

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.

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation MyPDB

Refinements Clear All

SCIENTIFIC NAME OF SOURCE ORGANISM
☒ Homo sapiens (107) Clear

TAXONOMY
☐ Eukaryota (107) Clear

EXPERIMENTAL METHOD
☐ X-RAY DIFFRACTION (107) Clear

POLYMER ENTITY TYPE
☐ Protein (107) Clear

REFINEMENT RESOLUTION (Å)
☐ 1.5 - 2.0 (23)
☐ 2.0 - 2.5 (63)
☐ 2.5 - 3.0 (19)
☐ 3.0 - 3.5 (2) Clear

RELEASE DATE
☐ 2000 - 2004 (2)
☐ 2005 - 2009 (6)
☐ 2010 - 2014 (24)
☐ 2015 - 2019 (74)
☐ 2020 - 2024 (1) Clear

ENZYME CLASSIFICATION NAME
☐ Hydrolases (107) Clear

SYMMETRY TYPE
☐ Cyclic (105) Clear

Summary Gallery Compact -- Tabular Report -- Score

Download Files ☒ All ☐ Selected

Displaying 1 to 25 of 107 Structures Page 1 of 5 -- Previous Next --

Display 25 per page

3I28 Download File View File ☒

Crystal Structure of soluble epoxide Hydrolase
 Farrow, N.A.
 (2009) J Med Chem **52**: 5880-5895

Released 2009-10-13
Method X-RAY DIFFRACTION 1.95 Å
Organisms Homo sapiens
Macromolecule Epoxide hydrolase 2 (protein)
Unique Ligands 34N

3I1Y Download File View File ☒

Crystal Structure of soluble epoxide Hydrolase
 Farrow, N.A.
 (2009) J Med Chem **52**: 5880-5895

Released 2009-10-13
Method X-RAY DIFFRACTION 2.469 Å
Organisms Homo sapiens
Macromolecule Epoxide hydrolase 2 (protein)
Unique Ligands 33N


3KOO Download File View File ☒

Crystal Structure of soluble epoxide Hydrolase
 Farrow, N.A.
 (2010) Bioorg Med Chem Lett **20**: 571-575

Released 2010-04-28

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.



D U D • E
A Database of Useful Decoys: Enhanced

[Home](#)
[Targets](#)
[Subsets](#)
[Generate](#)
[Other](#)
[FAQ](#)
[Revisions](#)
[Thanks](#)

Welcome to DUD•E, an enhanced and rebuilt version of DUD, a directory of useful decoys. DUD-E is designed to help benchmark molecular docking programs by providing challenging decoys. It contains:

- 22,886 active compounds and their affinities against 102 targets, an average of 224 ligands per target
- 50 decoys for each active having similar physico-chemical properties but dissimilar 2-D topology.

mol2 and SDF format now available in all packages for actives and decoys. [July 14]

DUD-E is provided by the [Shoichet Laboratory](#) in the [Department of Pharmaceutical Chemistry](#) at the [University of California, San Francisco \(UCSF\)](#). To cite DUD-E, please reference **Mysinger MM, Carchia M, Irwin JJ, Shoichet BK J. Med. Chem., 2012, Jul 5. doi 10.1021/jm300687e**.

We thank [NIGMS](#) for financial support (GM71896 to BKS and JJI). For correspondence about DUD-E, please write John Irwin jji at cgl dot ucsf dot edu.

DUD-E may be downloaded [target-by-target](#), organized by [subset](#) such as GPCR and kinase, or [all at once](#). You may also [generate your own](#) decoys.

DUD-E is a research tool which we have tried to make as useful and as correct as we know how. Anticipating that problems will undoubtedly be found, we have set up a [DUD-E wiki page](#) and a [DUD-E Facebook page](#) to allow the community to share problems or observations. We will endeavor to put right any problems promptly, as best we can.

[Browse](#)
 the DUD•E targets

[Download](#)
 the default DUD•E package

[Make Decoys](#)
 for your own ligands

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.

Protein-Ligand Interaction Profiler

Welcome to the PLIP web tool!
Easy and fast identification of non-covalent interactions between biological macromolecules and their ligands.

enter PDB ID (e.g. 1xdn) or Drop PDB file here (max. 10 MB)
Find PDB ID using our search tool Choose file...

Advanced Options
ANALYZE

Accepts

- PDB IDs from the PDB
- PDB files in .pdb format
- Custom files in .pdb format

Input and search tool

Identifies

- Protein-ligand interactions (**default**)
- DNA/RNA-ligand interactions (**NEW**)
- Interactions within one chain (**NEW**)

Detectable interactions
Treat DNA/RNA as receptor

Provides

- Atom-level binding characterisation
- Parsable output files
- Publication-ready visualizations

Output and downloadable results

Using PLIP in your work? Please cite: Adasme et al. PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA. NAR 2021

pharmAI biotec TECHNISCHE UNIVERSITÄT DRESDEN

Home About Help Download PLIP Privacy Legal Notice

Example – SEH (6HGV, Talinolol)

G3Q

G3Q-A-602

Interacting chains: A

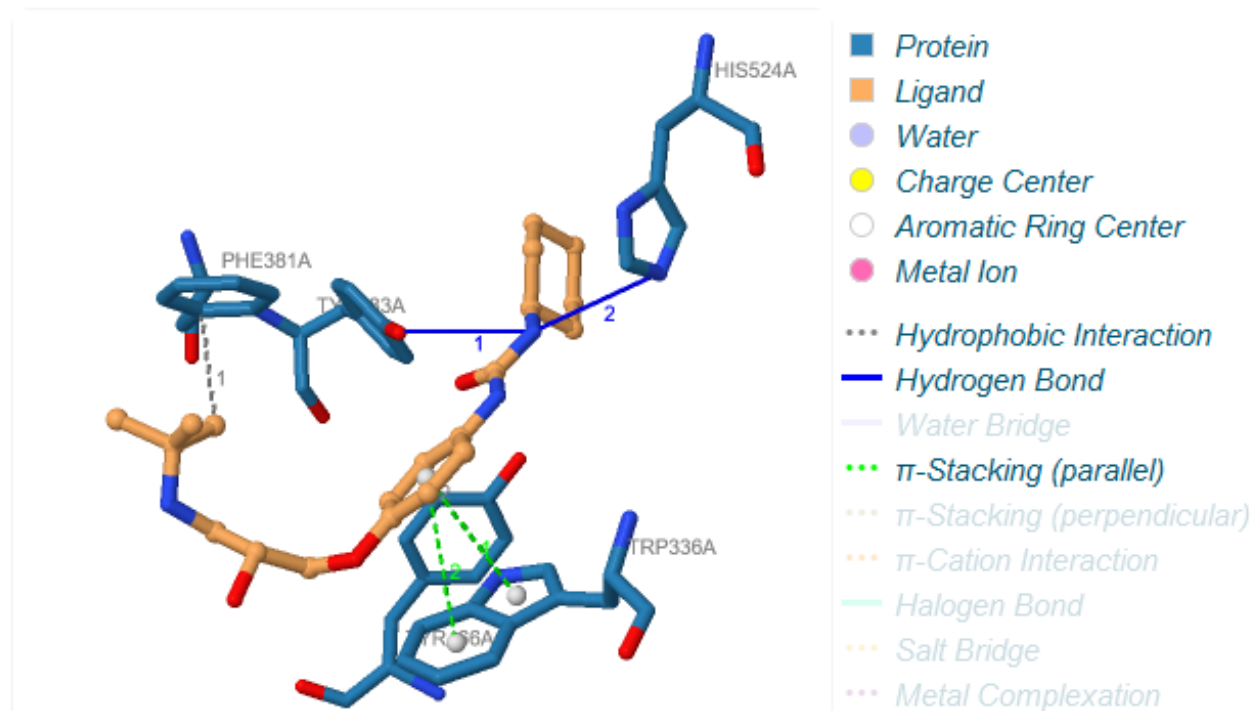


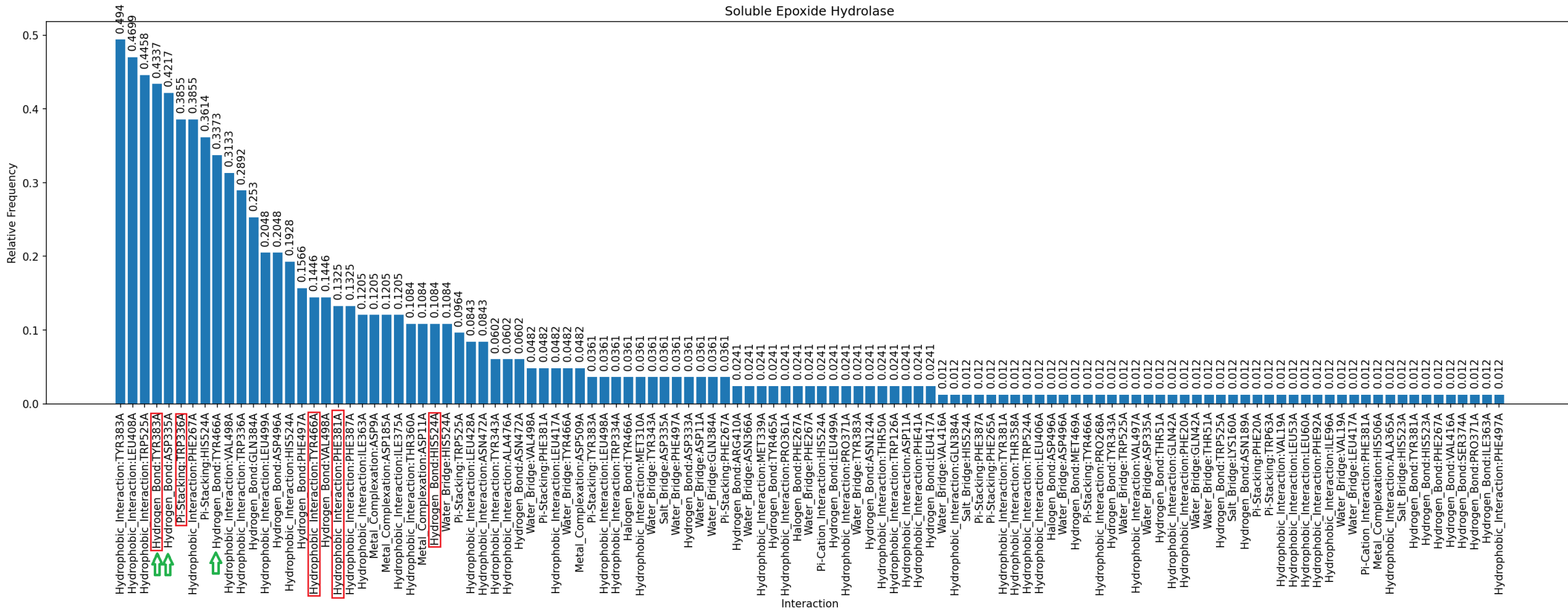
Image generated with PLIP web service. Provided by TU Dresden/PharmAI.

JSmol

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.

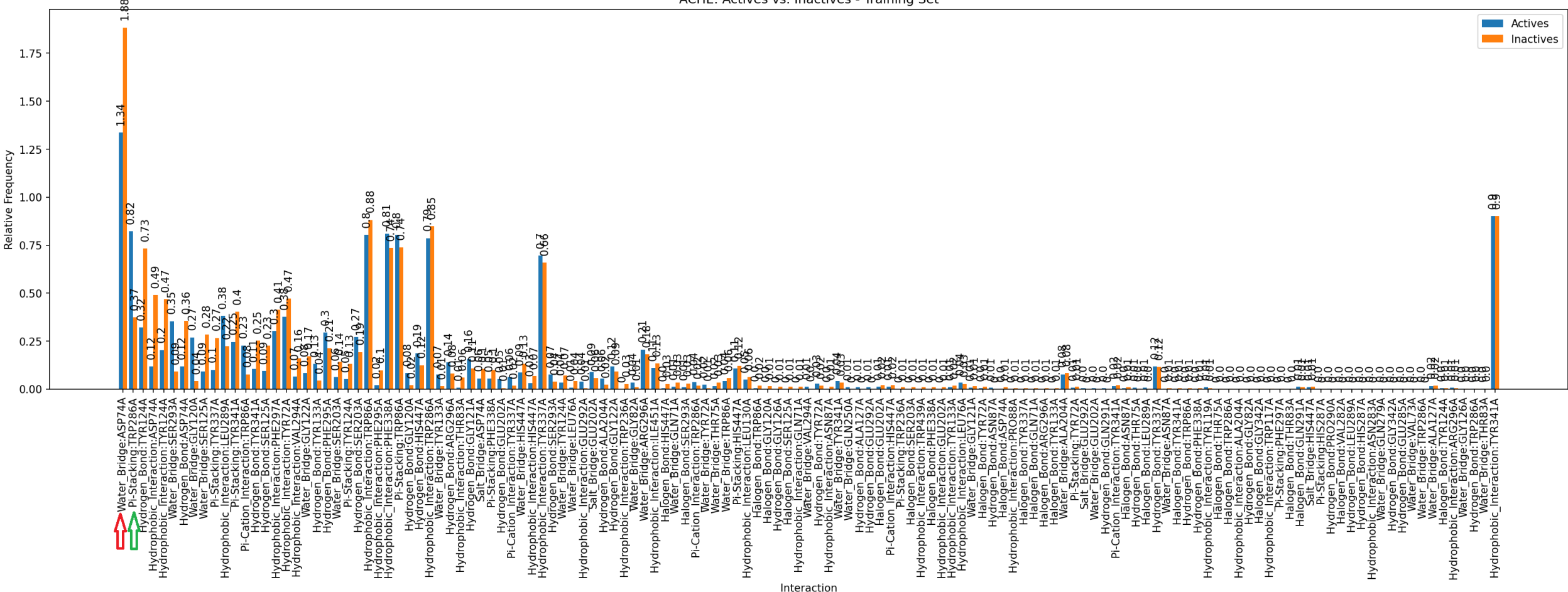
ACHE – Actives vs Inactives 1/2

- ▶ 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.

12

ACHE – Actives vs Inactives 2/2

ACHE: Actives vs. Inactives - Training Set



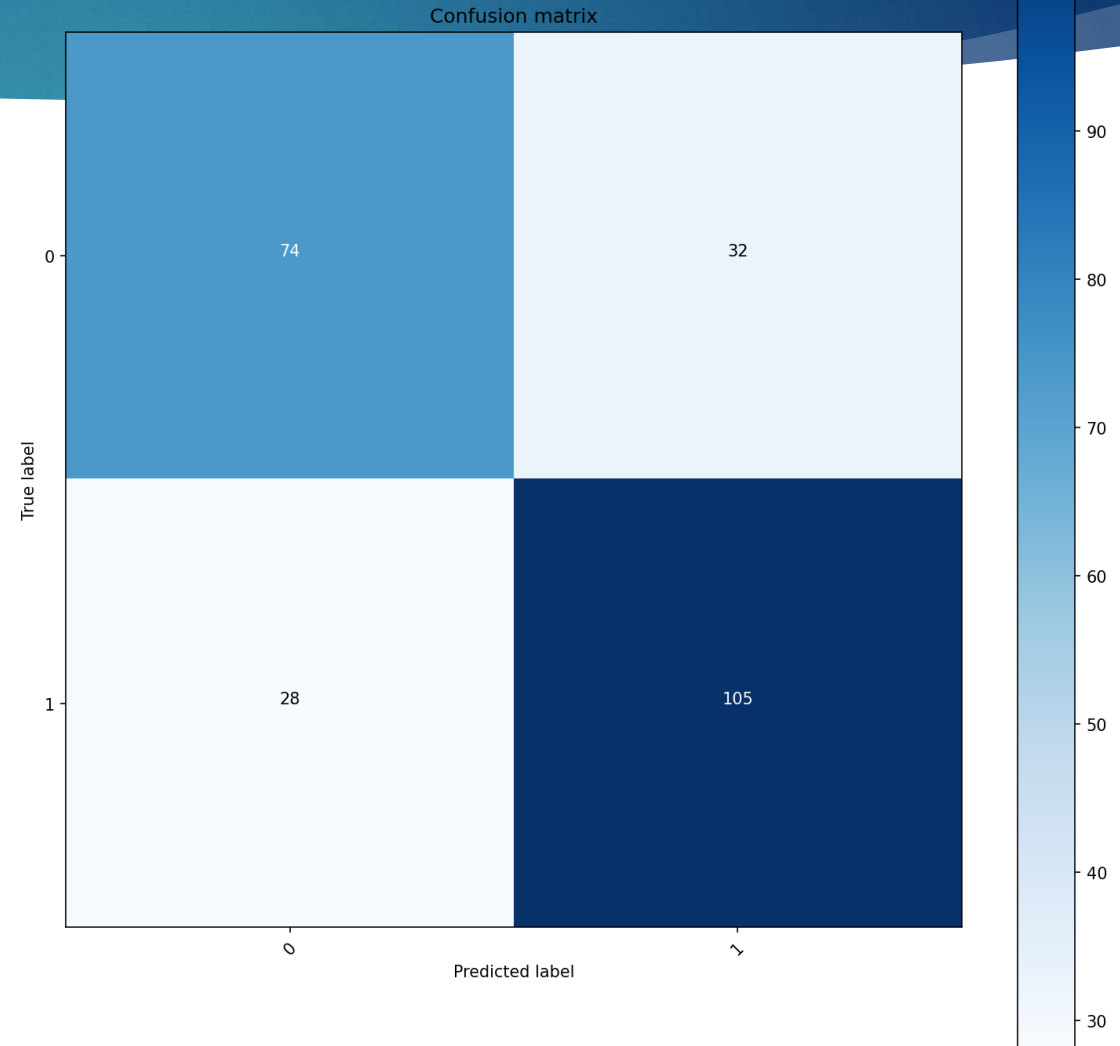
Red: Negative Interaction; **Green:** Positive Interaction.

ACHE – Scoring 1/3

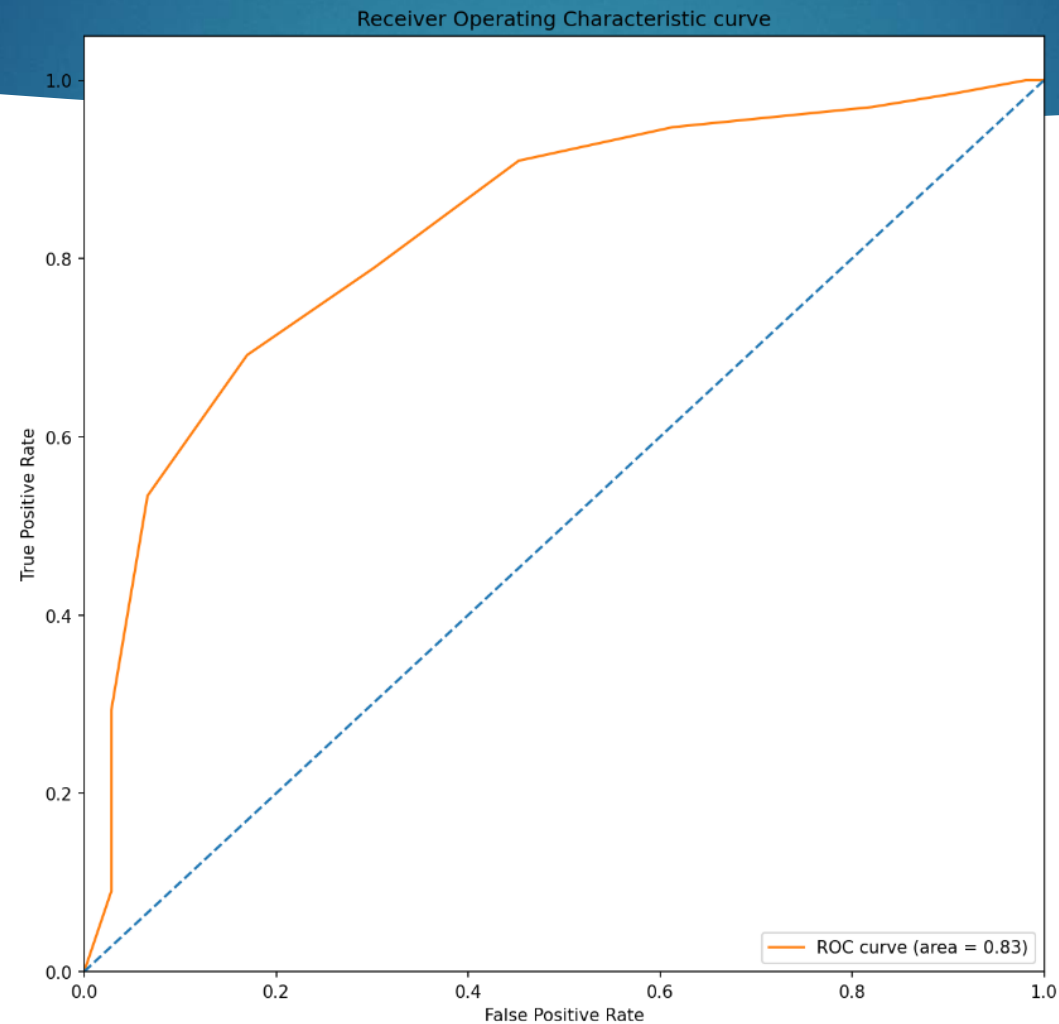
- ▶ Features: Select positive and negative interactions that have frequencies greater than m and differ more than n between actives and inactives.
 - ▶ Parameters m & n determined by grid search.
- ▶ Four different scoring strategies:
 - ▶ Score is the number of positive interactions in a protein-ligand 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 (Y_a ; 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	Y _a	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

Interaction Frequency Analyzer

Calculate interaction frequencies across different protein-ligand complexes based on PLIP.

Upload a PDB base structure!

Upload a PDB File:



Drag and drop file here

Limit 200MB per file • PDB

Browse files

Upload ligands in the form of a SDF file!

Upload a SDF File:



Drag and drop file here

Limit 200MB per file • SDF

Browse files

Analyze!

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/acsmchemlett.9b00075