Print ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Current approaches and tools for binding energy prediction in computer-aided drug design

Sergii Vakal*

Biomolecular Modeling Group, Chair of Biochemistry, Educational and Scientific Center, Institute of Biology and Medicine, Taras Shevchenko National University of Kyiv, 2 Academica Hlushkova avenue, building 12, Kyiv, 03022, Ukraine

***Corresponding author: E-Mail: serxio88@gmail.com, Tel: +380500545859**

ABSTRACT

Computer methods can now be used on almost every stage of drug development, but the most common areas of computers application are virtual screening and lead generation/optimization stages. Accurate prediction of the protein-ligand binding affinities is a crucial step in the structure-based drug design approach. Current algorithms and tools for binding energy calculation that are used upon the development of new drug candidates with an emphasize on underlying principles, advantages and limitations, software and general considerations in the selection of specific methods are discussed in the paper.

Four main classes of currently available physics-based computer methods (molecular docking, end point / approximate free energy, relative binding free energy, and absolute binding free energy) are reviewed in details. Molecular docking approaches are the method of choice to filter out compounds-nonbinders, but they are not accurate enough to predict binding affinity. The end point methods are more physically rigorous and closer to real free energy calculations, but they are more computationally-intensive and not predictive for some types of proteins. Relative binding free energy methods take into account conformational and entropic contributions, thus offering more accurate predictions. However, they have high computational requirements and can be used only to compare related ligands or receptors. The extremely computational-dependent method of absolute binding free energy calculation is the most powerful approach, giving predictions with good correlations to experimental binding affinities.

KEY WORDS: Binding Energy, *In Silico* Tools, Computer-Aided Drug Design, Free Energy Calculation, Ligand Binding Thermodynamics.

1. INTRODUCTION

Rational structure-based computer-aided modeling of protein-ligand interactions is now a key component in modern drug discovery paradigm (Charifson,1997). It is widely accepted that computational methods have played an extremely important role in the design process for a growing number of marketed drugs, and in the development of new drug candidates (Mobley & Dill 2010). Moreover, by the aid of computer-aided drug design (CADD), the cost of drug development could be reduced by up to 50% (Tan, 2010).

Computer methods can now be used on almost every stage of drug development, but the most common areas of computers application are virtual screening and lead generation/optimization stages (Xiang, 2012). Virtual screening methods, which are designed for searching large libraries of compounds *in silico*, are widely used within the drug R&D industry and play an indispensable role in modern CADD efforts. These methods usually give a much higher hit rate than the traditional high throughput screening (HTS) (Tang, 2006) and the hits from VS appear more drug-like than the ones from HTS (Shekhar, 2008). At the same time, there is concern that VS methods may have reached a limit in effectiveness (Schneider, 2010). Current virtual screening methods are not very effective in selecting molecules that are actually active against the selected target molecule, although they are undoubtedly useful in eliminating some inactive compounds (Chodera, 2012). Limitations of the VS methods come from a variety of approximations used to allow large numbers of compounds to be screened quickly, often neglecting statistical mechanical and chemical effects for computational efficiency (Chodera, 2012), thus leading to the inaccuracies in the estimation of protein-ligand binding energy.

Lead optimization is another crucially important step (Keseru & Makara, 2006) among all of the stages of drug discovery process. From the computational side, the key step in lead optimization process is an accurate prediction of the protein-ligand binding affinities (Jorgensen, 2009), since it is currently accepted that the biological activity of a compound is closely related to the affinity of the compound to macromolecular receptor (Gohlke & Klebe, 2002). Unfortunately, available methods for binding affinity estimation do not possess enough balance between calculation efficiency and reliability, and in a typical situation the most accurate methods are the most time consuming, while the fastest algorithms usually are not very rigorous and accurate (Xiang, 2012).

In this review we are going to discuss current approaches and tools for binding energy calculation that are used upon the development of new drug candidates with an emphasize on underlying principles, advantages and limitations, software and general considerations on the selection of specific methods for different users.

Computational Approaches to Binding Energy Prediction: Currently available physics-based computer methods can be grouped in at least four different classes. Below are listed from the fastest to slowest, and from the least

physical to most physical (Fig.1) (Mobley & Dill 2010).

Fast molecular docking methods, including those implemented in common software like Auto Dock, DOCK, Glide, FlexX, GOLD etc.

Approximate free energy / end point methods, in which the motions of protein and solvent are taken into consideration with less number of approximations. Such methods as MM-PBSA and MM-GBSA belong to this group of methods.

Relative binding free energy (RBFE) approaches, which include full protein and solvent motions, but which require prior knowledge of the similar structure of a complex of the protein with a ligand ('template').

Absolute binding free energy (ABFE) methods, which are the most computationally-expensive, but which include the physics in the most rigorous way that is currently achievable (Worth, 2009). Free energy perturbation (FEP) algorithm can be classified into this group.

Below we will discuss and take a closer look at each of the above mentioned groups of methods.

Molecular Docking: Molecular docking is probably the most popular method used in a structure-based drug design. As an illustration it can be mentioned that 88% of drug discovery-related publications for 2008 year cite different docking tools (Mobley & Dill, 2010). The term 'docking' emerged in the late 1970s, and on that time it was treated as a method for refinement of complex structure models through optimization of the relative orientation of fixed binding partners (de Ruyck, 2016). At the moment, docking can be defined as a technique aimed at finding the correct conformation (so called 'pose') of a ligand and its' receptor. The idea behind this technique is to generate a comprehensive set of conformations of the receptor complex, assess them quantitatively, and then to rank them according to their stability and affinity (Lopez-vallejo, 2011).

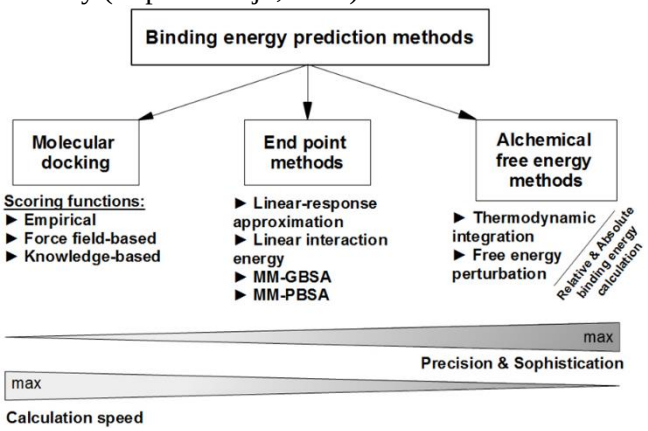

Figure.1. Schematic representation of the available methods for binding energy calculation in computer aided drug design

It should be noted that reconstruction of the conformational space available in a receptor-ligand complex is a sophisticated task, which needs some degree of approximation in order to have adequate computational requirements. On the basis of approximation level all methods of docking can be grouped in three categories (Holtje, 2008), rigid docking – the simplest approach, when both protein and its' ligand are treated as rigid bodies; semi flexible docking – protein treated as a rigid body, while ligand is conformationally flexible; flexible docking – the most complex approach, when the conformational flexibility of macromolecule (at least part of it) and ligand is taken into account.

Molecular docking programs use special scoring functions to estimate the binding energetics of the predicted ligand-receptor complexes (Ferreira, 2015). Prediction of the binding energy is performed through evaluation of the most important physical-chemical phenomena involved in protein-ligand binding, such as intermolecular interactions, desolvation and entropic effects. So, it can be said that a greater number of physical-chemical parameters corresponds to the greater accuracy of the scoring function (Jain 2006). Unfortunately, computational costs increase proportionally to the number of variables, so ideal scoring function should render a balance between accuracy and speed (Ferreira, 2015).

Docking scoring functions are usually classified in three groups: empirical, force field-based, and knowledge-based (Kitchen, 2004).

Empirical scoring functions use different dependencies of properties that are important for ligand binding (such as ionic and a polar interactions, hydrogen bonds, desolvation and entropic effects) to construct a complex equation predicting energy of protein-ligand binding. A series of protein-ligand complexes with known binding affinities is used as a training set to perform a multiple linear regression analysis. Then statistical model generates weight constants, which are used as coefficients adjusting the terms of the equation.

Limitations: high dependence on the quality of initial data used to develop the model; empirical functions are

expected to work better for proteins similar to those used in the training group.

Examples of empirical scoring functions (Ferreira 2015; Li, 2014): Auto Dock Score, Chem PLP, Chem Score, Fresno, F_Score, GlideScore-SP/XP, HYDE, Jain, LigScore2, London-dG, LUDI1/LUDI2/LUDI3, SFC score, PLP, X_Score. Common software using empirical scoring functions: Auto Dock, FlexX, Glide, Surflex etc.

Force field-based scoring functions estimate the binding energy by summing the contributions of non-bonded (van der Waals and electrostatic interactions described by Lennard-Jones potential and Coulomb law, respectively) and bonded (angle bending, bond stretching, dihedral variation) terms in a complex equation. The classical mechanics equations are used to calculate the energy related to each of the above mentioned terms.

Limitations: entropic contribution to the binding energy is not taken into consideration or assessed inaccurately, since there is no appropriate physical model to describe it. Moreover, there is a problem with desolvation energy estimation because the solvent is not explicitly considered. Large and polar molecules usually get the best score for enthalpic interaction (Holtje, 2008).

Examples of force field-based scoring functions (Ferreira, 2015; Li, 2014): Auto Dock, DOCK, Gold Score, ICM, Ligand Fit, Medusa Score, Molegro function, SYBYL G- and D-Score. Common software using empirical scoring functions: Auto Dock, DOCK, Gold, Molegro Virtual Docker etc.

Knowledge-based scoring functions use pairwise energy potentials of known ligand-receptor complexes to obtain a master equation. These potentials are formed by taking into account the frequency with which two different atoms are found within a given distance in the structural dataset. The different types of interactions observed in the dataset are classified and weighted according to their frequency of occurrence. The final score is given as a sum of these individual interactions. Solvation and entropic contributions are considered in implicit form. This type of scoring functions offers a good balance between accuracy and speed (Gohlke, 2002; Hendlich, Klebe 2000).

Examples of knowledge-based scoring functions (Ferreira, 2015; Li, 2014): Drug Score, Motif Score, PMF_Score, PESD_SVM, Pose Score, RF_Score, SMoG. Common software using knowledge-based scoring functions: SYBYL.

Due to limitations in each group of scoring functions, the simultaneous use of different scoring methodologies has been increasingly employed as a way to obtain a consensus scoring (Charifson, 1999). This approach is very promising, since such scoring functions combine the advantages and have less limitations of each specific method. Although there is evidence that consensus scoring is beneficial (Oda, 2006), in the case of solvated and highly flexible macromolecules such functions may not improve accuracy (Englebienne and Moitessier 2009). Examples of consensus scoring functions (Ferreira, 2015). CONSENSUS-DOCK, GFscore, MultiScore, SeleX-CS, SCS etc. Common software using consensus scoring functions: ConsDock, VoteDock etc.

Comparative assessment of scoring functions: Due to the popularity of molecular docking methods, many comparative assessments of scoring methods have been carried out in the past years (Li 2006; Jaeger, 2005; Humblet 2009; Skolnick, 2008; Rognan 2004). The most frequently tested software has been DOCK, GLIDE, GOLD, FLEXX, and ICM. The most recent and comprehensive assessment of the scoring functions implemented in common molecular docking software was performed in 2013 by Li, within the frame of so called 'critical assessment of scoring functions' (CASF) project (Li, 2014). This project was initiated in 2007, when 16 popular scoring functions were tested on 195 diverse protein−ligand complexes from the PDB bind database. Lately, in the CASF-2013, 20 scoring functions were tested, including LigScore2, PLP1/PLP2, PMF, Jain, LUDI1/2/3, GoldScore, ChemScore, GlideScore-SP/XP, ChemPLP, ASP, G-Score, D-Score, ChemScore, ΔSAS etc. (Li, 2014). These scoring functions were tested in four aspects: scoring power (the ability to produce binding scores in a linear correlation with experimental binding data), ranking power (the ability to correctly rank the known ligands of the same target protein by their binding affinities when the precise binding poses of these ligands are given), docking power (the ability to identify the native binding pose among computer- generated decoys), and screening power (the ability to identify the true binders to a given target protein among a pool of random molecules) (Li, 2014).

Briefly, it was shown that scoring functions perform well upon the prediction of native binding pose (two-thirds of the scoring functions were able to produce success rates over 60%, and few of them – around 80%), but have problems with ranking (35-58% success rates) and binding energy estimation (correlation coefficients between 0.22 and 0.61) of ligands. The following scoring functions were able to rich the correlation coefficient with experimentally-determined binding energy more than 0.5: X-Score, ΔSAS, Chem Score, Chem PLP, PLP1, G-SCORE, ASP, ASE, D-SCORE, and Alpha-HB (Li, 2014).

As a conclusion, it can be said that docking algorithms are not enough accurate to predict binding affinities. Typically, they are not able to discriminate between drugs that differ by less than one order of magnitude in affinity (by < 6 kJ/mol in ΔG_{bind}) (Genheden & Ryde 2015). But because of its speed, docking approaches are the method of choice to filter out compounds that are likely nonbinders and to identify native-like poses (Mobley & Dill 2010).

Approximate Free Energy Methods: These methods are based on sampling of the end states only, that is, the complex and the free protein and ligand. These methods are therefore called end point methods (Genheden & Ryde

2015). The most simple approximate method of binding free energy estimation is the linear-response approximation (LRA), in which the electrostatic free energy change is estimated on the basis of electrostatic interaction energy between the ligand and the surroundings (protein or solvent), standard simulations of the complex and the ligand, and simulations, in which the charges of the ligand have been zeroed (Tao, 2000). This method was used in the past to assess solvation free energies, but not binding free energies, since it lacks a non-polar part (Genheden & Ryde 2015).

In the 1990s Aqvist, developed more physically robust method on the basis of the LRA, which is currently known as linear interaction energy (LIE) method (Samuelsson, 1994). The main consideration of the LIE method is that only convergent averages of the interaction energies between the ligand and its surroundings need to be evaluated to obtain an estimate of binding free energies (de Amorim, 2008). The master LIE equation is based on the electrostatic and van der Waals interaction energies for the ligand, which are calculated from the molecular dynamics (MD) or Monte Carlo (MC) simulations. Two separate MD/MC simulations need to be carried out, one with the ligand free in solution and one where it is bound to the solvated receptor. Soon after the implementation of LIE method, several linear response models for estimation of absolute free energies of binding based on the changes in electrostatic and van der Waals energies and size effects were proposed (de Amorim, 2008).

Schutz and Warshel (2001), have devised a similar algorithm called semi-macroscopic protein-dipoles Langevin-dipoles method within a LRA (PDLD/s-LRA/b), in which the polar part is taken from the LRA and the non-polar part is calculated from LIE (Tao, 2000).

The most widespread end point methods currently are molecular mechanics with Poisson-Boltzmann (or Generalized Born) and surface area salvation (MM-PBSA / MM-GBSA). These algorithms use a continuum solvent model to replace the water by treating it as a continuous medium (implicit solvent model), thus reducing computational cost of the original LIE model (de Amorim, 2008). This approach allows us to obtain the average solvation properties of water without averaging over the interactions of thousands of real water molecules, which in simulations is connected to large fluctuations in solute-solvent and solvent-solvent energies (Aqvist 2006). Ligand-solvent interaction energies can be calculated accurately through solution of the Poisson–Boltzmann equation, or approximately by using the generalized Born theory. Since the generalized Born model is less computationally-intensive, it is more popular for MD simulations (Khandogin, 2008). Fortunately, a lot of improvements have been applied to generalized Born models, and now they are capable of reaching the same level of accuracy as Poisson-Boltzmann models (Aqvist 2003). These methods have been used in a range of settings, including protein design, protein-protein interactions, conformer stability and re-scoring (Genheden & Ryde 2015).

In MM-PBSA approach, the free energy of a state, that is, P (free protein), L (free ligand) or PL (complex) is estimated from the following sum:

$$G = E_{bond} + E_{el} + E_{vdW} + G_{pol} + G_{npol} - TS,$$

Where E_{bond}, E_{el} and E_{vdW} are standard molecular mechanics energy terms from bonded, electrostatic and van der Waals interactions, G_{pol} and G_{npol} are the polar and non-polar contributions to the salvation free energies. Polar term is calculated by solving of Poisson-Boltzmann (MM-PBSA) or generalized Born equation (MM-GBSA), while nonpolar term is obtained from a linear relation to the solvent accessible surface area (Genheden & Ryde 2015). The methods involve several huge approximations, for example, a questionable entropy, lacking the conformational contribution and missing effects from binding-site water molecules. Moreover, the methods often overestimate differences between sets of ligands. However, since MM-PBSA & MM-GBSA invest more effort in sampling and entropies, they are closer to a true free energy calculation than docking (Mobley & Dill 2010).

Average free energy of unbound ligand (G_L), unbound protein (G_P) and the complex (G_{PL}) is usually estimated from the separate MD or MC simulations for each of them. This approach is called three-average MM-PBSA (3A-MM-PBSA). However, it is more common to simulate only the complex (PL) and create the ensemble average of the unbound receptor and ligand by simply removing the appropriate atoms; such approach is called one-average MM-PBSA (1A-MM-PBSA). In a typical scenario, the simulations used to estimate the energy terms employ explicit solvent models, but since implicit solvent models (GBSA/PBSA) are used, later all solvent molecules are deleted from each trajectory snapshot. It was also suggested that MM-PBSA calculations can be based only on single minimized structures instead of a large number of MD/MC-trajectory snapshots (Genheden & Ryde 2015). And in practice, minimized structures often give results comparable with those obtained with MD/MC-simulations (Shen, 2014). At the same time, the results of such calculations are strongly dependent on the starting structure and ignore the dynamic effects.

Software implementation of the MM-PBSA & MM-GBSA methods. The MM-PBSA approach was originally developed for the AMBER software, and currently is available for free in the Amber Tools. During the past decade, automatic scripts were also created for popular free simulation packages Desmond, NAMD and GROMACS, as well as for APBS software (Genheden & Ryde 2015).

Assessment of the accuracy of end point methods: LIE and MM/PBSA have been compared several times, but the

results depend on the tested system. Genheden & Ryde showed that LIE is 2-7 times more efficient than MM-PBSA, owing to the time-consuming entropy estimate (Genheden and Ryde 2011). At the same time, MM-PBSA was shown to have better overall performance than MM-GBSA (Homeyer, 2014). However, surface generalized Born model (SGB) of continuum solvent with LIE algorithm was able to reach high correlation (r^2 = 0.72-0.81) with experimental binding data (Moro 2007). It is also recommended to use Onufriev, model of GB when exactly MM-GBSA calculations should be conducted (Onufriev, 2004). Schutz and Warshel compared the PDLD/s-LRA/b, LIE and MM-PBSA methods, and found that the former was the most accurate of them (Schutz, Warshel 2001). MM-PBSA was found to be more accurate and less computationally-demanding than LRA (Genheden, Ryde 2012). Upon comparison with alchemical perturbation methods, MM-PBSA is comparable (Guimaraes 2011; Laitinen, 2004), worse (Gouda, 2003) or even better (Bea, 2001), depending on the study and the system.

Poor precision is one of the main problems of the MM-PBSA & MM-GBSA methods, thus sometimes making them useless upon comparison of ligands with similar affinities. For example, the standard deviation of ΔG_{bind} over the 20 snapshots is 47-62 kJ/mol for MM-PBSA method (Genheden & Ryde 2015). MM-PBSA has severe convergence problems, requiring many independent simulations to yield a good precision. The problem with the precision is usually solved by calculating only interaction energies, studying as many MD-snapshots as possible, and using several independent simulations (Genheden & Ryde 2015).

Generally, the accuracy of end point methods (correlation coefficients compared with experiments of r^2 = 0.0-0.9, depending on the protein) is usually better than for molecular docking, but worse than for alchemical perturbation algorithms (Genheden & Ryde 2015). According to an expert opinion of Genhenden & Ryde, end point methods (particularly, MM-PBSA) may be useful to improve the results of docking and virtual screening or to understand observed affinities and trends. However, they are not accurate enough for later states of predictive drug design.

Relative and Absolute Binding Free Energy Methods: All methods described in previous sections of the paper rely upon a variety of approximations, neglecting or considering in a simplified form statistical mechanical and chemical effects to increase calculation speed. Free energy methods offer a way to incorporate these effects to quantitatively compute accurate binding affinities (Chodera, 2012). Such algorithms as thermodynamic integration (TDI) and free energy perturbation (FEP) are usually the most accurate, but are also more time consuming than the end point or docking approaches (Homeyer, 2014). These methods are usually called alchemical methods, since instead of simulating the binding/unbinding processes directly, which would require a simulation many times the lifetime of the complex, the ligand is alchemically transmuted into either another chemical species or a noninteracting molecule through intermediate, possibly nonphysical stages (Chodera, 2012).

Alchemical free energy methods can be used to compute either relative binding affinities (a difference between two or more related ligands) or absolute binding affinities (for an individual ligand to a receptor). Free energy calculations that use straightforward MD simulations generally suffer from slow exploration along many conformational degrees of freedom, which introduces difficult sampling issues for both absolute and relative free energy calculations. Absolute free energies cover a much larger dynamic range of binding affinities, so that experimental error is a much smaller fraction of this range. Interpretation of failure is also easier, as it is clear which compounds differ from experiment; with relative free energies, it is often not clear whether the calculation for one or both compounds suffer from pathologies (Chodera, 2012).

TDI has the advantage that the precision of binding free energy predictions can be increased by subsequently including additional intermediate states (Michel & Essex, 2010), so this approach offers the opportunity to start calculations at the lower level of precision and only perform sampling where necessary. This is valuable because of the inverse relationship between accuracy and required computing time, which requires one to find an optimal balance between prediction quality and computational demand (Homeyer, 2014). TDI transformations of one ligand into another are usually conducted through simulations at discrete *n* steps. The free energy difference ΔG for the transformation is calculated by integration over the average potential functions of the two states at each n step. To determine the difference in the binding free energy $\Delta\Delta G$ between two ligands, transformations are performed for both the complex-bound ligands and the solvated ligands. $\Delta\Delta G$ is calculated as the difference between the respective free energies: $\Delta\Delta G = \Delta G_{bound} - \Delta G_{solvated}$ (Homeyer, 2014).

FEP simulations rooted in statistical mechanics provide an avenue to incorporate missing effects into the calculations, e.g., conformational sampling, explicit solvent, and the shift of protonation states upon binding, but they generally require extensive computational resources and expertise (Christ, 2010). FEP methodology has been known for more than 20 years currently, but its impact on drug discovery is being recognized (Acevedo, 2012). The main problem for implementing FEP as a routine technique in CADD is obtaining reliable ΔG estimates for complex bimolecular systems within a reasonable computational time. FEP approach uses the classic Zwanzig expression to relate the free energy difference by constructing a nonphysical path connecting the desired initial and final state of a system. For relative free energies of binding, single or double topology perturbations can be made to convert one

ligand to another (Kollman,1993). In principle, both topology methods should provide results of equal precision; however, it has been reported that the single topology approach is more efficient than the dual topology for all but very long simulations (Pearlman 1994).

Configurations of ligands are generated using MD or MC simulations with the appropriate Boltzmann weights (Acevedo, 2012). The total free energy change is computed by summing the incremental free energy changes in each FEP window (window – MD or MC simulation at one point along the mutation coordinate, which transforms two ligands). The difference in free energies of binding for the ligands X and Y then is calculated from $\Delta\Delta G_{bind} = \Delta G_X - \Delta G_Y = \Delta G_F - \Delta G_C$. Two series of mutations are performed to transform X to Y unbound in water and complexed to the protein which yields ΔG_F and ΔG_C. A successful FEP protocol in drug discovery is to replace each hydrogen of a promising scaffold with 10 small groups that have been selected for difference in size, electronic character, and hydrogen-bonding patterns (Jorgensen, 2009).

Software implementation of the RBFE & ABFE methods. All popular software packages for molecular simulations are currently equipped with the tools for alchemical free energy calculations. Among them are AMBER, CHARMM, GROMACS, Desmond, NAMD etc. Given that free energy calculations rely on sophisticated algorithms requiring a lot of scripts, configuration files and separate simulations, software for automation and simplification of such calculations is of a big need. Fortunately, there are such tools as CHARMM-GUI Ligand Binder (Jo, 2013) and FESetup (Loeffler, 2015) that provide users a suitable interface and examples of necessary files to perform convenient FEP/MD simulations, thereby permitting an accelerated throughput of RBFE/ABFEE computations while decreasing the possibility of human errors.

A major problem with free energy methods is their high computational cost. It is nearly impossible to perform such class of calculations (even for small proteins) without access to modern supercomputers. Moreover, some programs (like Desmond) even won't run on clusters without a GPU. So, researchers are strongly encouraged to get an access to modern GPU-equipped supercomputing facilities before starting the projects involving RBFE / ABFE calculations.

Assessment of the accuracy of alchemical free energy methods: These methods were treated as very promising in the late 1980s and early 1990s following their introduction, but this enthusiasm was quickly diminished when it became evident that some of the early successes were due either to luck or bias. However, during past years, numerous methodological advances have sparked a new wave of enthusiasm (Chodera, 2012).

Common practice for estimation of the performance of alchemical free energy calculations is a comparison of the predicted binding energies with experimental affinity measurements, obtained by 'wet-lab' (biophysical or enzymatic) assays (Holdgate, Anderson, Edfeldt, Geschwinder 2010). However, a lot of issues should be taken into account upon comparison of predicted energy with experimental data (Brown, 2009).

Assessments of the alchemical free energy methods have largely focused on a few model receptor systems (Chodera, 2012). FK506 binding protein 12 (FKBP12) is one of the most intensively studied models (Fujitani, Tanida, Matsuura 2009). Predicted binding affinities vary between studies by up to 2–3 kcal/mol, and even minor details such as the need for an inhomogeneous dispersion correction to account for the differing density of van der Waals sites in the protein and solvent can result in deviations of up to 1 kcal/mol (Shirts, 2007).

The serine protease trypsin has also been investigated in a number of recent RBFE and ABFE calculation studies. Predicted free energies relative to unsubstituted benzamidine ligands ranged from -2.1 to +0.17 kcal/mol, while experimentally determined free energies ranged from -0.64 to +0.91 kcal/mol (Villa, 2003). More recent small study involving AMOEBA polarizable force field revealed an improved agreement with experiment, with an average error less than 0.5 kcal/mol (Jiao, 2009).

Hydrophobic cavity mutant (L99A) of T4 lysozyme is the most popular model system for ABFE calculations in recent years that has been nontrivial for free energy methods to quantitatively predict affinity (Mobley, 2007). Current force fields allowed researchers to obtain RMSE in computed binding free energies around 1–2 kcal/mol. Calculations for a double mutant (L99A + M102Q) gave similar accuracy with RMSE of 1–2 kcal/mol (Gallicchio, 2010).

In one of the largest recent studies the efficiency of nearly-automated free energy calculations on 92 ligands binding to five different targets was assessed (Christ & Fox 2013). Agreement with experiment was found to be system-dependent ranging from excellent (RMSE around 0.9 kJ/mol) to mediocre (RMSE = 4.7-8.7 kJ/mol). When analyses were restricted to free energy calculations with sample standard deviations below 1 kJ/mol, agreement with experiment improved (RMSE = 0.8-6.9 kJ/mol).

Unfortunately, despite many studies dealing with the improvement of current methodology, we still have a very limited idea about when alchemical free energy methods can currently be expected to perform well (Chodera, 2012). For example, conformational changes slow enough to present sampling difficulties can affect calculated binding affinities to a significant degree, and it is nearly impossible to know when these issues will appear (Boyce, 2009). Even in favorable cases, care must be taken to sample all relevant ligand binding modes, as these can

sometimes change in unexpected ways upon scaffold modification during lead optimization stage (Mobley & Dill 2010). At the same time, errors can be in the 1–2 kcal/mol range in computed binding free energies (Mobley, 2007; Boyce, 2009; Gallicchio, 2010) or sometimes even better (Jiao, 2009).

As a conclusion, it can be said that alchemical free energy methods are the most physically rigorous and relatively more accurate than other methods described in previous chapters. However, high computational demands, complexity of performance and need for high expertise do not allow these methods to supply the place of molecular docking. At the same time, RBFE & ABFE methods can be utilized in lead optimization in drug discovery efforts (Zeevaart, 2008).

2. CONCLUDING REMARKS

Initially, drug discovery was carried out using trial and error experimental techniques. But recent advances in CADD allow the targeted design of drugs for a chosen proteins, shortening the development cycle of new drugs and reducing the overall expenditures on development (Lionta, 2014). Extensive involvement of modern methods not only will reduce the number of false positive compounds, it may facilitate more rapid completion of difficult drug design projects, with potentially superior molecules as an end result (Wang, 2015).

As was mentioned in the text, accurate prediction of the protein-ligand binding affinities is a crucial step in the structure-based drug design approach. Technically speaking, accurate calculation of binding affinity differences and absolute free energy of binding for CADD purposes is a feasible and realistic task. It is important to keep in mind that the free energy of binding is not driven by a single conformation, but rather by the free energy landscape, and the entropic contribution to binding thermodynamics is not observable in single bound structures (Mobley & Dill 2010). Different approaches described above take this fact into account with a various number of physical details. Molecular docking methods are usually not enough accurate to predict binding energy, but because of the high speed it is the most widely and commonly used approach. It is especially effective for filtering out compounds that are likely non-binders and for identifying native-like poses of ligands. End point methods are able to perform more extensive conformational sampling, but they are much more computationally-intensive. These methods are not suitable for virtual screening task because of low speed, and are less physically rigorous than alchemical methods to accurately predict binding energy. RBFE & ABFE prediction methods are the most sophisticated currently, since they treat fully free energies associated with conformational changes and entropies. Up to date, the most powerful and accurate approach (with good correlations to experimental binding affinities and with lowest RMSE) is the alchemical calculation of absolute binding free energy. At the same time, this approach has extremely high computational costs. For example, in order to compute the ABFE for just one ligand to the target protein can cost dozens of CPU days.

According to the literature data and own experience of the author, the best approach in CADD currently is a combination of different methods. Virtual screening of large libraries of small molecules can still be effectively performed by molecular docking only. At the same time, lead optimization task should be carried out with more physically rigorous methods, of which alchemical calculations (FEP or TDI) are the most accurate and sophisticated. Such combined approach can dramatically reduce the number of false-positives and, thus, to shorten the total number of compounds that should be tested in ‘wet-lab’ conditions to just a hundred or two.

At the same time, there are still a lot of limitations and drawbacks in each of the methods that were described in the text. First of all, none of the methods can be used in automatic mode as a ‘black box’; each of them requires some expertise, experience and manual intervention. And this is especially the case for alchemical free energy methods, which in principle cannot be run automatically, even with the most up-to-date software tools. Novel programs for automating the preparation of systems using ‘best practices’ methodology are needed (Chodera, 2012). Secondly, the most rigorous and precise methods are highly computationally-dependent and need modern supercomputers for effective work. Fortunately, it is not a problem for the most of laboratories. At the same time, high computational cost of calculations hinders the use of best practices in a high throughput manner (e.g. for virtual screening). Thirdly, there are a lot of chemical and statistical mechanical effects in the protein-ligand binding that need to be taken into account upon calculations. Among them is conformational entropy, protein flexibility, multiple conformations, change of protonation states, differences in solvation etc. (Chodera, 2012; Mobley & Dill 2010). Fourthly, there are numerous examples where sampling has lead to an improved understanding of the limits of force field accuracy. Thus, better and more precise force fields are needed. Fortunately, there are several novel polarizable force fields (AMOEBA, CHARMM) with promising initial results for proteins (Jiao, 2009). Moreover, force field parameters for all of the chemical species present in the complex (protein, ligand, cofactors, solvent) must be generated or assigned from a database, and this process is complex and time-consuming, thus limiting its adoption in pharma (Chodera, 2012).

Unfortunately, it is impossible to cover all the issues related to computer-aided prediction of binding free energy in one review. For further details one can read some excellent recent papers (Mobley & Dill 2010; Christ & Fox 2013; Xiang, 2012; Homeyer, 2014; Wang, 2015; Chodera,. 2012).

REFERENCES

Acevedo O, Identification of HIV inhibitors guided by free energy perturbation calculations, Current pharmaceutical design, 18 (9), 2012, 1199–1216.

Aqvist J, Medina C, Samuelsson J.E, A new method for predicting binding affinity in computer-aided drug design, Proteins Eng, 7, 1994, 385–391.

Bea I, Cervello E, Kollman PA, Molecular recognition by beta- cyclodextrin derivatives, FEP vs MM/ PBSA study, Comb Chem High Through Screen, 4, 2001, 605–611.

Bortolato A, Moro S, In silico binding free energy predictability by using the linear interaction energy (LIE) method, bromobenzimidazole CK2 inhibitors as a case study, J Chem Inf Model, 47 (2), 2007, 572–582.

Boyce S, Mobley D, Rocklin G, Graves A, Dill K.A, Shoichet B.K, Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site, J Mol Biol, 294, 2009, 747–763.

Brandsdal B.O, Osterberg F, Almlof M, Feierberg I, Luzhkov V.B, Aqvist J, Free energy calculations and ligand binding, Adv Protein Chem, 66, 2003, 123–158.

Brown S.P, Muchmore S.W, Hajduk P.J, Healthy skepticism, assessing realistic model performance, Drug Discovery Today, 14, 2009, 420–427.

Carlsson J, Ander M, Nervall M, Aqvist J, Continuum solvation models in the linear interaction energy method, J Chem Phys B, 110 (24), 2006, 12034–12041.

Charifson P, Practical Application of Computer Aided Drug Design, CRC Press, 1997, 564.

Charifson P.S, Corkery J.J, Murcko M.A, Walters W.P, Consensus scoring, A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins, J Med Chem, 42, 1999, 5100–5109.

Chen H, Lyne P.D, Giordanetto F, Lovell T, Li J, On evaluating molecular-docking methods for pose prediction and enrichment factors, J Chem Inf Model, 46, 2006, 401–415.

Chen J, Brooks C.L, Khandogin J, Recent advances in implicit solvent-based methods for biomolecular simulations, Curr Opin Struct Biol, 18 (2), 2008, 140–148.

Cheng T, Li X, Li Y, Liu Z, Wang R, Comparative assessment of scoring functions on a diverse test set, J Chem Inf Model, 49, 2009, 1079–1093.

Chodera J.D, Alchemical free energy methods for drug discovery, progress and chanllenges, Current Opinion in Structural Biology, 21 (2), 2012, 150–160.

Christ C & Fox T, Accuracy assessment and automation of free energy calculations for drug design, Journal of chemical information and modeling, 2013.

Christ C.D, Mark A.E, Van Gunsteren W.F, Basic Ingredients of Free Energy Calculations, A Review, J Comput Chem of computational chemistry, 31, 2010, 1569–1582.

de Amorim H.L.N, Caceres R.A & Netz P.A, Linear interaction energy (LIE) method in lead discovery and optimization, Current drug targets, 9 (12), 2008, 1100–1105.

de Ruyck J, Molecular docking as a popular tool in drug design, an in silico travel, Adv. Appl. Bioinform. Chem, 9, 2016, 1–11.

Englebienne P, Moitessier N, Docking ligands into flexible and solvated macromolecules, 4 Are popular scoring functions accurate for this class of proteins? J Chem Inf Model, 49 (6), 2009, 1568–1580.

Ferreira L, Molecular Docking and Structure-Based Drug Design Strategies, Molecules, 20, 2015, 13384-13421.

Fujitani H, Tanida Y, Matsuura A, Massively parallel computation of absolute binding free energy with well-equilibrated states, Phys Rev E, 79 (2), 2009, 21914.

Gallicchio E, Lapelosa M, Levy R.M, Binding energy distribution analysis method (bedam) for estimation of protein ligand binding affinities, J Chem Theory Comput, 6, 2010, 2961–2977.

Genheden S & Ryde U, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, Expert Opin Drug Dis, 10 (5), 2015, 449–461.

Genheden S, Ryde U, Comparison of end-point continuum-solvation methods for the calculation of protein--ligand

binding free energies, Proteins, 80, 2012, 1326–1342.

Genheden S, Ryde U, Comparison of the efficiency of the LIE and MM/ GBSA methods to calculate ligand- binding energies, J Chem Theory Comput, 7, 2011, 3768–3778.

Gohlke H & Klebe G, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, Angewandte Chemie - International Edition, 41 (15), 2002, 2644–2676.

Gohlke H, Hendlich M, Klebe G, Knowledge-based scoring function to predict protein-ligand interactions, J Mol Biol, 295, 2000, 337–356.

Gouda H, Kuntz ID, Case DA, Free energy calculations for theophylline binding to an RNA aptamer, comparison of MM-PBSA and thermodynamic integration methods, Biopol, 68, 2003, 16–34.

Guimaraes CR, A direct comparison of the MM-GB/SA scoring procedure and free-energy perturbation calculations using carbonic anhydrase as a test case, strengths and pitfalls of each approach, J Chem Theory Comput, 7, 2011, 2296–2306.

Holdgate G.A, Anderson M, Edfeldt F, Geschwinder S, Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins, Matching high information content with high throughput, J Struct Biol, 172, 2010, 142–157.

Holtje H.D, Molecular Modeling, Basic Principles and Applications, Third Ed, Wiley-VCH, 2008, 320.

Homeyer N, Binding free energy calculations for lead optimization, Assessment of their accuracy in an industrial drug design context, Journal of Chemical Theory and Computation, 10 (8), 2014, 3331–3344.

Jain A, Scoring functions for protein-ligand docking, Curr Protein Pept Sci, 7 (5), 2006, 407–420.

Jiao D, Zhang J, Duke R.E, Li G, Schnieders M.J, Ren P, Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential, J Comput Chem, 30 (11), 2009, 1701–1711.

Jo S, CHARMM-GUI ligand binder for absolute binding free energy calculations and its application, Journal of Chemical Information and Modeling, 53 (1), 2013, 267–277.

Jorgensen W.L, Effcient Drug Lead Discovery and Optimization, Accounts of Chemical Research, 42 (6), 2009, 724–733.

Kellenberger E, Rodrigo J, Muller P, Rognan D, Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy, Proteins, 57, 2004, 225–242.

Keseru G.M & Makara G.M, Hit discovery and hit-to-lead approaches, Drug Discovery Today, 11 (15-16), 2006, 741–748.

Kitchen DB, Decornez H, Furr JR, Bajorath J, Docking and scoring in virtual screening for drug discovery, methods and applications, Nat Rev Drug Discov, 3 (11), 2004, 935–949.

Kollman P.A, Free energy calculations, Applications to chemical and biochemical phenomena, Chem Rev, 93, 1993, 2395–2417.

Laitinen T, Kankare J.A, Perakyla M, Free energy simulations and MM-PBSA analyses on the affinity and specificity of steroid binding to antiestradiol antibody, Proteins, 55, 2004, 34–43.

Li Y, Comparative Assessment of Scoring Functions on an Updated Benchmark, II. Evaluation Methods and General Results, Journal of chemical information and modeling, 2014.

Lionta E, Structure-based virtual screening for drug discovery, principles, applications and recent advances, Current topics in medicinal chemistry, 14 (16), 2014, 1923–1938.

Loeffler H.H, Michel J & Woods C, FESetup, Automating Setup for Alchemical Free Energy Simulations, Journal of Chemical Information and Modeling, 55 (12), 2015, 2485–2490.

Lopez-vallejo F, Integrating virtual screening and combinatorial chemistry for accelerated drug discovery, Comb Chem High Throughput Screen, 14 (6), 2011, 475-487.

Michel J & Essex J.W, Prediction of protein-ligand binding affinity by free energy simulations, Assumptions, pitfalls and expectations, Journal of Computer-Aided Molecular Design, 24 (8), 2010, 639–658.

Mobley D.L & Dill K. A, Binding of Small-Molecule Ligands to Proteins,What You See, Is Not Always,What You

Get, Structure, 17 (4), 2010, 489–498.

Mobley D.L, Graves A.P, Chodera J.D, McReynolds A.C, Shoichet B.K, Dill K.A, Predicting absolute ligand binding free energies to a simple model site, J Mol Biol, 371 (4), 2007, 1118–1134.

Oda A, Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes, Journal of Chemical Information and Modeling, 46 (1), 2006, 380–391.

Onufriev A, Bashford D, Case D.A, Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model, Proteins, 55, 2004, 383–394.

Pearlman DA, A Comparison of Alternative Approaches to Free Energy Calculations, J Phys Chem, 98, 1994, 1487–1493.

Schneider G, Virtual screening, an endless staircase? Nature reviews, Drug discovery, 9 (4), 2010, 273–276.

Schutz C.N, Warshel A, What are the dielectric,constants, of proteins and how to validate electrostatic models, Proteins, 44, 2001, 400–417.

Sham YY, Chu ZT, Tao H, Examining methods for calculations of binding free energies, LRA, LIE, PDLD- LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease, Proteins, 39, 2000, 393–407.

Shekhar C, *In Silico* Pharmacology, Computer-Aided Methods Could Transform Drug Development, Chemistry and Biology, 15 (5), 2008, 413–414.

Shirts M.R, Mobley D.L, Chodera J.D, Pande V.S, Accurate and efficient corrections for missing dispersion interactions in molecular simulations, J Phys Chem B, 111, 2007, 13052–13063.

Sun H, Li Y, Shen M, Assessing the performance of the MM/PBSA and MM/GBSA methods, Improved docking performance using high solute dielectric constant MM/GBSA and MM/ PBSA rescoring, Phys Chem Chem Phys, 16, 2014, 22035–22045.

Tan J.J, Therapeutic strategies underpinning the development of novel techniques for the treatment of HIV infection, Drug Discovery Today, 15 (5-6), 2010, 186–197.

Tang Y, New technologies in computer-aided drug design, Toward target identification and new chemical entity discovery, Drug Discovery Today, Technologies, 3 (3), 2006, 307–313.

Villa A, Zangi R, Pieffet G, Mark A.E, Sampling and convergence in free energy calculations of protein-ligand interactions, the binding of triphenoxypyridine derivatives to factor xa and trypsin, J Comput Aided Mol Des, 23, 2003, 673–686.

Wang L, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field, Journal of the American Chemical Society, 137 (7), 2015, 2695–2703.

Worth C.L, Kleinau G & Krause G, Comparative Sequence and Structural Analyses of G-Protein-Coupled Receptor Crystal Structures and Implications for Molecular Models, PLoS One, 4 (9), 2009, e7011.

Xiang M, Computer-Aided Drug Design, Lead Discovery and Optimization, Combinatorial Chemistry & High Throughput Screening, 15 (4), 2012, 328–337.

Zeevaart J.G, Wang L, Thakur V.V, Leung C.S, Tirado-Rives J, Bailey C.M, Domaoal R.A, Anderson K.S, Jorgensen W.L, Optimization of azoles as Anti-Human immunodeficiency virus agents guided by Free-Energy calculations, J Am Chem Soc, 130 (29), 2008, 9492–9499.