# *Improvement of Methods for the Comparison of Protein Binding Sites*

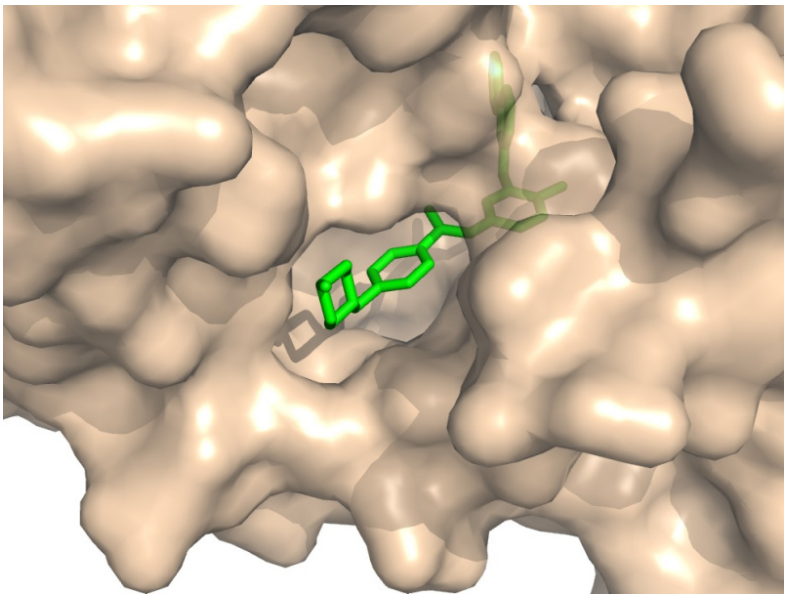

Fig. 1. Illustration of the drug imatinib (brand name Gleevec®, shown in green) bound in a deep cavitiy of the tyrosine-kinase ABL2. By its binding mode, the drug is able to inhibit the protein and is therefore being used for the treatment of chronic myeloid leukemia.

Researchers
Dr. Timo Krotzky, Thomas Rickmeyer, Thomas Fober and Eyke Hüllermeier

Principal Investigator
Prof. Dr. Gerhard Klebe

Project Term
2012 - 2014

Project Areas
Biological Chemistry and Food Chemistry

Clusters
MaRC2 Cluster Marburg

University
University of Marburg

Group of Prof. Dr. Klebe
http://pc1664.pharmazie.uni-marburg.de/

## Introduction

The detection of structural similarities between protein binding sites is an important challenge in medicinal drug discovery. Drug molecules most often bind in depressions on the surface of a protein, as pictured in Figure 1. Comparing such cavities of different proteins can thus help to understand selectivity considerations or to predict unexpected cross-reactivity of drugs as similar binding pockets will be able to bind the same drug molecules.

To this end, Cavbase was developed as a tool for the automated detection and storage of protein surface cavities. The database entries (currently about half a million binding pockets) can also be mutually compared in 3D space, which is computationally very demanding if large amounts of similarities have to be calculated. With the help of the Marburg HPC cluster MaRC2, we were able to exhaustively test different parameterizations of cavity comparison methods newly developed.

## Methods

In this context, we have developed a novel graph-based comparison method which employs additional node properties



JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

that contains local information about the shape of the protein surface [1]. This is to compensate the main drawback of most other graph-based models which are solely based on physicochemical properties and hence result in a loss of information about the surface shape. Furthermore, the detection of the maximum common substructure could be accelerated, which is the most time-consuming step during the comparison of two binding sites. Thus, we were able to propose a novel and efficient modeling formalism that does not increase the size of the graph model compared to previous approaches, but leads to equal-sized graphs that contain considerably more information assigned to the nodes and allows for much shorter runtimes.

In a following study, another method for large-scale mining of similar protein binding sites was developed. We proposed RAPMAD (RApid Pocket MAtching using Distances), a new evaluation formalism for Cavbase entries that allows for ultrafast similarity comparisons [2]. Here, protein binding sites are represented by sets of distance histograms that are both generated and compared very quickly. Since RAPMAD attains a speed of more than 20,000 comparisons per second, screenings across large data sets and even entire databases become easily tractable. We demonstrated the discriminative power and the short runtime by performing several classification and retrieval experiments. RAPMAD reaches better success rates than the comparison formalism originally implemented into Cavbase and several alternative approaches developed in recent time, while requiring only a fraction of their runtime. The practical use of the method was finally proven by a successful prospective virtual screening study that aims for the identification of novel inhibitors of the NMDA receptor, an important target protein for the treatment of various neurological disorders, such as depression, Alzheimer's disease, or Parkinson's disease.

## Results

Finally, it could be shown that the extraction of protein binding sites in close neighborhood of bound molecules ("ligands") makes comparisons simple due to inherent shape similarity [3]. Methods for comparing binding sites are frequently validated on data sets of pockets that were obtained simply by extracting the protein area next to the bound ligands. With this strategy, any unoccupied cavity will remain unconsidered. Furthermore, a large amount of ligand-biased intrinsic shape information is predefined, inclining the subsequent comparisons as rather trivial. We presented the results of a very simplistic and shape-biased comparison approach, which stress that unrestricted and ligand-independent cavity extraction is essential to enable unexpected cross-reactivity predictions among proteins and function annotations of orphan proteins. Such an unrestricted pocket extraction can be achieved by using one of the many published automated binding site detection algorithms. Also in this case, MaRC2 was utilized for both preparing the half a million pockets and computing similarities with the different evaluated comparison methods.



JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN TECHNISCHE UNIVERSITÄT DARMSTADT UNIKASSEL VERSITÄT Philipps Universität Marburg

## Reference

[1] T. Krotzky, T. Fober, E. Hullermeier, and G. Klebe (2014), Extended Graph-Based Models for Enhanced Similarity Search in Cavbase. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 11(5): 878-890. https://doi.org/10.1109/tcbb.2014.2325020

[2] T. Krotzky, C. Grunwald, U. Egerland, and G. Klebe (2014), Large-scale Mining for Similar Protein Binding Pockets: With RAPMAD Retrieval on the Fly Becomes Real. Journal of Chemical Information and Modeling. https://doi.org/10.1021/ci5005898

[3] T. Krotzky, T. Rickmeyer, T. Fober, and G. Klebe (2014), Extraction of protein binding pockets in close neighborhood of bound ligands makes comparisons simple due to inherent shape similarity. Journal of Chemical Information and Modeling, 54(11): 3229-3237. https://doi.org/10.1021/ci500553a

## *Last Update:* 2022-09-18 14:25