Teach Open CADD Update

RDKit UGM 2021

Dominique Sydow
Volkamer Lab
Charité, Berlin
www.volkamerlab.org

TeachOpenCADD in a nutshell

What? Pipelines for common CADD* tasks using open resources

How? Jupyter Notebooks (Python) & KNIME workflows

For what? Teaching, learning & starting point for research projects

Who? Beginners & advanced users

*CADD = computer-aided drug design

TeachOpenCADD links

Jupyter Notebooks

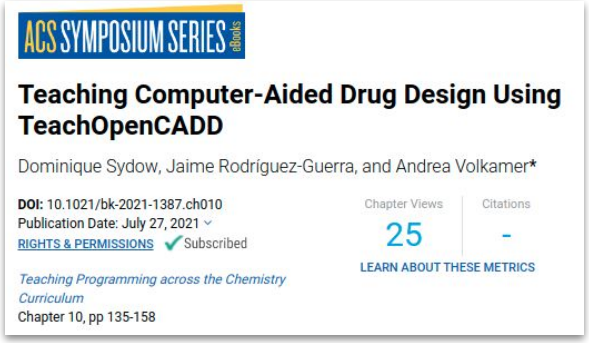
- https://github.com/volkamerlab/teachopencadd
- https://projects.volkamerlab.org/teachopencadd/
- Sydow et al. *J. Chem.* (2019)

KNIME Workflows

- https://hub.knime.com/volkamerlab/spaces/Public/latest/TeachOpenCADD
- Sydow and Wichmann et al. *JCIM* (2019)

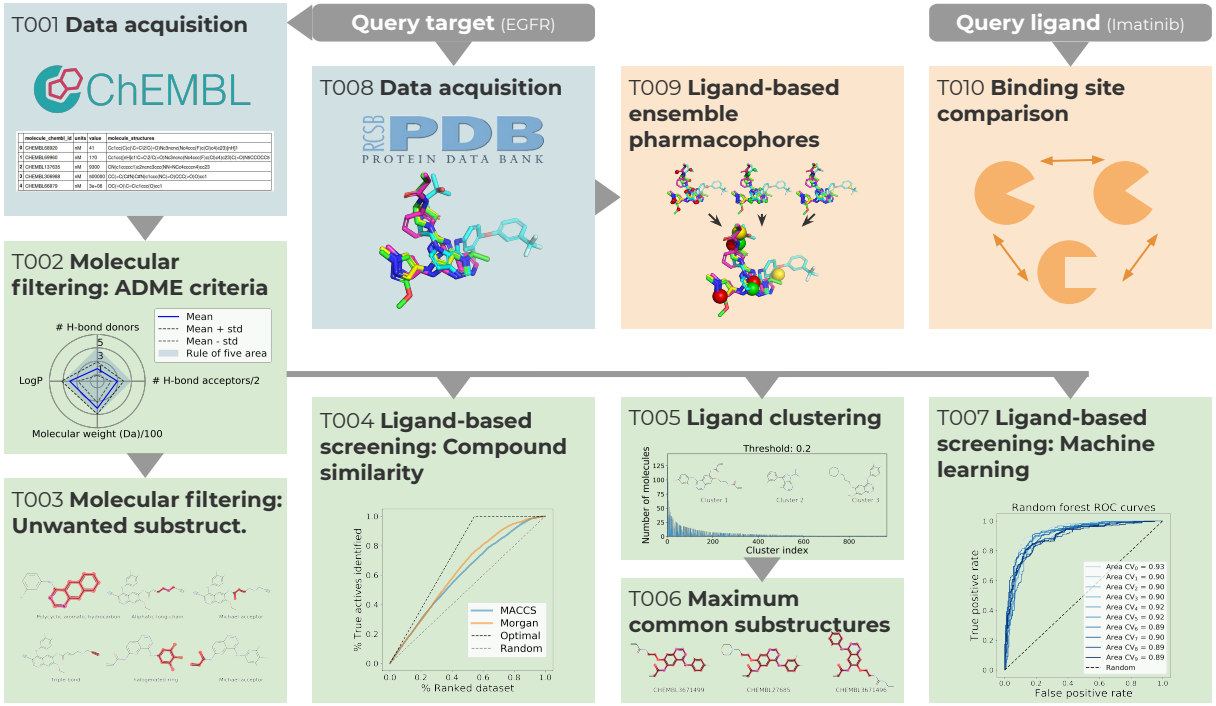
Teaching with TeachOpenCADD

- Sydow, Rodríguez-Guerra, Volkamer. ACS Symposium Series (2021)



TeachOpenCADD 2019

Cheminformatics | Structural Bioinformatics | Online API Webservices

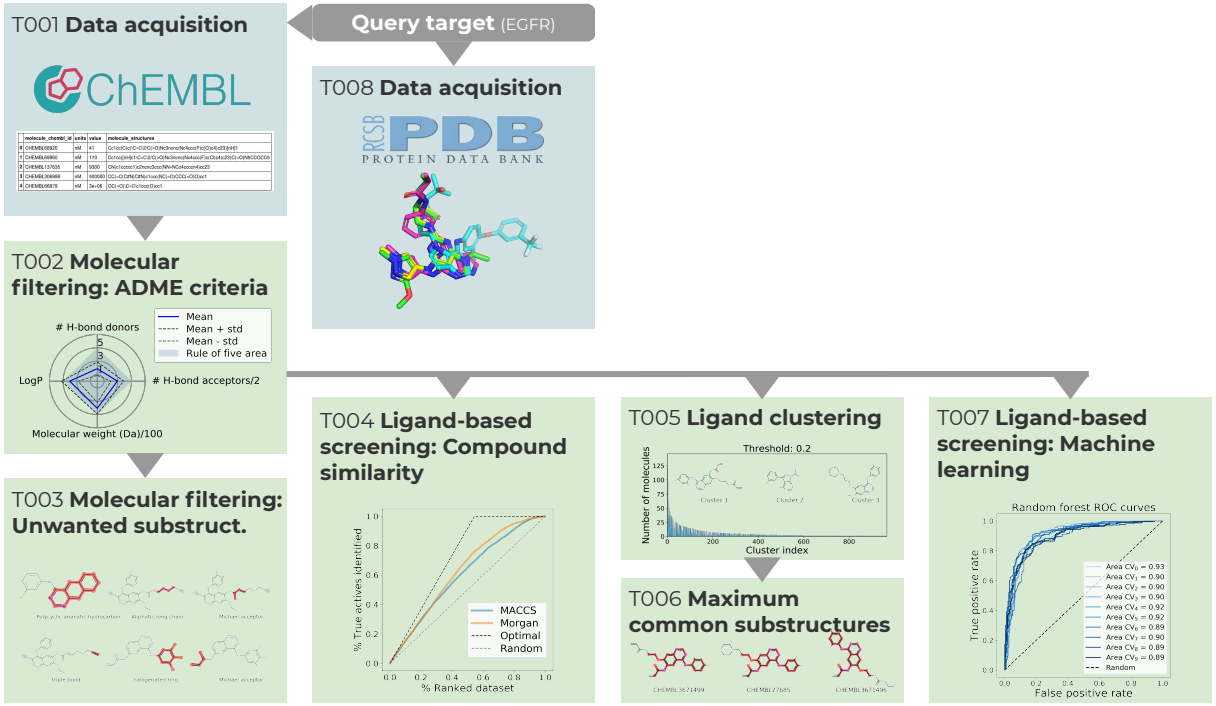


RDKit
Open-Source Cheminformatics and Machine Learning

Especially the cheminformatics-focused topics are powered by the RDKit

TeachOpenCADD 2019

Cheminformatics | Structural Bioinformatics | Online API Webservices



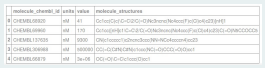
Open for Innovation ® KNIME

These topics are not only available as Jupyter Notebooks but also as KNIME Workflows

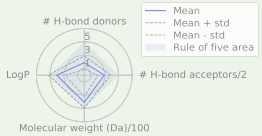
TeachOpenCADD 2021: New topics

Cheminformatics | Structural Bioinformatics | Online API Webservices

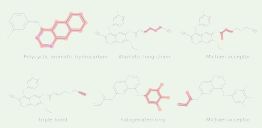
T001 **Data acquisition**



T002 **Molecular filtering: ADME criteria**



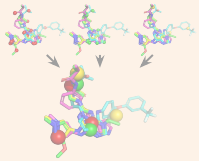
T003 **Molecular filtering: Unwanted substruct.**



Query target (EGFR)

T008 **Data acquisition**

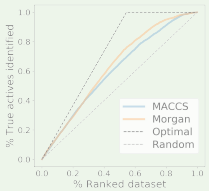
T009 **Ligand-based ensemble pharmacophores**



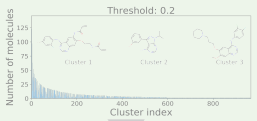
Query ligand (Imatinib)

T010 **Binding site comparison**

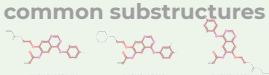
T004 **Ligand-based screening: Compound similarity**



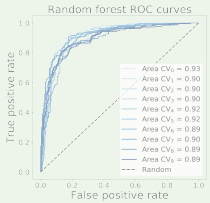
T005 **Ligand clustering**



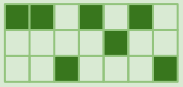
T006 **Maximum common substructures**



T007 **Ligand-based screening: Machine learning**



T021 **One-hot encoding**



T022 **Ligand-based screening: Neural Networks**

T011 **Query online API webservices**

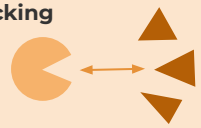
T012 **Query KLIFS**

T013 **Query PubChem**

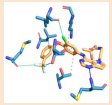
T018 **Fully automated pipeline for drug design**
Work-in-progress

T014 **Binding site detection**

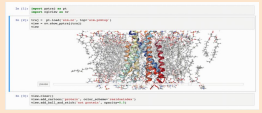
T015 **Protein-ligand docking**



T016 **Protein-ligand inter-actions**



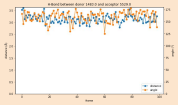
T017 **NGLView**



T019 **Molecular dynamics simulations**

OpenMM

T020 **Molecular dynamics analysis**



New talktorials!

All tasks rely on external open source Python packages and databases/webservers

External resources

Python packages

- Cheminformatics and structural bioinformatics: `rdkit`, `openbabel`, `mdanalysis`, `biopython`, `biopandas`, `opencadd`, `plip`, `openff`, `openff-toolkit`, `openmmforcefields`, `pdbfixer`, `mdanalysis`, `biotite`, `smina`
- Data science (PyData stack): `numpy`, `pandas`, `scikit-learn`, `keras`, `jupyter`, `ipywidgets`
- Data visualization: `matplotlib`, `mpl_toolkits`, `matplotlib_venn`, `seaborn`, `nglview`
- Web services clients: `pypdb`, `chembl_webresource_client`, `requests`, `bravado`, `beautifulsoup4`
- Utilities: `tqdm`, `requests_cache`, `redo`, `google-colab`, `condacolab`
- Continuous integration: `pytest`, `nbval`
- Documentation: `sphinx`, `nbsphinx`
- Code style: `black-nb`

Databases and webservers

- ChEMBL
- RCSB PDB
- KLIFS
- PubMed
- ProteinsPlus

If we are using your resource and forgot to add it here, please contact us so that we can rectify this, thank you!

Maintaining pipelines performing diverse tasks & using many different resources can be quite a challenge!

Especially with growing content we need automatized testing!

- Nightly tests
- Notify us if notebooks throw errors
- Notify us if notebooks' output is changing / differing between operating systems

https://github.com/volkamerlab/teachopencadd/tree/update-readme-docs#external-resources

TeachOpenCADD 2021: New website!

TeachOpenCADD (WIP 🚧) | Search | TeachOpenCADD 254 Stars · 89 Forks

TeachOpenCADD (WIP 🚧)
Our talktorials
Talktorials by collection
Complete list of talktorials
Run locally
Installing
Contributors
For contributors
API Documentation
External resources
Python programming introduction
Cheminformatics resources collections
Packages and webservers used in TeachOpenCADD

TeachOpenCADD

Open source programming packages for cheminformatics and structural bioinformatics are powerful tools to build modular, reproducible, and reusable pipelines for computer-aided drug design (CADD). While documentation for such tools is available, only few freely accessible examp ... especially users new to the field.

TeachOpenCADD is a teaching platform developed by students for students, which provides teaching material for central CADD topics. Since we cover both ... ddresses students and researchers with a biologica

For each topic, an interactive ... ython packages `rdkit`, `pypdb`, `biopandas`, `nglview`, and `mdanalysis`. Topics are continuously expanded and open for contributions from the community. Beyond their teaching purpose, the T ... ected modifications and extensions.

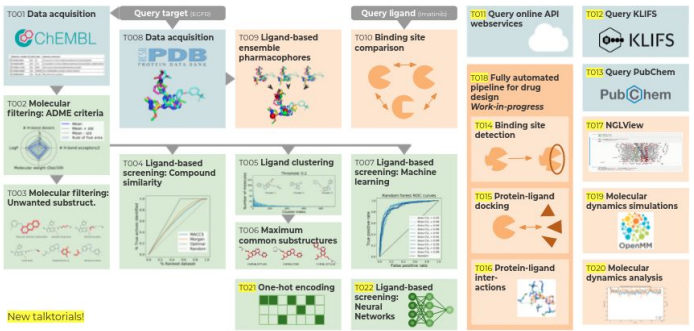
Contents
TeachOpenCADD
- Table of contents
- Citation
- Funding
- License

Show Source

Explore notebooks online

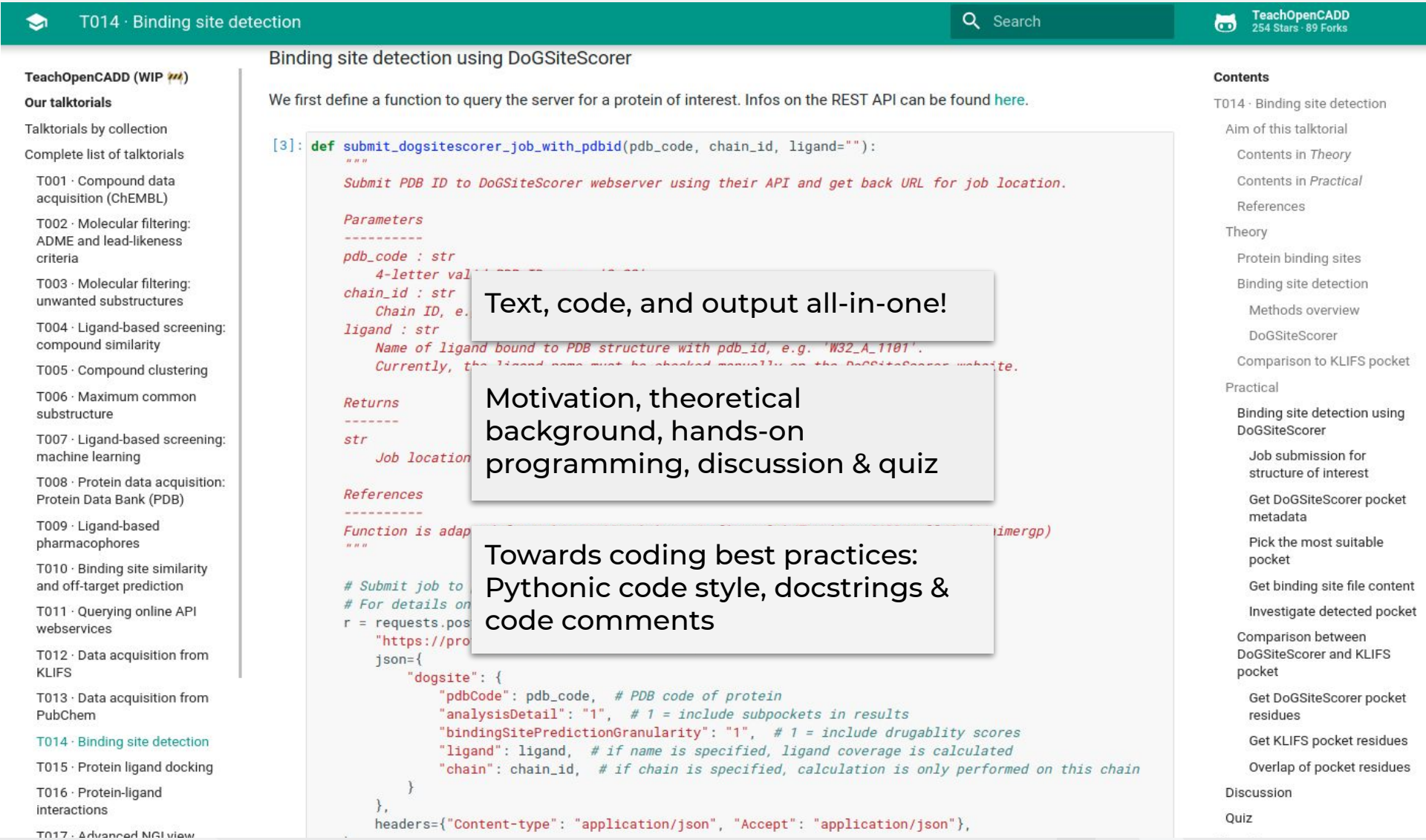
Find usage instructions (Binder & conda installation)

Find information on how to contribute to TeachOpenCADD



https://projects.volkamerlab.org/teachopencadd

TeachOpenCADD talktorials (talk + tutorial)



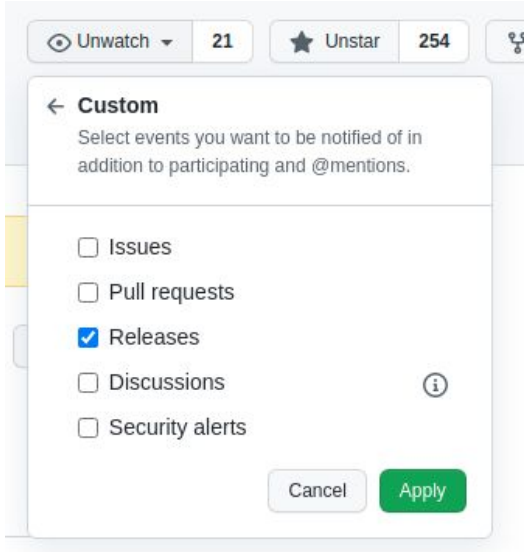
New official release coming VERY soon!

Open TODOs

- Get TeachOpenCADD on conda-forge
- Finish last pending talktorial T018 (see progress in PR #125)
- Fix a few website rendering artifacts

Stay tuned

- Watch new releases



- Or follow me on Twitter for TeachOpenCADD updates

  @dominiquesydow

Acknowledgements

Talktorials T001-T010

Students from CADD course 2017/18!!
Dominique Sydow
Andrea Morger
Maximilian Driller
Andrea Volkamer

KNIME workflows W1-8

Michele Wichmann
Dominique Sydow
Jaime Rodríguez-Guerra
Greg Landrum
Daria Goldmann
Andrea Volkamer

RDKit UGM 2019 Hackathon

Richard Gowers
Floriane Montanari
Jaime Rodríguez-Guerra
Dominique Sydow
Talia Kimber
Andrea Morger

TeachOpenCADD v2

Andrea Volkamer
Jaime Rodríguez-Guerra
Dominique Sydow
Talia Kimber
David Schaller
Corey Taylor
Mareike Leja
Sakshi Misra
Yonghui Chen
Michele Wichmann
Armin Ariamajd

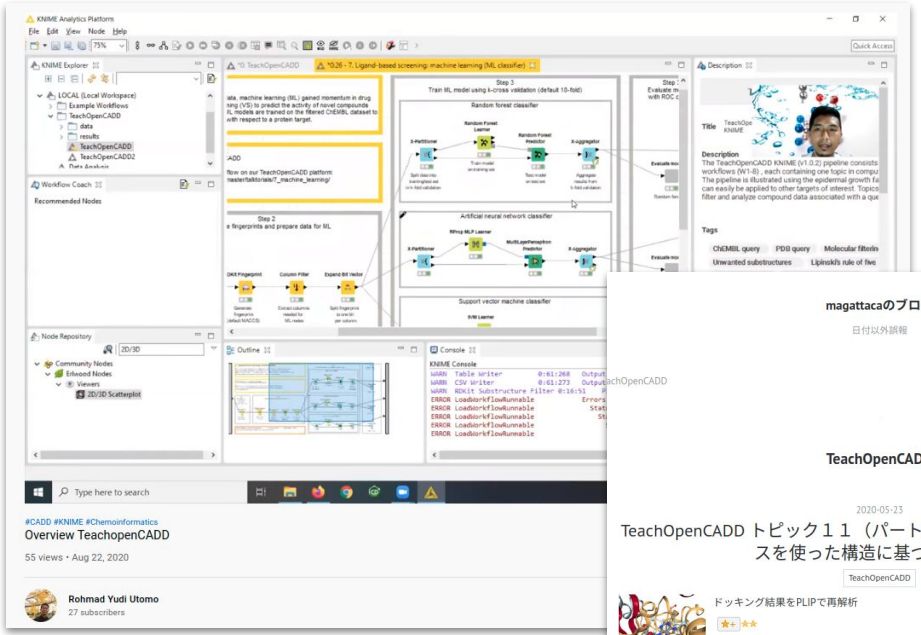
Additional contributors

Greg Landrum
Jeff Wagner
Piedro Gerletti
Ahmed Atta
Melanie Vogel
Abishek Laxmanan Ravi Shankar
Maria Trofimova
Ross Johnstone
Thomas Holder
Andrew Wilkinson

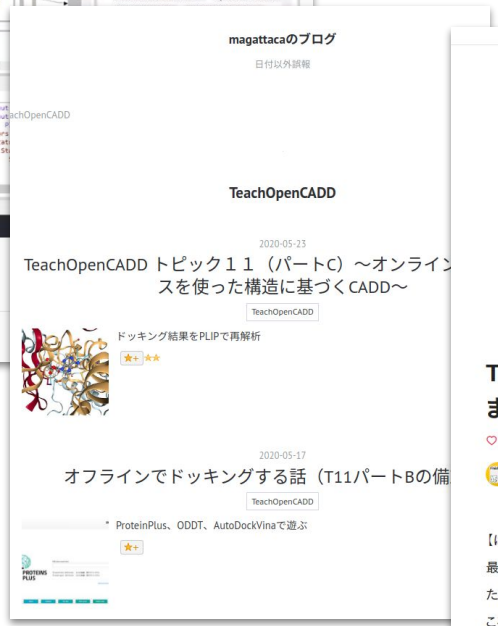
TeachOpenCADD is being used!

https://github.com/volkamerlab/teachopencadd

Star 254 Fork 89



https://www.youtube.com/watch?v=AaqOaNiD_94



https://magattaca.hatenablog.com/archive/category/TeachOpenCADD

https://note.com/knimesupportteam/n/n330111da83b2