Automatic identification of important interactions and interaction-frequency-based scoring 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 experimental data and domain knowledge.
- GOAL 1: Automatic identification of important interactions based on their frequency across multiple structures of the same target protein.
- GOAL 2: Development of novel, interaction-frequency-based scoring function(s) 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, center is negatively charged & the ring positively charged).
- Pi-cation interactions: Interaction between a pi-system (aromatic ring) and a cation.
- Metal complexations: Metal ion bound to multiple other ions or molecules.

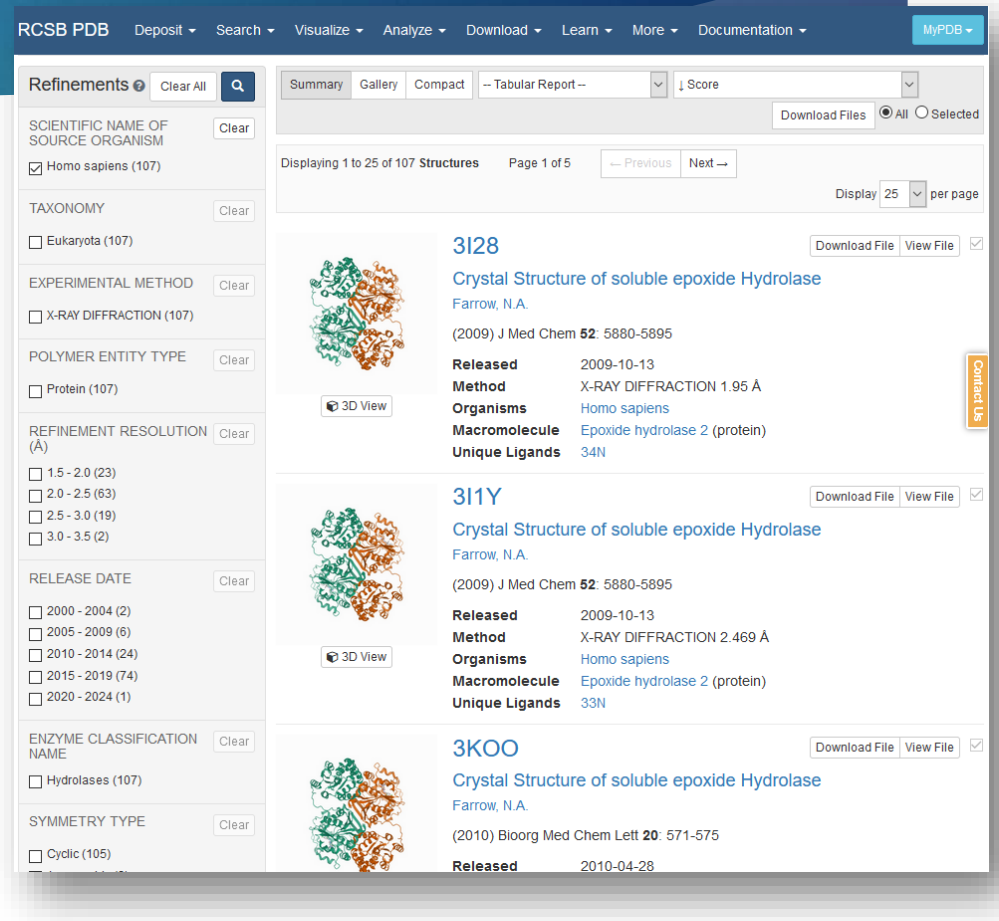
Data – Targets 1/2

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
  - Associated disease: Alzheimers
    ACHE regulates neurotransmission by hydrolysing the neurotransmitter acetylcholine.
  - Drugs: Donepezil
- Coagulation Factor XA (FXA)
  - Organism: Homo sapiens — Structures: 129
- Cyclooxygenase 1 & 2 (COX1/COX2)
  - COX1 - Organism: Ovis aries (sheep) — Structures: 25
  - COX2 - Organism: Homo sapiens — Structures: 7
    Mus musculus (mouse) — 44

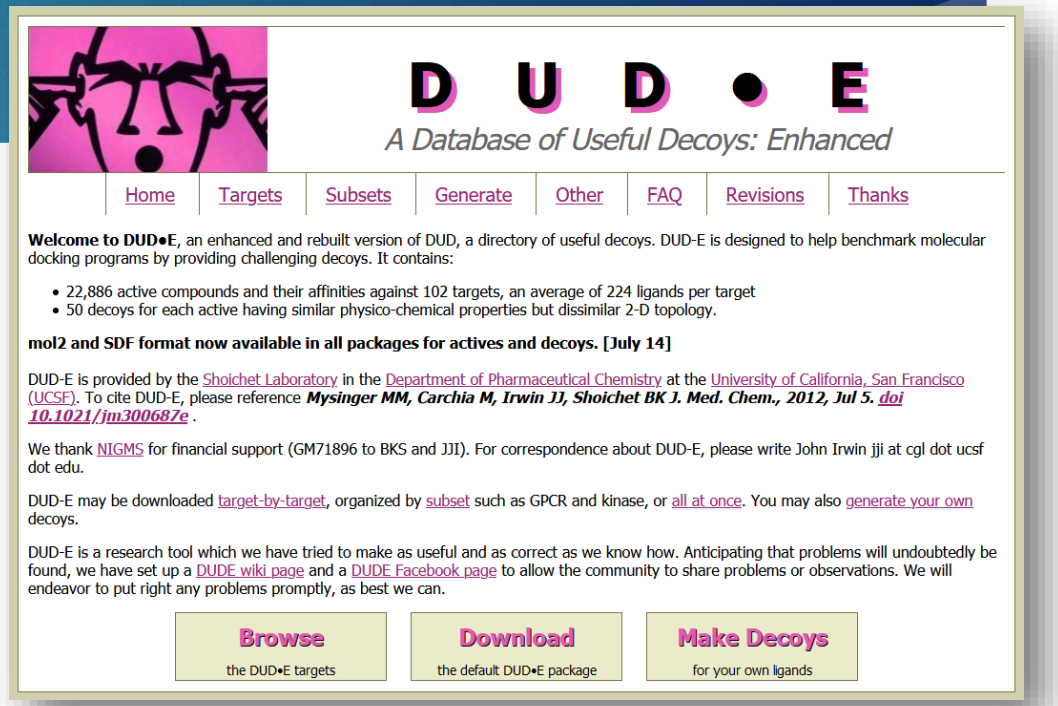
Data – Targets 2/2

- 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. SEH metabolizes arachidonic acid epoxides that play roles in blood pressure, cell growth, inflammation and pain.
  - Drugs: Ebselen (Diabetes), mostly experimental.



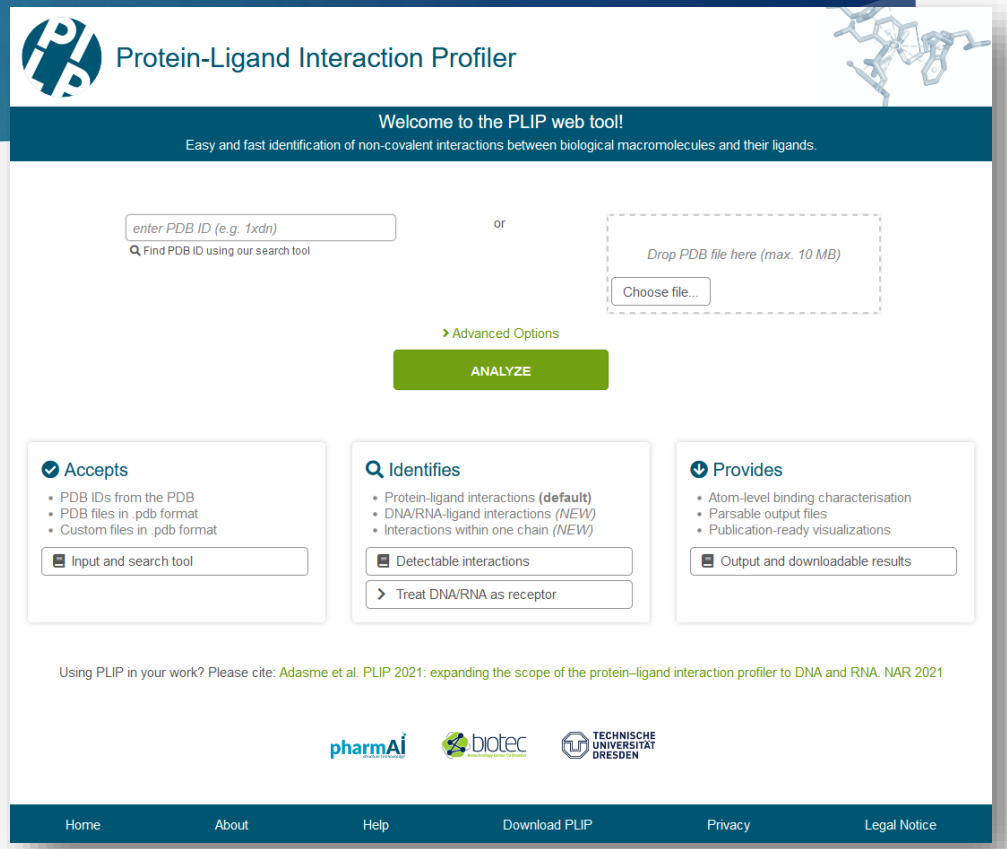
Data - Ligands

- 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.

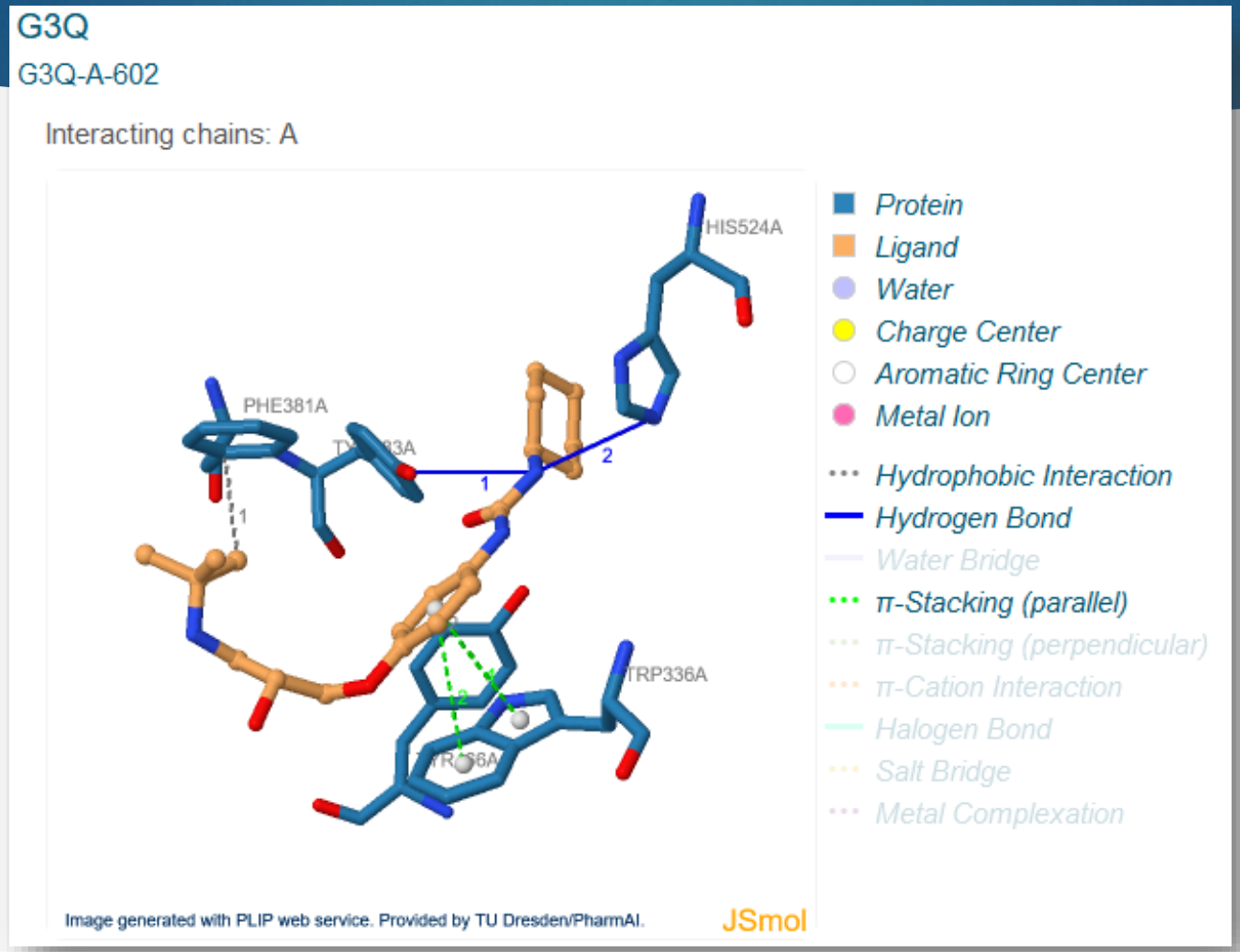


PLIP – Detection of Interactions

- 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.



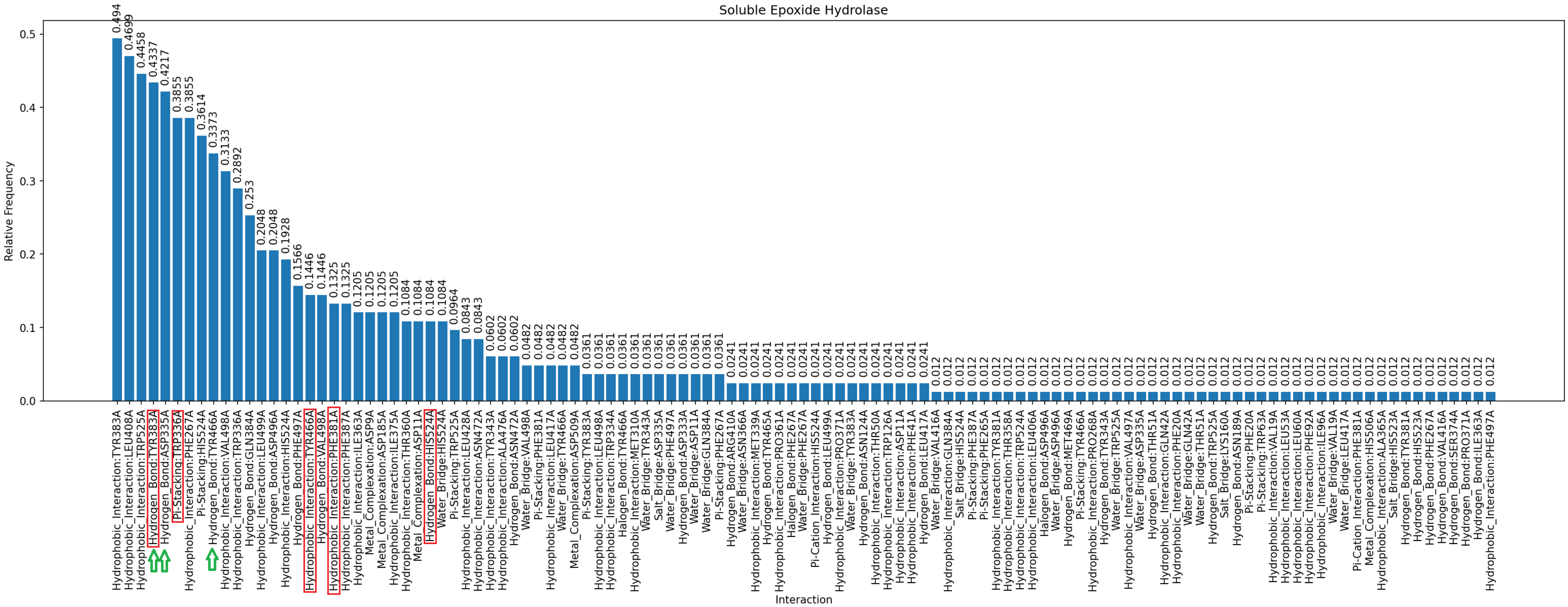
Example – SEH (6HGV, Talinolol)



Filtering of Interactions

- Not all interactions are necessarily of interest e.g. interactions with cofactors or small molecules that were needed for crystallization are unwanted – therefore filtering:
  - Removing any cofactor interactions.
  - Removing any interactions with suspicious ligands or artifacts (BioLiP blacklist).
  - Removing redundant hydrophobic interactions.

SEH Interaction Frequencies

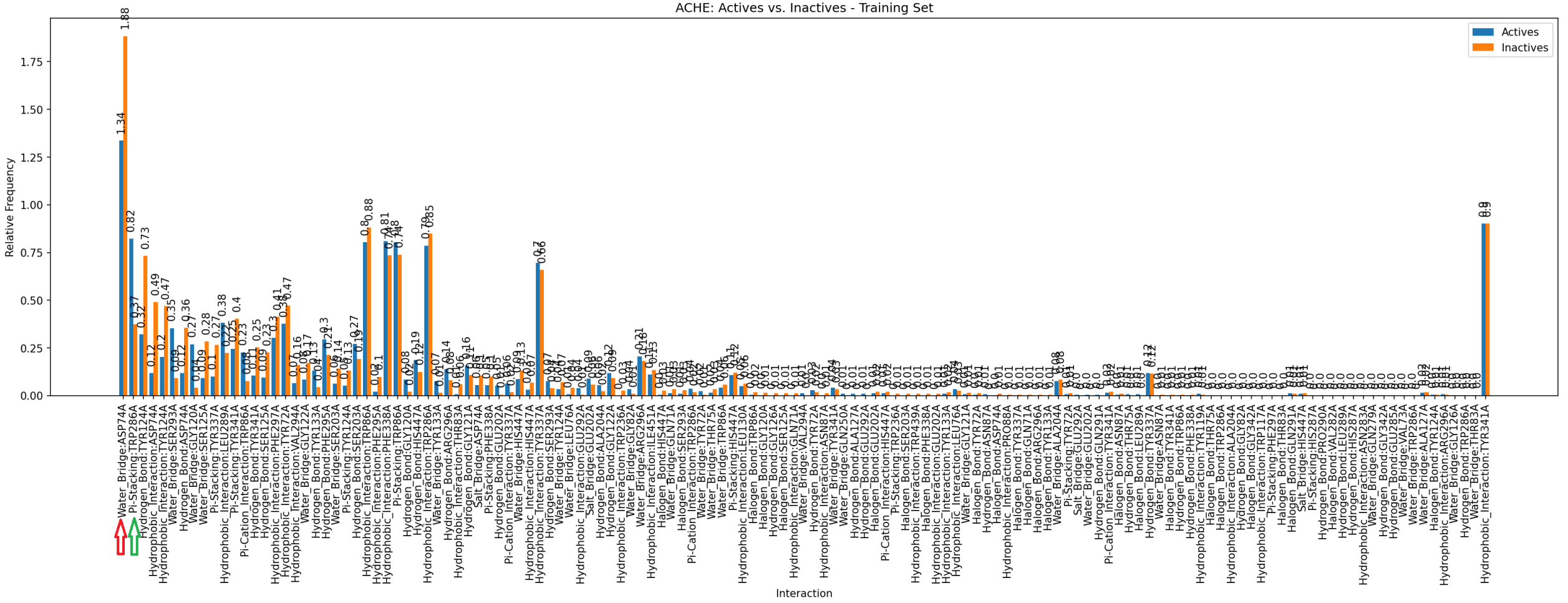


Interactions from 6HGV shown in **red**, known important interactions denoted by **green** arrow.

ACHE – Actives vs Inactives 1/2

- Docking of active and inactive ligands into ACHE structure (4EY7).
- Docking yielded 10 poses for every ligand.
- Selection of „best“ pose where best is denoted as that pose that yields the most interactions.
- Calculate frequencies for active and inactive ligands and compare them.

ACHE – Actives vs Inactives 2/2



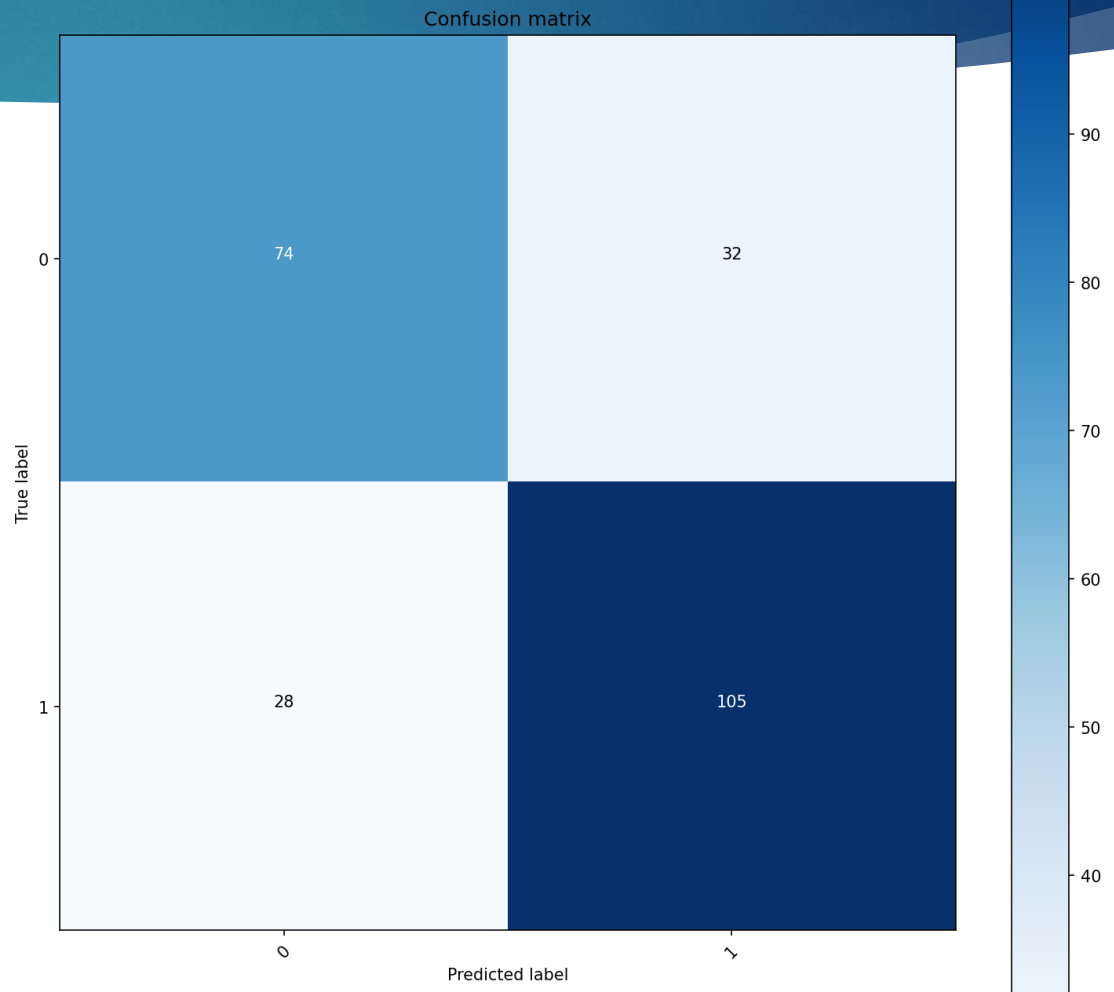
Red: Negative Interaction; **Green:** Positive Interaction.

ACHE – Scoring 1/3

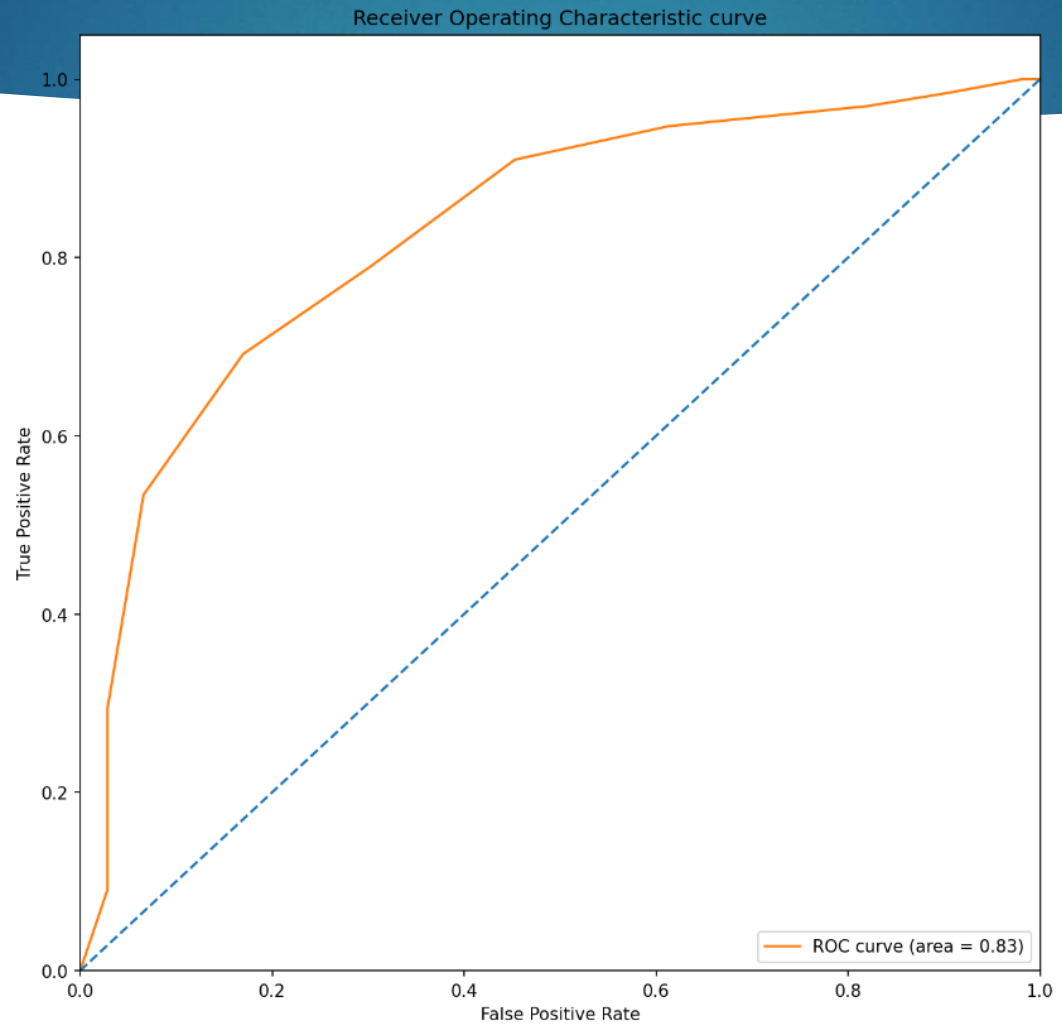
- Features: Select positive and negative interactions that have frequencies greater than *m* and differ more than *n* between actives and inactives.
  - Parameters m & n determined by grid search.
- Four different scoring strategies:
  - Score is the number of positive interactions in a protein-ligand complex.
  - Score is the number of positive interactions minus the number of negative interactions in a protein-ligand complex.
  - Score is the sum of the frequencies of the positive interactions in a protein-ligand complex.
  - Score is the sum of the frequencies of the positive interactions minus the sum of the frequencies of the negative interactions in a protein-ligand complex.
- Results:
  - Evaluated in terms of accuracy (ACC), area under the ROC curve (AUC), false positive rate (FPR), enrichment factor (EF; how much more frequent are active molecules in the predicted actives compared to the original dataset) and yield of actives (Ya; fraction of true actives among predicted actives).

ACHE – Scoring 2/3

- Baseline accuracy: 0.556
- Results on the test dataset:
  - ACC: 0.749
  - AUC: 0.831
  - FPR: 0.302
    - TP: 105
    - TN: 74
    - FP: 32
    - FN: 28
  - EF: 1.377
  - Ya: 0.766



ACHE – Scoring 3/3



Scoring – General Overview

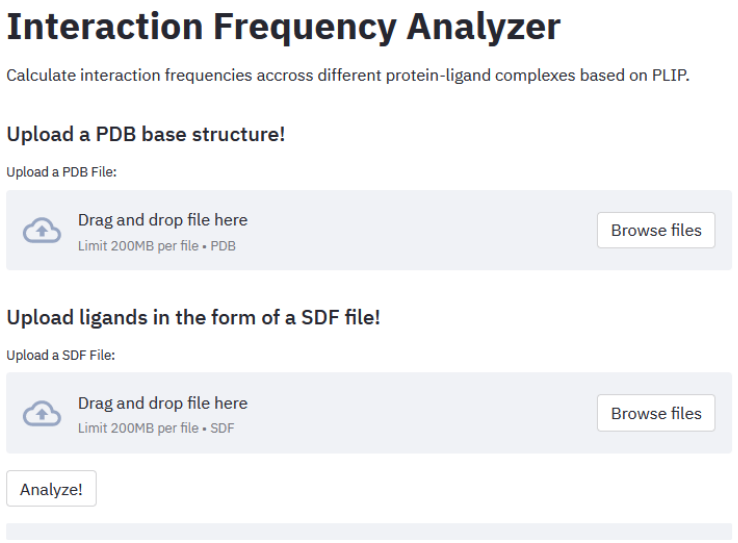
Target	Baseline	ACC	FPR	Ya	EF
ACHE	0.556	0.749	0.302	0.766	1.377
COX1	0.711	0.702	0.034	0.5	1.676
DPP4	0.517	0.651	0.330	0.675	1.299
MAOB	0.62	0.73	0	1	2.871
SEH	0.754	0.771	0.162	0.5	2.182

Conclusion & Outlook

- Interaction-frequency-based scoring "works", how well depends on the target.
- Approach should therefore be seen as supporting tool in virtual screening rather than a standalone solution.
- Interaction-frequency-based scoring - The more sophisticated approach:
  - Using ML methods to predict activeness.
  - Weighting different interaction types.
  - Preferably explainable AI approaches to get more insight on importance of different interactions.

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