Cell Biochem Biophys (2017) 75:15–23
DOI 10.1007/s12013-016-0769-y

REVIEW PAPER

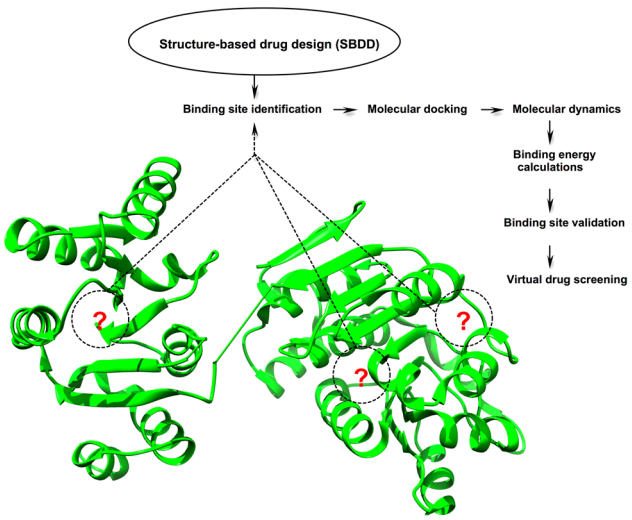
Can We Rely on Computational Predictions To Correctly Identify Ligand Binding Sites on Novel Protein Drug Targets? Assessment of Binding Site Prediction Methods and a Protocol for Validation of Predicted Binding Sites

Neal K. Broomhead[1] · **Mahmoud E. Soliman**[1]

Received: 18 April 2016 / Accepted: 19 October 2016 / Published online: 31 October 2016

Abstract In the field of medicinal chemistry there is increasing focus on identifying key proteins whose biochemical functions can firmly be linked to serious diseases. Such proteins become targets for drug or inhibitor molecules that could treat or halt the disease through therapeutic action or by blocking the protein function respectively. The protein must be targeted at the relevant biologically active site for drug or inhibitor binding to be effective. As insufficient experimental data is available to confirm the biologically active binding site for novel protein targets, researchers often rely on computational prediction methods to identify binding sites. Presented herein is a short review on structure-based computational methods that (i) predict putative binding sites and (ii) assess the druggability of predicted binding sites on protein targets. This review briefly covers the principles upon which these methods are based, where they can be accessed and their reliability in identifying the correct binding site on a protein target. Based on this review, we believe that these methods are useful in predicting putative binding sites, but as they do not account for the dynamic nature of protein–ligand binding interactions, they cannot definitively identify the correct site from a ranked list of putative sites. To overcome this shortcoming, we strongly recommend using molecular docking to predict the most likely protein–ligand binding site(s) and mode(s), followed by molecular dynamics simulations and binding thermodynamics calculations to validate the docking results. This protocol provides a valuable platform for experimental and computational efforts to design novel drugs and inhibitors that target disease-related proteins.

Graphical Abstract



✉ Mahmoud E. Soliman
soliman@ukzn.ac.za

[1] Molecular Modelling & Drug Design Research Group, School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

Keywords Protein–ligand binding · Druggability · Binding site identification · Molecular docking · Molecular dynamics simulations · Binding free energy and entropy

Abbreviations

PDB	Protein data bank
SBDD	Structure-based drug design
MD	Molecular dynamics
3D	Three-dimensional
MM-PBSA/GBSA	Molecular Mechanics-Poisson Boltzmann Surface Area/Generalised Born Surface Area
HIV	Human immunodeficiency virus
LIE	Linear interaction energy
FEP	Free energy perturbation
TI	Thermodynamic integration
ns	Nanoseconds

Introduction

The binding of molecules to various binding sites on the surface of a protein represents the biochemical functions of the protein. If a particular biochemical function and an associated binding site on the protein surface can be related to a particular disease, such as cancer, then that binding site can be targeted to treat the disease. It can be targeted with small drug-like molecules that treat or halt the disease through therapeutic action or by blocking the disease-related protein biochemical function [1]. If the drug-like molecule simply blocks the biologically active site, it is considered an inhibitor, whereas if it modulates a therapeutic action on the protein, it is considered a drug. If the drug binds with a high affinity at the binding site, and the drug delivers an effective therapeutic action, the binding site is considered as druggable [2]. In this review, we define small drug-like molecules, either drugs or inhibitors, as per the Lipinski "rule of five" [3]. They are organic molecules with molecular weight less than 500 atomic mass units and are herein referred to as ligands. The binding pockets, also referred to as binding sites, are defined as the sites on the protein structure where the ligands bind. Binding pockets are typically surface concavities of proteins. Pockets where small drug-like ligands bind are typically located in deep cavities. Studies have shown that the ligand binding sites are found in large and deep pockets on the protein surface, but there are reported cases of ligands binding to exposed shallow clefts [4].

Challenges in Binding Site Identification Using Bioinformatics

Although the number of 3D crystal structures appearing on the Protein Data Bank (PDB) [5] is growing rapidly, a lot of experimental protein–ligand binding data for novel protein targets and novel drugs/inhibitors remains unavailable. In many cases, either the protein structure is available or the protein–ligand complex is available, but it is not the ligand of interest. To address this challenge, a structure-based drug design (SBDD) approach is adopted. Identification, description and understanding of potential binding pockets on the 3D structure of the protein is a critical starting point and if, performed accurately, can successfully pave the way for development of novel drugs or inhibitors. It is vital that binding site prediction methods, used in predicting the most likely ligand-binding site, need to be accurate to ensure the reliability of results obtained from further computational efforts concerned with predicted sites. If the real binding site is not identified initially, then any efforts to validate it, using techniques like molecular docking and molecular dynamics that calculate the binding energy, would be rendered futile.

Computational Methods Used for Identification of Binding Sites

There are many computational methods available that can be used to identify protein–ligand binding sites. The methods are generally classified into two groups, either sequence-based or structure-based methods.

Sequence-Based Binding Site Identification Methods

Sequence-based prediction methods operate under the assumption that the protein residues involved in ligand binding are functionally important and are conserved [6]. The protein sequence is scanned and conserved residues are identified as potential binding sites. The fundamental limitation of sequence-based methods is that sequence conservation alone is not a specific criterion to identify binding residues, as many non-binding residues can have a high degree of conservation. Another large limitation of these methods is they do not account for specific structural and physicochemical attributes of the binding sites [7].

Structure-Based Binding Site Identification Methods

Structure-based binding site identification methods analyse three-dimensional (3D) structural models of proteins. These methods are considered important as structure plays a major role in predicting protein–ligand binding sites. The most

common are classified into two groups: template-based and pocket-based methods. Some less common structure-based methods that have been reported include a de-solvation based free energy model [8], solvent mapping [9], molecular docking [10, 11], machine learning [12, 13] and molecular dynamics (MD) [14, 15] methods.

Template-based methods

Template-based methods infer known binding sites from known template protein structures, which have similar structure to the query protein, to predict the query protein binding sites [16, 17]. These methods are limited to query proteins that have a structural similarity to the template protein.

Pocket-based methods

Pocket-based methods are able to find binding pockets by searching for surface cavities on the 3D structure of the query protein. They use predictors to assess the protein surface for cavities that facilitate ligand binding. Predictors vary according to the principles upon which they are based with the most common classified as geometry-based, energy-based or physicochemical-based. Physicochemical predictors considered important include pocket size [18, 19], pocket shape [19] and pocket residue hydrophobicity [18–21]. Pocket-based consensus methods consist of any combination of geometry-based, energy-based or physicochemical predictors. An advantage of consensus methods is that a combination of these different predictors is likely to show a better performance in accurately predicting binding sites than if individual predictors are used [18]. Various reports detail the principles upon which geometry-based, energy-based and consensus methods are established [4, 7, 22–25]. These pocket-based methods take a protein 3D structure and apply the predictor (s) to identify potential binding pockets. Each method uses their particular scoring algorithm to score and rank the identified binding pockets and finally reports a ranked list of the top predicted binding sites and their residues (Fig. 1). These methods are available on web-based servers or as stand-alone software packages. Table 1 provides a list of some commonly used pocket-based prediction methods, which are available on the internet.

Structure-Based Methods to Assess Druggability of Protein Targets

Over the last decade, structure-based methods that aim at assessing the druggability of a protein target have emerged. A protein-binding site is defined as druggable if it has the ability to bind a drug molecule with high affinity [2].

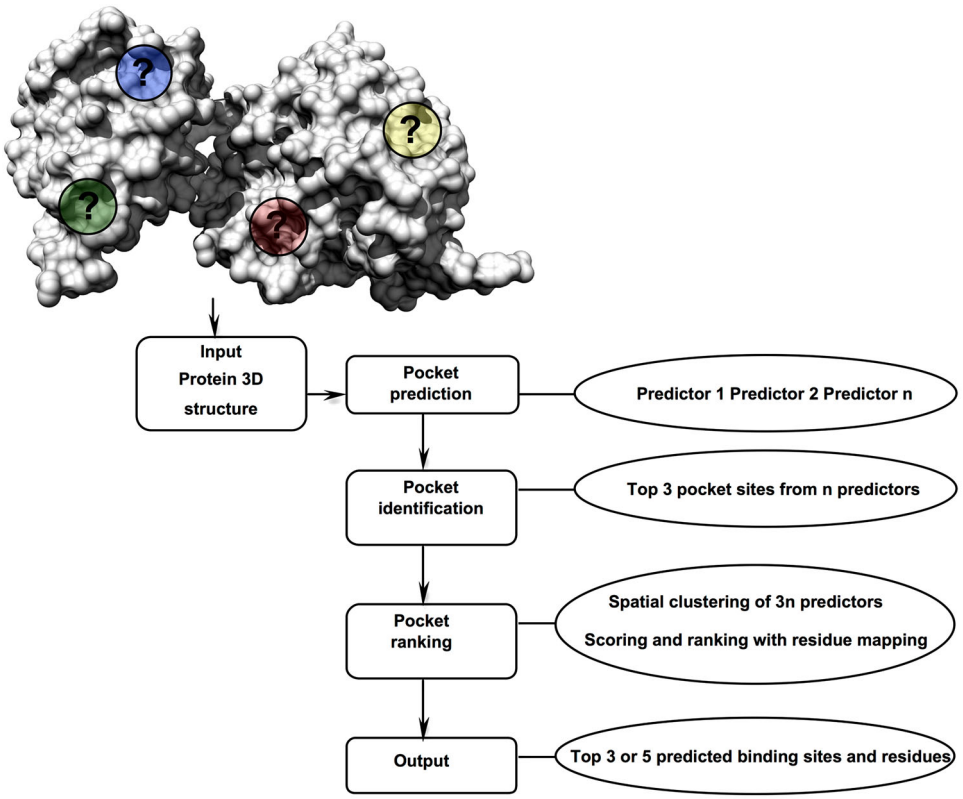


Fig. 1 Illustration of the procedure for structure-based pocket binding site prediction methods Based on the MetaPocket 2.0 method [22]. For a single predictor method, $n = 1$. For the Metapocket 2.0 consensus method which has 8 predictors, $n = 8$. Molecular graphics image was produced using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081) [26]

Table 1 Examples of commonly used structure-based pocket prediction methods available on the internet

Predictor type	Method	Availability
Geometry-based	KV Finder [27]	http://inbio.cnpem.br/bioinformatics/main/software
	MSPocket [28]	http://appserver.biotec.tu-dresden.de/mspocket
Energy-based	Sitehound [29]	http://sitehound.sanchezlab.org/download.html
	QSiteFinder [30]	http://www.bioinformatics.leeds.ac.uk/qsitefinder
Consensus	Metapocket [22]	http://projects.biotec.tu-dresden.de/metapocket
Hydrophobicity	Fuzzy oil drop [20]	http://bioinformatics.cm-uj.krakow.pl/activesite

Binding of the drug to the protein target must allow modulation of the drug in vivo and achieve the therapeutic action of the drug. Structure-based methods for assessing protein target druggability, like structure-based ligand binding site prediction methods, predict binding pockets using either geometry-based or energy-based predictors. They then analyse the predicted binding pockets in order to assess whether the physicochemical properties of the pockets are complementary to the properties of drug-like molecules. A combination of physicochemical descriptors is employed; for example the binding pocket dimensions [19], surface hydrophobicity [19] and surface polarity [31]. Finally, a druggability scoring function is applied based on the pocket predictor and physicochemical descriptor results. The druggability scoring function is based on a mathematical function or algorithm; for example linear regression, random forest classification or machine learning. The authors would like to refer the readers to recent reviews that discuss in detail this area of structure-based druggability assessment [19, 25, 32, 33]. In one of these reviews, Fauman et al. stated that these druggability assessment methods are accepted as standard procedure in target selection [19]. Table 2 lists some examples of structure-based druggability assessment methods that are currently available.

Table 2 Currently available structure-based druggability assessment methods

Druggability method	Availability
PockDrug-Server [34]	http://pockdrug.rpbs.univ-paris-diderot.fr
DoGSiteScorer [35]	http://dogsite.zbh.uni-hamburg.de/
DrugEBIlity [36]	https://www.ebi.ac.uk/chembl/drugebility/
SiteMap [37]	http://www.schrodinger.com/
DrugPred [38]	Contact authors
DLID [39]	Contact authors
SCREEN [40]	http://interface.bioc.columbia.edu/screen
MAP_{POD} [41]	Contact authors

Are the Current Binding Site Prediction Methods and Druggability Assessment Methods Reliable?

There is often not enough or no experimental data available on the Protein Data Bank (PDB) for a chosen protein–ligand complex especially as new proteins and ligands fall into the spotlight of research. Researchers must therefore rely on using structure-based protein–ligand binding site prediction methods (Table 1) and druggability assessment methods (Table 2) to identify potential binding sites on the target protein and to assess the druggability of the protein. How reliable are these methods? As the various pocket prediction methods are classified according to different predictors or parameters, each may introduce bias to the prediction result. A geometry-based method may find the most favourable pocket for the binding of a ligand purely on the geometric features of the pocket. On the other hand, an energy-based method, after calculating the Van der Waals interaction energies of the pocket residues, may find another protein pocket to be more favourable. To this end, we recommend that a consensus-based method such as MetaPocket be used. MetaPocket 2.0 [22], the latest version, combines eight different geometry-based and energy-based predictors. The limitation of Metapocket, however, is that it is predominantly geometric-based and it only recognizes pockets if they are surface concavities on the protein. If the real binding site is too small, too flat or at the interface of two protein domains, MetaPocket will not identify it [22]. Thus, in addition to MetaPocket, researchers could also use methods that predict with physicochemical predictors such as residue hydrophobicity. Although most of the available binding site prediction methods have been "validated" against training sets of proteins with known ligand-binding, a comparative study of ten representative prediction methods developed over the last 15 years, suggests that, for the top three performing methods, approximately 30 % of known biologically active binding sites were not detected [42]. In an another study, Krivak et al. evaluated the accuracy of binding site prediction by the ability of such methods to predict the true (experimentally confirmed) binding site [4]. It was concluded that scoring and ranking of the predicted pockets is a fundamental contributor to the overall accuracy of the method. They added that the ranked list of putative binding sites would contain false positive binding sites. Consequently, the method must ensure that

the order of the ranking is correct otherwise the method could be inaccurate [4]. The structure-based druggability assessment methods that have become available in recent years (see Table 2) are a good option to use as they not only have energy-based, geometry-based and physicochemical predictors built in but they also incorporate sophisticated scoring algorithms. Although these druggability assessment methods are designed to assess the druggability of a protein pocket they may equally be used to assess the ligandability of the pocket. Druggability indicates that the ligand is a therapeutic drug and ligandability indicates the ligand is an inhibitor molecule. Fauman et al. stated, however, that druggability models do not adequately account for protein–ligand binding interactions being dynamic [19]. This evidence suggests that current structure-based protein–ligand binding site prediction methods and druggability assessment methods are useful in identifying putative binding sites or druggable sites respectively, but they have limitations that do not allow them to confirm the true site in a ranked list of putative sites. Moreover, they do not take into account the dynamic nature and the chemical environment of real in vivo protein–ligand binding.

Molecular Docking as an Alternative Stand-Alone Approach to Prediction of Protein–Ligand Binding

Molecular docking programs allow fast and computationally affordable prediction of the binding modes and binding affinities of ligands at putative binding sites on the protein. Currently there are, however, significant challenges in docking that affect accuracy of the docking results. Major challenges are (a) the sampling approaches and (b) the scoring functions used by the different docking programs available. Protein and ligand flexibility allows for so many different binding conformations that exhaustive sampling is not possible. This is especially evident with rigid docking programs. Although flexible docking is widely adopted, it remains computationally expensive. Elokely et al. compared various docking programs that use a variety of approaches to the protein flexibility question and found that the docking accuracy varied in the range of 1 to 84 % [43]. The scoring function estimates the binding affinity of the ligand to the protein binding site and is comparable to an estimation of the free energy of binding. Current scoring algorithms do not adequately account for the entropic contributions and the contributions of structural water and surrounding ions to the free energy of binding [44–46]. Although current docking scoring functions are not considered able to accurately determine the absolute free energy of binding, they may be used to estimate relative binding affinities of ligands.

Molecular Dynamics Simulations and Binding Thermodynamics Calculations are Important tools to Validate Protein–Ligand Binding Site Predictions

Current binding site prediction methods and druggability assessment methods do not take into account the dynamics, flexibility and chemical environment of real in vivo protein–ligand binding. Although most current molecular docking programs can quickly and cheaply predict protein–ligand binding modes and affinities, there are still significant challenges in the accuracy of the docking predictions (see previous chapter). Furthermore, we have observed in many instances, that when the docking predictions are followed up by short molecular dynamics (MD) simulations, the ligand "flies off" from the predicted binding site(s). This ligand "fly off" event has been reported in literature and indicates unstable docking poses [47].

MD simulations allow the dynamics, flexibility and timescale of protein–ligand binding interactions to be examined at the atomistic level under conditions of temperature, pressure, solvation and pH that simulate real conditions in vivo. Conventional molecular dynamics (cMD) allows simulation in the nanosecond timescale. The dynamics of protein–ligand binding often lends itself to the possibility of conformational changes, especially in the protein, that occur on the millisecond timescale. In such cases it is necessary to use an advanced sampling technique such as accelerated molecular dynamics (aMD) to ensure that all conformational states are sampled [48–50]. Thermodynamic calculations obtained from the MD trajectories allow for measurement of the protein–ligand binding energetics which predict the protein–ligand binding affinity and the stability of the protein–ligand complex. The thermodynamics calculations should include binding free energy and entropy (Fig. 2). Methods for calculation of free energy are numerous but we will mention two broad categories. Alchemical transformation methods such as free energy perturbation (FEP) and thermodynamic integration (TI) [51] are theoretically rigorous, accurate and reliable for calculation of absolute free energy [52, 53], although computationally expensive.

In comparison end-point methods such as molecular mechanics-Poisson Boltzmann surface area/generalised Born surface area (MM-PBSA/GBSA) [54, 55] and linear interaction energy (LIE) [56] although less vigorous, are more computationally affordable and are acceptable for calculation of the free energy of protein–ligand binding [57].

The MM-PBSA/GBSA approach to free energy calculations is arguably the most popularly used method even though it is comparable to the LIE method [58]. MM-PBSA/GBSA is widely employed as a standard

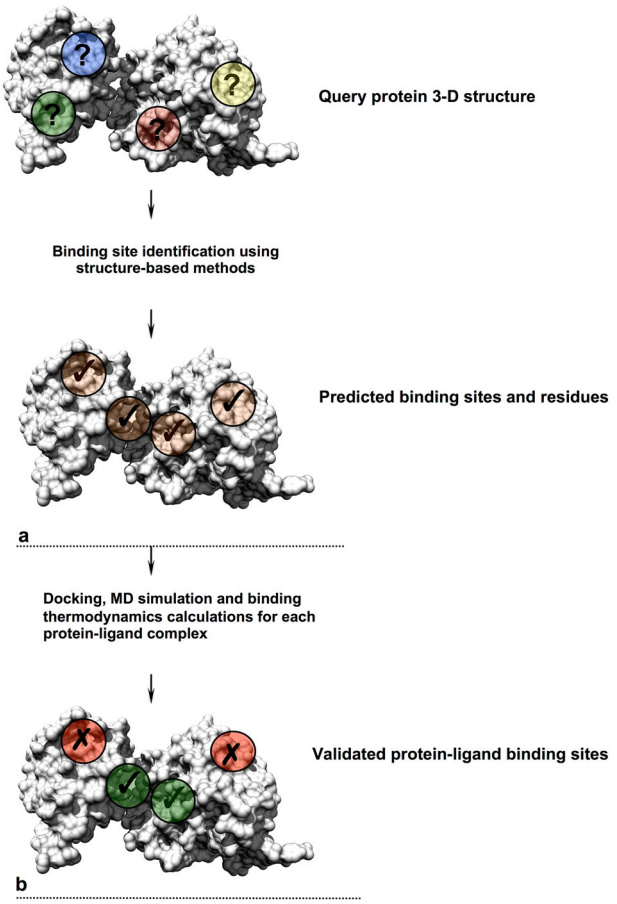


Fig. 2 Schematic representation of the recommended protocol for binding site identification and validation. **a** The 3-D structure of the query protein is inputted into a structure-based pocket binding site prediction method or structure-based druggability assessment method. The method reports its binding site predictions by listing the residues for each identified binding site. The identified binding sites are depicted on the protein structure by ✓. **b** The drug or inhibitor molecule (ligand) is docked separately into each of the predicted binding sites. The molecular docking method estimates the binding affinity and binding mode of each protein–ligand complex. The complexes are subject to validation using MD simulation and binding thermodynamics calculations (binding free energy and entropy). The results show which of the protein–ligand binding sites are valid ✓ or invalid ✗. Molecular graphics images were produced using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081) [26]

computational method in a range of ligand–receptor studies [59–61]. The results are sensitive to certain parameters in the method such as dielectric constant [62] and the continuum solvation method [63]. Typically this approach suffers from poor precision, making it important to run multiple independent simulations and to take a large number of snapshots in order to achieve statistically valid results [58, 64]. In addition the entropy calculation methods utilise significant approximations that affect the precision of the results [62, 65]. These limitations may affect the ability of MM-PBSA/GBSA based methods to reliably predict absolute free energies but they are effective for comparison of relative free energies (applies mostly to structurally similar ligands) although, if the calculated binding free energy difference between ligands is <12 kJ/mol the comparisons may be unreliable [61]. For structurally diverse ligands the comparisons are less reliable [62]. Recently it has been reported that the MM-PBSA/GBSA approach is unable to successfully predict the binding free energies of complexes with highly charged ligands [66, 67]. As the MM-GBSA method is an approximation of the original MM-PBSA method, it is computationally cheaper and faster. Srivastrava et al. reported that MM-PBSA/GBSA binding free energies correlated well with experimental half maximal inhibitory concentrations (IC_{50}) for a series of HIV protease inhibitors. This correlation was achieved in a shorter MD simulation time for MM-GBSA (5ns) than MM-PBSA (10ns) [68]. In an extensive study, involving more than 1800 protein–ligand crystal structures, Sun et al. reported that MM-PBSA prediction accuracies are more sensitive to variations in the molecular system than MM-GBSA. They suggest that MM-PBSA is suited to the ranking of binding free energy of ligands within an individual protein family, with MM-GBSA better suited to comparisons between different protein families [67]. In a large study, involving 59 ligands and 6 proteins, Hou et al. reported that MM-PBSA is more accurate in predicting absolute binding free energies whilst MM-GBSA is more accurate for predicting relative binding free energies for systems without metals [62]. Of the six energy terms used in MM-PBSA/GBSA based methods to calculate the free energy, the first three are for the standard molecular mechanics (MM) interaction energy between the protein and ligand, the fourth and fifth for the solvation free energy of the protein–ligand complex and the sixth for the entropy of binding [58]. Based on the literature references used in this review the free energy is predominantly referred to as binding free energy but in some instances it is referred to as interaction energy [59]. If the term "interaction energy" is used it appears to represent the protein–ligand interaction only (the first three energy terms). In our view it is more appropriate to use the term "binding free energy" as it represents the entire process of protein–ligand binding i.e., interaction between protein and ligand followed by formation of the bound complex (all six energy terms). Recently our research group has reported using conventional molecular dynamics (cMD) simulations and post-cMD thermodynamics calculations to extensively evaluate putative binding sites that were predicted by a structure-based consensus method. Chetty et al. used cMD simulations and MM-GBSA free energy and entropy calculations to compare the binding modes and binding thermodynamics of two predicted putative sites with a series of

ligand analogues [69]. The MD simulations revealed differences in the ligand flexibility within the binding sites, the hydrophilic or hydrophobic nature of the sites and in the forces of interaction between the ligand atoms and the binding site residues. The calculated binding free energies were not significantly different which reflected what was seen in the docking results. The entropic contribution to the binding revealed a significant difference in the conformational change upon binding between the two sites, possibly indicating that one site was larger than the other [69]. This in-depth knowledge of the binding mode, affinity and entropy offers further value to computational protein–ligand binding site prediction.

Conclusions

Protein–ligand binding site identification is a critical requirement to any structure-based drug design. Which would be the most effective (SBDD) protocol aimed at targeting novel proteins with novel drugs or inhibitors? We strongly recommend the following protocol:

(i) identification of the putative ligand binding site(s) on the protein using a consensus binding site prediction method or a druggability assessment method (ii) use of molecular docking to predict the most likely binding site(s) and mode(s) by estimation of the relative binding affinities and (iii) validation of the docking results by running molecular dynamics (MD) simulations and performing binding thermodynamics calculations (obtained from the MD trajectories) which should include binding free energy and entropy (Fig. 2). Critically the binding free energy and entropy calculations could distinguish which ligand-binding site is most energetically stable and therefore most favourable. In addition, MD simulation and post-MD analysis offer valuable knowledge about the binding site landscapes and the possible ligand binding modes. If the binding conformational changes are known to occur on a timescale beyond the limits of conventional molecular dynamics (cMD) then accelerated molecular dynamics (aMD) simulations should be run instead. We suggest using the MM-PBSA/GBSA approach to calculate binding free energy and entropy as it is computationally affordable and has performance that is accepted as being strong enough for validation of docking predictions of relative binding affinities. The method performance is better for ligands that are (i) structurally similar (ii) have binding affinities that are not <12 kJ/mol and (iii) not highly charged. The performance of MM-PBSA vs. MM-GBSA varies depending on the systems studied but MM-GBSA has the advantage of being faster and computationally less expensive. If the user requirement is for accurate determination of absolute binding free energies then alchemical transformation methods such as FEP and TI should be used. We intend to follow up this review with a research article that utilizes this protocol in a real case study. This protocol provides a valuable platform for further experimental and computational efforts to design novel drugs and inhibitors that effectively target disease-related proteins and that are aimed at arresting or eliminating disease.

Acknowledgments The authors would like to acknowledge The School of Health Sciences, University of KwaZulu-Natal, Westville campus for providing the infrastructure for this work. Neal K. Broomhead would like to acknowledge the National Research Foundation of South Africa for financial support.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no financial or intellectual conflicts of interest.

References

1 Zheng, X., Gan, L., Wang, E., & Wang, J. (2012). Pocket-based drug design: exploring pocket space. *The AAPS Journal*, *15*(1), 228–241.
2 Hopkins, A. L., & Groom, C. R. (2002). The druggable genome. *Nature Reviews Drug Discovery*, *1*, 727–730.
3 Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *23*, 3–25.
4 Krivák, R., & Hoksza, D. (2015). Improving protein-ligand binding site prediction accuracy by classification of inner pocket points using local features. *Journal of Cheminformatics*, *7*, 12.
5 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Databank. *Nucleic Acids Research*, *28*, 235–242.
6 Capra, J. A., & Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics (Oxford, England)*, *23*(15), 1875–1882.
7 Ghersi, D., & Sanchez, R. (2011). Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures. *Journal of Structural and Functional Genomics*, *12*, 109–117.
8 Coleman, R. G., Salzberg, A. C., & Cheng, A. C. (2006). Structure-based identification of small molecule binding sites using a free energy model. *Journal of Chemical Information and Modeling*, *46*(6), 2631–2637.
9 Landon, M. R., Lancia, D. R., Yu, J., Thiel, S. C., & Vajda, S. (2007). Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *Journal of Medicinal Chemistry*, *50*(6), 1231–1240.
10 Fukunishi, Y., & Nakamura, H. (2011). Prediction of ligand-binding sites of proteins by molecular docking calculation for a random ligand library. *Protein Science*, *20*(1), 95–106.
11 Heo, L., Shin, W. H., Lee, M. S., & Seok, C. (2014). GalaxySite: ligand-binding-site prediction by using molecular docking. *Nucleic Acids Research*, *42*, W210–W214.
12 Komiyama, Y., Banno, M., Ueki, K., Saad, G., & Shimizu, K. (2016). Data and text mining. Automatic generation of bioinformatics tools for predicting protein—ligand binding sites. *Bioinformatics (Oxford, England)*, *32*(6), 901–907.

13 Wang, K., Gao, J., Shen, S., Tuszynski, J. A., Ruan, J., & Hu, G. (2013). An accurate method for prediction of protein-ligand binding site on protein surface using SVM and statistical depth function. *BioMed Research International*, *2013*, 409658.
14 Ivetac, A., & McCammon, J. A. (2012). A molecular dynamics ensemble-based approach for the mapping of druggable binding sites. *Methods in Molecular Biology*, *819*, 3–12.
15 Seco, J., Luque, F. J., & Barril, X. (2009). Binding site detection and druggability index from first principles. *Journal of Medicinal Chemistry*, *52*(8), 2363–2371.
16 Skolnick, J., & Brylinski, M. (2009). FINDSITE: a combined evolution/structure-based approach to protein function prediction. *Briefings in Bioinformatics*, *10*(4), 378–391.
17 Wass, M. N., Kelley, L. A., & Sternberg, M. J. E. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, *38*(Suppl 2), W469–W473.
18 Gao, J., Liu, Q., Kang, H., Cao, Z., & Zhu, R. (2012). Comparison of different ranking methods in protein-ligand binding site prediction. *International Journal of Molecular Sciences*, *13*(7), 8752–8761.
19 Fauman, E. B., Rai, B. K., & Huang, E. S. (2011). Structure-based druggability assessment-identifying suitable targets for small molecule therapeutics. *Current Opinion in Chemical Biology*, *15*, 463–468.
20 Bryliński, M., Prymula, K., Jurkowski, W., Kochańczyk, M., Stawowczyk, E., Konieczny, L., & Roterman, I. (2007). Prediction of functional sites based on the fuzzy oil drop model. *PLoS Computational Biology*, *3*(5), 909–923.
21 Oda, A. (2011). Development and validation of programs for ligand-binding-pocket search. *Yakugaku Zasshi*, *131*(10), 1429–1435.
22 Zhang, Z., Li, Y., Lin, B., Schroeder, M., & Huang, B. (2011). Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics*, *27*(15), 2083–2088.
23 Henrich, S., Salo-Ahen, O. M. H., Huang, B., Rippmann, F., Cruciani, G., & Wade, R. C. (2010). Computational approaches to identifying and characterizing protein binding sites for ligand design. *Journal of Molecular Recognition*, *23*(2), 209–219.
24 Leis, S., Schneider, S., & Zacharias, M. (2010). In silico prediction of binding sites on proteins. *Current Medicinal Chemistry*, *17*(15), 1550–1562.
25 Yuan, Y., Pei, J., & Lai, L. (2013). Binding site detection and druggability prediction of protein targets for structure-based drug design. *Current Pharmaceutical Design*, *19*(12), 2326–2333.
26 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*, 1605–1612.
27 Oliveira, S. H. P., Ferraz, F. A. N., Honorato, R. V., Xavier-Neto, J., Sobreira, T. J. P., & de Oliveira, P. S. L. (2014). KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics*, *15*, 197.
28 Zhu, H., & Pisabarro, M. T. (2011). MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics*, *27*, 351–358.
29 Ghersi, D., & Sanchez, R. (2009). EasyMIFs and SiteHound: a toolkit for the identification of ligand-binding sites in protein structures. *Bioinformatics*, *25*, 3185–3186.
30 Laurie, A. T. R., & Jackson, R. M. (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, *21*, 1908–1916.
31 Schmidtke, P., & Barril, X. (2010). Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *Journal of Medicinal Chemistry*, *53*, 5858–5867.
32 Nisius, B., Sha, F., & Gohlke, H. (2012). Structure-based computational analysis of protein binding sites for function and druggability prediction. *Journal of Biotechnology*, *159*, 123–134.
33 Trosset, J. Y., & Vodovar, N. (2013). Structure-based target druggability assessment. *Methods in Molecular Biology*, *986*, 141–164.
34 Hussein, H. A., Borrel, A., Geneix, C., Petitjean, M., Regad, L., & Camproux, A. C. (2015). PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Research*, *43*, 436–442.
35 Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). Dogsitescorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, *28*, 2074–2075.
36 Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., & Overington, J. P. (2014). The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, *42*, D1083–D1090.
37 Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. *Journal of Chemical Information and Modeling*, *49*, 377–389.
38 Krasowski, A., Muthas, D., Sarkar, A., Schmitt, S., & Brenk, R. (2011). DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *Journal of Chemical Information and Modeling*, *51*, 2829–2842.
39 Sheridan, R. P., Maiorov, V. N., Holloway, M. K., Cornell, W. D., & Gao, Y. D. (2010). Drug-like density: a method of quantifying the "bindability" of a protein target based on a very large set of pockets and drug-like ligands from the protein data bank. *Journal of Chemical Information and Modeling*, *50*, 2029–2040.
40 Nayal, M., & Honig, B. (2006). On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins Structure Function and Genetics*, *63*, 892–906.
41 Cheng, A. C., Coleman, R. G., Smyth, K. T., Cao, Q., Soulard, P., Caffrey, D. R., Salzberg, A. C., & Huang, E. S. (2007). Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnology*, *25*, 71–75.
42 Chen, K., Mizianty, M. J., Gao, J., & Kurgan, L. (2011). A critical comparative assessment of predictions of protein-binding sites for biologically relevant organic compounds. *Structure*, *19*(5), 613–621.
43 Elokely, K. M., & Doerksen, R. J. (2013). Docking challenge: protein sampling and molecular docking performance. *Journal of Chemical Information and Modeling*, *53*, 1934–1945.
44 Sousa, S. F., Ribeiro, A. J. M., Coimbra, J. T. S., Neves, R. P. P., Martins, S. A., Moorthy, N. S. H. N., Fernandes, P. A., & Ramos, M. J. (2013). Protein-ligand docking in the new millennium–a retrospective of 10 years in the field. *Current Medicinal Chemistry*, *20*, 2296–2314.
45 Grinter, S. Z., & Zou, X. (2014). Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules*, *19*, 10150–10176.
46 Ferreira, L., dos Santos, R., Oliva, G., & Andricopulo, A. (2015). Molecular docking and structure-based drug design strategies. *Molecules*, *20*, 13384–13421.
47 Chen, Y. C. (2015). Beware of docking!. *Trends in Pharmacological Sciences*, *36*, 78–95.
48 Hamelberg, D., Mongan, J., & McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *Journal of Chemical Physics*, *120*(24), 11919–11929.
49 Pierce, L. C. T., Salomon-Ferrer, R., Augusto, F., De Oliveira, C., McCammon, J. A., & Walker, R. C. (2012). Routine access to millisecond time scale events with accelerated molecular

dynamics. *Journal of Chemical Theory and Computation*, *8*(9), 2997–3002.
50 Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, *9*, 71.
51 Beveridge, D. L., & DiCapua, F. M. (1989). Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annual Review of Biophysics and Biophysical Chemistry*, *18*, 431–492.
52 Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S., & Biggin, P. C. (2016). Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science*, *7*(1), 207–218.
53 Shirts, M. R., Mobley, D. L., & Chodera, J. D. (2007). Chapter 4 Alchemical free energy calculations: ready for prime time?. *Annual Reports in Computational Chemistry*, *3*, 41–59.
54 Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *Journal of the American Chemical Society*, *120*(37), 9401–9409.
55 Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., & Cheatham, T. E. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research*, *33*(12), 889–897.
56 Aqvist, J., Medina, C., & Samuelsson, J. E. (1994). A new method for predicting binding affinity in computer-aided drug design. *Protein Engineering*, *7*, 385–391.
57 Lill, M. A., & Thompson, J. J. (2011). Solvent interaction energy calculations on molecular dynamics trajectories: increasing the efficiency using systematic frame selection. *Journal of Chemical Information and Modeling*, *51*(10), 2680–2689.
58 Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449–461.
59 Srivastava, H. K., Chourasia, M., Kumar, D., & Sastry, G. N. (2011). Comparison of computational methods to model DNA minor groove binders. *Journal of Chemical Information and Modeling*, *51*(3), 558–571.
60 Srivastava, H. K., & Sastry, G. N. (2013). Efficient estimation of MMGBSA-based BEs for DNA and aromatic furan amidino derivatives. *Journal of Biomolecular Structure & Dynamics*, *31*(5), 522–537.
61 Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Molecular Informatics*, *31*(2), 114–122.
62 Hou, T., Wang, J., Li, Y., & Wang, W. (2010). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, *51*, 69–82.
63 Genheden, S., Mikulskis, P., Hu, L., Kongsted, J., Söderhjelm, P., & Ryde, U. (2011). Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *Journal of the American Chemical Society*, *133*(33), 13081–13092.
64 Genheden, S., & Ryde, U. (2010). How to obtain statistically converged MM/GBSA results. *Journal of Computational Chemistry*, *31*, 837–846.
65 Genheden, S., & Ryde, U. (2012). Will molecular dynamics simulations of proteins ever reach equilibrium?. *Physical Chemistry Chemical Physics*, *14*, 8662–8677.
66 Henriksen, N. M., Hayatshahi, H. S., Davis, D. R., & Cheatham, T. E. (2014). Structural and energetic analysis of 2-aminobenzimidazole inhibitors in complex with the hepatitis C virus IRES RNA using molecular dynamics simulations. *Journal of Chemical Information and Modeling*, *54*(6), 1758–1772.
67 Sun, H., Li, Y., Tian, S., Xu, L., & Hou, T. (2014). Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Physical Chemistry Chemical Physics*, *16*, 16719–16729.
68 Srivastava, H. K., & Sastry, G. N. (2012). Molecular dynamics investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches. *Journal of Chemical Information and Modeling*, *52*(11), 3088–3098.
69 Chetty, S., & Soliman, M. E. S. (2015). Possible allosteric binding site on Gyrase B , a key target for novel anti-TB drugs : homology modelling and binding site identification using molecular dynamics simulation and binding free energy calculations. *Medicinal Chemistry Research*, *24*, 2055–2074.