



KLIFS: Making Kinase Structures Work

RCS CICAG Workshop

Albert J. Kooistra, University of Copenhagen

Dominique Sydow & Andrea Volkamer,
Charité – Universitätsmedizin Berlin

www.volkamerlab.org

Outline

FAIR and OpenSource developments

- VolkamerLab
- TeachOpenCADD
- KLIFS-based kinase work
 - OpenCADD - KLIFS
 - KiSSim
 - TeachOpenCADD - kinase edition
 - KinFraglib

- Hands-on code demonstration

- If you'd like to follow along the Jupyter notebook, please already open it in [Google Colab](#) and run the "Installation (Google Colab)" code cell

- Installation (Google Colab)

```
[ ] # If the notebook is run on Google Colab
# install condacolab and kissim
try:
    import google.colab
    !pip install condacolab
    import condacolab
    condacolab.install()
    !mamba install -yq kissim
except ModuleNotFoundError:
    pass
```

Volkamer Lab - Open Source Developments



Structural bioinformatics & *in silico* toxicology

Method development & application in CADD and predictive toxicology. Focus on integration of structural and chemical information, machine learning & cancer research (kinases)



<https://github.com/volkamerlab>

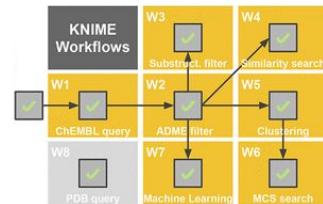
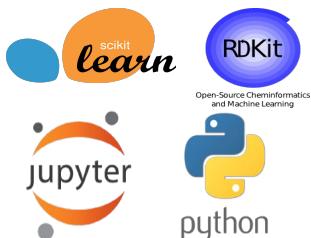
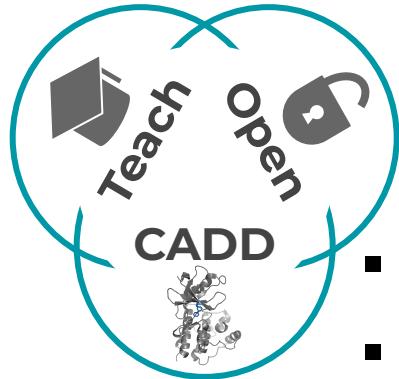


- Teaching material
 - Teaching platform for computer-aided drug design: [TeachOpenCADD](#)
 - Basics in python programming with focus on [AI in medicine](#) (together with Ritter lab)
 - Python library for structural bio- and cheminformatics: [OpenCADD](#)
 - Machine learning-based activity, property and toxicity prediction
 - SMILES augmentation for accurate molecular prediction: [maxsmi](#)
 - Conformal prediction (CP) framework: [KnowTox](#), [CP-Tox21](#), [CP-recalibration](#)
 - Kinase related work: [Kissim](#), [KinFragLib](#), [openkinome](#) (together with Chodera lab, MSKCC, NY)

TeachOpenCADD

Using Open Source Packages and Data

- Pack of **28** introductory lessons for central cheminformatics and structural bioinformatics topics
 - Database queries
 - Chemical properties, compound similarity, virtual screening
 - Docking, molecular dynamics simulations, ...
- Conceived as “from students for students”
- Starting material to solve diverse research questions
- Each lesson is comprised of a Jupyter Notebook
 - Theory + code in a single executable document
- Availability: <https://github.com/volkamerlab/TeachOpenCADD>
Web: <https://projects.volkamerlab.org/teachopencadd/>



[TeachOpenCADD: a teaching platform for computer-aided drug design using open source packages and data.](#) Sydow, Morger, Driller, Volkamer. *J Cheminf* (2019)

[TeachOpenCADD-KNIME: A Teaching Platform for Computer-Aided Drug Design Using KNIME Workflows.](#) Sydow, Wichmann, Rodríguez-Guerra, Goldmann, Landrum, Volkamer. *JCIM* (2019)

[Teaching Computer-Aided Drug Design Using TeachOpenCADD.](#) Sydow, Rodríguez-Guerra, Volkamer. *Teaching Programming Across the Chemistry Curriculum ACS* (2021)

[TeachOpenCADD 2021: Open Source and FAIR Python Pipelines to Assist in Structural Bioinformatics and Cheminformatics Research.](#) Sydow and Rodríguez-Guerra, Kimber, Schaller, Taylor, Chen, Leja, Misra, Wichmann, Ariamajd, Volkamer. *NAR* (2022)

Structure and Content of Each Talktorial

T002 · Molecular filtering: ADME and lead-likeness criteria

Authors:

- Michele Wichmann, CADD seminars 2017, Charité/FU Berlin
- Mathias Weinberg, CADD seminars 2018, Charité/FU Berlin
- Dominique Sydow, 2018-2020, Volkamer lab, Charité
- Andrea Volkamer, 2018-2020, Volkamer lab, Charité

Talktorial T002: This talktorial is part of the TeachOpenCADD pipeline described in the [first TeachOpenCADD paper](#), comprising of talktutorials T001-T010.

ADME - absorption, distribution, metabolism, and excretion

Pharmacokinetics are mainly divided into four steps: Absorption, Distribution, Metabolism, and Excretion. These are summarized as **ADME**. Often, ADME also includes Toxicology and is thus referred to as ADMET or ADMETox. Below, the ADME steps are discussed in more detail ([Wikipedia](#) and [Mol Pharm. \(2010\). 7\(5\): 1388-1405](#)).

Absorption: The amount and the time of drug-uptake into the body depends on multiple factors which can vary between individuals and their conditions as well as on the properties of the substance. Factors such as (poor) compound solubility, gastric emptying time, intestinal transit time, chemical (in-)stability in the stomach, and (in-)ability to permeate the intestinal wall can all influence the extent to which a drug is absorbed after e.g. oral administration, inhalation, or contact to skin.

Distribution: The distribution of an absorbed substance, i.e. within the body, between blood and different tissues, and crossing the blood-brain barrier are affected by regional blood flow rates, molecular size and polarity of the compound, and binding to serum proteins and transporter enzymes. Critical effects in toxicology can be the accumulation of highly apolar substances in fatty tissue, or crossing of the blood-brain barrier.

Metabolism: After entering the body, the compound will be metabolized. This means that only part of this compound will actually reach its target. Mainly liver and kidney enzymes are responsible for the break-down of xenobiotics (substances that are extrinsic to the body).

Excretion: Compounds and their metabolites need to be removed from the body via excretion, usually through the kidneys (urine) or in the feces. Incomplete excretion can result in accumulation of foreign substances or adverse interference with normal metabolism.

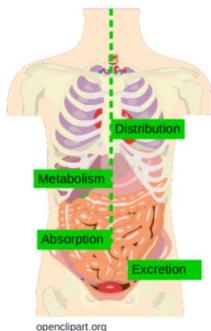


Figure 1: ADME processes in the human body (figure taken from [Openclipart](#) and adapted).

Discussion

In this talktorial, we have learned about Lipinski's Ro5 as a measure to estimate a compound's oral bioavailability and we have applied the rule on a dataset using rdkit. Note that drugs can also be administered via alternative routes, i.e. inhalation, skin penetration and injection.

In this talktorial, we have looked at only one of many more ADME properties. Webservers such as [SwissADME](#) give a more comprehensive view on compound properties.

Aim of this talktorial

In the context of drug design, it is important to filter candidate molecules by e.g. their physicochemical properties. In this talktorial, the compounds acquired from ChEMBL (Talktorial 001) will be filtered by Lipinski's rule of five to keep only orally bioavailable compounds.

Contents in Theory

- ADME - absorption, distribution, metabolism, and excretion
- Lead-likeness and Lipinski's rule of five (Ro5)
- Radar charts in the context of lead-likeness

Contents in Practical

- Define and visualize example molecules
- Calculate and plot molecular properties for Ro5
- Investigate compliance with Ro5
- Apply Ro5 to the EGFR dataset
- Visualize Ro5 properties (radar plot)

References

- ADME criteria ([Wikipedia](#) and [Mol Pharm. \(2010\). 7\(5\): 1388-1405](#))
- [SwissADME](#) webserver
- What are lead compounds? ([Wikipedia](#))
- What is the LogP value? ([Wikipedia](#))
- Lipinski et al. "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." ([Adv. Drug Deliv. Rev.](#) (1997), 23: 3-25)
- Ritchie et al. "Graphical representation of ADME-related molecule properties for medicinal chemists" ([Drug Discov. Today](#) (2011), 16: 65-72)

Talktorial sections

- Aim, content & references
- Theory
- Practical
- Discussion
- Quiz

```
In [16]: molecules["molecular_weight"] = molecules["ROMol"].apply(Descriptors.ExactMolWt)
molecules["n_hba"] = molecules["ROMol"].apply(Descriptors.NumHAcceptors)
molecules["n_hbd"] = molecules["ROMol"].apply(Descriptors.NumHDonors)
molecules["logP"] = molecules["ROMol"].apply(Descriptors.MolLogP)
# Colors are used for plotting the molecules later
molecules["color"] = ["red", "green", "blue", "cyan"]
# NBVAL_CHECK_OUTPUT
molecules[["molecular_weight", "n_hba", "n_hbd", "logP"]]
```

```
Out[16]:
```

	molecular_weight	n_hba	n_hbd	logP
0	1201.841368	12	5	3.26900
1	306.184447	4	1	1.68492
2	536.438202	0	0	12.60580
3	314.224580	2	2	5.84650

```
In [8]: # Full preview
molecules
```

```
Out[8]:
```

	name	smiles	ROMol	molecular_weight	n_hba	n_hbd	lc
0	cyclosporine	CCC1C(=O)N(CC(=O)N(C(C(=O)NC(C(=O)N(C(C(=O)NC(=O)C1)C2=CC=C(C=C2)C=C3=NC4=C(N2)C=C(C=C4)C		1201.841368	12	5	3.
1	clozapine	CN1CCN(CC1)C2=C3C=CC=CC3=NC4=C(N2)C=C(C=C4)C		306.184447	4	1	1.
2	beta-carotene	CC1=C(C(CCC1)(C)C)C=CC(=CC=CC=C(C)C=CC=...		536.438202	0	0	1.
3	cannabidiol	CCCCCC1=CC(=C(C=C1)O)C2C=C(CCC2C(=C)C)O		314.224580	2	2	5.

Explanatory text
Executable code
Code output

All-in-one
Jupyter Notebook

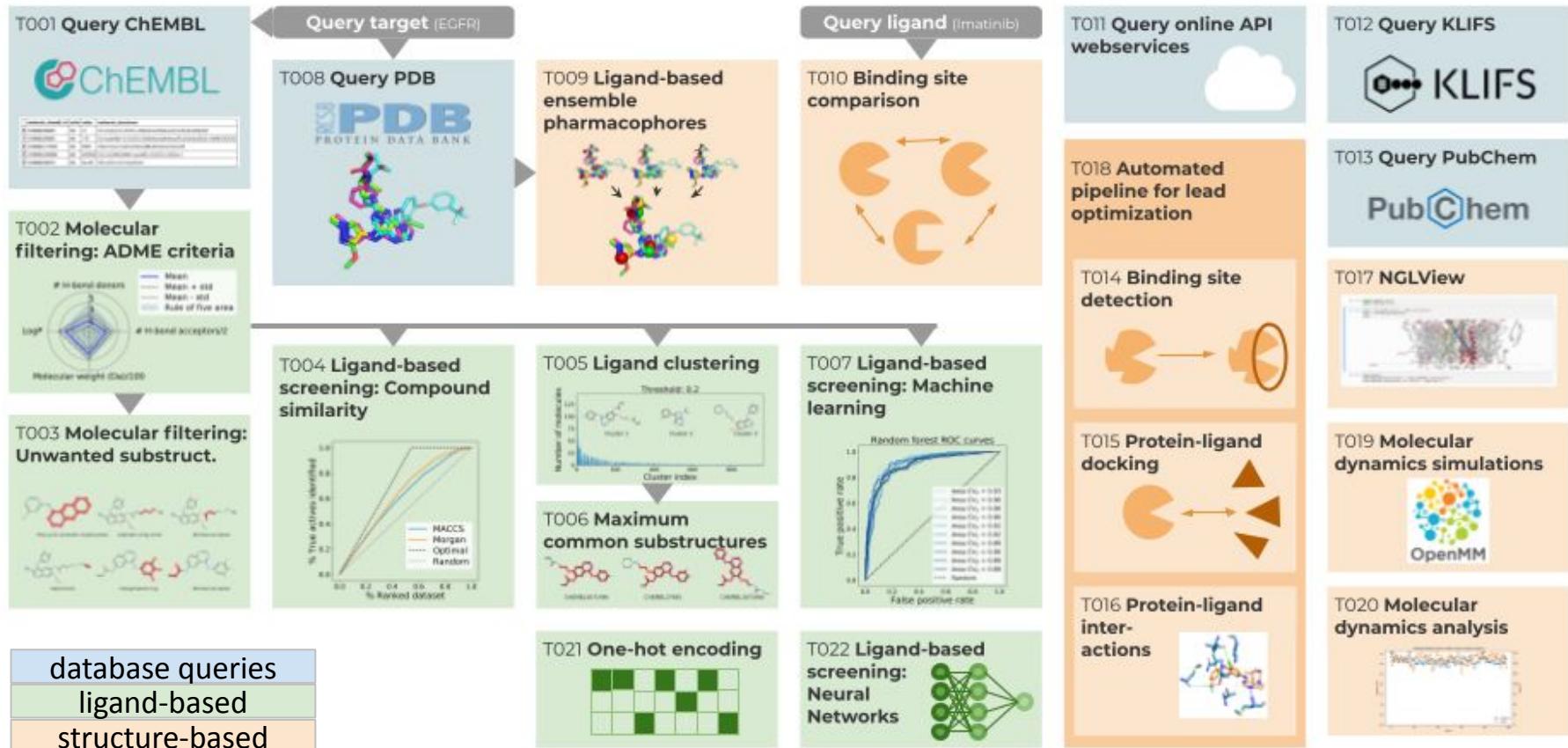


Quiz

- In what way can the chemical properties described by the Ro5 affect ADME?
- Find or design a molecule which violates three or four rules.
- How can you plot information for an additional molecule in the radar charts that we have created in this talktorial?

TeachOpenCADD - Version 2021.12.1

Open Source and FAIR Python Pipelines [\(Sydow, et al., NAR, 2022, qkac267\)](#) for Structural Bioinformatics and Cheminformatics Research



Github: <https://github.com/volkamerlab/TeachOpenCADD>
Web: <https://projects.volkamerlab.org/teachopencadd/>

New [kinase similarity edition](#): T23-28!

Challenges in Kinase Research

Kinases as drug targets

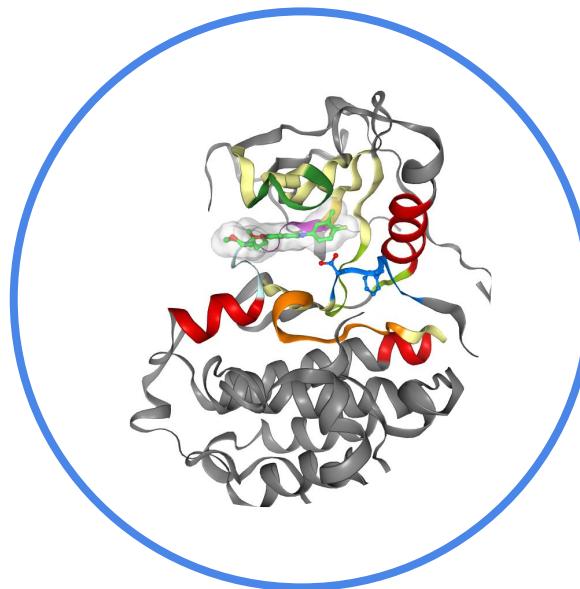
- Kinase-mediated phosphorylation is critical for signal transduction
- Dysregulation is linked to cancer, inflammation, metabolic, autoimmune, and neurological conditions

Advantages

- Well-studied
- High structural coverage

Challenges

- Competition with ATP
- Selectivity
- Physicochemical properties
- Intellectual property



Computational tools

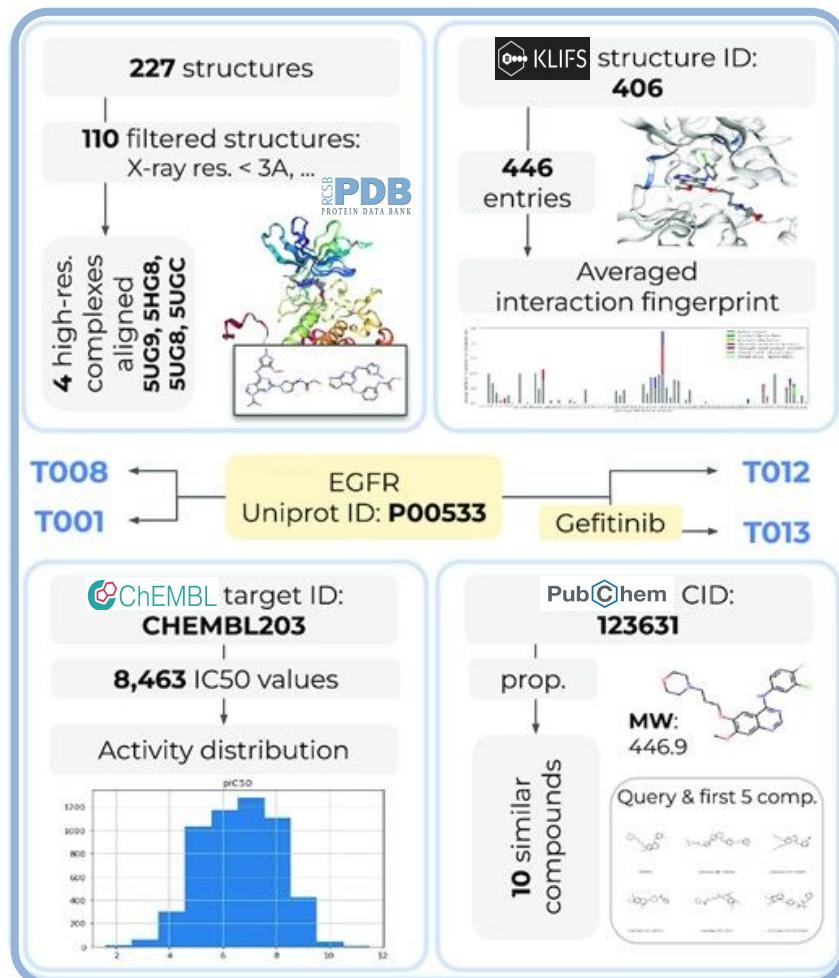
- Kinase data acquisition
- Virtual screening
- Kinome-wide off-target prediction
- Fragment-based drug design

Mortenson et al. *Methods in Enzymol.* 2014

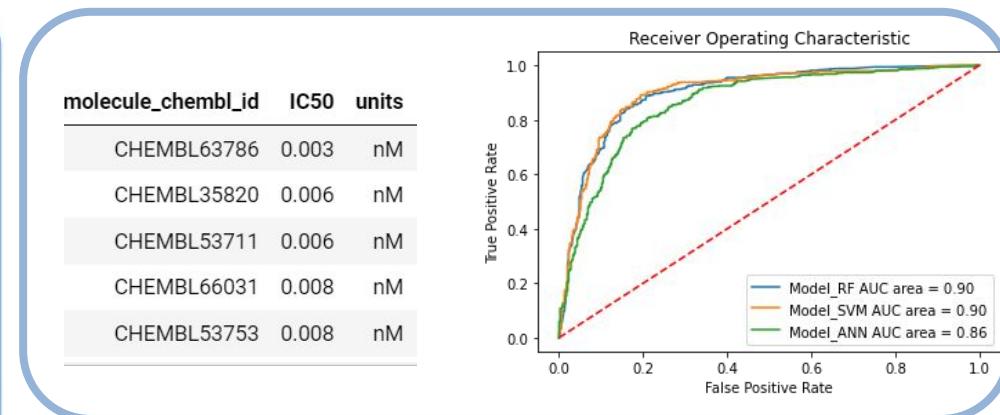
TeachOpenCADD Kinase Applications

Data Queries and Virtual Screening

Query kinase data by Uniprot ID or name



EGFR activity prediction: ML model (T007)



Lead optimization pipeline (**T018**): EGFR case study

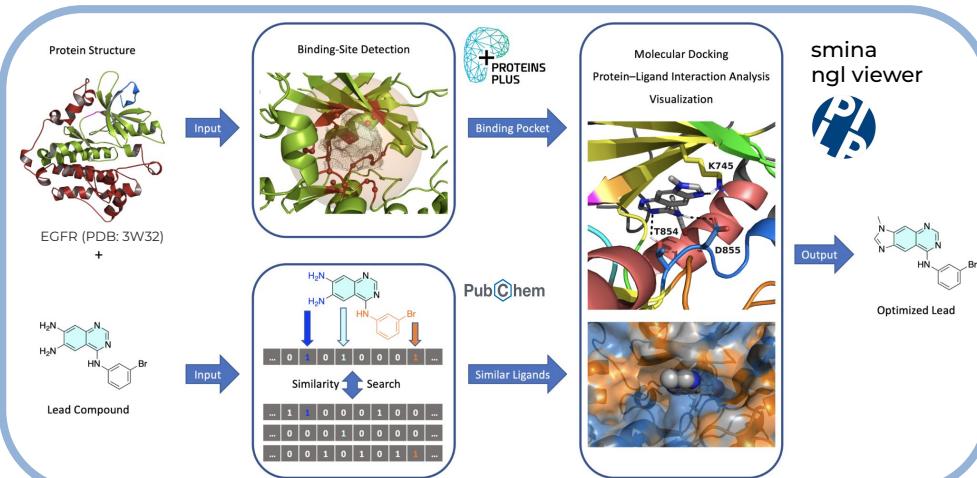
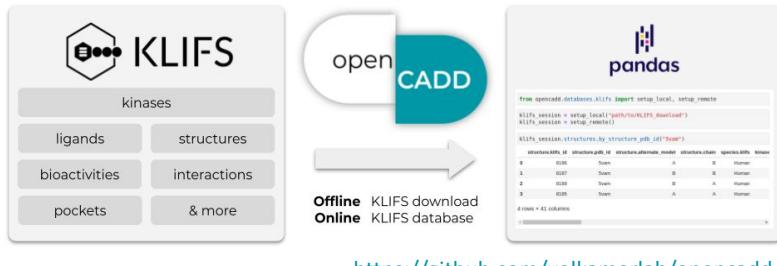


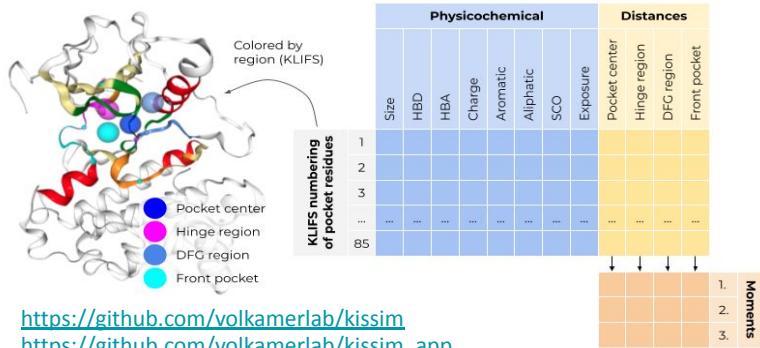
Figure taken from Sydow D, et al. NAR (2022), gkac267, DOI:

FAIR and Open Source Prediction & Analysis Tools for the Kinome

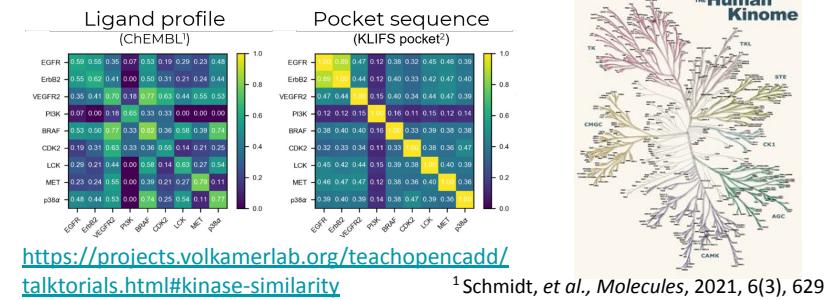
OpenCADD-KLIFS: Integration of the KLIFS data into workflows to facilitate computational kinase research, Sydow et al., JOSS, 7(70), 3951, 2021, [DOI](#)



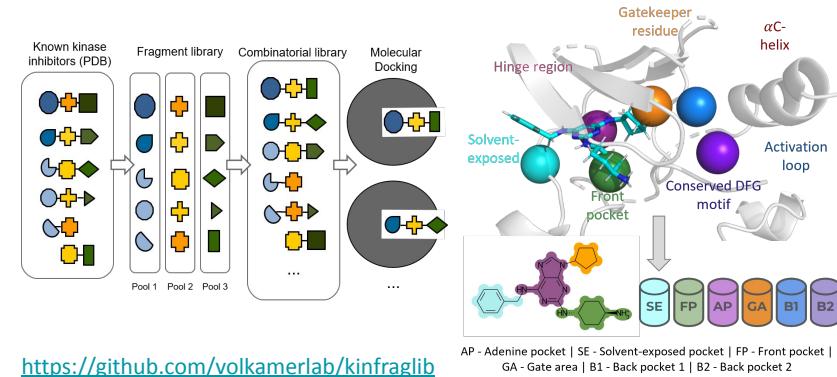
KiSSim: Predicting off-targets from structural similarities in the kinome.
Sydow et al., JCIM, 2022, epub, [DOI](#)



TeachOpenCADD-Kinase Edition:
Kinase-similarity from different angles (Sequence, structure, interaction fingerprints, profiling data)¹



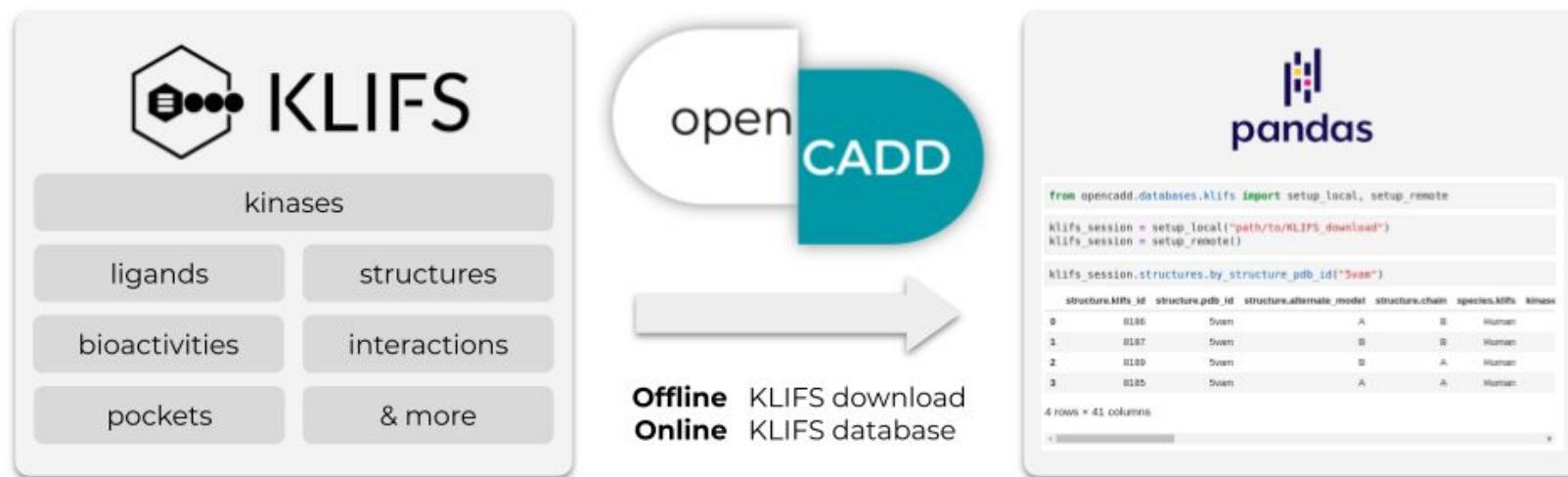
KinFragLib: Exploring the kinase inhibitor space using subpocket-focused fragmentation and recombination. Sydow et al., JCIM, 60(12), 6081, 2020, [DOI](#)



OpenCADD and OpenCADD-KLIFS

Python package for structural bio- and cheminformatics
to streamline common tasks across projects

- `databases.klifs`: utilities to query the KLIFS database, offline or online
- `structure.pocket`: define and visualize protein (sub)pockets
- `structure.superposition`: superimpose macromolecules using sequence & structure



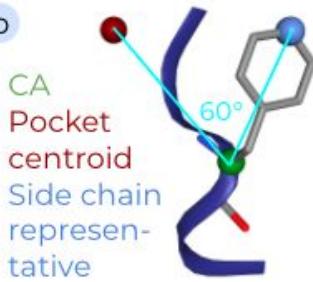
OpenCADD-KLIFS: A Python package to fetch kinase data from the KLIFS database. Sydow D, Rodríguez-Guerra J, Volkamer A. JOSS, 7(70), 3951, 2021, DOI: <https://github.com/volkamerlab/opencadd>

KiSSim: Kinase (Sub)Pocket Fingerprint

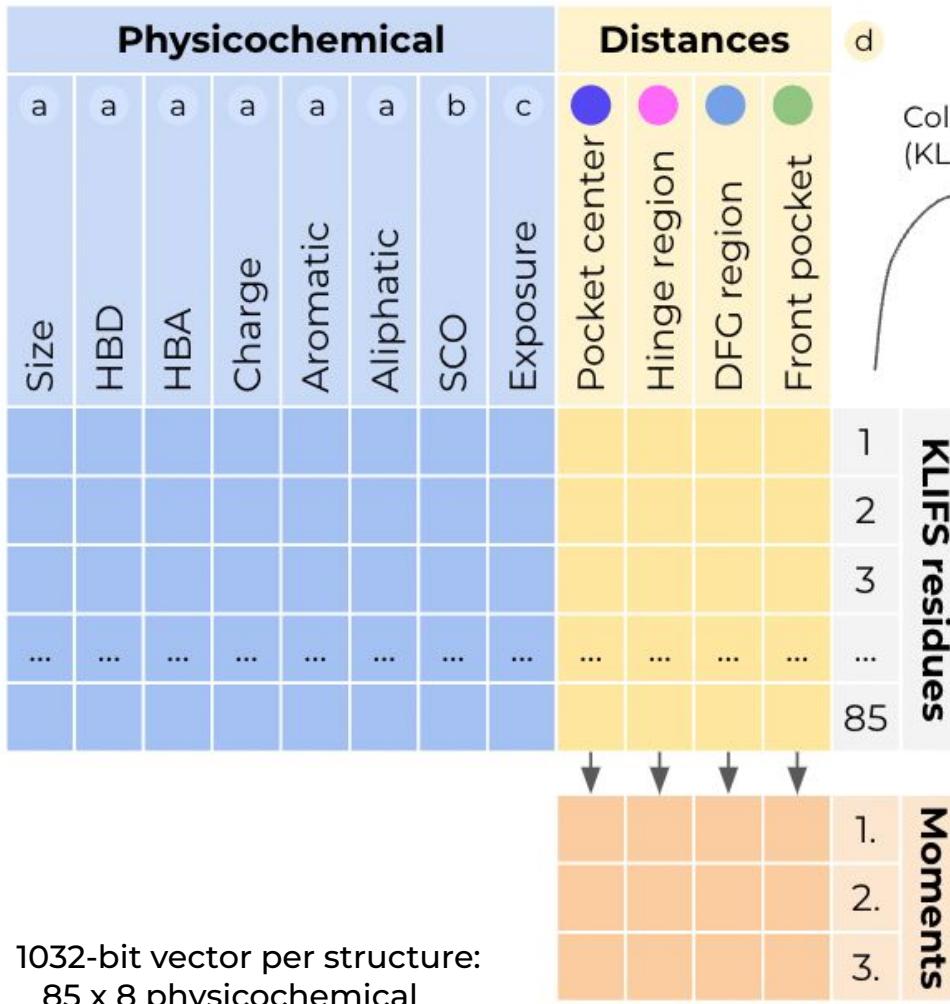
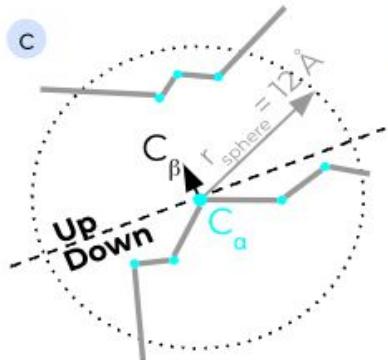
a

PHE_{SiteAlign}
 Size 1 2 3
 HBD 0 1 3
 HBA 0 1 2
 Charge -1 0 1
 Aromatic 0 1
 Aliphatic 0 1

b

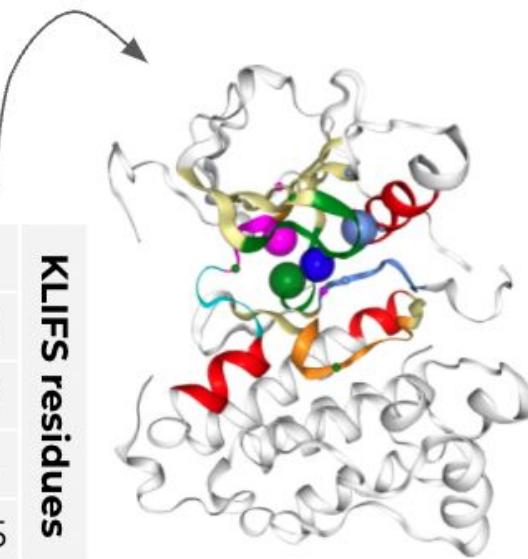


c



d

Colored by region
 (KLIFS)



e

mean,
 standard deviation,
 and skewness

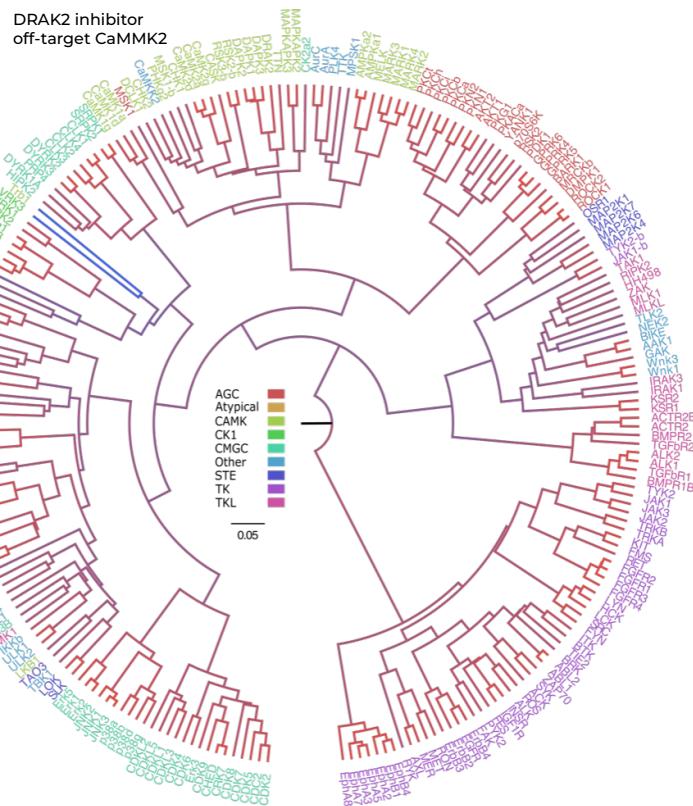
^{1,2}Schalon, et al. Proteins (2008)

³ Hamelryck, Proteins (2005)

KiSSim Detects Non-intuitive Off-targets

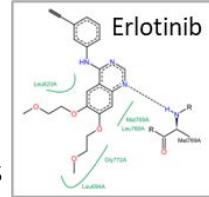
Structure-based kinase tree

- Structure holds complementary information
 - Rationalize ‘non-intuitive’ off-targets effects
 - Guide selectivity/polypharmacology studies

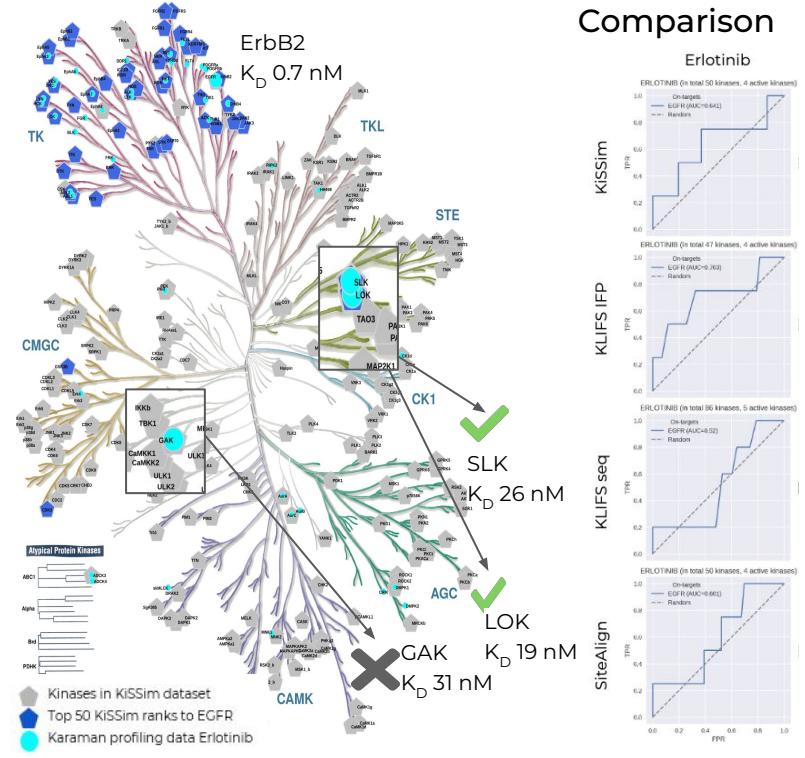


Erlotinib on/off-targets

- Key target: EGFR
 - Karaman profiling data
 - Binds 49 of ~300 tested kinases



Comparison



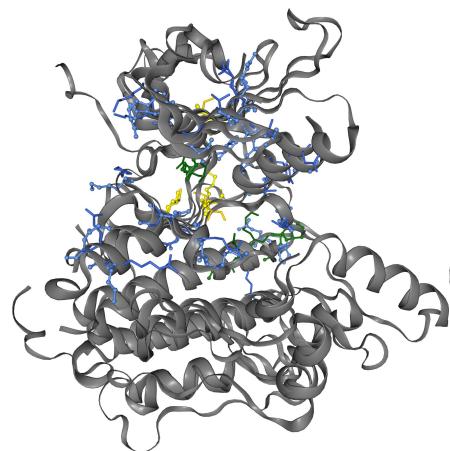
KinMap - Eid et al., *BMC Bioinf* (2017). Profiling - Karaman et al. and Davis et al.

KiSSim: Predicting off-targets from structural similarities in the kinome. Sydow et al., JCIM, 2022, epub, DOI: <https://doi.org/10.1002/jcim.12500>

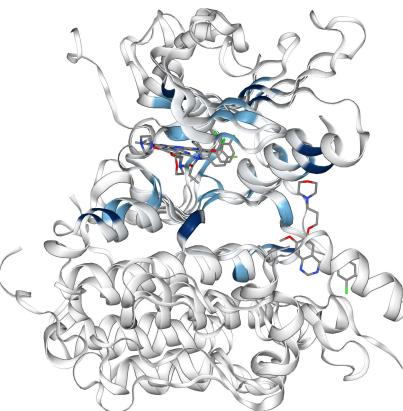
KiSSim: Visualise (Dis)Similarities

EGFR vs. GAK
Residues colored by...

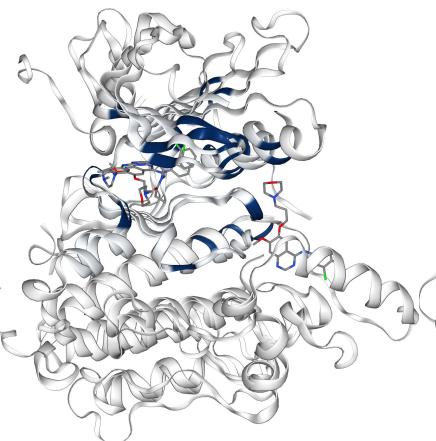
(a) **large** feature differences



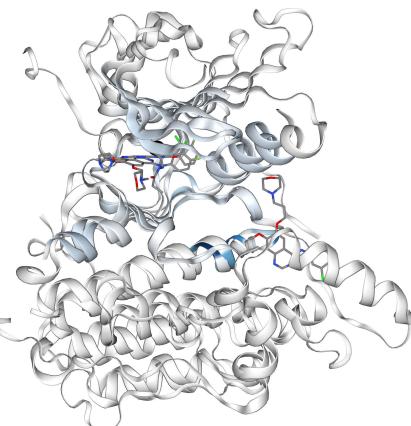
(b) **HBA** feature differences



(c) **aliphatic** feature differences



(d) **hinge region** feature differences



- Physicochemical
- Spatial
- Physicochemical + spatial



Kinase Similarity Meets TeachOpenCADD

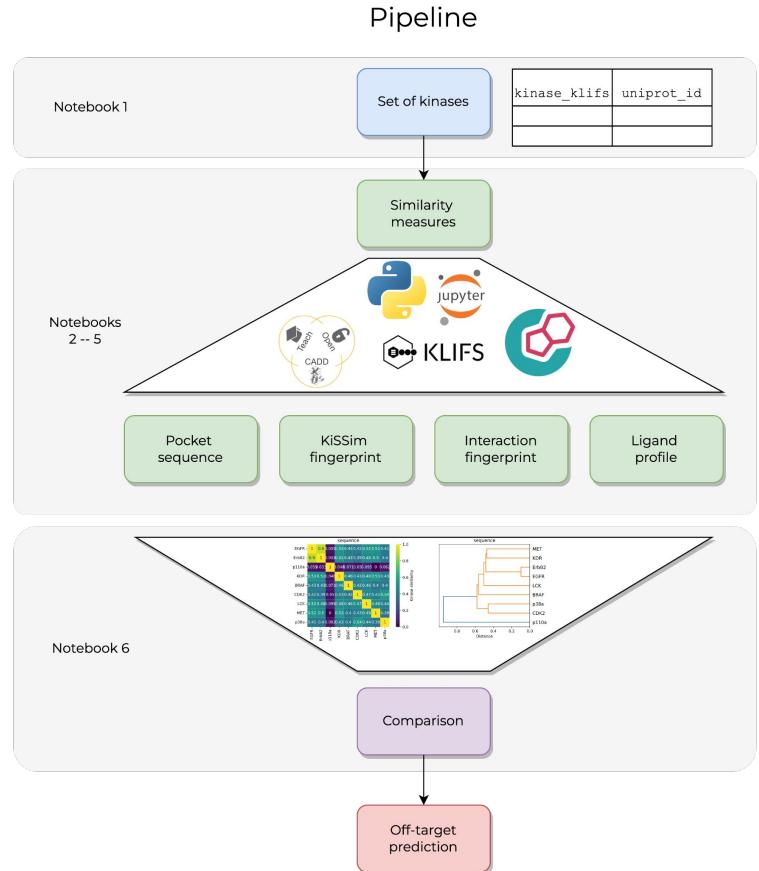
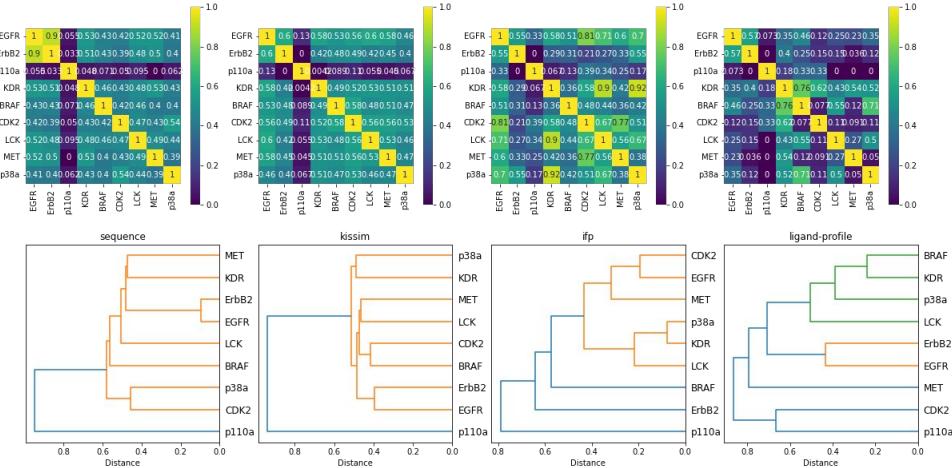
Set of introductory notebooks that discuss kinase similarity

T023: What is a kinase?

Different similarity perspectives

- **T024:** Pocket sequence identity/similarity
- **T025:** Pocket structure (KiSSim)
- **T026:** Pocket-ligand interactions (KLIFS IFP)
- **T027:** Ligand profiling

T028: Compare all perspectives!



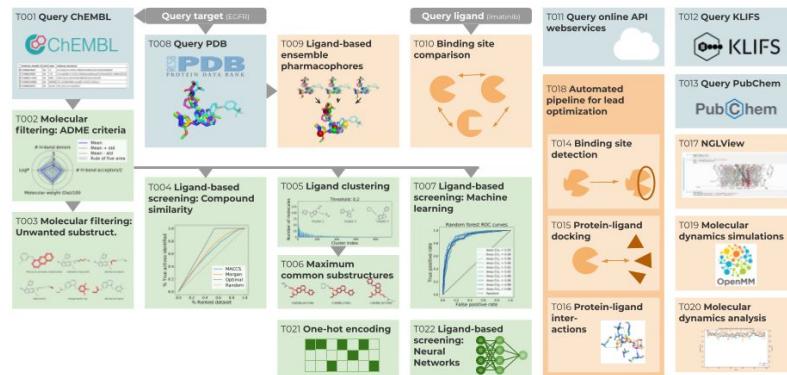
Acknowledgements

- Pharmacophore perception (Mortier, Dhakal, Wichmann, Sydow)
- Target similarity & identification, kinases (Sydow (DFG), Aßmann, Schmiel, Kimber, Leo)
- Toxicity prediction (Morger (BASF), Kimber, Webel)
- Kinomi (Kimber & Rodríguez-Guerra/Taylor (Einstein), Schaller (Bayer))
- And collaboration partners!
- **Find out more about us:** <https://volkamerlab.org>



**Thank you for your attention!
Enjoy the hands-on workshop!**

TeachOpenCADD: Teaching Platform for CADD Using Open Source Packages and Data



<https://github.com/volkamerlab/TeachOpenCADD>

Funding

