CDCS
CENTER FOR DATA AND COMPUTING
IN NATURAL SCIENCES

OPENING SYMPOSIUM 2022

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.

TUHH

Contribution ID: **83** Type: **Poster**

MicroMiner: Mining Mutations & more from PDB/AlphaFoldDB using Residue 3D Micro-Environments

Experimental protein structures can provide valuable insights into the structural consequences and therefore the biological effect of a mutation. However, often structures for both the wild-type and the mutant are not available for comparison. To address this issue we developed a new tool for database search of mutant structures for a given wild-type structure. Using MicroMiner we can provide a wealth of wild-type/mutant structure pairs from structure databases like the PDB and the AlphaFoldDB exemplifying the local effects of mutations.

MicroMiner focuses on the local 3D micro-environment of the mutation position and is based on the SIENA [1] technology. For each residue R in a given protein structure the local 3D micro-environment of R is used to search for similar environments in a protein structure database. MicroMiner uses a fast index-based k-mer look-up to retrieve protein candidates. These are verified by an approximate string matching algorithm and a fast similarity score of the 3D arrangement. In this way, highly similar 3D environments can be quickly identified that differ only at the position of the query residue R. Finally, found micro-environments are superposed to the query environment and provided to the user as structural ensemble.

MicroMiner recovers 90% of known wild-type/mutant structure pairs from ProTherm [2], ThermoMutDB [3] and Platinum [4]. The tool can be applied to annotate experimental mutant structures to thermodynamic mutation data, e.g. stability or affinity changes upon mutation. Here, we are able to increase the annotation coverage 2.4-fold. The result illustrate that there are thousands of wild-type/mutant structure pairs readily available from the PDB that can be mined in a few seconds. The time of a single search for a whole protein structure in the PDB takes <10.5 sec for 50% of the proteins contained in ProTherm.

MicroMiner is available at https://proteins.plus/.

[1] Bietz et al. (2016), J. Chem. Inf. Model., 56, 248-259
[2] Kumar et al. (2006), Nucleic Acids Res., 34, D204-D206
[3] Xavier et al. (2021), Nucleic Acids Res., 49, D475-D479
[4] Pires et al. (2015), Nucleic Acids Res., 43, D387-D391

Primary author: SIEG, Jochen (University Hamburg)

Co-author: Prof. RAREY, Matthias (Universität Hamburg)

Presenter: SIEG, Jochen (University Hamburg)

Session Classification: Poster session with buffet

Track Classification: CDL3 (Systems Biology)