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

Andrea Volkamer\textsuperscript{1}, Daniel Kuhn\textsuperscript{2}, Friedrich Rippmann\textsuperscript{2} and Matthias Rarey\textsuperscript{1,*}

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

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://d ogsite.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 attractiveness 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 Pocket 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 (Brlvish and Skolnick, 2000), SplitPocket (Lewis et al., 2009), iPOP (Lazar et al., 2009), ProBiS (Kone and Jameson, 2010) and SiteComp (Kim et al., 2010). 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 (Batt et al., 2008), Fpocket (Schmidtke and Bari, 2010) and DLID (Sheridan et al., 2011) 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 Bari, 2010) but only for precalculated data points.

DoGSiteScorer\textsuperscript{1} 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 Bari, 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.
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).

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 of interest and e.g. investigate potential neighborhood of a pocket of interest and e.g. investigate potential druggability predictions.

REFERENCES


