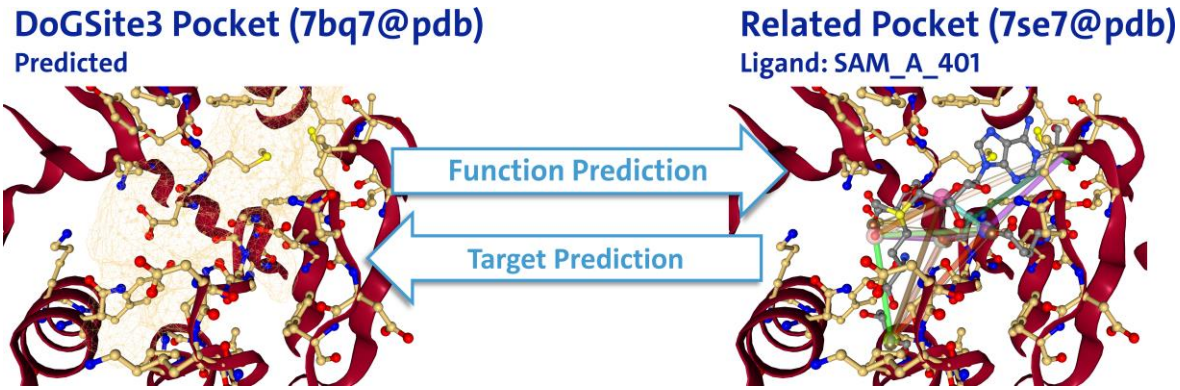
Binding Site Prediction and Characterization with DoGSite3 – Method and Applications

Christiane Ehrt, Joel Graef, Konrad Diedrich, Matthias Rarey

Universität Hamburg, ZBH - Center for Bioinformatics, Hamburg, Germany

The prediction and characterization of protein binding sites is a non-trivial endeavor. [1] However, it is a crucial methodology given the increasing numbers of predicted and experimentally determined protein structure models [2] and their impact on structure-based design applications such as function or off-target prediction. Given a structure of interest, the DoGSite algorithm predicts binding sites for yet uncharacterized proteins, enabling protein function prediction and binding site druggability estimation. [3,4]

Based on the first implementation of the grid-based DoGSite methodology, we developed DoGSite3. [5] The new implementation is characterized by more robust binding site boundaries, improved prediction accuracy, and a considerably lower run time. Its integration on the Proteins*Plus* web server (https://proteins.plus) enables easy access to predicted protein pockets and their descriptors, even for non-experienced users.



In this contribution, we will outline the basic methodology and optimization of the DoGSite3 algorithm, focusing on binding site characterization and prediction accuracy. Based on selected application examples, we will demonstrate the broad applicability of DoGSite3 in various structure-based design approaches, such as pocket annotation and function prediction by known protein-ligand complexes and the elucidation of potential off-targets for known drugs.

[1] J. Zhao, Y. Cao, L. Zhang, *Comput Struct Biotechnol J*, **2020**, *18*, 417-26.
[2] N.V. Dokholyan, *J Proteomics*, **2020**, *220*, 103777.
[3] A. Volkamer, A. Griewel, T. Grombacher, M. Rarey, *J Chem Inf Model*, **2010**, *50*, 2041-52.
[4] A. Volkamer, D. Kuhn, T. Grombacher, F. Rippmann, M. Rarey, *J Chem Inf Model*, **2012**, *52*, 360-72.
[5] J. Graef, C. Ehrt, M. Rarey, *J Chem Inf Model*, **2023**, *63*, 3128-37.