



# Teach Open CADD News

RDKit UGM 2020

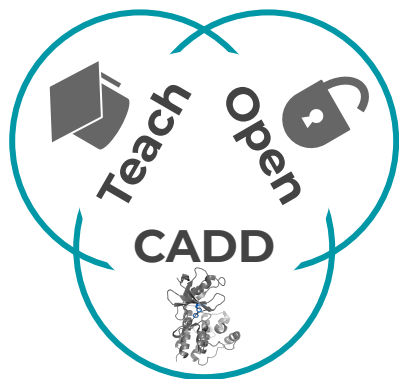
Dominique Sydow

Jaime Rodríguez-Guerra

[www.volkamerlab.org](http://www.volkamerlab.org)



# TeachOpenCADD - what is this?



**Teach common CADD tasks using open resources (packages, databases) for beginners and advanced users!**

[www.github.com/volkamerlab/TeachOpenCADD](https://www.github.com/volkamerlab/TeachOpenCADD)



## How?

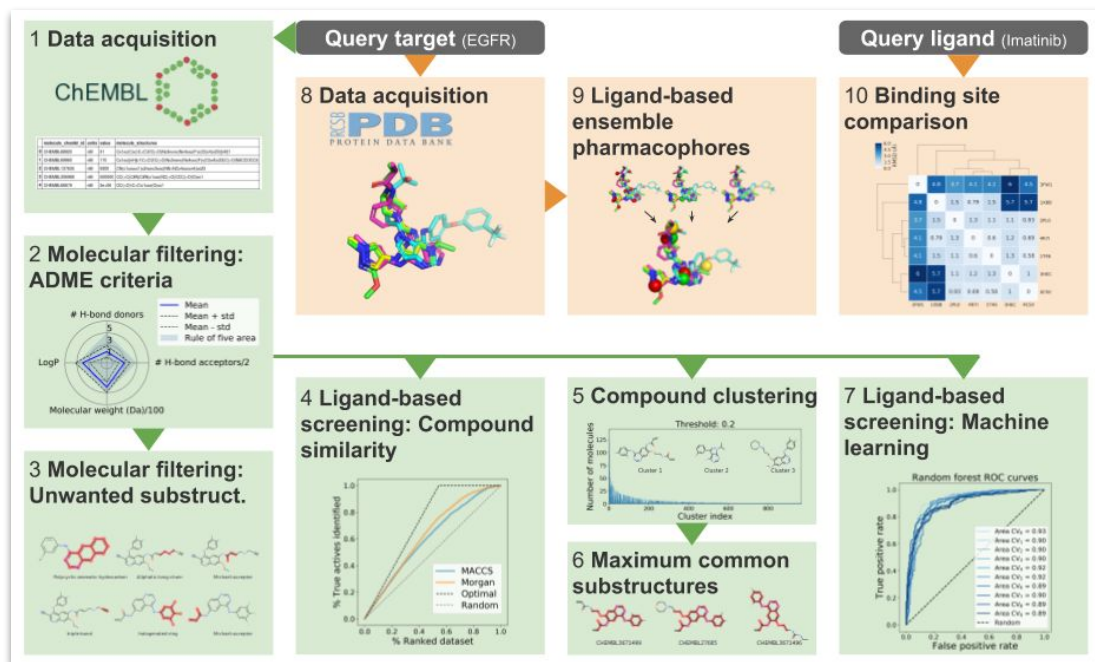
- Coding-based Jupyter notebooks<sup>1</sup>
- GUI-based KNIME workflows<sup>2</sup>
- Highly interactive
- All in one place: Theory + code/nodes + output

<sup>1</sup>Sydow et al. *J. Chem.* (2019), <sup>2</sup>Sydow and Wichmann et al. *JCIM* (2019)

# TeachOpenCADD - what is offered so far?

## Cheminformatics and structural bioinformatics

### Published content



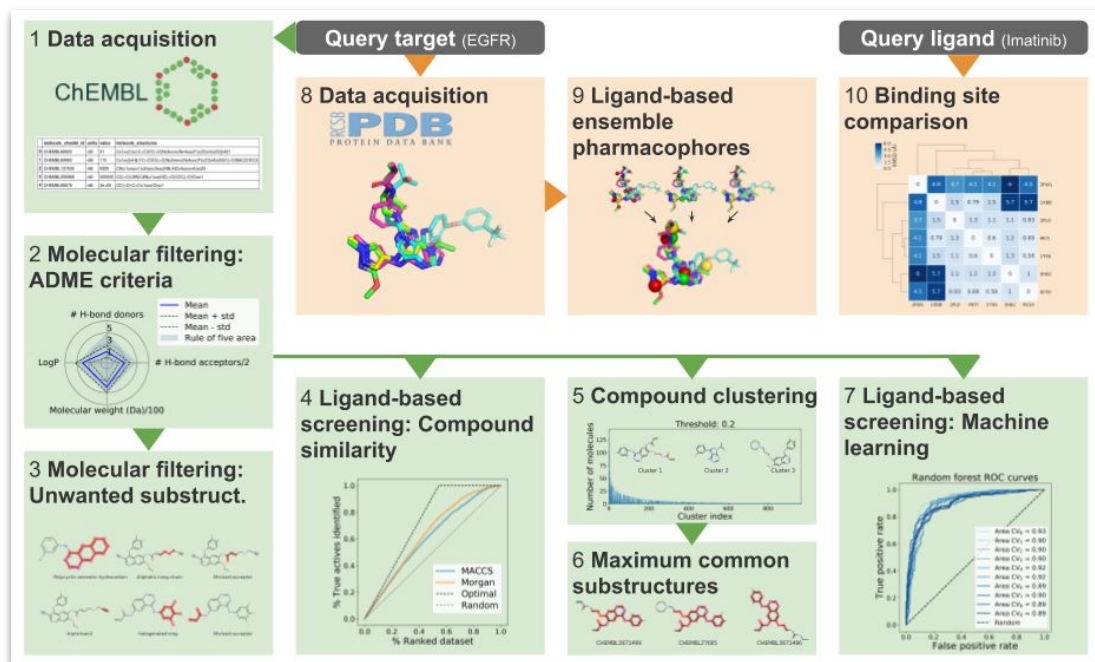
**WIP: Towards more pythonic code - refactoring 2020**

<https://github.com/volkamerlab/TeachOpenCADD/pull/42>

# TeachOpenCADD - what is offered so far?

## Cheminformatics and structural bioinformatics

### Published content



### New content / WIP

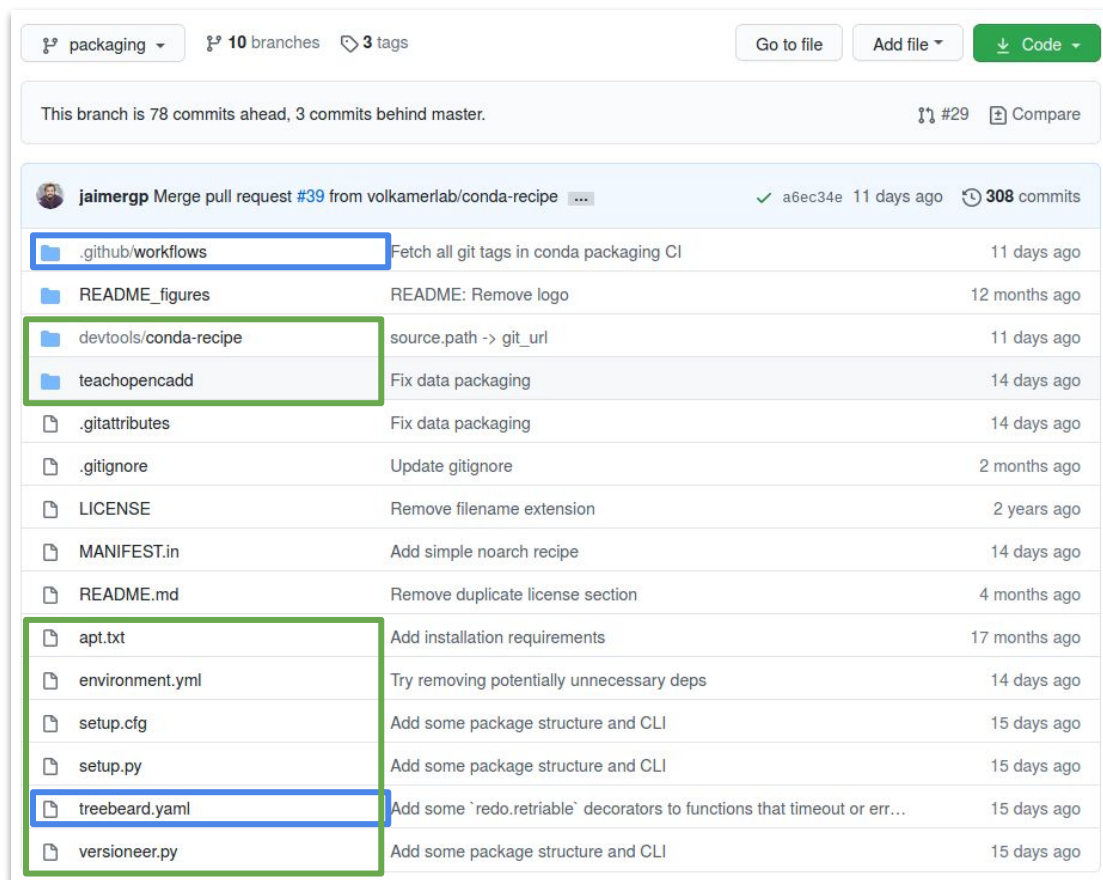
- Programmatic access to online databases
  - PubChem
  - KLIFS
- Protein-ligand docking
- Protein-ligand interactions
- Binding site detection
- Molecular dynamics
- Deep learning and one-hot encoding

WIP: Towards more pythonic code - refactoring 2020

<https://github.com/volkamerlab/TeachOpenCADD/pull/42>

# TeachOpenCADD - coming up...

## On our way to a teachopencadd conda package!



File	Commit Message	Time Ago
<code>.github/workflows</code>	Fetch all git tags in conda packaging CI	11 days ago
<code>README_figures</code>	README: Remove logo	12 months ago
<code>devtools/conda-recipe</code>	source.path -> git_url	11 days ago
<code>teachopencadd</code>	Fix data packaging	14 days ago
<code>.gitattributes</code>	Fix data packaging	14 days ago
<code>.gitignore</code>	Update gitignore	2 months ago
<code>LICENSE</code>	Remove filename extension	2 years ago
<code>MANIFEST.in</code>	Add simple noarch recipe	14 days ago
<code>README.md</code>	Remove duplicate license section	4 months ago
<code>apt.txt</code>	Add installation requirements	17 months ago
<code>environment.yml</code>	Try removing potentially unnecessary deps	14 days ago
<code>setup.cfg</code>	Add some package structure and CLI	15 days ago
<code>setup.py</code>	Add some package structure and CLI	15 days ago
<code>treebeard.yaml</code>	Add some `redo.retriable` decorators to functions that timeout or err...	15 days ago
<code>versioneer.py</code>	Add some package structure and CLI	15 days ago

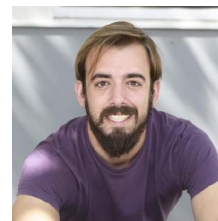
<https://github.com/volkamerlab/TeachOpenCADD/pull/29>

### Packaging!

- ✓ Apply package structure to repo
- ✓ Add CLI
- Make conda package

### Continuous integration!

- ✓ GH Action & Treebeard
- ✓ Automatically spot raised errors in notebooks
- Test notebook output (selected cells)




Driven  
by  
Jaime



# TeachOpenCADD - coming up...

## On our way to a sphinx-based TeachOpenCADD website!

 **TeachOpenCADD**  
145 Stars · 52 Forks

Molecular filtering: ADME and lead-likeness criteria

Search

**TeachOpenCADD**  
**Getting started**  
Installing  
Talktorials  
All talktorials  
Compound data acquisition (ChEMBL)  
Molecular filtering: ADME and lead-likeness criteria  
Molecular filtering: unwanted substructures  
Ligand-based screening: compound similarity  
Ligand-based screening: compound similarity  
Compound clustering  
Maximum common substructure  
Ligand-based screening: machine learning  
Protein data acquisition: Protein Data Bank (PDB)  
Ligand-based pharmacophores  
Binding site similarity and off-target prediction  
Structure-based CADD using online APIs/servers  
Structure-based CADD using online APIs/servers  
Structure-based CADD using online APIs/servers  
Structure-based CADD using online APIs/servers  
**Developers**

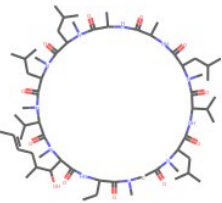
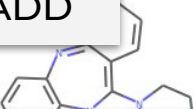
Define and visualize example molecules

Before working with the whole dataset retrieved from ChEMBL, we pick four example compounds to investigate their chemical properties. We import the necessary libraries and load the dataset.


```
[3]: smiles = [
    "CCC1C(=O)N(CC(=O)N(C)C)C(=O)N(C)C",
    "CN1CCN(CC1)C2=C3C=CC(=CC=C3N4C(=C(C=C4)C)C)C",
    "CC1=C(C(CCC1)(C)C)C=CC(=CC=CC(=CC=C(C)C)C=CC(=C(C)C)C2=C(CCC2(C)C)C)C",
    "CC(CCC1=CC(=C(C(=C1)O)C2C=C(CCC2C(=C)C)C)O"
]
names = ["cyclosporine"]
```

```
[4]: molecules = pd.DataFrame({ 'name': names, 'smiles': smiles, })
PandasTools.AddMoleculeColumnToFrame(molecules, "smiles")
molecules
```

```
[4]:
```

	name	ROMol
0	cyclosporine	<chem>CCC1C(=O)N(CC(=O)N(C)C)C(=O)N(C)C</chem> 
1	clozapine	<chem>CN1CCN(CC1)C2=C3C=CC(=CC=C3N4C(=C(C=C4)C)C)C</chem> 

Driven by Jaime



Find easy installation instructions (conda installation)

Explore notebooks online

Find detailed information on how to contribute to TeachOpenCADD

<https://github.com/volkamerlab/TeachOpenCADD/pull/40>

**Contents**  
Molecular filtering: ADME and lead-likeness criteria  
Aim of this talktorial  
Contents in *Theory*  
Contents in *Practical*  
References  
**Theory**  
ADME - absorption, distribution, metabolism, and excretion  
Lead-likeness and Lipinski's rule of five (Ro5)  
Radar charts in the context of lead-likeness  
**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)  
Calculate statistics on Ro5 properties  
Define helper functions to prepare data for radar plotting  
Generate radar plots, finally!  
**Discussion**

# Acknowledgements

## Talktorials T1-10

Students from CADD course 2017 & 2018!!

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

...

## Refactoring 2020

Andrea Volkamer  
Jaime Rodríguez-Guerra  
Dominique Sydow  
Talia Kimber  
Mareike Leja  
Sakshi Misra  
Yonghui Chen  
David Schaller

## Additional contributors

Greg Landrum  
Jeff Wagner  
Ross Johnstone  
Thomas Holder  
Andrew Wilkinson

