BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 15 2012, pages 2074–2075
doi:10.1093/bioinformatics/bts310

Structural bioinformatics

Advance Access publication May 23, 2012

DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment

Andrea Volkamer[1], Daniel Kuhn[2], Friedrich Rippmann[2] and Matthias Rarey[1,*]

[1]Center for Bioinformatics, University of Hamburg, Bundesstr and [2]Merck KGaA, Merck Serono, Global Computational Chemistry, Frankfurter Str. 250, 64293 Darmstadt, Germany

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Many drug discovery projects fail because the underlying target is finally found to be undruggable. Progress in structure elucidation of proteins now opens up a route to automatic structure-based target assessment. DoGSiteScorer is a newly developed automatic tool combining pocket prediction, characterization and druggability estimation and is now available through a web server.

Availability: The DoGSiteScorer web server is freely available for academic use at http://dogsite.zbh.uni-hamburg.de

Contact: rarey@zbh.uni-hamburg.de.

Received on March 31, 2012; revised on May 11, 2012; accepted on May 19, 2012

1 INTRODUCTION

Rating the attractiveness of a drug target is one of the major challenges in the early stages of drug discovery. Besides attractivity assessment based on medical rationale and commercial viability, the properties of the target and its ability to be modulated by small drug-like compounds (further referred to as druggability) have to be analyzed. Due to the large amount of available crystal structures, the automatic collection of target information gains importance.

In a first step, binding pockets have to be detected on the protein surface. Some methods fulfilling this task are available through web services, e.g. QSite-Finder, CASTp, SCREEN, PocketDepth, MetaPocket and Fpocket (servers are referenced in Schmidtke *et al.* 2010). The next step on the path toward target classification or druggability prediction is the annotation and comparison of target-specific pocket properties. Some servers exist that allow—besides binding site prediction—for their analysis and functional classification, e.g. FINDSITE (Brylinski and Skolnick, 2008), SplitPocket (Tseng *et al.*, 2009), fPOP (Tseng *et al.*, 2010), ProBis (Konc and Janezic, 2010) and SiteComp (Lin *et al.*, 2012). Many of these approaches search for structural similarities, which can help to predict side effects of known drugs or to identify the role of yet uncharacterized proteins.

Although methods for fully automatic structure-based druggability predictions such as SiteMap (Halgren, 2009), Fpocket (Schmidtke and Barril, 2010) and DLID (Sheridan *et al.*, 2010) exist, none of these methods is available online for predictions on new targets. Fpocket allocates a web service where druggability scores and information can be requested (Schmidtke and Barril, 2010) but only for precalculated data points.

DoGSiteScorer (Volkamer *et al.*, 2012) provides the functionality to detect potential binding pockets and subpockets of a protein of interest. Subsequently, it analyzes the geometric and physico-chemical properties of these pockets and estimates druggability with aid of a support vector machine (SVM). DoGSiteScorer has been evaluated on a large dataset containing 1069 structures and shows prediction accuracies of 88%. Thus, the method provides valuable information for target assessment and can now be accessed through a web server.

2 METHODS

The first step in the DoGSiteScorer procedure is the prediction of potential pockets on the protein surface solely based on the protein heavy atom coordinates. A grid is spanned around the protein and grid points are labeled depending on their spatial overlap with any protein atom. Subsequently, a difference of Gaussian (DoG) filter is applied to the grid. With this operation, positions on the protein surface are identified where the location of a sphere-like object is favorable. Based on a density threshold, these positions are clustered to potential subpockets. Finally, neighboring subpockets are merged to pockets. (Volkamer *et al.*, 2010)

Several geometric and physico-chemical properties are automatically calculated for the predicted pockets and subpockets. Pocket volume and surface are calculated by counting the grid points constituting the pocket volume or its surface and multiplying this number with the grid box volume or surface, respectively. A breadth-first search is used for pocket depth computation, starting from the solvent exposed pocket parts toward the most deeply buried regions. Ellipsoids fitted into the pocket volume reflect the overall pocket shape. The pocket enclosure is derived from the ratio between pocket hull and surface grid points. Each atom within 4 Å of any pocket point is considered a pocket atom. Pocket atom counts or functional groups and amino acid compositions describe the physico-chemical features of the pocket. Furthermore, the lipophilic character of the pockets is addressed by the lipophilic surface and the overall hydrophobicity ratio. In addition, if a ligand is provided, the overlap between ligand and pocket volume is computed. Moreover, a SimpleScore is calculated by a linear combination of the three properties pocket volume, enclosure and lipophilic character.

For druggability predictions, a supervised machine learning technique-more precisely a SVM-is incorporated. Based on a discriminate analysis, a subset of descriptors best suited to separate druggable from undruggable pockets has been selected. The model has been trained and tested on the non-redundant version of the druggable dataset (Schmidtke and Barril, 2010). External cross validation, randomly taking one half of the data as training and the other half as test set, showed a mean accuracy of 90%. (Volkamer *et al.*, 2012)

For each input structure, the method predicts potential pockets, describes them through descriptors and queries the SVM model for druggability estimations. A druggability score between 0 and 1 is returned. The higher the score the more druggable the pocket is estimated to be.

*To whom correspondence should be addressed.

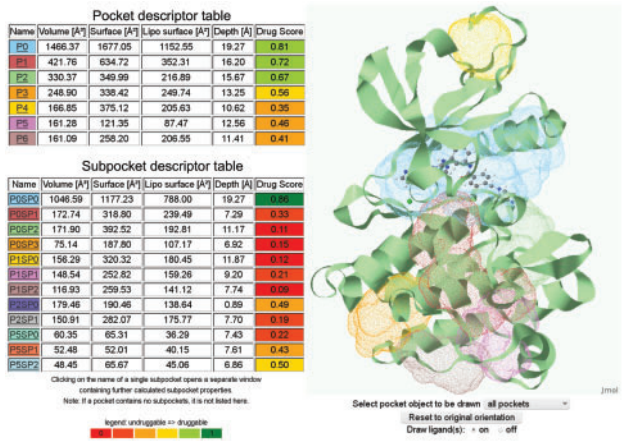

Fig. 1. Server result page. Left: table containing the main shape descriptors for all detected pockets and subpockets. Right: the seven detected pockets together with a cartoon representation of chain B of cAbl (PDB code 1iep)

3 USAGE AND OUTPUT

The DoGSiteScorer server requires a PDB code or a user-specified PDB file as input. To calculate the ligand/pocket overlap, the ligand can be extracted from the structure or provided as mol2 file. Further, format specifications can be obtained through an info button. After a validity check, the user can select if the entire protein or a selected chain should be used. DoGSiteScorer can be customized to work on the pocket and subpocket level. In addition, the druggability estimation for pockets can be switched on. Clicking the 'calculate and analyze pockets' button leads to the result page (Fig. 1).

On the left side, a table containing volume, surface, lipophilic surface and depth of each detected pocket or subpocket is shown. The rows are sorted by descending pocket volume. The last table column either holds the calculated SimpleScore or the druggability score. The druggability score fields are colored according to the druggability character of the pockets based on a traffic light coloring scheme (green: high druggability score, red: low druggability score). Note that the subpocket druggability scores do not add up to the respective pocket score. Both pocket and subpocket scores are calculated with individual druggability models, trained on pockets and subpockets, respectively. For a quick visual analysis, the predicted pockets and subpockets as well as protein and ligand are shown in a Jmol applet (http://www.jmol.org/). The color code used for the pocket representation corresponds to the background color of the respective entry in the pocket descriptor table. Pockets can individually be displayed or hidden. Thus, the user can explore the neighborhood of a pocket of interest and e.g. investigate potential extensions toward highly ranked (sub)pockets.

Further, pocket properties can be accessed in a separate table by clicking on the name of a pocket. If a ligand is contained in the pocket, the respective ligand and pocket coverage are annotated. The descriptor information is provided in five tables, partitioned based on the descriptor category. The shape and size table contains volume, surface, lipophilic surface and depth of the pocket, ratios between ellipsoid main axes and pocket enclosure. Information about functional groups such as number of hydrogen bond donors and acceptors, metals and hydrophobic interactions as well as the hydrophobicity ratio is presented in the next table. Pocket atom and specific element counts are provided in the third table. Finally, information about amino acids is spread over the last two tables containing the ratio of apolar, polar, positive and negative amino acids as well as the occurrence of each of the 20 amino acids, respectively. All information and data (descriptor and PDB files for pocket lining residues) can be received through email.

As a use case, the tyrosine protein kinase Abl1 (cALB) with its inhibitor Gleevec (imatinib) is exemplarily shown (PDB code 1iep-chain B). DoGSiteScorer detects seven potential pockets on the surface of the protein structure, listed in the result table and visualized with Jmol (Fig. 1). The first three pockets are estimated to be druggable. The selection of the pocket with the highest druggability of 0.81 opens a detailed descriptor page. The anti-cancer drug imatinib is completely contained in the detected pocket and fills the pocket to 35.7%. A large pocket volume of 1466 $Å^3$, a high depth of 19.27 Å as well as a high apolar amino acid ratio of 0.5 are only three properties of this pocket that generally characterize druggable pockets. Since the detected pocket is rather large, the analysis based on subpocket level may provide a more realistic view on the ATP-binding site. The largest subpocket P0SP0 achieves a druggability score of 0.89, contains the drug completely and is 50% covered. The other three subpockets are rather small pieces that are cutoff at the edges of the original pocket and get druggability scores below 0.43.

4 CONCLUSION

DoGSiteScorer web server provides an easy to use interface to predict pockets and subpockets of a protein structure of interest. Furthermore, key properties characterizing the pocket and druggability estimations are supplied.

Funding: The project is part of the Biokalayse2021 cluster and jointly funded by the German Federal Ministry of Education and Research (BMBF, grant 0315292A) and Merck KGaA, Darmstadt.

Conflict of Interest: none declared.

REFERENCES

Brylinski,M. and Skolnick,J. (2008) A threading-based method (findsite) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. USA*, **105**, 129–134.

Halgren,T. (2009) Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.*, **49**, 377–389.

Konc,J. and Janezic,D. (2010) Probis: a web server for detection of structurally similar protein binding sites. *Nucleic Acids Res.*, **38**, W436–W440.

Lin,Y. *et al.* (2012) Sitecomp: a server for ligand binding site analysis in protein structures. *Bioinformatics (Oxford, England)*. **28**, 1172–1173.

Schmidtke,P. and Barril,X. (2010) Understanding and predicting druggability. a high-throughput method for detection of drug binding sites. *J. Med. Chem.*, **53**, 5858–5867.

Schmidtke,P. *et al.* (2010) fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.*, **38**, W582–W589.

Sheridan,R. *et al.* (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. *J. Chem. Inf. Model.*, **50**, 2029–2040.

Tseng,Y. *et al.* (2009) Splitpocket: identification of protein functional surfaces and characterization of their spatial patterns. *Nucleic Acids Res.*, **37**, W384–W389.

Tseng,Y. *et al.* (2010) fpop: footprinting functional pockets of proteins by comparative spatial patterns. *Nucleic Acids Res.*, **38**, D288–D295.

Volkamer,A. *et al.* (2010) Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J. Chem. Inf. Model.*, **50**, 2041–2052.

Volkamer,A. *et al.* (2012) Combining global and local measures for structure-based druggability predictions. *J. Chem. Inf. Model.*, **52**, 360–372.