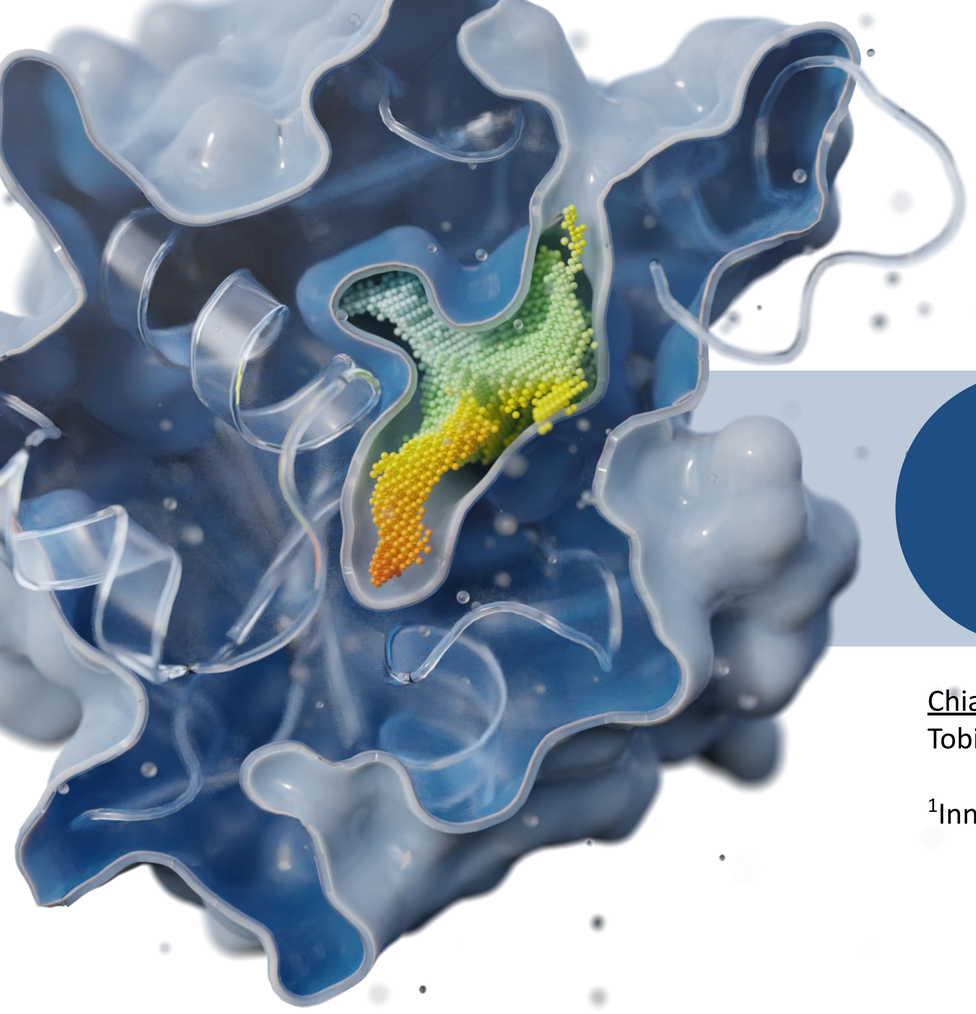

Innophore

AI in support of drug off-target effect research: the CavitOmix Copilot

Chiara Gasbarri[1], Katharina Rauchenwald-Köchl[1], Lena Parigger[1] , Haris Rudalija[1] , Michael Hetmann[1] , Markus Fleck[1] , Tobias Schopper[1], Georg Steinkellner[1], Christian C. Gruber[1]

[1]Innophore GmbH, 8010, Graz, Austria

Explore our Copilot!

From human proteome sequences to cavities

Safety concerns account for approximately 17% of failures in phase 3 clinical trials, typically resulting from therapeutic agents interacting with unintended protein targets with similar binding pockets to the desired targets. [1] Predicting similarities between protein pockets has the potential to mitigate the failure rate observed in drug discovery programs. To address this issue, we introduce CavitOmiX Copilot, a novel platform developed by Innophore, which employs an AI-based matching approach to identify structural similarities among protein pockets.
Through comprehensive modeling of the human proteome using homology modeling and advanced AI techniques leveraging NVIDIA's BioNeMo service, we generated multidimensional point clouds referred to as *catalophores*. [2]

These point clouds represent unique fingerprints of protein binding pockets by encoding their physicochemical properties, which are subsequently embedded into high-dimensional vectors. This framework facilitates ultra-rapid comparisons and matching of diverse cavities, allowing for the detection of similarities between proteins without bias introduced by protein sequence or structure.

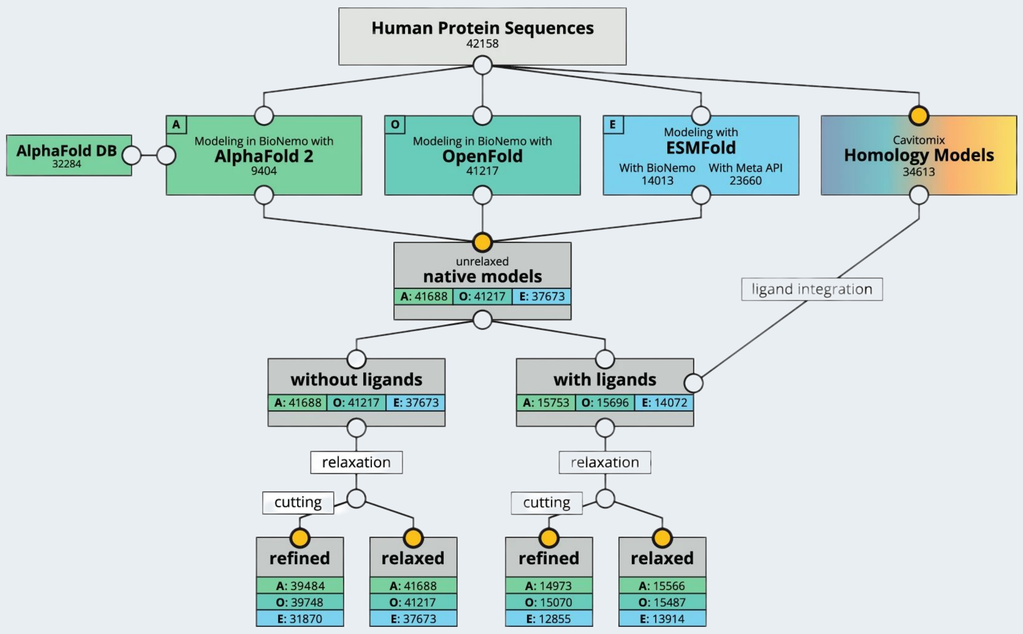

Figure 1: Flowchart illustrating the methodology in generating the dataset and structure of the dataset. The numbers indicate the quantity of models included in each respective data records. Yellow circles denote data records that are available for download. "Relaxation" refers to energy minimization with the Amber03 force field, while "cutting" refers to the removal of low confidence regions. A: AlphaFold 2; O: OpenFold; and E: ESMFold.[2]

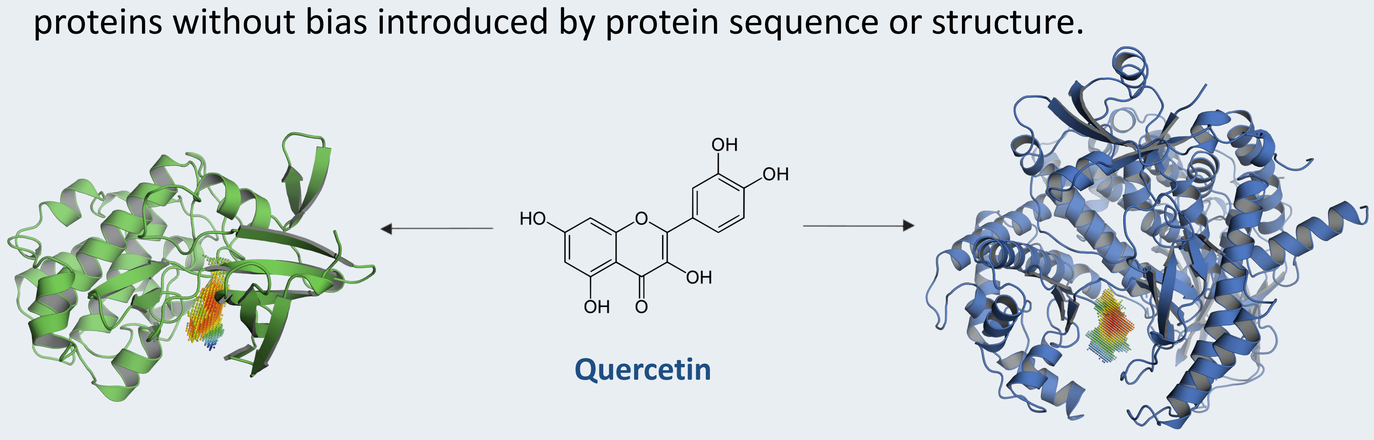

Figure 2: Example of cavity similarity between quercetin binding sites of PI3KCG (blue) and PIM1 (green). These two proteins share only 8% sequence identity and a TM-score of 0.46, indicating minimal correlation between their sequences and structures. Despite of these evidences, quercetin exhibits the activity to bind both kinases. The binding pockets are illustrated using our proprietary point cloud technology. [3]

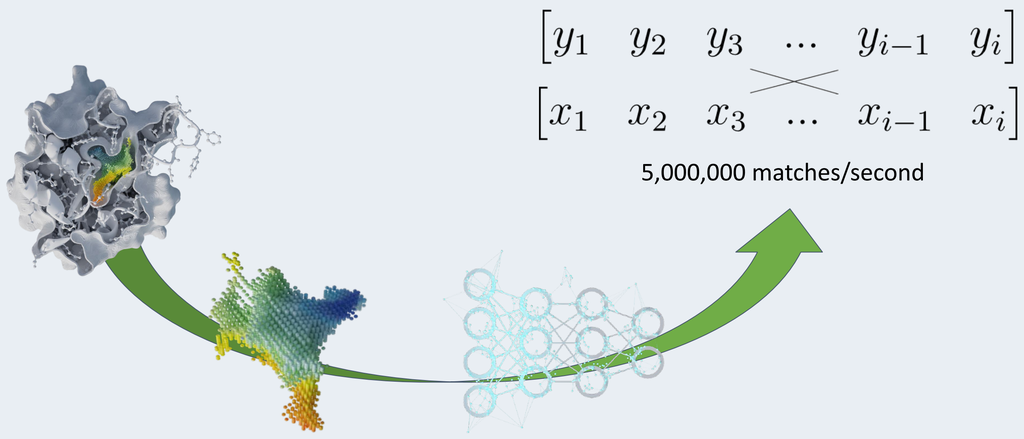

Figure 3: Cavity matching algorithm scheme. Each cavity is embedded into high dimensional vectors. These vectors are then used to match cavities with a speed of up to 5 millions matches per second.

From cavities to the CavitOmiX Copilot

All the structures and cavities are accessible through a web application at https://cavitomix.catalophore.com. This application operates on a token-based access, enabling users to explore the entire human proteome and conduct real-time cavity matchings. Results are available for download in multiple file formats, also shareable via e-mails. The insights gained from these matchings offer valuable potential in predicting drug side effects by examining binding pocket similarities between the target and other human proteins.

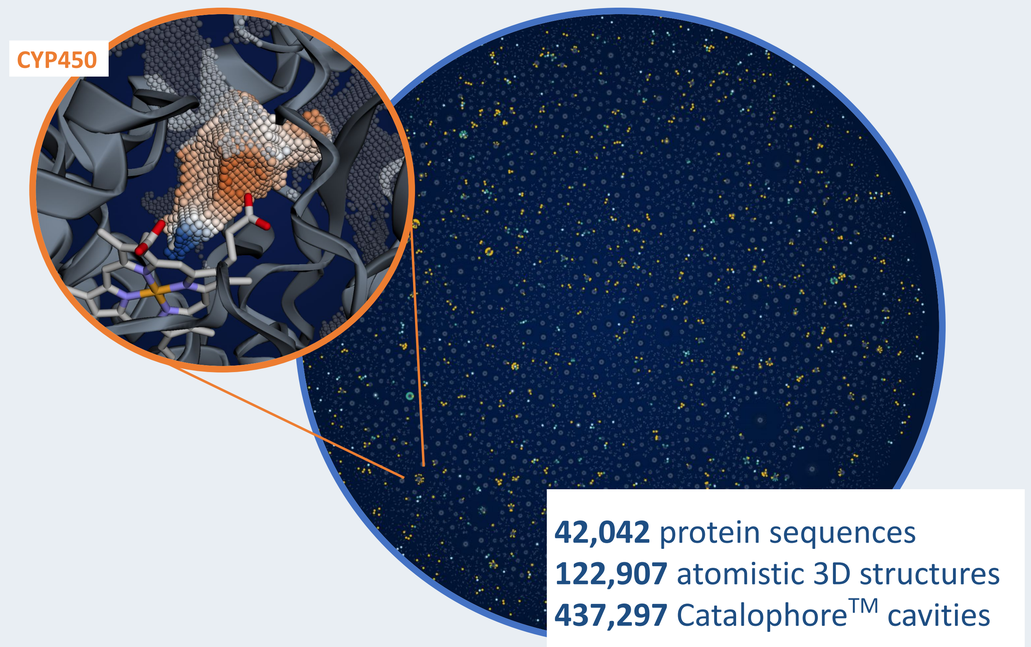

Figure 4: Example view of the CavitOmiX Copilot. This interactive tool allows users to navigate based on a specific UniProt identifier or to select a protein class of interest. After choosing a cavity, the application efficiently searches the entire proteome to identify cavities that share similarities with the reference cavity.

References

[1] David B. Fogel, Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review, Contemporary Clinical Trials Communications, Volume 11, 2018, Pages 156-164, ISSN 2451-8654, https://doi.org/10.1016/j.conctc.2018.08.001.
[2] Hetmann, M., Parigger, L., Sirelkhatim, H. et al. Folding the human proteome using BioNeMo: A fused dataset of structural models for machine learning purposes. Sci Data 11, 591 (2024). https://doi.org/10.1038/s41597-024-03403-z
[3] Haupt VJ, Daminelli S, Schroeder M Drug Promiscuity in PDB: Protein Binding Site Similarity Is Key. PLoS ONE 8(6): e65894 (2013). https://doi.org/10.1371/journal.pone.0065894.

Acknowledgements

The computational results presented have been achieved using HPC resources provided by NVIDIA and Innophore.