomb electrostatic energies are the weakest types of interactions among atoms (non-bonded interactions). The bonded and non-bonded energy terms are said to make up a *force field* [22]. However, the force field is not absolute, i.e., not all the energy terms in Eq. (2) may be necessary to predict the steric energy of a molecule. On the other hand, many force fields use additional energy terms [22].

Nevertheless, in general, there are three types of variables in any energy term in a force fields equation, e.g., Eq. (2): (1) empirical parameters, e.g., A_{ij}, B_{ij} determined by curve fitting the 3D structure of well-known compounds or predicted using parametric ML techniques according to the atom types i and j, (2) raw geometrical features, e.g., R_{ij} the interatomic distance between atoms of types i and j. The 3D structure of the underlying molecule(s), e.g., the coordinates of the atoms in the bound molecules can be retrieved from molecular databases, e.g., the PDB database [16] and are measured experimentally using either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, and (3) induced energy terms, e.g., VdW and electrostatic energy terms which are mainly derived from the first two kinds of terms (1) and (2).

2.4. Molecular biology

Molecular biology introduces important concepts such as *binding affinity* which characterizes the stability of the protein–ligand (enzyme-inhibitor) complex molecule after the docking process. In Section 3, we discuss the computational prediction of binding affinities using ML techniques.

The binding affinity can be experimentally represented by either one of the following constants: (1) the inhibition constant of the enzyme-inhibitor complex K_i, (2) the dissociation constant of the enzyme-inhibitor complex K_d, or (3) the concentration at 50% inhibition IC_{50}. The unit of these constants is the molar (mass unit), usually in nM; 1 molar is the quantity of enzyme (protein) that can digest 1 gram of DNA. Any of these constants can be experimentally determined in the lab. Since the binding affinities uniformly span many orders of magnitude, they are typically log-transformed: $pK_i = -\log_{10}K_i$, $pK_d = -\log_{10}K_d$, and $pIC_{50} = -\log_{10}IC_{50}$.

There are relationships among the three constants, e.g., IC_{50} value can be converted to the absolute inhibition constant K_i using the Cheng–Prusoff equation [27]. IC_{50} is not a direct indicator of the binding affinity, however, K_i and IC_{50} can be related at least for competitive agonists and antagonists by the Cheng–Prusoff equation [27]:

$$K_i = \frac{IC_{50}}{1 + (|S|/K_m)} \tag{5}$$

where $|S|$ is the fixed concentration of the substrate,[2] and K_m is the concentration of the substrate at which the enzyme activity is at *half maximal*. The change in the *binding free energy* in kcal/mol calculated *experimentally* (in lab) is given by:

$$\Delta G_{\text{binding}} = -R \times T \ln K_{\text{binding}} = R \times T \ln K_i, \tag{6}$$

where R is the ideal gas constant (1.9871917 cal/K/mol) and T is the room temperature in Kelvin (298.15 K).

Consequently, Eq. (7) is used to derive the estimated binding affinity K_i (in nM) given the estimated binding free energy $\Delta G_{\text{binding}}$ (kcal/mol) which is calculated using an MM force field; this is to be verified with the experimentally measured binding affinity (e.g., available in the protein data bank bind (PDBbind) database [28]).

$$K_i = \exp^{(\Delta G_{\text{binding}} \times 1000/(R \times T))} \tag{7}$$

As mentioned in Ref. [29], the experimentally discovered binding free energies are calculated from IC_{50} using the following relationships:

$$\begin{aligned}\Delta G_{\text{binding}} &= R \times T \ln K_d = R \times T \ln(IC_{50} + 0.5C_{\text{enzyme}}) \\ &\approx R \times T \ln IC_{50}\end{aligned} \tag{8}$$

where C_{enzyme} is the concentration of the enzyme (protein) (it is a very small number after equilibration and can be omitted). The negative sign in Eq. (6) is due to the fact that binding and inhibition occur in opposite directions, so we lose the minus sign: $-\ln K_{\text{binding}} = \ln K_i$.

The 3D structure of the *biomolecular* complexes (e.g., proteins, ligands) are represented in the molecular databases with specific file format, e.g.,.mol2,.pdb,.sdf, etc. These files mainly include data like atom features, position, connectivity, etc. The.PDB format is commonly used for proteins but can be used for other types of molecules as well [30]. One of the most widely used industry standards are *chemical table file* formats, e.g., structure data file (SDF)

[2] The substrate is similar to the ligand such that it can bind to the target protein. There are two types of substrates: competitive and non-competitive. Competitive: the substrate competes with the ligand on the same *active site* of the protein. Non-competitive: the substrate and ligand each binds with a different active site.

format [30]. These latter file types follow strict formats for representing multiple chemical structure records and the associated data fields [30]. For a complete list of chemical file formats, the reader is referred to [30]. From the computer science perspective, and in particular when applying ML techniques, understanding these formats is crucial for parsing them and extracting the necessary information, feature extraction, and for training purposes.

2.5. *Combining MM and molecular biology*

In Ref. [31], the authors predicted the binding free energy of 20 inhibitors of the 2009 H1N1 influenza (PBD ID: 3NSS). The correlation coefficient between the predicted binding free energies (using MM force field, Eq. (9)) versus the experimentally determined binding free energies (using the biologically measured binding affinity IC_{50}, Eq. (8)) of the 20 inhibitors is $R^2 = 0.75$. The authors in Ref. [31] calculated the binding free energy using the solvated interaction energies (SIE) method [32], Eq. (9). Note that Eq. (9) is one form of MM force field other than the total steric energy force field expressed by Eq. (2).

$$\begin{aligned}\Delta G_{\text{binding}}(\rho, D_{\text{in}}, \alpha, \gamma, C) = \alpha \times [&E_c(D_{\text{in}}) \\ &+ \Delta G^R_{\text{bind}}(\rho, D_{\text{in}}) + E_{VdW} \\ &+ \gamma \Delta MSA(\rho)] + C \qquad (9)\end{aligned}$$

where the calculated energy terms are the intermolecular Coulomb E_c and the VdW E_{VdW} interaction energies in the bound state [31]. These values are calculated using the AMBER MM force field (FF99) with an optimized dielectric constant [31]. In addition, the optimal values of the SIE model parameters $\rho, D_{\text{in}}, \alpha, \gamma, C$ were calculated by adjusting the MM force field Eq. (9) to comply with the experimental binding free energy Eq. (8) for a set of known protein–ligand complexes.

The optimized values of these parameters are [31]: $\rho = 1.1$, $D_{\text{in}} = 2.25$, $\alpha = 0.1048$, $\gamma = 0.0129$ kcal/(mol Å^2), and $C = -2.89$ kcal/mol.

3. Scoring function

3.1. *Introduction*

The docking process consists of two essential components: pose generation and scoring of the resulting complex [33]. Pose generation is estimating the proper conformation and orientation of the ligand when bound to the target protein, whereas scoring is predicting how strongly the ligand binds to the target protein. As mentioned in Ref. [33], there are relatively accurate algorithms for pose generation, but imperfections of scoring functions continue to be the major limiting factor for the reliability of computational docking. Hence, the most important step in the docking process is scoring the *conformations of a ligand* in the corresponding *binding site* of the receptor protein by using a *scoring function*. In this section, we will talk exclusively about scoring functions, their design and their essential role in computational docking. The binding affinity prediction using a *scoring function* determines which binding mode is considered the best; a *scoring function* determines which ligand is considered the *most effective drug*.

There are generally three main capabilities a reliable computational scoring function should satisfy [6]: (1) the scoring power: the ability to produce scores for the different binding poses; these scores are supposed to be *linearly correlated* with the experimentally determined binding affinities of the protein–ligand complexes of known 3D structures, (2) the ranking power: the ability to correctly *rank* a given set of ligands with known binding poses when bound to a common protein, and (3) the docking power: the ability to identify the *best binding pose* of a given ligand from a set of computationally generated poses when bound to a specific protein.

These three performance attributes were referred to by Cheng et al. [34] as scoring power, ranking power and docking power. In Ref. [35], the authors present also the *screening power* which is the ability of a scoring function to identify the true binders to a given target protein among a pool of random molecules. Cheng et al. [34] conducted an extensive test of 16 scoring functions which are either used in docking tools or used by some researchers in academia. Using different evaluation criteria and test data sets, they concluded that no single scoring function was better than the others in every aspect. As mentioned in Ref. [6], the best scoring function in predicting the binding constants that are most correlated with the experimentally determined ones was not even in the top five when the goal is to identify the *best binding pose* of the ligand.

3.2. *Scoring power*

3.2.1. *Classical approaches*

Most scoring functions in use today can be categorized as either [6]: (1) force field-based, (2) empirical-based, or (3) knowledge-based. Force field scores are approximate MM interaction energies, consisting of VdW and electrostatic components, e.g., Eq. (10) of the DOCK software [36]. The parameters that define the intermolecular interactions are derived from experimental data and *ab initio* simulations.

$$E_{\text{binding}} = \sum_{i=1}^{\text{ligand}} \sum_{j=1}^{\text{protein}} \left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332 \frac{q_i q_j}{D r_{ij}} \right) \qquad (10)$$

Empirical scoring functions adopt a different trend in calculating the binding free energy of the protein–ligand complex. The whole energy is assumed to be composed of weighted energy terms, e.g., Eq. (11) of the X-Score software [37]. Linear regression methods are used to learn the coefficients of the model. This can be done by fitting the known experimental binding energies to a *training data set*.

$$\begin{aligned}\Delta G_{\text{binding}} &= w_0 + w_1 \Delta G_{VdW} + w_2 \Delta G_{\text{h-bond}} \\ &\quad + w_3 \Delta G_{\text{rotor}} + w_4 \Delta G_{\text{hydrophobic}} \qquad (11)\end{aligned}$$

Finally, a knowledge-based scoring function, e.g., Eq. (12) of the potential of mean force (PMF) score [38], is based on the theory that large databases of protein–ligand complexes can be statistically mined to deduce rules and models that are implicitly embedded in the data.

$$\text{PMF} = \sum_{\text{protein}} \sum_{\text{ligand}} A_{ij}(d_{ij}) A_{ij}(d_{ij}) = -k_B T \ln [f^{j}_{\text{Vol\_corr}}(r) \frac{\rho^{ij}_{\text{seg}}(r)}{\rho^{ij}_{\text{bulk}}}] \qquad (12)$$

For instance, the AutoDock4 software [25] uses semi-*empirical free energy force field* to evaluate the conformations generated during the docking simulation. The force field is parameterized using a large number of protein–ligand complexes for which K_i is known. The weighting constants W of the empirical force field are optimized to calibrate the empirical free energy based on a set of experimentally determined binding constants.

As mentioned in Ref. [21], the physical-based and knowledge-based scoring functions are *weak predictors* for the binding free energy (and consequently for the binding affinity). This means that the *correlation coefficient* between the predicted and the experimentally determined binding affinity is very low.

Traditional docking scoring functions assign a common set of weights to the individual energy terms that contribute to the overall energy score. However, as mentioned in Ref. [21], the weights assigned to the individual energy terms that contribute to the *overall energy score* are in reality protein-family dependent. Thus, in order to estimate a more accurate binding free energy (and consequently estimate the binding affinity), a scoring function must be trained to derive a unique set of weights for each individual protein-family.

Moreover, traditional docking scoring functions improperly assume that the individual energy terms contribute towards the total binding free energy in an additive manner. Therefore, they predict the binding free energy from a *linear combination* of the individual energy terms. However, this is not theoretically sound [21], since it fails to consider the cooperative effects of the non-covalent interactions. Thus, the scoring function have to model the non-linear relationships among the individual energy terms.

3.2.2. ML approaches

In this subsection, we survey the most recent and prominent ML techniques used in the design of scoring functions. ML is a paradigm shift that has proved itself in the context of virtual screening witnessed by the following. First, improving the prediction ability of the binding affinity than traditional scoring functions (i.e., force field, empirical, etc.), e.g., see [21]. The authors used a support vector machine (SVM) to improve the binding affinity prediction of the electronic high throughput screening (eHiTS) molecular docking software [39] (which is an empirical knowledge-based scoring function) This is done by deriving a unique set of weights for each individual protein family – the w_i's in Eq. (11). Similarly, a force field scoring function can be trained to derive a unique set of parameters for each individual protein family, e.g., the A_{ij}'s and B_{ij}'s in Eq. (10). Second, the ML approach predicts the binding affinity based on some features of the protein–ligand complex molecule which are naturally available in the literature (e.g., geometric features, physical force field energy terms, pharmacophore features, etc.). In particular, the goal is to learn the relationship between these features and the corresponding experimentally measured binding affinity given some training set of complex molecules. Afterwards, we can make use of this learned function to predict the binding affinity of new complexes whose features are known but their experimental binding affinity is still unknown.

Recently, there are some non-parametric ML techniques used to model the *functional form* of scoring functions given molecular databases, i.e., data-driven (not knowledge-based), e.g., [40]. Particularly, some ML-based techniques help in learning the *non-linear dependency* of the structure of the protein–ligand complex and accordingly give more accurate prediction of the binding affinity [40].

In these models, each complex structure is represented as a set of features that are relevant in predicting the complex binding affinity. In the work presented in Ref. [40] each feature represents the number of occurrences of a particular protein–ligand atom type pair interacting within a certain distance range. The authors in Ref. [40] use random forest (RF) [41] to learn how the *atomic-level description* of the complex relates to the experimental binding affinity; they merge the K_d and K_i measurements in a single binding constant *K*. This increases the amount of data that can be used to train the ML algorithm. RF is a large ensemble of diverse decision trees. RF introduces the following modifications in the tree training algorithm: (1) a new set of N complexes (*bootstrap sample*) is randomly selected with replacement from the same N training complexes and (2) select the best split (e.g., with lowest variance) at each node of the tree of randomly chosen features instead of using all features. The challenging point is to characterize the protein–ligand complex as a set of *numerical features*. In Ref. [40] the authors consider 9 common elemental atom types for both the protein *P* and the ligand *L*.

$$\begin{aligned}\{P(j)\}_{j=1}^{9} &= \{C, N, O, F, P, S, Cl, Br, I\},\\ \{L(i)\}_{i=1}^{9} &= \{C, N, O, F, P, S, Cl, Br, I\}\end{aligned} \tag{13}$$

The occurrence count of a particular $j-i$ atom type pair is defined as:

$$x_{Z(P(j)),Z(L(i))} \equiv \sum_{k=1}^{K_j}\sum_{l=1}^{L_i}\theta(d_{\text{cutoff}} - d_{kl}) \tag{14}$$

where d_{kl} is the Euclidean distance between the *k*th protein atom of type *j* and the *l*th ligand atom of type *i*. This distance is calculated from the protein–ligand 3D structure available in the PDBbind database [28]. K_j is the total number of protein atoms of type *j* and L_i is the total number of ligand atoms of type *i* in the considered complex. θ is the Heaviside step function that counts the contacts within a d_{cutoff} = 12Å neighborhood of the given ligand atom.

The authors in Ref. [40] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.774 on the PDBbind v2007 core set (195 complexes).

The features chosen by the authors are essentially of geometric nature. They ignore the energy terms induced by the interactions of the ligand and the protein. In addition, the authors ignore the interaction fingerprint features of the protein–ligand complex, e.g., hydrogen bonds with ligand acceptor/donor, ionic interactions with ligand cation/anion, hydrophobic interactions, etc.

The first kind of features (energy terms) are of physical nature whereas the second type (fingerprints) are of chemical nature. In general, a more advanced feature selection process should be conducted in order to choose the minimal set of non-correlated features.

In the work presented in Ref. [42], the authors show that replacing the linear regression used by Cyscore [43] by RF can improve prediction performance. In addition, the authors find that given sufficient training samples, RF comprehensively captures the non-linearity between structural features and measured binding affinities. Moreover, the authors prove that incorporating more structural features and training with more samples can both boost RF performance. The authors in Ref. [42] use three sets of features: Cyscore [43], AutoDock Vina [44], and RF-Score [40]. Cyscore compromises four numerical features: $\Delta G_{\text{hydrophobic}}$, ΔG_{VdW}, ΔG_{hbond} and $\Delta G_{\text{entropy}}$. AutoDock Vina compromises six numerical features: $Gauss_1$, $Gauss_2$, *Repulsion*, *Hydrophobic*, *HBonding* and N_{rot}. RF-Score compromises 36 features defined as the occurrence of intermolecular contacts between two elemental atom types. The authors in Ref. [42] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.803 on the PDBbind v2007 core set (*N* = 195 complexes).

In the work presented in Ref. [45], the authors investigate the impact of the chemical description of the complex on the predictive power of the resulting scoring function. The authors in Ref. [45] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.803 on the PDBbind v2007 core set (*N* = 195 complexes). The authors found that a more precise chemical description of the protein–ligand complex does not generally lead to a more accurate prediction of binding affinity. RF-Score was shown to be inversely correlated to the physical relevance of the descriptors on which the model was trained on (element-dependent distance counts are better than atom type-dependent distance counts which are better than true interaction descriptors). The authors discuss four factors that may contribute to this result: modeling assumptions, codependence of representation and regression, data restricted to the bound state, and conformational heterogeneity of data.

In Ref. [46], the authors presents SFCscoreRF, an RF-based scoring function for improved affinity prediction of protein–ligand complexes. The authors in Ref. [46] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.779 on the PDBbind v2007 core set (N = 195 complexes). In Ref. [47], the authors present a comparative assessment of scoring and ranking powers of machine-learning-based scoring functions on the PDBbind v2013 and compared versus the conventional scoring functions [35]. The use of RF and more generally ensemble-based approaches, e.g., boosted regression trees (BRT) which are resilient to over-fitting with utilizing as many relevant features as possible is so far the best known technique for predicting the binding affinity (as mentioned in Ref. [6]).

Despite the superiority of RF-based scoring functions to predict the experimental binding constants from protein–ligand X-ray structures of the PDBbind dataset, the ranking/docking/screening powers of RF-based scoring functions should also be examined. In Ref. [48], the authors present RF-based scoring function trained on simple descriptors that outperforms a prototype scoring function in predicting binding constants from atomic coordinates (scoring power test). The authors in Ref. [48] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.791 on the PDBbind v2007 core set (195 complexes). However, the proposed RF-based scoring function in Ref. [48] does not discriminate actives of the directory of useful decoys-enhanced (DUD-E) [49] from decoys in docking experiments (virtual screening power test). Moreover, the proposed RF-based scoring function in Ref. [48] is insensitive to docking pose accuracy (docking power test). Thus, the authors in Ref. [50] propose a deep learning-based scoring function (used in both of scoring and ranking powers) which competes the RF-based scoring function trained on the same training complexes of PDBbind v2013 using the same panel of features and tested on the same testing complexes.

The work presented in Ref. [21] improves the correlation between the known binding affinities of the input complexes and those predicted by the eHiTS molecular docking software [39] (that originally uses empirical-knowledge based scoring function). This work opens the door for improving the prediction ability of other traditional scoring functions in other docking software, e.g., AutoDock4 [25] (that originally uses semi-empirical fore field scoring function). The scoring function is trained to derive a unique set of weights for the individual energy terms (features) specific to the target protein. The authors use essentially physical features of the molecular complexes. There are 20 individual energy terms retrieved from eHiTS [39]. However, not all of them are included; two energy terms are always zeros and other two energy terms are *protein-specific*. In Ref. [21], the authors present a regression SVM model trained using the experimentally determined IC_{50} values of 80 different inhibitors (ligands) of the M.tb protein from the BindingDB database [51]. For each complex, this model uses the experimentally determined IC_{50} value to weight the energy terms according to their likely contribution to the overall empirical binding free energy. The scoring function must be trained to derive a unique set of weights for each individual protein family. This is unlike the model presented in Ref. [40] that trains on the experimentally determined $K_{d/i}$ of 1105 complexes in the refined set of the PDBbind v2007 database [28] which includes various protein-families.

In Ref. [21] each combination of features is divided into two groups: 9 mandatory and 7 optional resulting in $2^7 = 128$ combination of features. Data partitioning is used to overcome the over-fitting problem. A 5-fold cross-validation was used to compare the relative accuracies of the 128 different combination of features. The best combination resulted in a significant improvement in the correlation coefficient ($\rho = 0.607$) compared to the one of eHiTS ($\rho = 0.117$). In Ref. [40], the correlation coefficient was 0.776 on 195 test complexes of the core set of the PDBbind v2007 database [28] with no overlap between the testing and training complexes. In case of 65 complexes overlap, the correlation coefficient is 0.827, i.e., it deals better with over-fitting; the learner not only gives good prediction results on the training data, but also on any testing data (i.e., better generalization ability) [21].

In Ref. [52], the authors use ML algorithms for feature selection and model generation of customized docking scoring functions for known inhibitors of the M.tb enoyl acyl carrier protein (enoyl-ACP) reductase. The authors present customized target-specific docking score functions generated with multiple linear regression (MLR), partial least squares (PLS), SVM, and artificial neural networks (ANN). The SVM generates an optimal hyperplane from projections of input descriptors (features) separating active and inactive compounds, whereas ANN is used to place decision boundaries in spaces by using a system of transfer functions that map the input features to an output decision. The features include small molecule descriptors derived from the MOE [17] (290 descriptors), Accord [53] (29 descriptors), and Molegro [54] (14 descriptors), as well as in silico docking energies/scroes from the genetic optimization for ligand docking (GOLD) software [55] (27 descriptors) and AutoDock [25] (1 descriptor which is the binding energy). This generated a total of 361 descriptors (features).

Many descriptors contained invariant values over all the ninety compounds under consideration in Ref. [52], hence, they are removed from further analysis as being useless for prediction of the experimental pIC_{50} values ($-\log IC_{50}$). The authors then conducted linear regression over the remaining descriptors to identify those descriptors with highest correlation with the experimentally measured pIC_{50}. Ninety-six of the descriptors had $R^2 > 0.15$. The remaining descriptors with $R^2 \leq 0.15$ were removed from further analysis given their very week correlation with the pIC_{50}. The authors then performed stepwise MLR over the remaining 96 descriptors to determine a best linear model for predicting the pIC_{50}. From this analysis, 3 descriptors were identified that are used for all subsequent model generation: PLS, SVM, and ANN trained using back-propagation and evolutionary computation.

The MLR, PLS, and SVM models appeared nearly identical in performance with a slight improvement in the ANN models; however, all these models (using customized target-specific scoring functions) performed significantly better than the generic docking scores: GOLDscore [56], ChemScore [57], AutoDock [25], and MolDockScore [54]. The resulting models can be used to identify key descriptors for the ligand under study and are useful for HTS of novel ligands. This paper [52] evaluates and contrasts several strategies for model generation for quantitative structure–activity relationships (QSAR); this is achieved by developing customized target-specific scoring functions using ML techniques to combine QSAR models and scores from *in silico* docking experiments. In general, QSAR models are used to analyze experimental data and build numerical models of the data for prediction and interpretation purposes (e.g., using the QuaSAR suite of applications in MOE [17]).

In Ref. [58], the authors present CScore a data-driven scoring function using a modified cerebellar model articulation controller (CMAC) learning architecture [59]. CMAC is a type of neural network based on the modeling of cerebellum, a region in the human brain which is known for its importance in precise motor control and cognition. It is a kind of associate memory that learns and stores information for the entire input space during training and can be directly used for future prediction. The authors in Ref. [58] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.7668 on the PDBbind v2009 core set (N = 219 complexes). The target-specified CScore achieves an even better result with Pearson correlation coefficient of 0.8237 trained on a much smaller but more relevant dataset for each target.

In Ref. [60], the authors present β contacts and B factor scoring function (B2BScore). B2BScore integrates two physico-chemical properties for protein–ligand binding affinity prediction. The first is the property of β contacts. A β contact between two atoms requires no other atoms to interrupt the atomic contact and assumes that the two atoms should have enough direct contact area. The second is the property of B factor to capture the atomic mobility in the dynamic protein–ligand binding poses. The authors in Ref. [60] identified several important contact descriptors of protein–ligand binding through the RF learning in B2BScore. Some of these descriptors are closely related to contacts between Carbon atoms without covalent-bond Oxygen/Nitrogen, preferred contact of metal ions, interfacial backbone atoms from proteins, or π rings. Some others are negative descriptors relating to those contacts with Nitrogen atoms without covalent-bond hydrogens or non-preferred contacts of metal ions. These descriptors can be directly used to guide protein–ligand docking. The authors in Ref. [60] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.746 on the PDBbind v2009 core set (N = 219 complexes).

In Ref. [61], the authors present ID-Score, a new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions; these descriptors cover nine categories: VdW, hydrogen-bonding interaction, electrostatic interaction, π-system interaction, metal-ligand bonding interaction, desolvation effect, entropic loss effect, shape matching, and surface property matching. A total of 2278 complexes were used as the training set and a modified support vector regression (SVR) algorithm was used to fit the experimental binding affinities. The authors in Ref. [61] achieve Pearson correlation coefficient between the predicted and experimentally determined binding affinities of 0.753 on the PDBbind v2007 core set (N = 195 complexes).

3.3. *Ranking power*

One of the most difficult challenges is the problem of properly ranking the ligands based on their binding affinity to a common protein [6]. The ranking power means to rank the potential solutions of the docking process using a scoring function. This can be achieved by evaluating the potential poses of the protein–ligand complex. This evaluation represents the measure of ligand fitness into the active-site of the receptor protein. This process requires physics-based energy calculations in order to estimate the energy of each binding pose.

3.3.1. *Classical approaches*

The test set used in Ref. [35] consists of 65 clusters of the PDBbind v2013 core set (N = 195 complexes). Each cluster has three complexes formed by the same target protein where the binding affinity of the best complex is required to be at least 100 times higher than that of the poorest; if a scoring function correctly ranked the three complexes in a specific cluster as "the best > the median > the poorest", one point is recorded for this scoring function, i.e., "high-level" ranking; an overall success rate is computed accordingly over the entire test set [35]. In order to provide an additional index, a "low-level" success rate is considered when a scoring function is able to only rank the best complex as the top one in the cluster regardless of the ranking of the median and the poorest complexes in the same cluster [35]. From a panel of 20 conventional scoring functions [35], the best scoring function X-Score^{HM} ranks the ligands bound to fixed target protein with accuracy 58.5% for the high-level ranking (correctly ranking the three ligands bound to the same target protein in a cluster) and with accuracy 72.3% for the low-level ranking (correctly ranking the best ligand only in the cluster).

3.3.2. *ML approaches*

The work presented in Ref. [6] assesses the ranking accuracy of ML-based scoring functions and conventional scoring functions using both 2007 and 2010 PDBbind benchmark data sets [28], working on both diverse and protein-family specific test sets. The protein–ligand complexes are divided into clusters with different ligands bound to the same target protein. The best ML-based scoring function (RF-based) ranks the ligands correctly based on their experimentally determined binding affinities with accuracy 62.5% and identifies the top binding ligand with accuracy 78.1%.

The authors in Ref. [6] use diverse and homogenous (i.e., protein-family-specific) test sets. For each protein–ligand complex, the authors extracted features using: X-Score (a set of 6 features denoted by X) [37], AffiScore (a set of 30 features denoted by A) [62–64], and RF-Score (a set of 36 features denoted by R) [40]. By considering all the 7 combinations of these 3 types of features (i.e., X, A, R, X ∪ A, X ∪ R, A ∪ R, and X ∪ A ∪ R), the authors generated 7 versions of the training and testing data sets. They exploit 6 ML-based scoring functions: MLR, multivariate adaptive regression splines (MARS) [65], k-nearest neighbors (kNN) [66], SVM [67], RF [68], and BRT [69]. For each protein family, the authors consider 3 protein–ligand complexes in the refined set of PDBbind v2007 database [28]; those are the complexes with the highest, lowest, and median experimentally determined binding affinities among all complexes corresponding to a given protein.

Calculating the ranking power of a given scoring function is then straightforward [6]: (1) each protein–ligand complex is scored, (2) for each protein cluster (i.e., 3 protein–ligand complexes associated with a common protein), complexes are ordered according to their predicted binding affinities, and (3) any given cluster is considered properly ranked if its order based on the predicted binding affinities matches its order based on the experimentally determined binding affinities. The ranking performance is calculated for 64 protein families in the PDBbind v2007 core test set. The results of the experiments in Ref. [6] found that utilizing as many relevant features as possible in conjunction with ensemble-based approaches like BRT and RF (which are resilient to over-fitting) is the best option.

The authors in Ref. [70] propose a non-parametric ML approach to build target-specific scoring functions by using features derived from the X-ray structures and binding affinities of known ligands to a particular protein. In particular, the proposed method is very useful in lead optimization (i.e., predicting the best novel candidate ligand) in which support vector inductive logic programming (SVILP) has two functionalities: (1) predicting the binding affinity of a series of novel ligands to a particular protein and (2) learning the binding rules: which interactions of ligands within the protein binding site are important in determining the affinity, see Fig. 14, hence, may be thought of as analogous to the QSAR techniques. The concept of *pharmacophore type* has also been employed to assist the scoring of ligands to a specific target protein.

As mentioned in Ref. [70], inductive logic programming (ILP) is a *qualitative* ML method that learns logical rules according to observations (known binding affinities of the ligands) and background knowledge (distances between fragments of the ligand and protein atoms). In this approach a central non-hydrogen atom and all of the atoms directly bonded to it (including hydrogen atoms) are defined as a *fragment*. Then, by combining the ILP with SVM regression, a quantitative set of rules can be obtained. The Tanimoto coefficient of molecule 1 in the dataset of protein–ligand

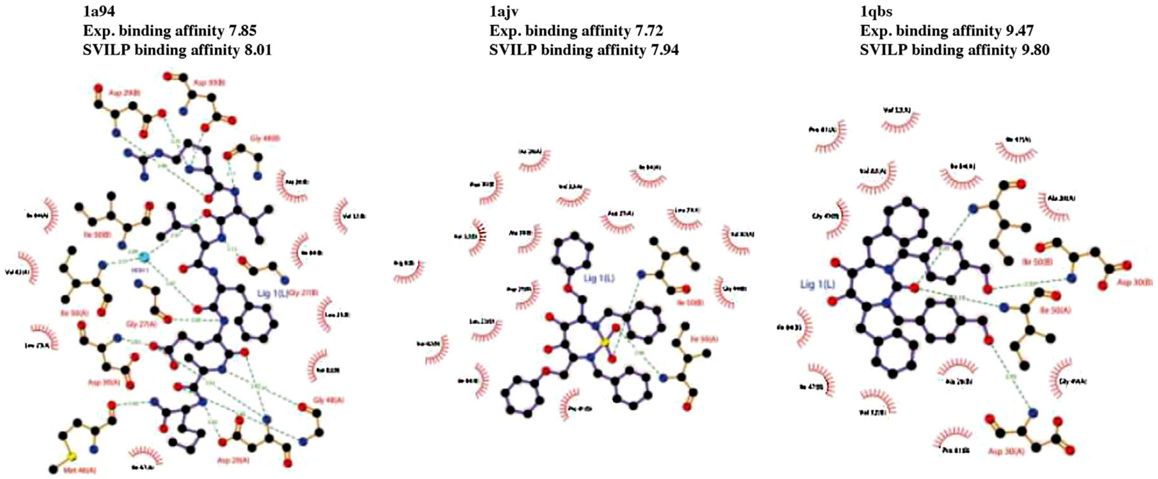

1a94
Exp. binding affinity 7.85
SVILP binding affinity 8.01
1ajv
Exp. binding affinity 7.72
SVILP binding affinity 7.94
1qbs
Exp. binding affinity 9.47
SVILP binding affinity 9.80
Lig 1(L)
Lig 1(L)
Lig 1(L)

Fig. 14. Binding interactions of ligands; hydrogen bonds indicated by dashed lines and hydrophobic interactions by red arcs [70]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

complexes is measured against all the other molecules based on the number of similar fragments the two molecules share, Eq. (15):

$$T_c = \frac{nc}{(nq + nt - nc)}, \tag{15}$$

where *nc* equals the number of fragments common to the two molecules, *nq* is the total number of fragments in molecule 1, and *nt* is the total number of fragments in molecule 2. The diversity of the ligands in each dataset is calculated using a similarity index (*T*) based on the Tanimoto coefficient, Eq. (16):

$$T = 100 \times \frac{N_1}{N_t}, \tag{16}$$

where N_1 equals the number of molecules in the dataset that have at least one similar molecule (with Tanimoto coefficient greater than 0.8) in the set and N_t is total number of molecules.

In Ref. [71], the authors use the community structure–activity resource (CSAR) datasets (http://www.csardock.org/) to train and test 2 SVR-based scoring functions; SVR-knowledge based (SVR-KB) and SVR-empirical descriptors based (SVR-EP). The features used to train SVR-KB are knowledge-based pairwise potentials while SVR-EP is based on physico-chemical properties. SVR-KB trained using PDBbind v2010 outperformed X-Score by nearly 0.1.

3.4. Docking power

In Ref. [35], the docking power refers to the ability of a scoring function to identify the native binding pose among computer-generated decoys. Ideally, the native binding pose is identified as the one with the best binding score.

3.4.1. Classical approaches

In Ref. [35] a set of decoy binding poses (up to 100) was generated for each protein–ligand complex in the PDBbind v2013 core set (195 complexes) by using several molecular docking programs. Each scoring function (in a panel of 20 scoring functions [35]) was applied to score the decoy set of each protein–ligand complex and finds consequently the best-scored binding pose. The decoy set of each complex includes the native binding pose to ensure that there exists at least one correct binding pose. If the property-matched root-mean-square-deviation (RMSD) [35] value between the native binding pose and the best-scored binding pose (among all decoys plus the native one) fell below predefined cutoff, e.g., RMSD <2.0Å, it is recorded as a successful prediction. Once this analysis was completed over the entire test set, an overall success rate was computed for every scoring function. ChemPLP@GOLD [72] and Chemscore@GOLD [57] achieved success rates above 80%.

3.4.2. ML approaches

The docking power is the ability to identify the *best binding pose* of a ligand from a set of *computationally generated poses*. The docking simulation process includes searching the conformational space of the protein active site for optimal ligand binding poses. In this process, the protein–ligand state is represented by the orientation (static features) and conformation (dynamic features). The goal is to search for the most stable state of the protein–ligand complex, i.e., the state with *minimum energy configuration*. This process of matching two molecules together is combinatorially intractable. It is even harder when considering the protein flexibility (change in conformation). Examples of used techniques for orientation/conformational search space are: steepest descent optimization, Monte Carlo simulation, simulated annealing, MD, genetic algorithms, and distance geometry methods.

For instance, in the AutoDock4 software [25], the docking process is carried out using several methods including traditional genetic algorithms, simulated annealing, and lamarckian genetic algorithm which is the most efficient of them [73]. These algorithms are based on using the maps generated by AutoGrid [25] to evaluate the ligand–protein interaction at each point in the docking simulation.

In Ref. [48], the authors propose RF-based and SVM-based scoring functions that are insensitive to docking pose accuracy (docking power test). Thus, this work [48] highlights the need of checking any novel ML-based scoring function versus the best conventional scoring functions with respect to their docking powers. In Ref. [74], the authors present a comparative assessment of docking and screening powers of machine-learning-based scoring functions on the PDBbind v2013 and compare versus the best conventional scoring functions [35].

3.5. Screening power

In Ref. [35], the screening power of a scoring function is defined as the ability of a scoring function to identify the true binders to a given target protein among a pool of random molecules (decoys).

3.5.1. Classical approaches

In Ref. [35] the screening power was evaluated in a cross-docking trial. The test set includes 65 clusters of PDBbind v2013 core set (195 complexes). Each cluster consists of three complexes formed by a certain protein. For each protein, the three known ligands were taken as positives; whereas the other 195 − 3 = 192 ligands were taken as negatives. For each of the 65 proteins, all 195 ligands were docked into its binding site, resulting in a total of 65 × 195 = 12,675 protein–ligand pairs. Since each protein has three different structures, the structure of the best complex in each cluster was selected to be the cluster representative. Up to 50 representative ligand binding poses were selected for each protein–ligand pair. Each scoring function (in a panel of 20 scoring functions [35]) was applied to score the binding poses of all 195 ligand molecules (including true binders and negatives) of each target protein. For any given ligand, the best-scored binding pose among all available poses was taken as the predicted binding pose and the corresponding binding score was taken as the predicted binding affinity by this scoring function. All 195 ligands were then ranked according to their binding scores in a descending order. The screening power of a scoring function is measured by counting the total number of true binders among the 1%, 5%, and 10% top-ranked ligands. Enrichment factor (EF) is computed using the following equation [35]:

$$\mathrm{EF}_{x\%} = \frac{\mathrm{NTB}_{x\%}}{\mathrm{NTB}_{\mathrm{total}} \times x\%} \tag{17}$$

where $\mathrm{NTB}_{x\%}$ is the number of true binders observed among the top x% (where $x = 1, 5$, or 10) candidates selected by a given scoring function. $\mathrm{NTB}_{\mathrm{total}}$ is the total number of true binders for the given target protein (which is typically 3 for each target protein if there are no cross-binders).

As another performance indicator for the screening power in Ref. [35]; if the best ligand was found among the 1%, 5%, and 10% top-ranked candidates, a point is counted for the scoring function under test; an overall success rates on the entire test set at the three different levels are computed accordingly. GlideScore-SP [75] is the best conventional scoring function in this test with an average EF near 20 and success rate of 60% at the top 1% level. EF at the top 5% and top 10% levels are considerably lower for all 20 scoring functions because the test set in Ref. [35] consists of a rather limited number of true binders (normally three) to each target protein.

3.5.2. ML approaches

In Ref. [76] the authors presented a virtual screening method based on ML regression. Recall that virtual screening techniques are employed to complement HTS by minimizing the number of ligands to be physically screened. The performance of HTS against antibacterial targets is generally unsatisfactory with high costs and low rates of hit identification; all known inhibitors are derivatives of the same core scaffold, i.e., little scaffold diversity. Scaffold hopping aims at identifying novel active ligands that are likely to be good candidates for drugs by replacing a portion of a known compound (the scaffold), while preserving the remaining chemical groups [17]. Thus, the authors in Ref. [76] proposed a virtual screening method that is able to identify a high proportion of structurally diverse inhibitors by searching large molecular databases in an efficient manner. In particular, this method exploits the structures (active site specifically) of two antibacterial targets M.tb and Streptomyces coelicolor (S.cl)) to identify novel inhibitors. State-of-the-art docking protocols (e.g., RF-Score) are combined hierarchically with ultrafast shape recognition method: a rapid descriptor-based shape similarity technique capable of quickly identifying innovative active scaffolds in very large molecular databases. Fig. 15 shows 3 known inhibitors used as templates for shape similarity with the active site of the 2 targets (M.tb and S.cl). The core scaffold is circled; this is the closest scaffold to the catalytic residues (amino acids in the target binding site).

The wealth of chemical diversity (9 million commercially available molecules in the ZINC database [77]) is a key component of the proposed virtual screening method in Ref. [76]; a smaller database generated with the same procedure will contain a lower number of innovative scaffolds. The proposed method identified 4379 diverse molecules (from the 9 million molecules initially considered) that are similar in shape to the 3 inhibitors in Fig. 15 and thus can fit into the active site of the target protein.

The authors evaluate the pose generation quality by re-docking the 3 largest co-crystallized ligands back to their respective targets in Fig. 15; two methods for scoring functions are used. The first is a consensus scoring function which considers the 3 sets of docking poses containing the 3 largest co-crystallized ligands in Fig. 15 and each set is sorted with the average rank of the pose according to ChemScore [57], GOLDscore [56], and Astex statistical potentials (ASP) [78]. Note that high ranked poses by the three different scoring functions (ChemScore [57], GOLDscore [56], and ASP [78]) represent a more reliable prediction than any of the three scoring functions alone. The consensus scoring has been generally found to improve virtual screening performance dramatically with respect

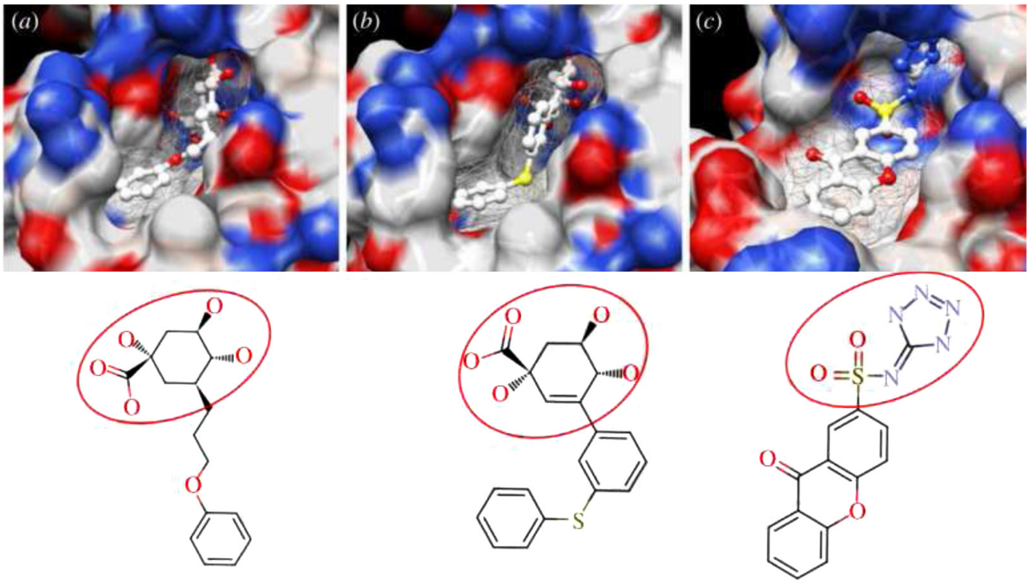


Fig. 15. Three ligands used as templates for shape similarity [76]: (a) CA2 inhibitor with S.cl target, (b) RP4 inhibitor with S.cl target, and (c) GAJ inhibitor with M.tb target.

to the individual scoring functions. The second scoring function is the RF-Score which uses non-parametric ML to build predictive models of the binding affinity in an entirely data-driven manner. The technique proposed in Ref. [76] identified 100 new inhibitors with calculated K_i ranging from 4 to 250 μ M with outstanding hit rates (identifying novel ligands that bind to the target protein) of about 60% and 62% against M.tb and S.cl respectively. Moreover, the authors discovered 50 new active molecular scaffolds.

In Ref. [21], the authors present a classification SVM which is trained using the experimentally determined active and decoy (inactive) complexes retrieved from the directory of useful decoys (DUD) database [79]. A multi-planar classification model is used to address the issue of data imbalance resulted from the negative data over-representation in the HTS data sets. This model is applied to a diverse set of targets (specifically 12). However, every protein-family is analyzed independently of the other protein-families. The classification SVM separates positive and negative ligands in a multi-dimensional space by drawing an optimal hyper-plane with a maximum distance between the two classes. The popular choices when addressing the issue of data imbalance is to either over-sample the minority class or under-sample the majority class. The classification accuracy is measured using the *F*-score [80]:

$$F\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{18}$$

where the precision [81] equals the number of correct positive results (true positives) divided by the total number of predicted positive results (true positives plus false positives), whereas the recall [81] equals the number of correct positive results (true positives) divided by the total number of results that should have been predicted as positive (true positives plus false negatives). The data imbalance results in a tendency of the model to classify ligands into the negative class, and therefore have high precision (due to small false positives) but low recall (due to large false negatives).

In order to compare the performance of the eHiTS scoring function (which uses the default energy terms) versus the SVM classifier [21] at discriminating actives from decoys, the receiver operating characteristic (ROC) curves [82] were plotted. ROC is a graphical plot that illustrates the performance of a binary classifier as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the positives, i.e., precision (horizontal axis) versus the fraction of false positives out of the negatives (vertical axis) at various threshold settings. The eHiTS curve is below the random classifier curve in the ROC graph due to the data imbalance towards the negative class. The use of imbalanced data in training the SVM classifier [21] pushes the optimal hyper-plane towards the positive class (due to a wider negative class). In order to directly compare the performance of the SVM classifier with that of the SVM regressor [21], the latter was used to rank the DUD InhA data set according to the predicted $\log IC_{50}$ values. The SVM classifier has been shown to be better than the SVM regressor in virtual screening due to the efficient handling of the data imbalance problem.

In Ref. [83], the authors did not predict the binding free energy, rather they proposed an ML classifier to discriminate between active and inactive ligands of a target from the public repository for chemical structures (PubChem) [84]. In Ref. [83] the authors use the docking poses of 2000 decoy compounds randomly selected from the PubChem database as the negative samples. For the positive samples of the test set, 100 known inhibitors of the target proteins were selected from the structure activity relationship (StARLITe)[3] database [86] using hierarchial clustering according to distances between their 2D structural molecular fingerprints; public molecules access system (MACCS) keys [87]. This work is similar to the work presented in Ref. [21] where an SVM classifier was trained using the DUD database [79]. Nevertheless, the work presented in Ref. [21] proposes as well an SVM regressor trained using the BindingDB database [51] that predicts the binding free energy based on the IC_{50} values.

In Ref. [83] the authors use the docking poses of active and inactive ligands of a target protein. These poses depend on the distances between atoms types in the ligand–protein atom pair from the 3D geometrical structure of the complex. Protein–ligand interactions are detected using the functions built in the MOE software [17]. On one hand, this work is similar to the work presented in Ref. [40] (using RF) that depends on distance-based features derived from the PDBbind database [28]. On the other hand, this work differs from the work presented in Ref. [21] (using SVM) that depends on energy terms retrieved from the eHiTS docking program [39].

The work presented in Ref. [83] proposes 5 target-specific classifiers; ANN, SVM, naïve Bayesian classifier, and RF. The proposed methods generally outperform the GlideScore::SP [75] in the Schrödinger software [88]. On one hand, this work is similar to the work presented in Ref. [21] that outperforms the results of the eHiTS docking program [39]. On the other hand, it differs from such work that uses only SVM models [21]. Finally, [83] uses two performance indices; ROC (like the work presented in Ref. [21]) and EF at 10% level.

In Ref. [48], the authors propose RF-based and SVM-based scoring functions that do not discriminate DUD-E [49] actives from decoys in docking experiments (virtual screening power test). Thus, this work [48] highlights the need of checking any novel ML-based scoring function versus the best conventional scoring functions with respect to their screening powers.

In Ref. [89], the authors present neural-network-based scoring function (NNScore). Aside from providing additional validation of the original NNScore function, the authors in Ref. [89] present a second version namely NNScore 2.0 that considers many more binding characteristics when predicting the binding affinity than the original NNScore does. In addition, the network output of NNScore 2.0 differs from that of NNScore 1.0; rather than a binary classification of ligand potency, NNScore 2.0 provides a single estimate of the pK_d. The authors measure the NNScore performance using both of the ROC and EF.

In Ref. [71], the high performance in ranking power did not translate into greater enrichment in virtual screening evaluated using 40 targets of the DUD database. To overcome this issue, a variant of SVR-KB was developed by following a target-specific docking strategy using randomly picked decoys (SVR-KBD); similar strategy was previously employed by the authors to derive SVM target-specific model (SVM-SP). SVR-KBD shows much higher enrichment outperforming all other scoring functions and is comparable in performance to the previously derived scoring function SVM-SP.

In Ref. [61], ID-Score can correctly differentiate structurally similar ligands indicating higher sensitivity to analogues. Thus, the high performance of ID-Score enables it as a useful tool not only in assessing protein–ligand affinity in structure-based drug design, but also in lead optimization.

4. Molecular software

In this section, we survey some of the most successful molecular docking software currently in the literature. One of the most famous

[3] StARLITe was a commercial database manually compiled from many journals then the European bioinformatics institute, an outstation of the European molecular biology laboratory (EMBL), bought StARLITe and it became an open-access database named ChEMBL [85].

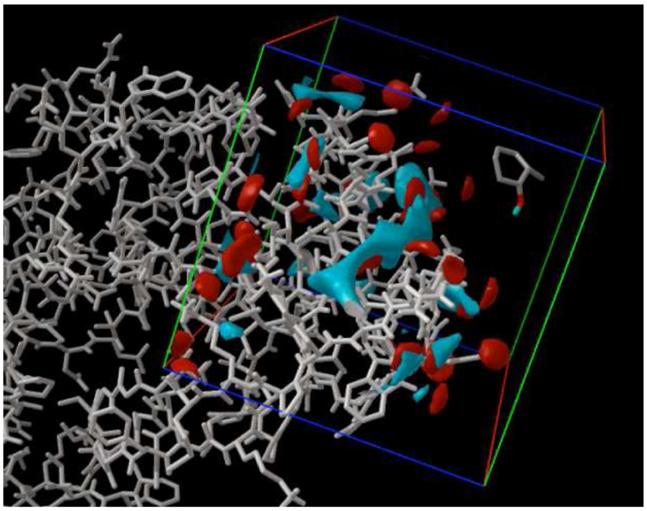


Fig. 16. Rendering a protein in AutoGrid.

and ancient such software is the AutoDock4 [25]. AutoDock4 employs MD simulation and is open-source. The ADT [25] is the GUI interface of AutoDock4. The ADT can be used for creating the coordinate file formate; protein data bank with partial charges and atom types (PDBQT), that are readily usable by AutoDock, from the traditional PDB files.

In AutoGrid [25], the protein is embedded in a 3D-grid and an atom is placed at each grid point, see Fig. 16.

Every atom type has a specific color, for example, Carbon has the gray color and Oxygen has the red color. Almost the whole left gray bonds represent the protein, whereas the rest of the complex including the binding site (pocket of the ligand) is at the upper right corner of the figure. The light blue color represents a cross-section of the Carbon atoms. Note that not all the residues (amino-acids) in a protein are equally important. The rectangular box to the right of the figure represents the highly important residues. The blue linear contours at the top and bottom right parts of Fig. 16 surround the area inside the grid box which are most favorable for binding of Carbon atoms. The red contours on the other hand show areas that favor Oxygen atoms. The AutoDock Vina [44], a protein–ligand docking software, is able to analyze the important part (active site) given the protein 3D structure.

The idock [90] is a free open-source multi-threaded virtual screening tool for flexible ligand docking. The RF-Score [40] is now integrated in idock where the docking poses are re-evaluated using the RF scoring function. The AMBER biomolecular simulation program [23] has MD module, however, it is not free. The Avogadro [91] is a free advanced semantic chemical editor, visualization, and analysis platform, however, it has no MD module, so it does not do molecular docking. The ChemOffice [92] is used to draw molecules and reactions in documents, search databases, and predict properties, however, it is commercial and has no MD module. The CHARMM [24] is a highly versatile and widely used molecular simulation program, however, it is not free. The eHiTS [39] is a flexible docking system producing accurate docking poses and individual energy terms, however, it is not free and unavailable on Windows OS. The MOE [17] can be used to detect pharmacophore protein–ligand interactions using built-in functions. In addition MOE has the following functionalities: molecular modeling & simulations, protein modeling & bioinformatics, structure-based design, cheminformatics & high throughput discovery, and development environment. However, MOE is commercial.

FlexX [93] is one of the most established protein–ligand docking in the literature. The technology behind FlexX is based on a robust incremental construction algorithm. The ligand is decomposed into pieces and then flexibly built-up in the active site of the protein using a variety of placement strategies. The poses are scored based on variety of different scoring functions. Nevertheless, FlexX is commercial.

In Ref. [94], the authors present the biochemical algorithms library (BALL) which is a comprehensive rapid application development for structural bioinformatics that has molecular modeling library. The key functionalities of BALL are: support for various file formats, molecular edit-functionality, new MM force fields, novel energy minimization techniques, docking algorithms, and support for cheminformatics. BALL is available for all major operating systems free of charge. In addition, BALL is available as source code and binary packages from the project web site at http://www.ball-project.org.

5. Molecular databases

Every molecular database has some specific raw data which can be processed for feature extraction. Examples of currently available molecular databases include: PDB archive [16], Binding mother of all databases (Binding MOAD) [95], ZINC [77], PubChem [84], PDBbind [28], DUD [79], BindingDB [51], etc. The same molecular complex can exist in more than one database, however, some databases may record different information about the same complex, e.g., some record the $K_{d/i}$ while others record the IC_{50}. Moreover, different databases (even different versions of the same database) may include different complexes.

The PDB archive [16] contains information about experimentally determined 3D structures of large biological molecules including proteins, nucleic acids, and complex structures. Using the PDB archive, one can perform advanced searches based on structure and function (e.g., cancer related proteins), visualize, download, and analyze molecules. For instance, the available features are the molecule name, number of polymers (i.e., polymer, dimer, etc.), molecule type (e.g., protein), length (i.e., number of amino-acids), organism from which the molecule is extracted (e.g., homo sapiens, i.e., a human), and whether the molecule is bound to a ligand.

Binding MOAD [95] is a subset of the PDB database, containing every high-quality example of protein–ligand binding. Thus, it is called mother of all databases. The last checked Binding MOAD contains 21,109 protein–ligand structures, 7284 binding data, and 10,156 different ligands. The creators of Binding MOAD [95] have searched the primary reference of each PDB entry looking for the experimental K_d, K_i, and IC_{50} data.

BindingDB [51] is a public web-accessible database of measured binding affinities (K_i, IC_{50}, K_d, EC_{50}), focusing mainly on the interactions of protein–ligand molecules. The last checked BindingDB contains 1,009,290 binding data for 6589 protein targets and 427,325 ligands. There are 2046 protein–ligand crystal structures with affinity measurements for proteins with 100% sequence identity and 5815 crystal structures for proteins with 85% sequence identity (structure diversity leading to non-redundant structure repository). BindingDB continually references a set of journals not covered by other public databases.

The PDBbind database [28] is a collection of experimentally measured binding affinity data (K_d, K_i, or IC_{50}) for the protein–ligand complexes in the PDB database (with known 3D structures). For instance, version 2013 is based on the contents of PDB officially released on Jan 1st 2013. This version provides binding affinity data and structural information for a total of 10,776 biomolecular complexes, including protein–ligand (8302), nucleic acid-ligand (83), protein–nucleic acid (587), and protein–protein complexes (1804). All of the binding affinity data are collected from more than 24,000 references.

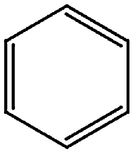
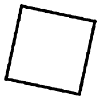
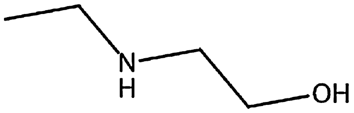
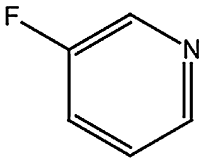
1		{ is-aromatic, has-ring, has-C }
2		{ has-ring, has-C }
3		{ has-C, has-N, has-O }
4		{ is-aromatic, has-ring, has-C, has-N, has-halogen }

Fig. 17. Possible fingerprints [17].

ZINC [77] is a free database of commercially available compounds for virtual screening. ZINC contains over 21 million purchasable compounds in ready-to-dock 3D formats. Huang et al. recently constructed the DUD database [79] for benchmarking virtual screening (help to test docking algorithms) with 2950 ligands for 40 different targets. Every ligand has 36 decoy molecules that are physically similar but topologically distinct, leading to a database of 98,266 compounds. DUD is derived from ZINC, so compounds in DUD are purchasable. Thus, DUD is useful for the validation of ligand-based screening. The DUD provides a useful resource for training SVM classification models, where the feature vector of each molecule can be labeled as either active (+1) or decoy (-1). The ratio of actives to decoys in the DUD is approximately 1:36, thus there is a strong bias towards negative examples in the training data.

The PubChem database [84] is a public repository for biological properties of small molecules hosted by the US national institutes of health. The last checked PubChem BioAssay database contains biological test results for more than 700,000 compounds; this information is easily accessible to biomedical and computer scientists researchers. PubChem has web-based services that provide tools for rapid data retrieval, integration, comparison of biological screening results, etc.

6. Feature extraction

Feature extraction is a key point in the ML approaches to protein–ligand docking. ML techniques can be used for advanced feature extraction [96]. In addition, the correlation between the various features affects the scoring-function accuracy, e.g., if two features are highly correlated then one feature might be sufficient to be considered. Feature extraction and studying the correlations among different features have been done effectively in other related areas; for example, in the prediction of biological hazards induced by the use of nano materials [97].

Features based on energy terms can be calculated from the molecules raw data available in molecular databases. The energy terms that define the intermolecular interactions are derived from experimental data and ab initio simulations. These include the VdW forces, bonding and non-bonding interactions, and electrostatic energies. Some scoring functions such as eHiTs [39] and X-Score [37] generate energy terms (VdW interactions, Hydrogen bonding, deformation effect, hydrophobic effect, etc). The energy-based features are more of dynamic properties that depend on MD. The work done in Ref. [21] is based on 20 energy terms extracted using the eHiTS docking software given the 3D structure of the protein–ligand complexes from the BindingDB molecular database.

Another kind of features, that are static in nature, are the geometry-based properties of the underlying complex. For example the RF scoring function [40] uses features such as the number of Carbon–Carbon (C–C) contacts and Nitrogen–Carbon (N–C) contacts etc., within a certain distance.

The work presented in Ballester et al. [40] and Ashtawy et al. [6] are similar in the sense that both are mainly not target-specific. Rather, they train on almost the whole PDBbind database (not only compounds related to a specific protein family).

Pharmacophore description of the bound complex molecule includes features such as the type of hydrogen atoms involved in the bond. These are essentially chemical features and used to create what is called the *interaction fingerprint* which is a distinguishing characteristic of the underlying molecule. Fig. 17 shows 4 possible fingerprints chosen from a universe of 8 features consisting of U={is-aromatic, has-ring, has-C, has-N, has-O, has-S, has-P, has-halogen}; the total number of possible fingerprints are 2^8 leading to 256 possible molecules [17]. Once a fingerprint is derived from a chemical structure, a metric is required to compare fingerprints and accordingly the molecules [17]. A common metric is the *Tanimoto coefficient* which is useful in ranking compounds in virtual screening experiments [17]. The *Tanimoto coefficient* can take values ∈[0, 1] where 0 means maximum dissimilarity and 1 means maximum similarity [17]. The *Tanimoto coefficient* between fingerprints is defined to be the number of common features divided by the number of total features in both fingerprints; for instance, the *Tanimoto coefficient* between molecules 1 and 2 in Fig. 17 equals the number of {has-ring, has-C} (which is 2) divided by the number of {is-aromatic, has-ring, has-C} (which is 3) equals 0.666 [17]. Moreover, the *Tanimoto coefficient* can be used as a threshold to define similarity [17]; for instance, 2 compounds having a *Tanimoto coefficient* greater than or equal to 0.6 can be considered as similar.

In Ref. [52], descriptors are derived from existing docking scoring functions and from calculated physical-chemical properties and energies to develop target-specific scoring functions for the prediction of the experimental pIC_{50} values for enoyl-ACP inhibitors. These descriptors include both 2D descriptors (based only on the atoms and their bonding interactions) and 3D descriptors (based on the conformation of the molecule).

Recently traditional 1D- and 2D-QSAR methodologies based on MLR, PLS, and principle component analysis (PCA) have been very useful in generating target-specific scoring functions with descriptors that are specific to the protein–ligand interactions [52]. MLR is often used for modeling the linear correlation between descriptors (features) and activities (e.g., binding affinities). PLS is useful for cases where the number of samples is small relative to the number of descriptors (as in CADD). PLS analysis consists of a linear regression model based on the transformation of the original descriptors into a new space composed of a smaller number of orthogonal descriptors. PCA also transforms a large number of correlated descriptors into smaller number of orthogonal uncorrelated descriptors but finds principle components of maximum variance; the first few principle components account for as much of the variability as possible, and each subsequent principle component accounts for additional variability. These three methods are commonly used in commercially available cheminformatics software packages, e.g., the MOE [17] offers the PCA method to reduce the dimensionality of a large set of molecular descriptors by linearly transforming the data.

7. Conclusion

CADD is the application of the computational science and technology in the life cycle of drug design. This new paradigm, which can be considered as virtual screening, has shown to be very effective and economical. Central to virtual screening is the notion of computational docking. Such process mainly docks the candidate drug into the active site of the target protein and evaluates the stability of the resulting pose. This can be done through the design and implementation of a proper scoring function. As expected such function must be both accurate enough and computationally feasible. Many scoring functions have been designed based on traditional techniques but generally the performance was not very satisfactory. In recent years there have been an abundant flow of biochemical and biophysical data into the literature based on much experimental work. This has led to a data-oriented approach to the design of scoring functions. Therefore, the techniques employed are based on the ML technology.

In this paper we have presented the basic biochemical and biophysical background in a manner that is accessible to all audience particularly computer scientists. We surveyed the state-of-the-art learning techniques used in the design of novel scoring functions. Some of the developed techniques are parametric such as the use of ANN and MLR. The scoring functions based on such techniques are good predictors particularly when the target protein is fixed, that is, they work well for specific classes of proteins. On the other hand, non-parametric methods such as RF have been successful on generic datasets with multiple classes of target proteins.

A crucial task in developing any ML technique is the extraction and selection of features which are relevant to the docking process. Generally, three types of features have been used in the design of predictive scoring functions: geometric features based on the 3D description of the underlying molecule, physical force field features, and chemical features based on the design of pharmacophore fingerprints.

As a future work, more analysis need to be done into the selection and extraction of an effective set of relevant features to the docking process. There might not be a unified such set, and it may depend much on the precise task and purpose of the docking process. For example, this may depend on the particular protein family under consideration, particular class of protein families, or generic class of protein families. Also, it may depend on the particular functionality(ies) required of the scoring function; whether just scoring, that is, predicting the binding affinity and/or ranking of potential drug candidates and/or a description of the docking pose. Whether classification or regression approach is taken might also affect the feature selection/extraction process. Utilizing the advanced ML techniques, e.g., ensemble-based methods as RF or advanced ANN based on deep learning, and the combination of the scores of such advanced models may further outperform the current state-of-the-art scoring functions in scoring, ranking, docking and screening powers.

The ML paradigm can provide insights into the workings of the docking process from the perspectives of biochemistry and biophysics. The work done so far focused almost exclusively on developing an accurate predictive scoring function, that is, a model that accurately predicts the stability of the docked complex molecule as well as whether or not an inhibitor is active. A potential research direction is using ML techniques in predicting the actual binding pose. This latter task is of more dynamical nature, for it might need the MD simulation which is computationally prohibitive. Much work and several research groups around the world have been devoted to the application of new hardware technologies, e.g., HPC, for the MD simulation. However, sophisticated ML techniques should also be employed for that matter using available time-series data. This may include the use of Gaussian processes and its variants. Another research direction could be directed towards the inclusion of quantum effects (as they provide rigorous description of molecular systems) into the design and implementation of scoring functions. Success in protein–ligand docking can also open the door for generalization of bio-molecular applications, e.g., protein–protein docking.

Acknowledgment

This work is mainly supported by the Information Technology Industry Development Agency (ITIDA) under the ITAC Program grant no. CFP#58 and in part by E-JUST Research Fellowship.

References

[1] Rupp B, Ruzsics Z, Buser C, Adler B, Walther P, Koszinowski UH. Random screening for dominant-negative mutants of the cytomegalovirus nuclear egress protein M50. J Virol 2007;81(11):5508–17.

[2] Klein C. On chance discovery in rational drug design. *Aspergillus fumigatus* and angiogenesis. Pharm Unserer Zeit 2006;36(6):450–1.

[3] Denny WA. The design and development of anti-cancer drugs. Tech. Rep. New Zealand Institute of Chemistry; 1998 http://nzic.org.nz/ChemProcesses/biotech/12J.pdf

[4] Marshall GR. Computer-aided drug design. Annu Rev Pharmacool Toxicol 1987;27(1):193–213.

[5] DockingWikipedia. Docking (molecular) – wikipedia; 2014 http://en.wikipedia.org/wiki/Docking_(molecular) (accessed on 28.12.14).

[6] Ashtawy HM, Mahapatra NR. A comparative assessment of ranking accuracies of conventional and machine-learning-based scoring functions for protein–ligand binding affinity prediction. IEEE/ACM Trans Comput Biol Bioinform 2012;9(5):1301–13.

[7] Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ. Virtual screening with AutoDock: theory and practice. Expert Opin Drug Discov 2010;5(6):597–607.

[8] Purohit R, Rajendran V, Sethumadhavan R. Studies on adaptability of binding residues flap region of TMC-114 resistance HIV-1 protease mutants. J Biomol Struct Dyn 2011;29(1):137–52.

[9] Purohit R, Sethumadhavan R. Structural basis for the resilience of Darunavir (TMC114) resistance major flap mutations of HIV-1 protease. Interdiscip Sci: Comput Life Sci 2009;1(4):320–8.

[10] Purohit R. Role of ELA region in auto-activation of mutant KIT receptor: a molecular dynamics simulation insight. J Biomol Struct Dyn 2014;32(7): 1033–46.

[11] Rajendran V, Purohit R, Sethumadhavan R. In silico investigation of molecular mechanism of laminopathy caused by a point mutation (R482W) in lamin A/C protein. Amino Acids 2012;43(2):603–15.

[12] Purohit R, Rajendran V, Sethumadhavan R. Relationship between mutation of serine residue at 315th position in *M. tuberculosis* catalase-peroxidase enzyme and isoniazid susceptibility: an in silico analysis. J Mol Model 2011;17(4):869–77.

[13] Kumar A, Purohit R. Use of long term molecular dynamics simulation in predicting cancer associated SNPs. PLoS Comput Biol 2014;10(4):e1003318.

[14] Rajendran V, Sethumadhavan R. Drug resistance mechanism of PncA in *Mycobacterium tuberculosis*. J Biomol Struct Dyn 2014;32(2):209–21.

[15] Sanner MF. Python: a programming language for software integration and development. J Mol Graph Model 1999;17:57–61.

[16] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, et al. The protein data bank. Nucleic Acids Res 2000;28(1):235–42 http://www.rcsb.org

[17] Chemical Computing Group Inc. Molecular operating environment (MOE), 2013.08; 2013.

[18] Peng L, Tan G, Kalia RK, Nakano A, Vashishta P, Fan D, et al. Scalability study of molecular dynamics simulation on Godson-T many-core architecture. J Parallel Distrib Comput 2013;73:1469.

[19] Richardson JL. Visualizing quantum scattering on the CM-2 supercomputer. Comput Phys Commun 1991;63:84–94.

[20] Vassiliev V. Introduction to amber: The theory and practice of biomolecular simulations using the amber suite of programs. In: Presentation; NCI national facility, the Australian National University; ACT 0200. 2011.

[21] Kinnings SL, Liu N, Tonge PJ, Jackson RM, Xie L, Bourne PE. A machine learning based method to improve docking scoring functions and its application to drug repurposing. J Chem Inform Model 2011;51(2):408–19, http://dx.doi.org/10.1021/ci100369f.

[22] Shattuck TW. Colby college molecular mechanics tutorial introduction. Maine: Molecular Mechanics Tutorial Introduction; Department of Chemistry Colby College Waterville; 2008 http://www.colby.edu/chemistry/CompChem/MMtutor.pdf

[23] Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The amber biomolecular simulation programs. J Comput Chem 2005;26(16):1668–88.

[24] Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30(10):1545–614.

[25] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 2009;30(16):2785–91.

[26] Wang YT. Molecular simulations workshop: Introductions/amber: how to set-up calculations. In: Presentation; Department of Biochemistry, College of Medicine, Kaohsiung Medical University. 2014.

[27] Yung-Chi C, Prusoff WH. Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. Biochem Pharmacol 1973;22(23):3099–108.

[28] Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. J Med Chem 2004;47(12):2977–80.

[29] Wang J, Morin P, Wang W, Kollman PA. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc 2001;123:5221–30.

[30] ChemicalFileFormatWikipedia. Chemical file format; 2014 http://en.wikipedia.org/wiki/Chemical_file_format (accessed on 28.12.14).

[31] Lu SJ, Chong FC. Combining molecular docking and molecular dynamics to predict the binding modes of flavonoid derivatives with the neuraminidase of the 2009 H1N1 influenza a virus. Int J Mol Sci 2012;13(4):4496–507.

[32] Naïm M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, et al. Solvated interaction energy (SIE) for scoring protein–ligand binding affinities. 1. Exploring the parameter space. J Chem Inform Model 2007;47(1):122–33.

[33] Ballester PJ. Machine learning approaches to predicting protein–ligand binding. In: Presentation; Cambridge Computational Biology Institute – European Molecular Biology Laboratory EMBL-EBI. 2013.

[34] Cheng T, Li X, Li Y, Liu Z, Wang R. Comparative assessment of scoring functions on a diverse test set. J Chem Inform Model 2009;49(4):1079–93.

[35] Li Y, Han L, Liu Z, Wang R. Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. J Chem Inform Model 2014;54:1717–36.

[36] Moustakas D, Lang P, Pegg S, Pettersen E, Kuntz I, Broojimans N, et al. Development and validation of a modular, extensible docking program: DOCK 5. J Comput Aided Mol Des 2006;20:601–9.

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

[38] Muegge I. PMF scoring revisited. J Med Chem 2006;49:5895–902.

[39] Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. eHiTS: a new fast, exhaustive flexible ligand docking system. J Mol Graph Model 2007;26(1):198–212. PubMed: 16860582.

[40] Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics 2010;26(9):1169–75.

[41] Breiman L. Random forests. Mach Learn 2001;45:5–32.

[42] Li H, Leung KS, Wong MH, Ballester PJ. Substituting random forest for multiple linear regression improves binding affinity prediction of scoring functions: Cyscore as a case study. BMC Bioinform 2014;15:291, http://dx.doi.org/10.1186/1471-2105-15-291 http://www.biomedcentral.com/1471-2105/15/291

[43] Cao Y, Li L. Improved protein–ligand binding affinity prediction by using a curvature-dependent surface-area model. Bioinformatics 2014;30(12):1674–80.

[44] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 2010;31(2):455–61.

[45] Ballester PJ, Schreyer A, Blundell TL. Does a more precise chemical description of protein–ligand complexes lead to more accurate prediction of binding affinity. J Chem Inform Model 2014;54:944–55.

[46] Zilian D, Sotriffer C. SFCscoreRF: a random forest-based scoring function for improved affinity prediction of protein–ligand complexes. J Chem Inform Model 2013;53(8):1923–33.

[47] Khamis MA, Gomaa W. Comparative assessment of scoring and ranking powers of machine-learning-based scoring functions on an updated benchmark PDBbind 2013; 2015. Unpublished.

[48] Gabel J, Desaphy J, Rognan D. Beware of machine learning-based scoring functions – on the danger of developing black boxes. J Chem Inform Model 2014;54:2807–15.

[49] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem 2012;55:6582–94.

[50] Khamis MA, Gomaa W, Galal B. Deep learning competes random forest in computational docking; 2015. Unpublished.

[51] Liu T, Lin Y, Wen X, Jorissen R, Gilson M. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Res 2007;35(D):198–201. PubMed: 17145705.

[52] Fogel G, Tran J, Johnson S, Hecht D. Machine learning approaches for customized docking scores: Modeling of inhibition of *Mycobacterium tuberculosis* enoyl acyl carrier protein reductase. In: Bi CC, editor. Proc. IEEE 2010 symposium on computational intelligence in bioinformatics and computational biology (CIBCB 2010). NJ, United States: IEEE at Piscataway; 2010. p. 1–6, http://dx.doi.org/10.1109/CIBCB.2010.5510700.

[53] Accelrys Inc. Accord; 2015, urlhttp://accelrys.comhttp://accelrys.com (accessed on 04.01.15).

[54] Thomsen R, Christensen M. MolDock: a new technique for high-accuracy molecular docking. J Med Chem 2006;49:3315–21.

[55] Jones G, Willett P, Glen R, Leach A, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–48.

[56] Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol 1995;245:43–53, http://dx.doi.org/10.1016/S0022-2836(95)80037-9.

[57] Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. 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 1997;11:425–45, http://dx.doi.org/10.1023/A:1007996124545.

[58] Ouyang X, Handoko SD, Kwoh CK. CScore: a simple yet effective scoring function for protein–ligand binding affinity prediction using modified CMAC learning architecture. J Bioinform Comput Biol 2011;09:1–14.

[59] Albus J. A new approach to manipulator control: the cerebellar model articulation 32 controller (CMAC). J Dyn Syst Meas Control 1975;97:220.

[60] Liu Q, Kwoh CK, Li J. Binding affinity prediction for protein–ligand complexes based on β contacts and B factor. J Chem Inform Model 2013;53(11):3076–85.

[61] Li GB, Yang LL, Wang WJ, Li LL, Yang SY. ID-Score: a new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions. J Chem Inform Model 2013;53(3):592–600.

[62] Zavodszky M, Sanschagrin P, Kuhn L, Korde R. Distilling the essential features of a protein surface for improving protein–ligand docking, scoring, and virtual screening. J Comput Aided Mol Des 2002;16:883–902.

[63] Schnecke V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. Perspect Drug Discov Des 2000;20(1):171–90.

[64] Zavodszky M, Kuhn L. Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis. Protein Sci 2005;14(4):1104–14.

[65] Milborrow S. Earth: multivariate adaptive regression spline models. Derived from Mda:mars by Trevor Hastie and R. Tibshirani. R package version 2.4-5; 2010.

[66] Schliep K, Hechenbichler K. kknn: weighted k-nearest neighbors. R package version 1.0-8; 2010.

[67] Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. e1071: miscellaneous functions of the department of statistics (e1071), TU Wien. R package version 1.5-24; 2010.

[68] Breiman L. Random forests. Mach Learn 2001;45:5–32.

[69] Ridgeway G. GBM: generalized boosted regression models. R package version 1.6-3.1; 2010.

[70] Amini A, Shrimpton PJ, Muggleton SH, Sternberg MJE. A general approach for developing system-specific functions to score protein–ligand docked complexes using support vector inductive logic programming. Proteins 2007;69:823–31, http://dx.doi.org/10.1002/prot.21782.

[71] Li L, Wang B, Meroueh SO. Support vector regression scoring of receptor-ligand complexes for rank-ordering and virtual screening of chemical libraries. Journal of Chemical Information and Modeling 2011;51(9):2132–8.

[72] Korb O, Stutzle T, Exner TE. Empirical Scoring Functions for Advanced Protein–Ligand Docking with PLANTS. Journal of Chemical Information and Modeling 2009;49:84–96.

[73] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19(14):1639–62.

[74] Khamis MA, Gomaa W, Galal B. Comparative assessment of docking and screening powers of machine-learning-based scoring functions on an updated benchmark PDBbind 2013; 2015. Unpublished.

[75] Friesner R, Banks J, Murphy R, Halgren T, Klicic J, Mainz D, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47(7):1739–49.

[76] Ballester PJ, Mangold M, Howard NI, Robinson RLM, Abell C, Blumberger J, et al. Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification. J R Soc Interface 2012;9:3196–207.

[77] Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. J Chem Inform Model 2012;52(7):1757–68.

[78] Mooij WT, Verdonk ML. General and targeted statistical potentials for protein–ligand interactions. Proteins 2005;61:272–87, http://dx.doi.org/10.1002/prot.20588.

[79] Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. J Med Chem 2006;49(23):6789–801.

[80] F_ScoreWikipedia. *F*-score; 2014 http://en.wikipedia.org/wiki/F1_score (accessed on 28.12.14).

[81] PrecisionRecallWikipedia. Precision and recall; 2014 http://en.wikipedia.org/wiki/Precision_(information_retrieval) (accessed on 28.12.14).

[82] ReceiverOperatingCharacteristicWikipedia. Receiver operating characteristic; 2014 http://en.wikipedia.org/wiki/Receiver_operating_characteristic (accessed on 28.12.14).

[83] Sato T, Honma T, Yokoyama S. Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening. J Chem Inform Model 2010;50:170–85.

[84] Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res 2009:4 (Epub ahead of print).

[85] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2012;40(Database issue):D1100–7.

[86] Inpharmatica Ltd. Starlite; 2007.

[87] MDL Information Systems, Inc. Maccs drug data report, release 2000.2. Tech. Rep. San Leandro, CA: MDL Information Systems, Inc.; 2000.

[88] Schrödinger L. The schrödinger software (version 8.0); 2005.

[89] Durrant JD, McCammon JA. NNScore 2.0: a neural-network receptor-ligand scoring function. J Chem Inform Model 2011;51(11):2897–903.

[90] Li H, Leung KS, Wong MH. idock. A multithreaded virtual screening tool for flexible ligand docking. In: Peng Y, editor. Proc. IEEE 2012 symposium on computational intelligence in bioinformatics and computational biology (CIBCB 2012). NJ, United States: IEEE at Piscataway; 2012. p. 77–84.

[91] Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 2012;4(1):1–17.

[92] Cambridge Soft Corporation. Chemoffice, cs, software package, version 8.0; 2004 http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemOffice/Default.aspx

[93] Kramer B, Rarey M, Lengauer T. Evaluation of the flexx incremental construction algorithm for protein–ligand docking. Proteins: Struct Funct Genet 1999;37:228–41.

[94] Hildebrandt A, Dehof AK, Rurainski A, Bertsch A, Schumann M, Toussaint NC, et al. BALL - biochemical algorithms library 1.3. BMC Bioinform 2010;11:531–5, http://dx.doi.org/10.1186/1471-2105-11-531 http://www.biomedcentral.com/1471-2105/11/531

[95] Hu L, Benson ML, Smith RD, Lerner MG, Carlson HA. Binding MOAD (mother of all databases). Proteins 2005;60(3):333–40, http://dx.doi.org/10.1002/prot.20512.

[96] Gorodetsky V, Samoilov V. Feature extraction for machine learning: logic-probabilistic approach. J Mach Learn Res-Proc Track 2010;10:55–65.

[97] Sayes C, Ivanov I. Comparative study of predictive computational models for nanoparticle-induced cytotoxicity. Risk Anal 2010;30(11):1723–34.