

Building a community-driven ecosystem for fast, reproducible and reusable molecular simulation analysis using MDAnalysis

Irfan Alibay^{1,2}, Jonathan Barnoud³, Oliver Beckstein⁴, Richard J. Gowers¹, Ian M. Kenney⁴, Philip R. Loche⁵,

Hugo MacDermott-Opeskin¹, Micaela Matta⁶, Fiona B. Naughton⁷, Tyler Reddy⁸, Lily Wang¹

¹Open Molecular Software Foundation, Irvine, CA, USA, ²Department of Biochemistry, University of Oxford, Oxfordshire, United Kingdom, ³CiTIUS, Universidade de Santiago de Compostela, Spain,

⁴Department of Physics, Arizona State University, Tempe, AZ, United States, ⁵Laboratory of Computational Science and Modelling, IMX, EPFL, Switzerland, ⁶Department of Chemistry, King's College London, London, United Kingdom, ⁷Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA, ⁸CCS-7 Applied Computer Science, Los Alamos National Laboratory, Los Alamos, NM, United States



Download this poster!



What is MDAnalysis?

An open-source Python library for handling and extracting features from molecular simulations

Why MDAnalysis?

• Easy to use and install

Simple Python API, familiar NumPy components.
Available on PyPI and conda-forge.

• Extensible

Frameworks and library components to create your own analysis, file I/O, etc...

• Interoperable

Support for over 40 file formats.

Features

• Selections

Intuitive keywords, geometry, connectivity, etc..

• Fast distance routines

C/C++ distances backend.

• Trajectory manipulations

On-the-fly transformations: centering, alignment, periodic un/wrapping.

• Built-in analyses

Distances, RMSD, hydrogen bonds, density, and more!

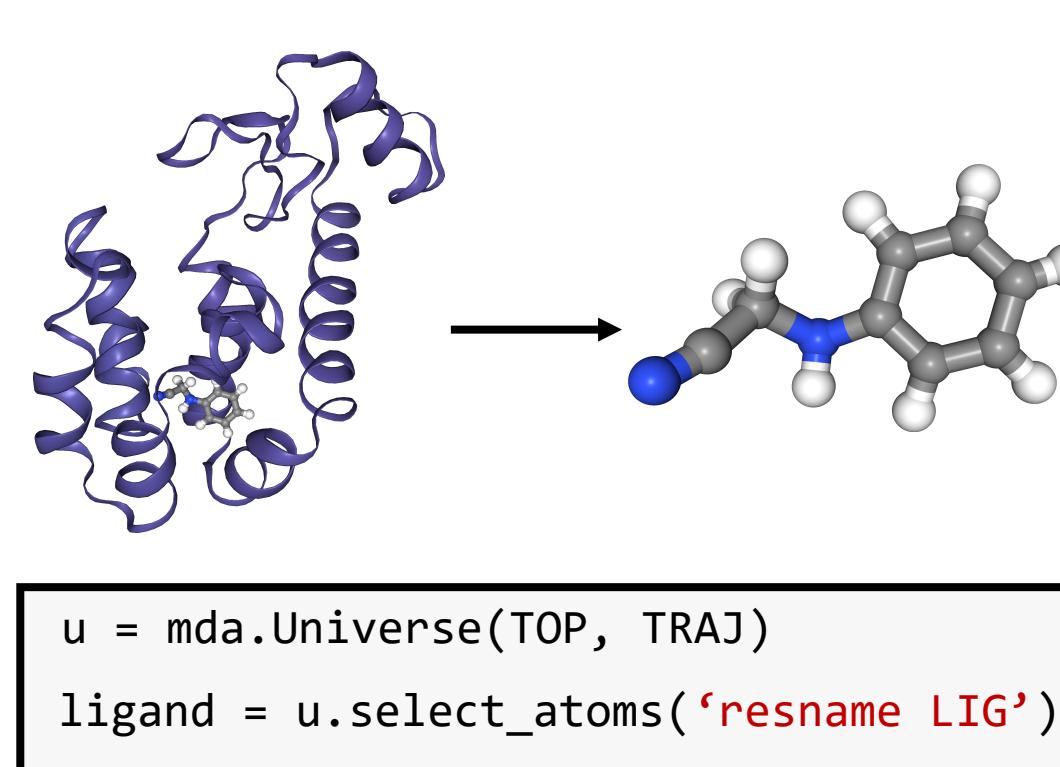
• Converters

Seamless conversion to/from other Python packages: Chemfiles, ParmEd, RDKit, OpenMM, and more to come soon!

Supported file formats

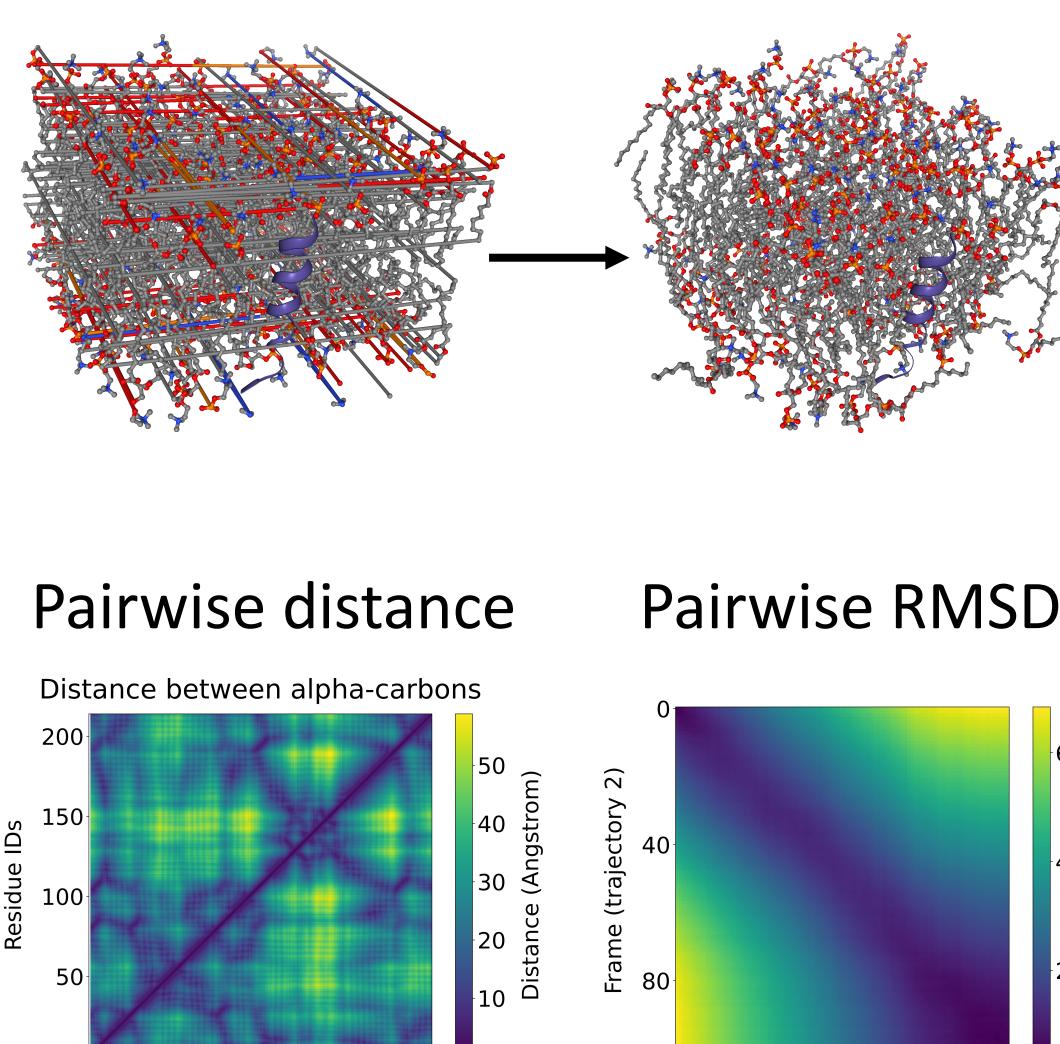
Software	File Type
AMBER	PRMTOP, RST7, TRJ, NETCDF
GROMACS	ITP, TPR, GRO, TRR, XTC
CHARMM	PSF, DCD, CRD
NAMD	DCD, COOR, NAMDBIN
LAMMPS	CONFIG, DATA, DUMP, DCD
DL_POLY	CONFIG, HISTORY
HOOMD	XML, GSD
GAMES	GMS
DESRES	DMS
Others	XYZ, TXZ, PDB, PDBQT, PQR, TRZ, MOL2, MMTF, FHI-aims, HSMDF, etc...

Intuitive atom selections

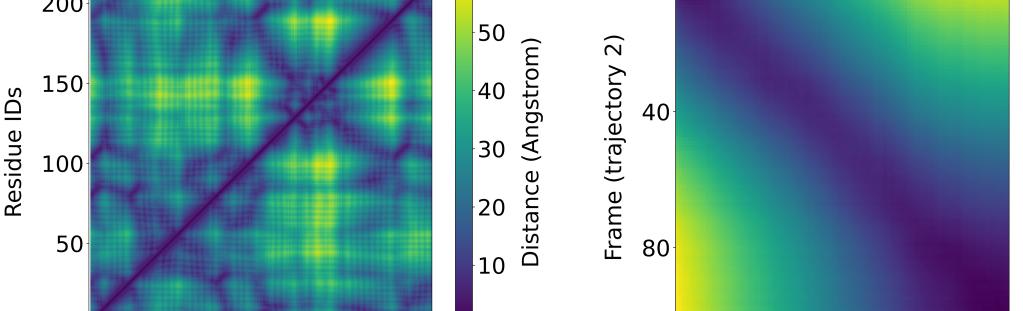


```
u = mda.Universe(TOP, TRAJ)
ligand = u.select_atoms('resname LIG')
```

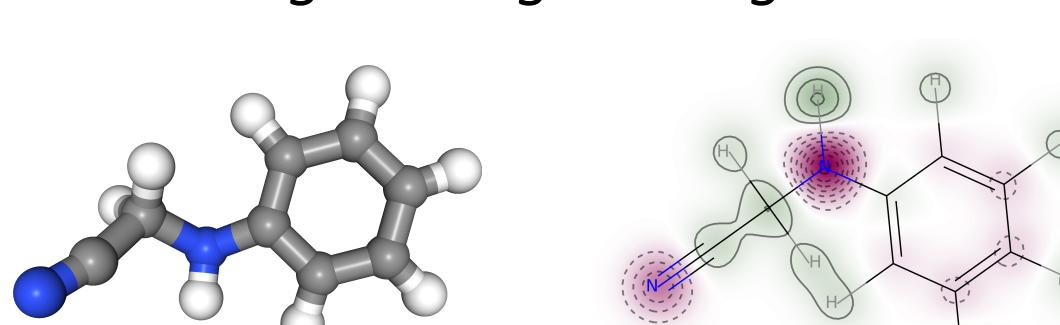
unwrap (make whole)



Pairwise distance



Calculating Gasteiger charges via RDKit



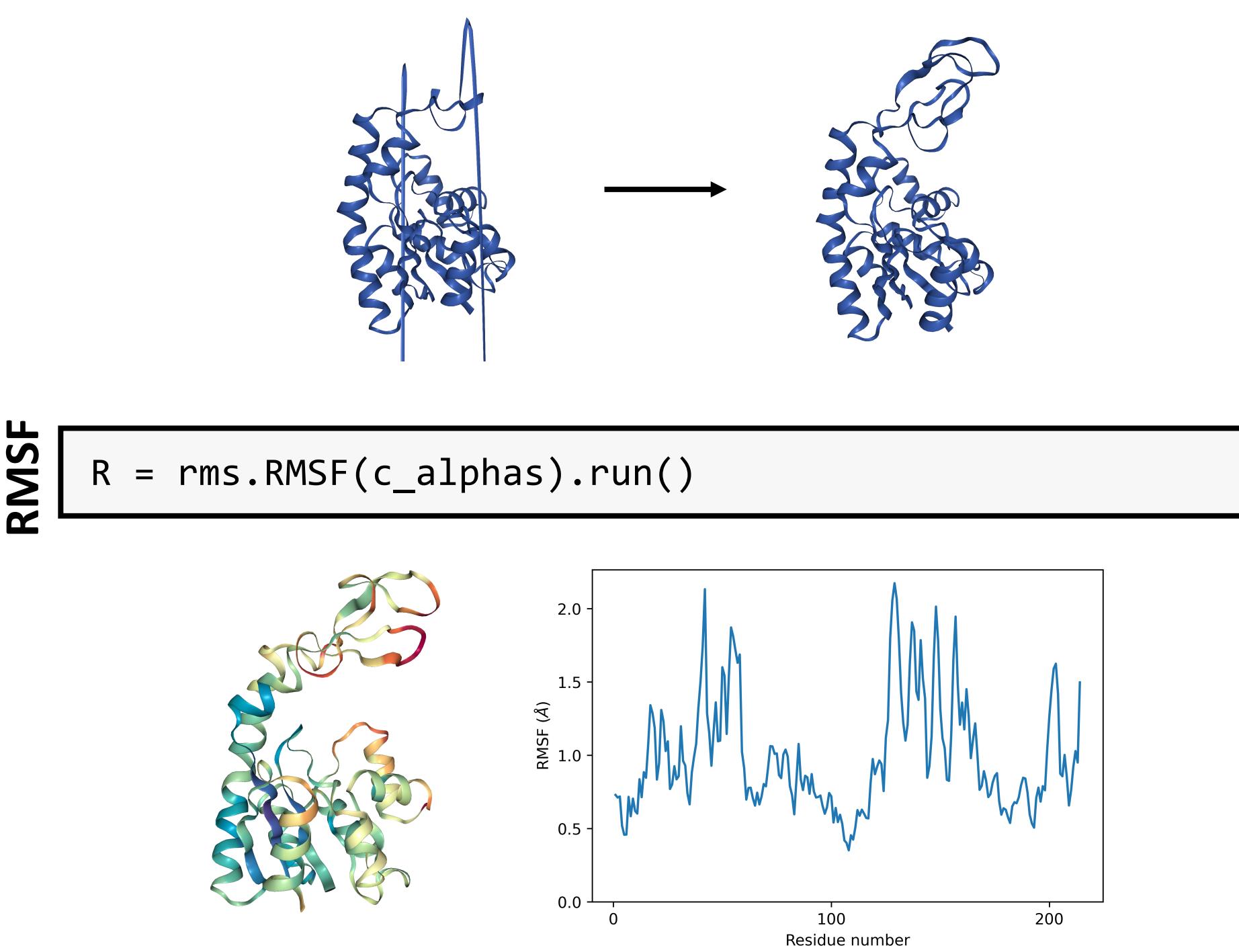
Example Analysis: RMSF

```
import MDAnalysis as mda
import MDAnalysis.transformations as transf
from MDAnalysis.analysis import rms, align

u = mda.Universe(TOP, TRAJ)
protein = u.select_atoms('protein')
not_protein = u.select_atoms('not protein')
c_alpha = u.select_atoms('protein and name CA')

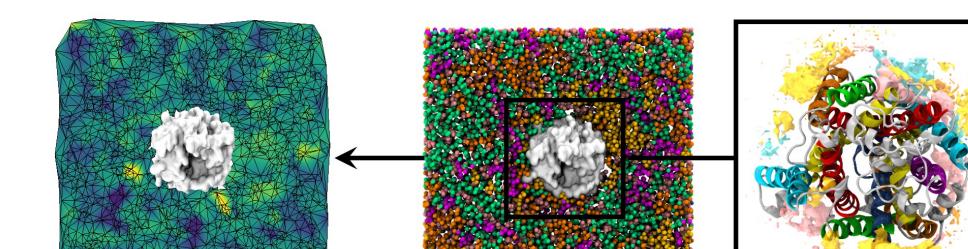
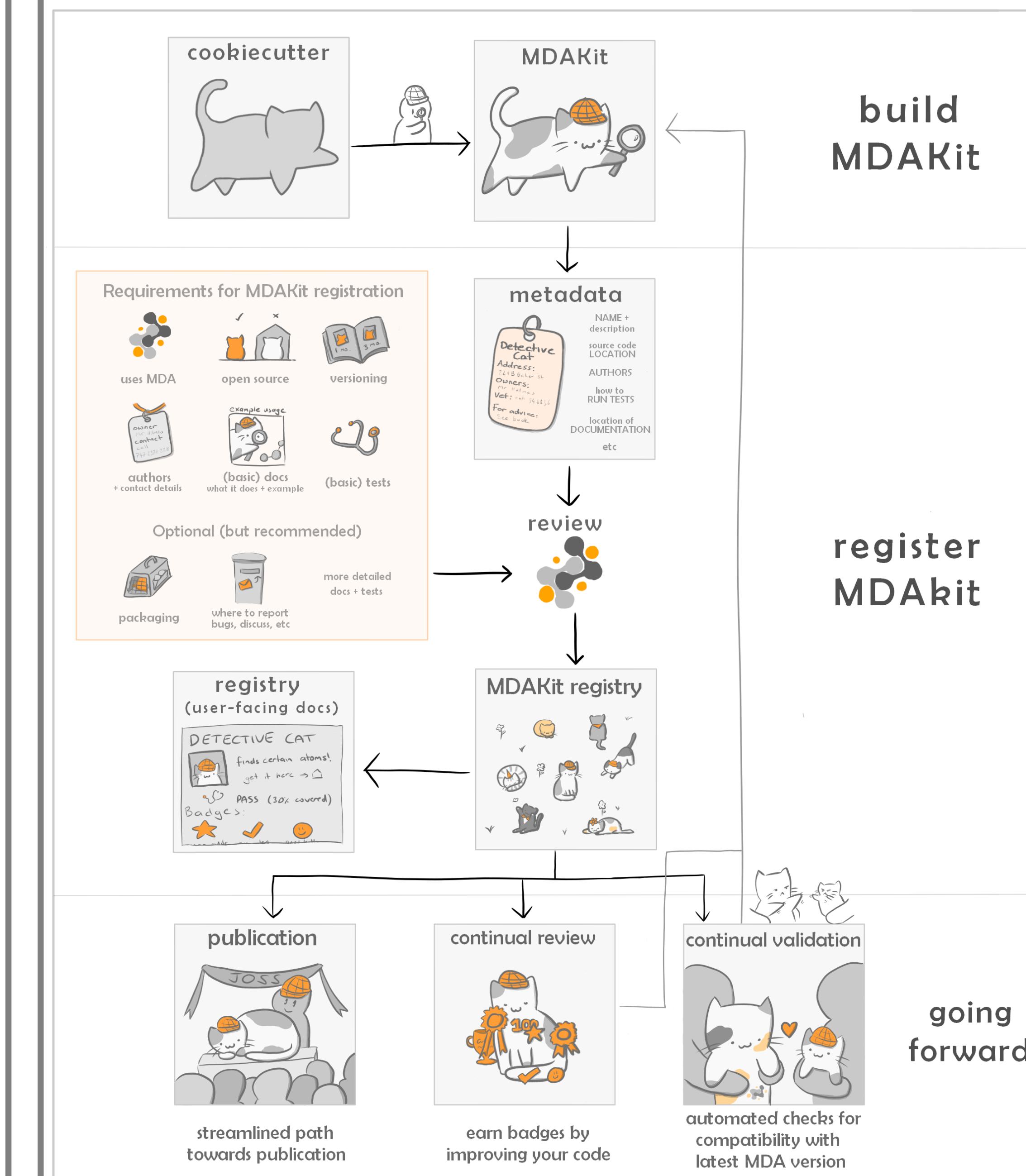
transform = [transf.unwrap(protein),
            transf.center_in_box(protein, wrap=True),
            transf.wrap(not_protein),
            transf.fit_rot_trans(
                c_alpha, c_alpha, weights='mass')]
u.trajectory.add_transformations(transform)

RMSF = rms.RMSF(c_alpha).run()
```

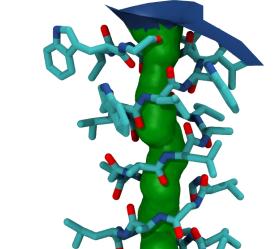


MDAKits: towards sustainable community-built tools

The MDKit framework



lipids
Leaflet-based membrane analysis
Authors: Lily Wang



hole2-mdkit
Pore Analysis of MD trajectories.
Authors: Ian Kenney et al.



MDKit Registry

And maybe even your MDKit!



Read our whitepaper to know more.

New Improvements & Future Directions

• Improved Cythonization (C/C++) of core components

Faster distance library (distopia), improved memory access for core objects.

• New file formats, analyses, and components

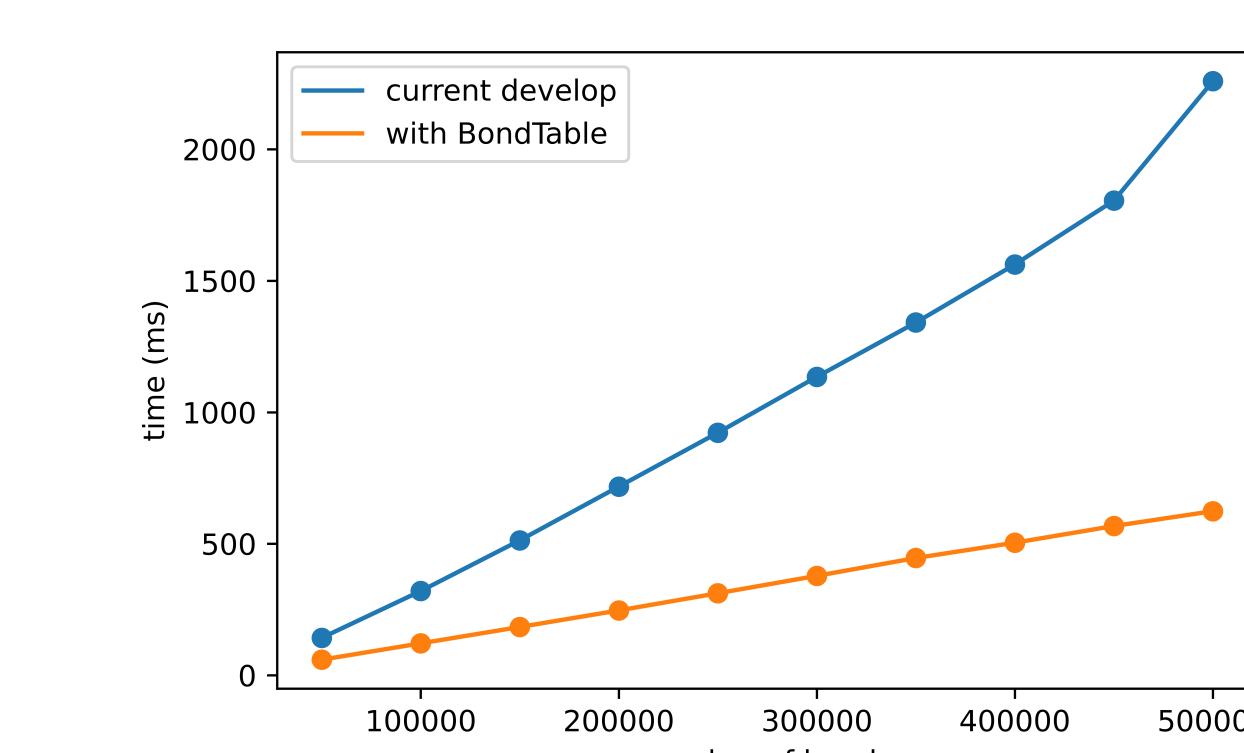
TNG and EDR readers, Analysis Collections.

• In progress addition of new converters

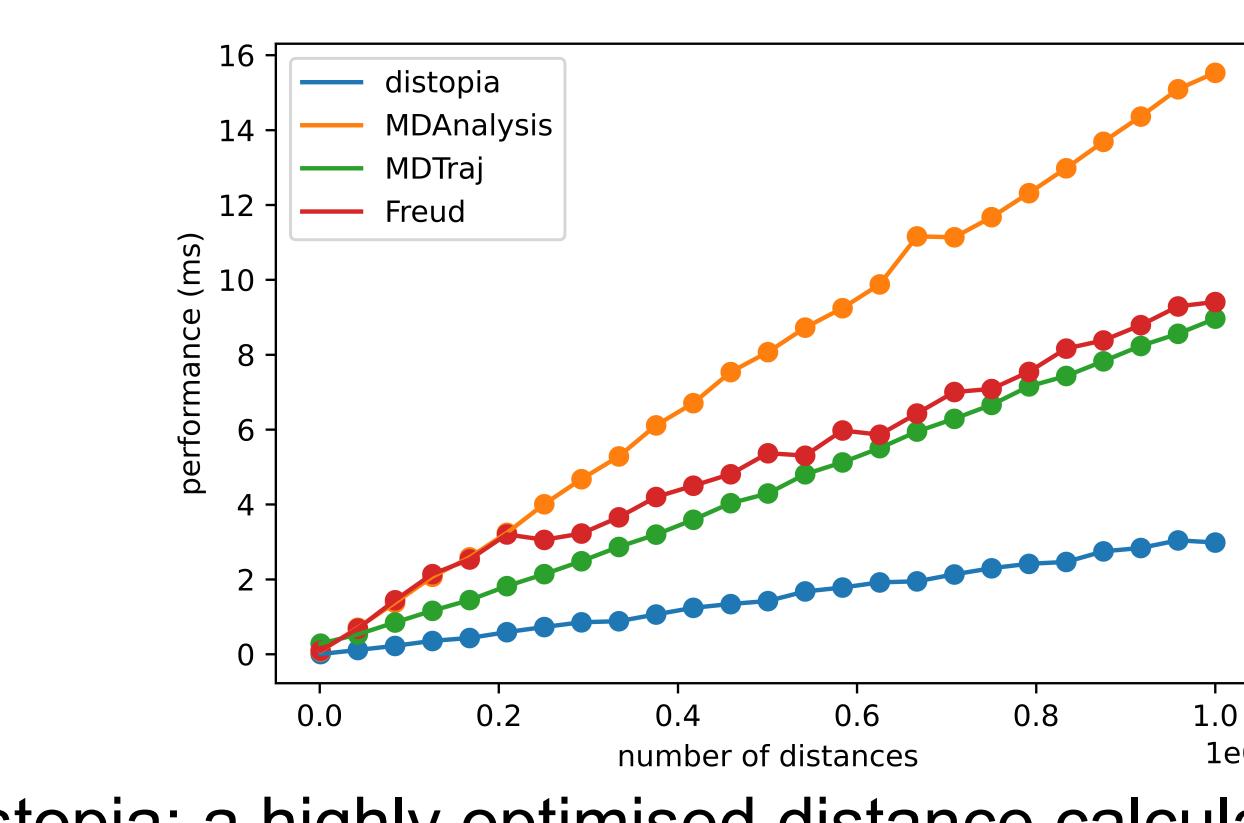
ASE, OpenBabel, LOOS, PyTraj, MDTraj.

• Looking to change of license to LGPLv2+

Removing downstream copyleft requirement.



Improved access of bond information via lookup table



Distopia: a highly optimised distance calculation library

Get started with MDAnalysis today!

GitHub



github.com/MDAnalysis

User Guide



userguide.mdanalysis.org

Acknowledgements

MDAnalysis would not exist without the continued input of the community, including its 172 authors and support from C2I EOSS4 [2021-237663(5022)] and EOSS5 funding.

