Binding Site Comparison with SiteMine for the Functional Annotation of Predicted Protein Structures

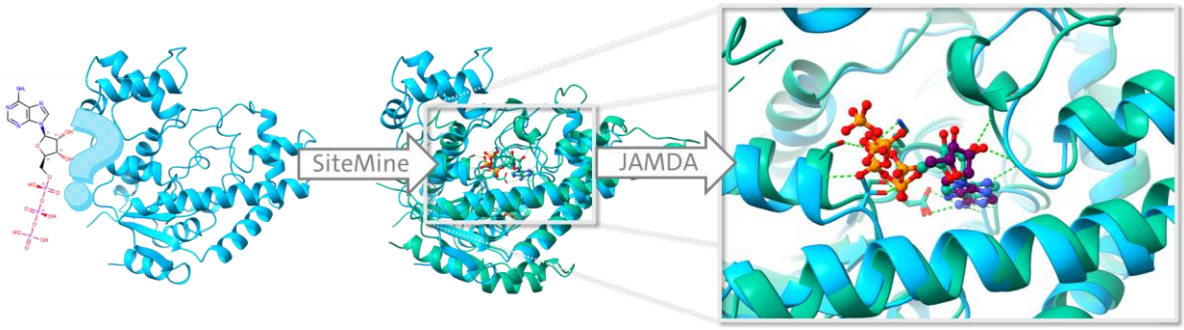
Christiane Ehrt, Joel Graef, Martin Poppinga, Matthias Rarey

University of Hamburg, ZBH – Center for Bioinformatics

Advances in the field of protein structure prediction enable unprecedented access to models of structurally uncharacterized proteins. Methods such as AlphaFold2[1] generate models using protein sequence information and knowledge of publicly available experimentally solved protein structures. However, these models lack the corresponding functional annotations that enable researchers to analyze these structures and utilize them in computational workflows such as virtual screening or molecular probe design. Although more recent co-folding approaches enable users to predict, for example, protein structures in complex with a known ligand, they are heavily biased toward available knowledge and the training dataset, and often lack the necessary performance.[2]

Considering these challenges, we propose SiteMine[3] as a viable alternative. It is an efficient and sequence-independent binding site comparison method, relying on the GeoMine[4] database technology that captures binding sites and their characteristics for all available PDB structures. Based on binding site similarity and the corresponding local alignment, users cannot only learn where known or unknown ligands might bind but also understand where the binding site prediction is derived from. The latter is especially beneficial to assess the quality and reliability of predictions.

In contrast to the AlphaFill[5] methodology, the method does not rely on sequence similarity, focuses on local rather than global similarities, and is not restricted to subsets of small molecules in the Protein Data Bank.



Based on a dataset of predicted protein models with known ligands, we will illustrate the applicability of SiteMine for annotating ligand binding sites and present a pipeline that enables users to predict the binding mode of the ligands in the corresponding binding site using the pose optimization functionalities of JAMDA.[6]

[1] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger *et al.*, *Nature*, **2021**, *596(7873)*, 583-589. DOI: https://doi.org/10.1038/s41586-021-03819-2
[2] E. Nittinger, Ö. Yoluk, A. Tibo, G. Olanders, C. Tyrchan, *Artif. Intell. Life Sci.*, **2025**, *8*, 100136. DOI: https://doi.org/10.1016/j.ailsci.2025.100136
[3] T. Reim, C. Ehrt, J. Graef, S. Günther, A. Meents, M. Rarey, *Arch. Pharm.*, **2024**, *357(5)*, e2300661. DOI: https://doi.org/10.1002/ardp.202300661
[4] J. Graef, C. Ehrt, K. Diedrich, M. Poppinga, N. Ritter, M. Rarey, *J. Med. Chem.*, **2022**, *65(2)*, 1384-1395. DOI: https://doi.org/10.1021/acs.jmedchem.1c01046
[5] M. L. Hekkelman, I. de Vries, R.P. Joosten, A. Perrakis, *Nat. Methods*, **2023**, *20*, 205–213. DOI: https://doi.org/10.1038/s41592-022-01685-y
[6] F. Flachsenberg, C. Ehrt, T. Gutermuth, M. Rarey, *J. Chem. Inf. Model.*, **2024**, *64(1)*, 219-237. DOI: https://doi.org/10.1021/acs.jcim.3c01573