Automatic identification of important interactions in protein-ligand complexes

MICHA JOHANNES BIRKLBAUER

SUPERVISORS: FH-PROF. MMAG. DR. GERALD LIRK, UNIV. PROF. DR. DANIELA SCHUSTER & MAG. VERONIKA TEMML, PHD

Motivation

- In silico protein-ligand docking is an important tool for finding new active agents for drug usage.
- Detection of interactions between protein and ligand can already be automated…
- BUT assessment of the importance of the different interactions in a protein-ligand complex is still a manual task that requires deep domain knowledge.
- GOAL 1: Automatic identification of important interactions based on their frequency across multiple structures of the same protein.
- GOAL 2: Development of novel scoring function to distinguish active ligands from inactive ones.

Interactions

- Hydrogen bonds: Hydrogen bond acceptor donates electrons to the hydrogen bond donor.
- Water bridges: Water-bridged hydrogen bonds
- Salt bridges: Attraction of ions of opposite charge.
- Halogen bonds: Halogen bond acceptor donates electrons to the halogen bond donor.
- Hydrophobic interactions: Aggregation of hydrophobic molecules to exclude water molecules.
- Pi stacking: Interaction between pi-systems (aromatic rings).
- Pi-cation interactions: Interaction between a pi-system (aromatic ring) and a cation.
- Metal complexations: Metal ion bound to multiple other ions or molecules.

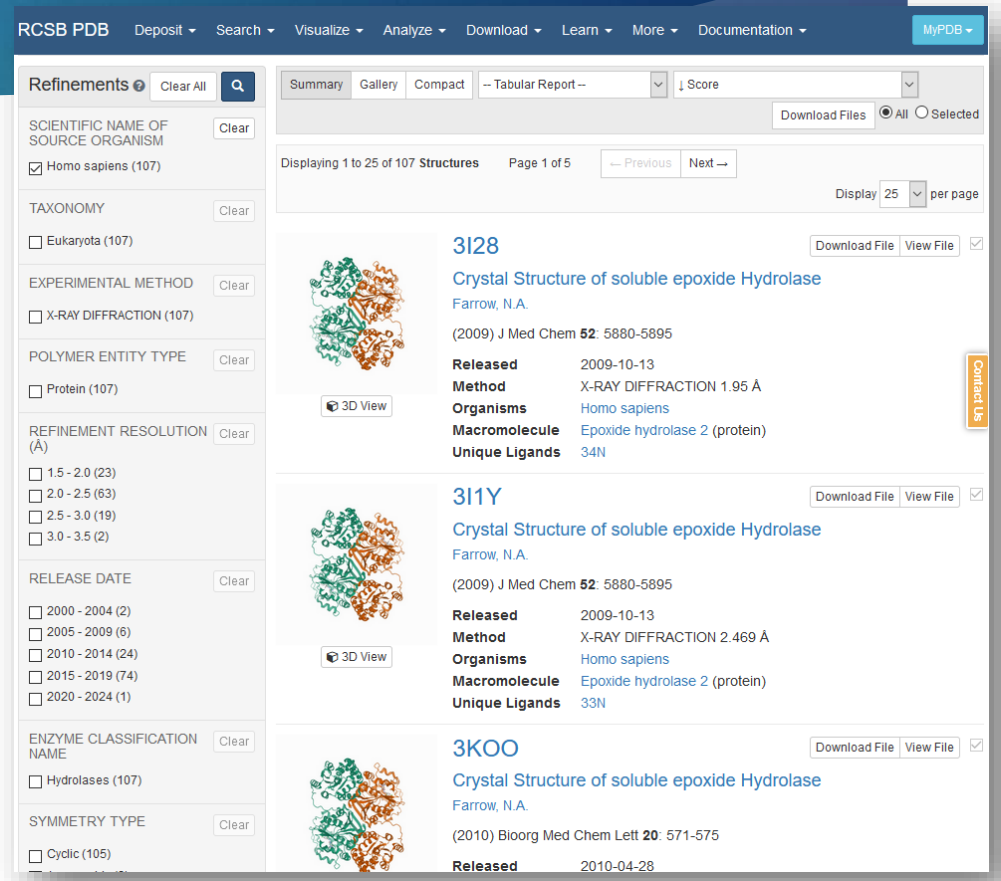
Data – Targets 1/2

11 Targets with a total of 868 manually selected structures from the PDB. Structures without ligands or mutated/chimeric/fusion proteins were discarded.

- 11-Beta-Hydroxysteroid Dehydrogenase 1 (HSD11B1)
  - Organism: Homo sapiens Structures: 28
- Acetylcholinesterase (ACHE)
  - Organism: Homo sapiens Structures: 53
- Coagulation Factor XA (FXA)
  - Organism: Homo sapiens Structures: 129
- Cyclooxygenase 1 (COX1)
  - Organism: Ovis aries (sheep) Structures: 25
- Cyclooxygenase 2 (COX2)
  - Organism: Homo sapiens Structures: 7
    Mus musculus (mouse) 44

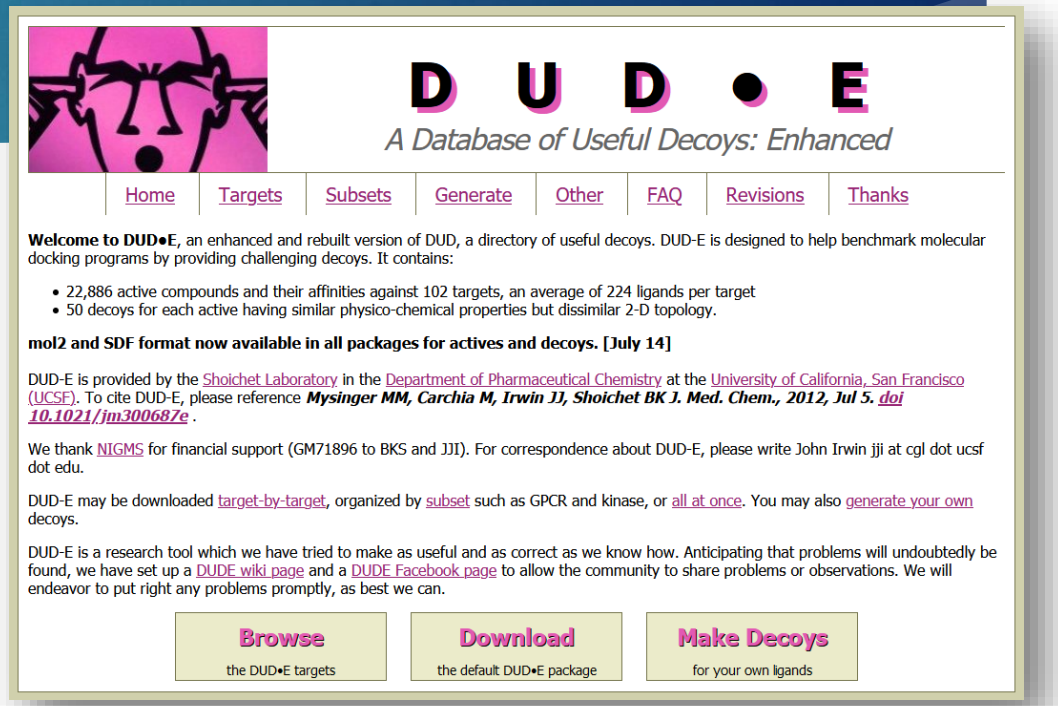
Data – Targets 2/2

- Dipeptidyl Peptidase IV (DPP4)
  - Organism: Homo sapiens Structures: 98
- Monoamine Oxidase B (MAOB)
  - Organism: Homo sapiens Structures: 47
- P38 Mitogen-Activated Protein Kinase 14 (MAPK14)
  - Organism: Homo sapiens Structures: 199
- Phosphodiesterase 5 (PDE5A)
  - Organism: Homo sapiens Structures: 32
- Protein-Tyrosine Phosphatase 1B (PTP1B)
  - Organism: Homo sapiens Structures: 102
- Soluble Epoxide Hydrolase (Epoxide Hydrolase 2, SEH)
  - Organism: Homo sapiens Structures: 104
  - Associated diseases: hypertension, cardiac hypertrophy, arteriosclerosis
  - Drugs: Ebselen (Diabetes), mostly experimental



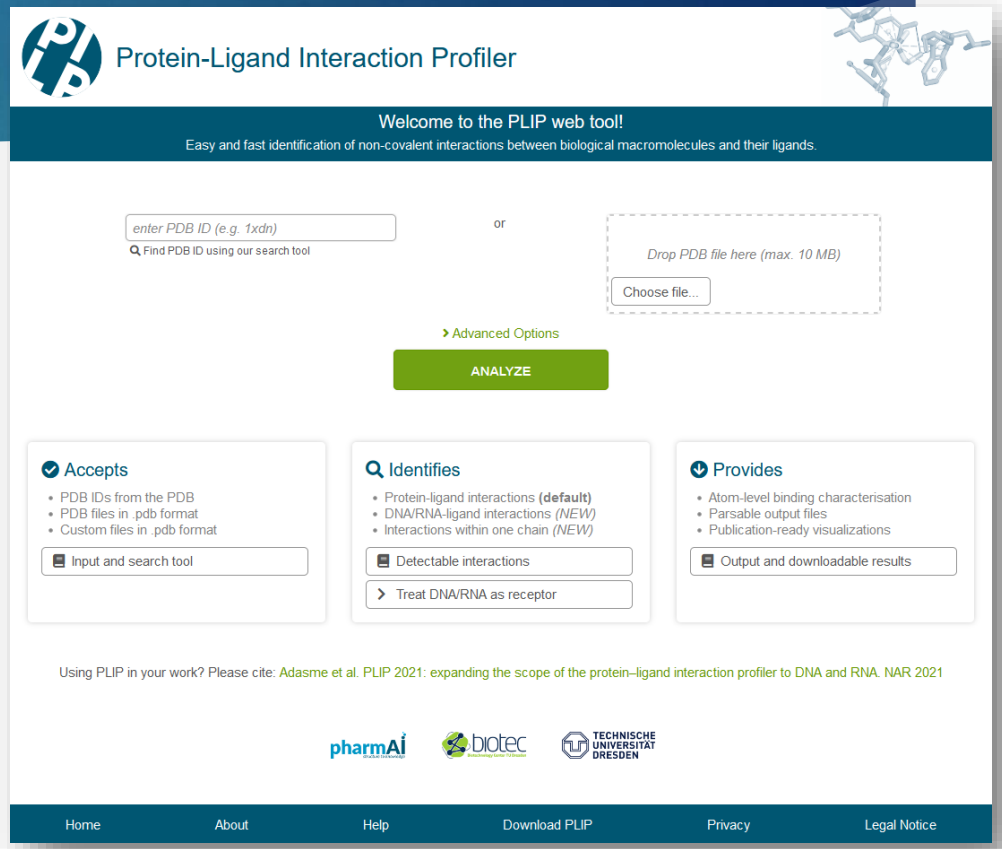
Data - Ligands

- 1. Cocrystallized ligands from the PDB
- 2. Active and inactive compounds from the „Database of Useful Decoys: Enhanced" (DUD-E) http://dude.docking.org/
  - Open-source benchmark dataset
  - Contains target-specific ligands and their binding affinity
  - Only subset that is relevant for afore mentioned targets is used

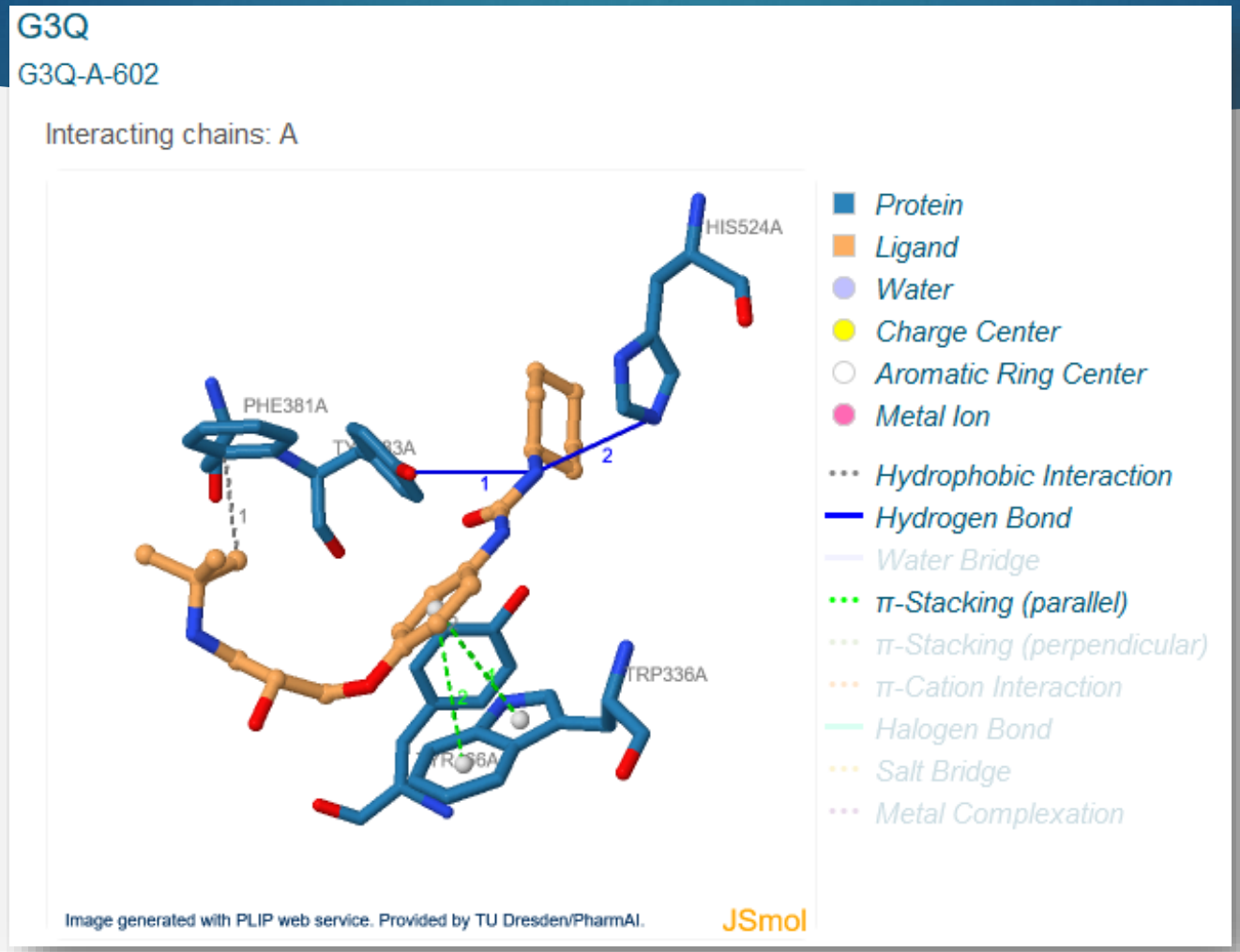


PLIP – Detection of Interactions

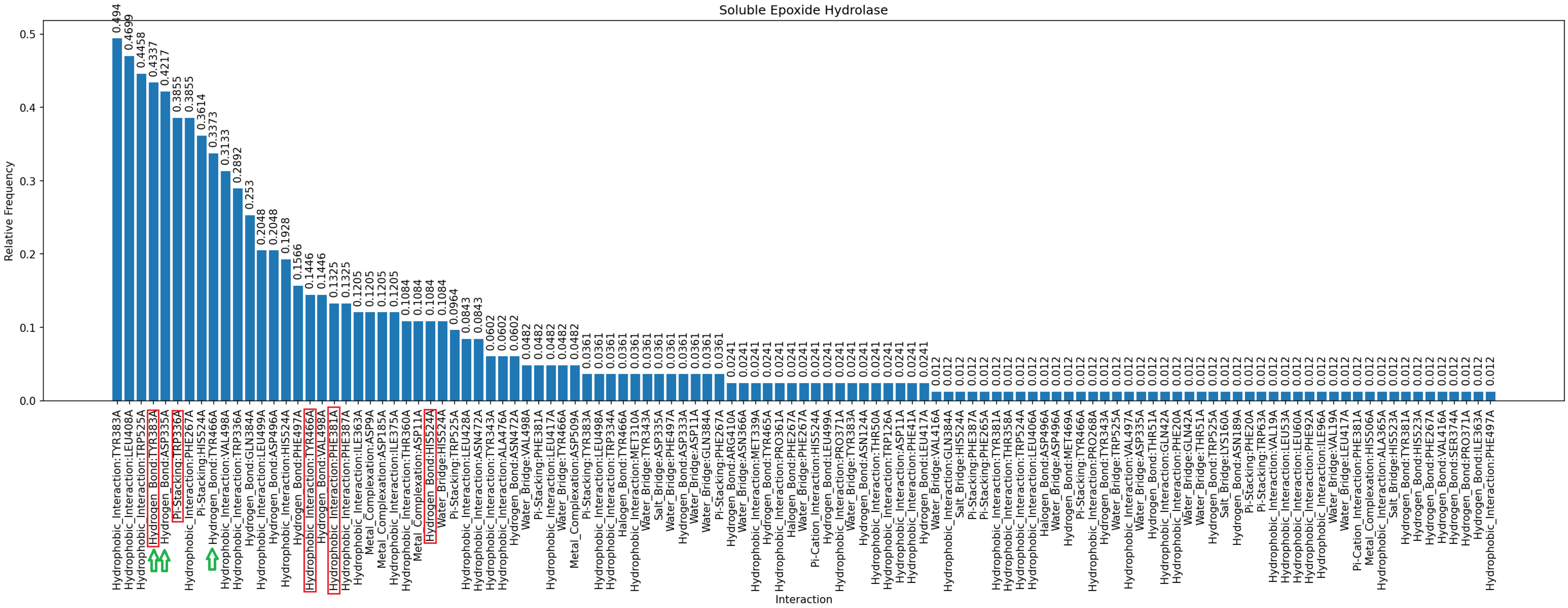
- Protein-Ligand Interaction Profiler by TU Dresden
- https://plip-tool.biotec.tu-dresden.de
- Web service available as well as a (undocumented) python package, docker and singularity image
- Detects interactions using a rule-based approach e.g. by atom properties and geometric information like distance and angle between functional groups.



Example – SEH (6HGV, Talinolol)



SEH Interaction Frequencies

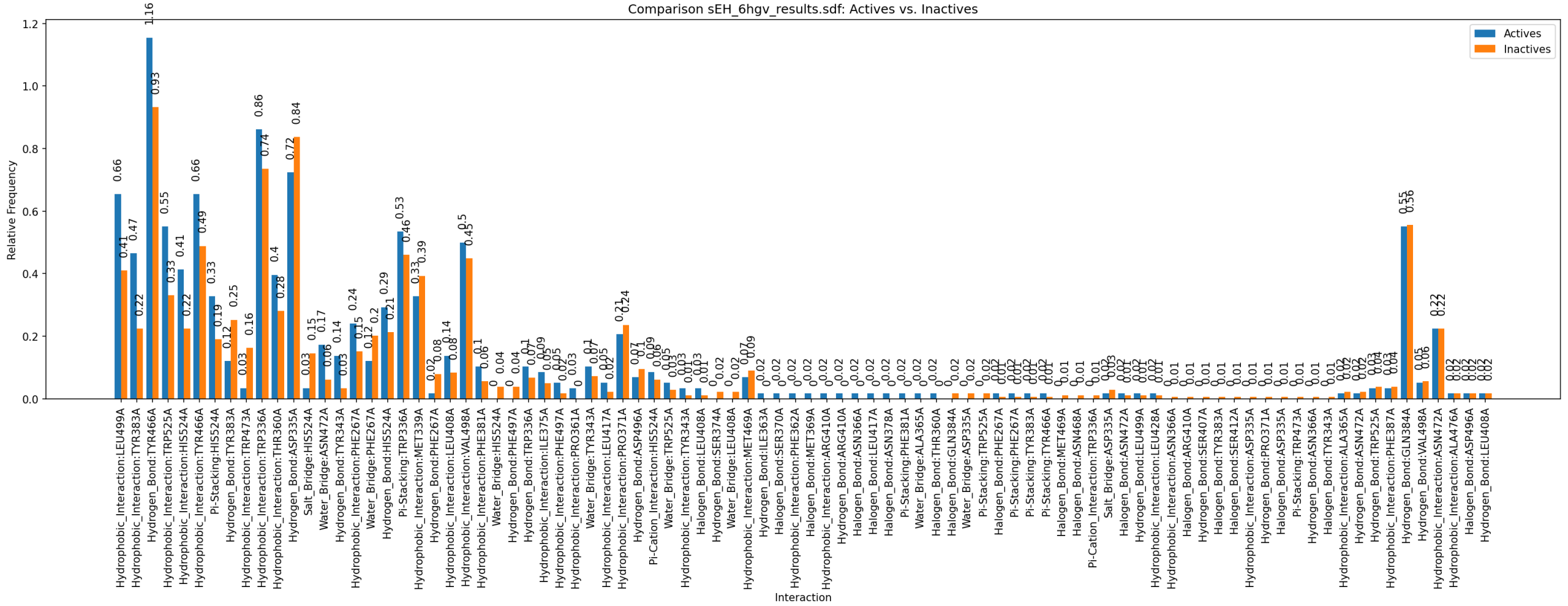


Interactions from 6HGV shown in **red**, known important interactions denoted by **green** arrow.

SEH – Actives vs Inactives 1/2

- Docking of active and inactive ligands into SEH structure (6HGV).
- Docking yielded 10 poses for every ligand
- Selection of „best“ pose where best is denoted as that pose that yields the most interactions excluding multitudes of hydrophobic interactions
- Calculate frequencies for active and inactive ligands and compare them

SEH – Actives vs Inactives 2/2



Scoring

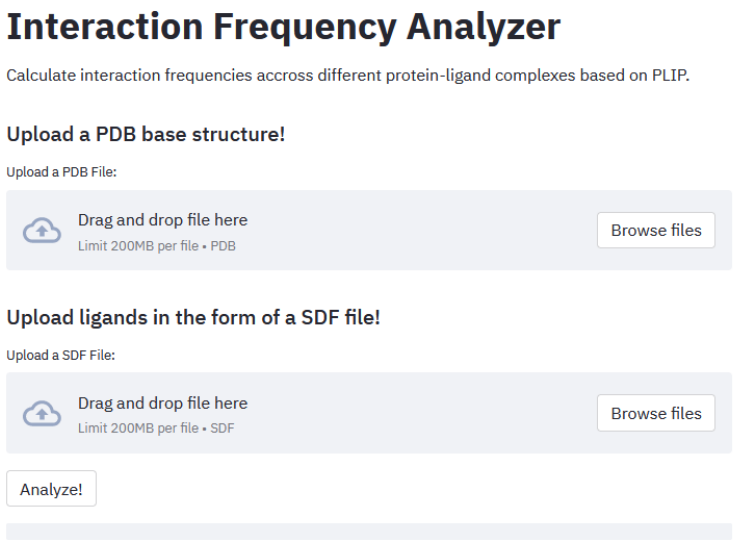
- Development of novel scoring function that is independent of the docking score and can predict if a ligand is active or inactive
- The simple case:
  - Filtering those interactions that appear more frequent than a certain threshold
  - Filtering those interactions whose difference exceeds a certain threshold
  - For a ligand for each interaction assign 0/1 if interactions is present or not
  - Score = sum of all fullfilled interactions
- Results:
  - To be continued…

The best is yet to come…

- Scoring – the more sophisticated approach:
  - Using ML methods to predict activeness
  - Preferably explainable AI approaches to get more insight
- Results:
  - Hopefully also still to come!

Try it yourself!

- GitHub repository: https://github.com/michabirklbauer/protein_docking
- DockerHub: https://hub.docker.com/r/michabirklbauer/protein_docking
- Web service: http://46.38.240.10/
- Example files:
  - /workflows/sdf/6hgv.pdb
  - /workflows/sdf/results_vs_6hgv_6A_Gold.sdf



Questions?

Thanks for your attention!

References:

- Morisseau, C. et al. "Development of an HTS assay for EPHX2 phosphatase activity and screening of nontargeted libraries." Anal. Biochem., 2013. doi: 10.1016/j.ab.2012.11.017
- Mysinger, M. et al. "A Database of Useful Decoys: Enhanced." Med. Chem., 2012. doi: 10.1021/jm300687e
- Salentin, S. et al. „PLIP: fully automated protein-ligand interaction profiler." Nucl. Acids Res., 2015. doi: 10.1093/nar/gkv315
- Karami, L. et al. „Investigation of the binding mode of 1, 3, 4-oxadiazole derivatives as amide-based inhibitors for soluble epoxide hydrolase (sEH) by molecular docking and MM-GBSA." Eur. Biophys., 2017. doi: 10.1007/s00249-016-1188-0
- Hiesinger, K. et al. „Computer-Aided Selective Optimization of Side Activities of Talinolol." ACS Med. Chem. Lett., 2019. doi: 10.1021/acsmedchemlett.9b00075