The Journal of Open Source Software

OpenCADD-KLIFS: A Python package to fetch kinase data from the KLIFS database

Dominique Sydow*[1], Jaime Rodríguez-Guerra[1], and Andrea Volkamer[1]

1 *In Silico* Toxicology and Structural Bioinformatics, Institute of Physiology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Augustenburger Platz 1, 13353 Berlin, Germany

DOI: 10.21105/joss.03951

Software

- Review
- Repository
- Archive

Editor: Richard Gowers

Reviewers:

- @ojeda-e
- @andrewtarzia
- @mcs07

Submitted: 09 November 2021
Published: 08 December 2021

Summary

Protein kinases are involved in most aspects of cell life due to their role in signal transduction. Dysregulated kinases can cause severe diseases such as cancer, inflammatory and neurodegenerative diseases, which has made them a frequent target in drug discovery for the last decades (Cohen et al., 2021). The immense research on kinases has led to an increasing amount of kinase resources (Kooistra & Volkamer, 2017). Among them is the KLIFS database, which focuses on storing and analyzing structural data on kinases and interacting drugs and other small molecules (Kanev et al., 2020). The OpenCADD-KLIFS Python module offers a convenient integration of the KLIFS data into workflows to facilitate computational kinase research.

Statement of need

OpenCADD-KLIFS is aimed at current and future users of the KLIFS database who seek to integrate kinase resources into Python-based research projects. This module offers access to KLIFS data (Kanev et al., 2020) such as information about kinases, structures, ligands, interaction fingerprints, and bioactivities. KLIFS thereby focuses especially on the ATP binding site, defined as a set of 85 residues and aligned across all structures using a multiple sequence alignment (MSA) (Linden et al., 2014). With OpenCADD-KLIFS, KLIFS data can be queried either locally from a KLIFS download or remotely from the KLIFS webserver. The presented module provides identical APIs for the remote and local queries for KLIFS data and streamlines all output into standardized Pandas DataFrames (The pandas development team, 2020) to allow for easy and quick downstream data analyses (Figure 1). This Pandas-focused setup is ideal to work with in Jupyter notebooks (Kluyver et al., 2016).

OpenCADD-KLIFS (`opencadd.databases.klifs`) is a part of the OpenCADD package, a collection of Python modules for structural cheminformatics.

*corresponding author

The Journal of Open Source Software

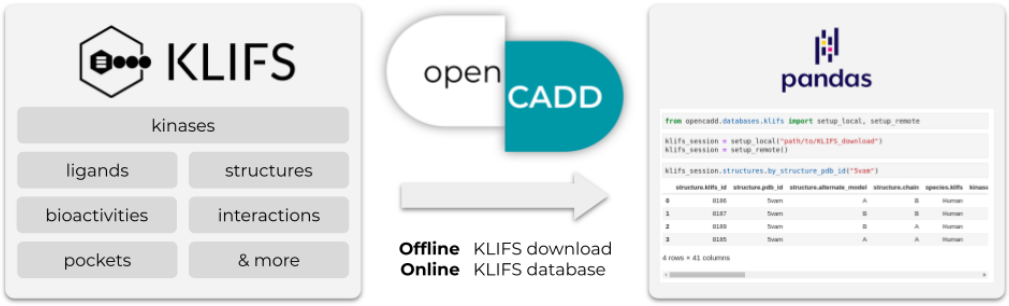

Figure 1: OpenCADD-KLIFS fetches KLIFS data (Kanev et al., 2020) offline from a KLIFS download or online from the KLIFS database and formats the output as user-friendly Pandas DataFrames (The pandas development team, 2020).

The KLIFS database offers a REST API compliant with the OpenAPI specification (KLIFS, 2021). Our module OpenCADD-KLIFS uses bravado (bravado, 2021) to dynamically generate a Python client based on the OpenAPI definitions and adds wrappers to enable the following functionalities:

- A session is set up, which allows access to various KLIFS *data sources* by different *identifiers* with the API `session.data_source.by_identifier`. *Data sources* currently include kinases, structures and annotated conformations, modified residues, pockets, ligands, drugs, and bioactivities; *identifiers* refer to kinase names, PDB IDs, KLIFS IDs, and more. For example, `session.structures.by_kinase_name` fetches information on all structures for a query kinase.
- The same API is used for local and remote sessions.
- The returned data follows the same schema regardless of the session type (local/remote); all results obtained with bravado are formatted as Pandas DataFrames with standardized column names, data types, and handling of missing data.
- Files with the structural 3D coordinates deposited on KLIFS include full complexes or selections such as proteins, pockets, ligands, and more. These files can be downloaded to disc or loaded via biopandas [Raschka:2017] or RDKit (RDKit, 2021).

OpenCADD-KLIFS is especially convenient whenever users are interested in multiple or more complex queries such as "fetching all structures for the kinase EGFR in the DFG-in conformation" or "fetching the measured bioactivity profiles for all ligands that are structurally resolved in complex with EGFR." Formatting the output as DataFrames facilitates subsequent filtering steps and DataFrame merges in case multiple KLIFS datasets need to be combined. OpenCADD-KLIFS is currently used in several projects from the Volkamer Lab (Volkamer Lab, 2021) including TeachOpenCADD (TeachOpenCADD, 2021), OpenCADD-pocket (OpenCADD, 2021), KiSSim (KiSSim, 2021), KinoML (OpenKinome, 2021), and PLIPify (PLIPify, 2021). For example, OpenCADD-KLIFS is applied in a TeachOpenCADD tutorial to demonstrate how to fetch all kinase-ligand interaction profiles for all available EGFR kinase structures to visualize the per-residue interaction types and frequencies with only a few lines of code.

Acknowledgements

We thank the whole KLIFS team for providing such a great kinase resource with an easy-to-use API and especially Albert Kooistra for his help with questions and wishes regarding the KLIFS database. We thank David Schaller for his feedback on the OpenCADD-KLIFS

module. We acknowledge the contributors involved in software programs and packages used by OpenCADD-KLIFS, such as bravado, RDKit, Pandas, Jupyter, and Pytest, and Sphinx.

References

bravado. (2021). bravado. In *GitHub repository*. GitHub. https://github.com/Yelp/bravado

Cohen, P., Cross, D., & Jänne, P. A. (2021). Kinase drug discovery 20 years after imatinib: Progress and future directions. *Nature Reviews Drug Discovery*, *20*(7), 551–569. https://doi.org/10.1038/s41573-021-00195-4

Kanev, G. K., de Graaf, C., Westerman, B. A., de Esch, I. J. P., & Kooistra, A. J. (2020). KLIFS: an overhaul after the first 5 years of supporting kinase research. *Nucleic Acids Research*, *49*(D1), D562–D569. https://doi.org/10.1093/nar/gkaa895

KiSSim. (2021). KiSSim: Subpocket-based fingerprint for kinase pocket comparison. In *GitHub repository*. GitHub. https://github.com/volkamerlab/kissim

KLIFS. (2021). *KLIFS OpenAPI*. https://dev.klifs.net. https://dev.klifs.net/swagger_v2/

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & team, J. development. (2016). Jupyter notebooks - a publishing format for reproducible computational workflows. In F. Loizides & B. Scmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press. https://eprints.soton.ac.uk/403913/

Kooistra, A. J., & Volkamer, A. (2017). Kinase-Centric Computational Drug Development. *Annu. Rep. Med. Chem.*, *50*, 197–236. https://doi.org/10.1016/BS.ARMC.2017.08.001

Linden, O. P. J. van, Kooistra, A. J., Leurs, R., Esch, I. J. P. de, & Graaf, C. de. (2014). KLIFS: A knowledge-based structural database to navigate kinase–ligand interaction space. *Journal of Medicinal Chemistry*, *57*(2), 249–277. https://doi.org/10.1021/jm400378w

OpenCADD. (2021). OpenCADD-Pocket: Identification and analysis of protein (sub)pockets. In *GitHub repository*. GitHub. https://github.com/volkamerlab/opencadd

OpenKinome. (2021). KinoML: Structure-informed machine learning for kinase modeling. In *GitHub repository*. GitHub. https://github.com/openkinome/kinoml

PLIPify. (2021). PLIPify: Protein-ligand interaction frequencies across multiple structures. In *GitHub repository*. GitHub. https://github.com/volkamerlab/plipify

RDKit. (2021). RDKit: Open-Source Cheminformatics. In *RDKit website*. RDKit. http://www.rdkit.org

TeachOpenCADD. (2021). TeachOpenCADD: a teaching platform for computer-aided drug design (CADD) using open source packages and data. In *GitHub repository*. GitHub. https://github.com/volkamerlab/teachopencadd

The pandas development team. (2020). Pandas-dev/pandas: pandas. In *Zenodo repository*. Zenodo. https://doi.org/10.5281/zenodo.3509134

Volkamer Lab. (2021). Volkamer Lab website. In *Volkamer Lab website*. Volkamer Lab. https://volkamerlab.org/