

MDAnalysis

A Python library for handling molecular simulation data

CZI Open Science 2022 Annual Meeting

Get these slides



GitHub



github.com/MDAnalysis



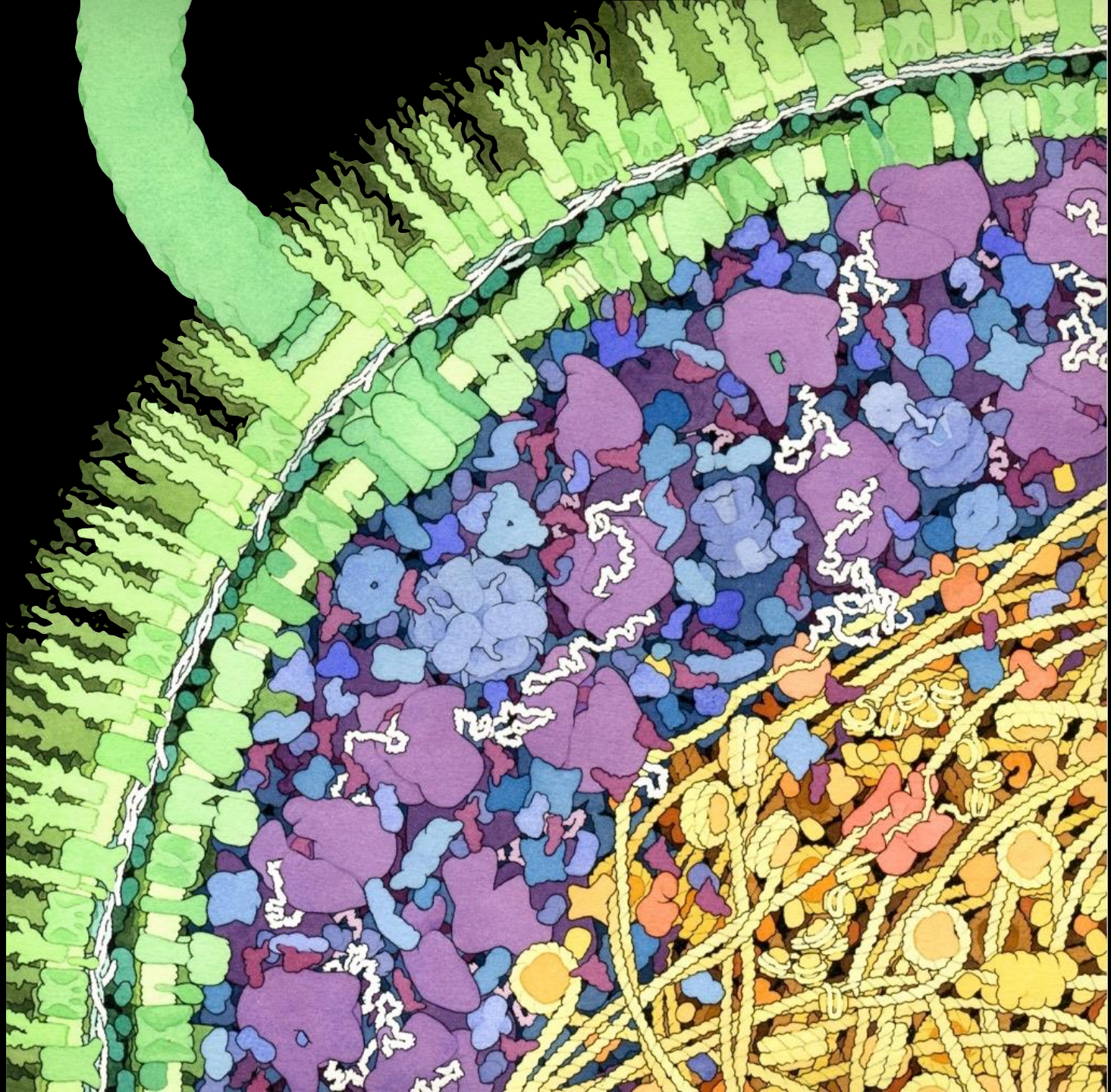
NUMFOCUS
OPEN CODE = BETTER SCIENCE

LIFE

cells

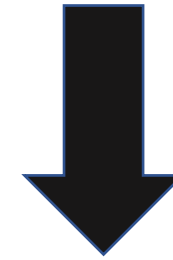
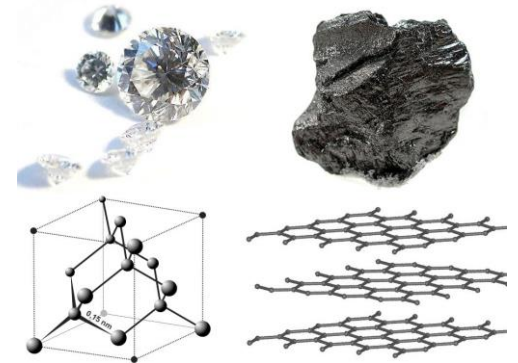
molecules

atoms



From experiment to simulation

- Atomic structural details from experiment
 - X-ray crystallography, Cryo-EM, NMR
 - Provide a limited view of molecular behaviour
- Insights from dynamics
 - Time-dependent re-arrangements
 - Interactions with environment
 - Solutes, macromolecular assembly, etc...
- Physics based models of molecular systems



RCSB **PDB**
PROTEIN DATA BANK

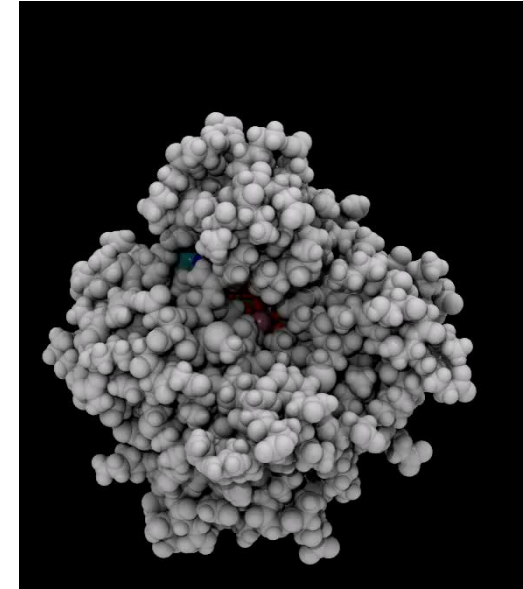
 **EMDataResource**
Unified Data Resource for 3DEM

WORLDWIDE
 **PDB**
PROTEIN DATA BANK

ndb NUCLEIC ACID
DATABASE

From experiment to simulation

- Atomic structural details from experiment
 - X-ray crystallography, Cryo-EM, NMR
 - Provide a limited view of molecular behaviour
- Insights from dynamics
 - Time-dependent re-arrangements
 - Interactions with environment
 - Solutes, macromolecular assembly, etc...
- Physics based models of molecular systems



Energy function (a.k.a “force field”)

$$U(\mathbf{r}_1, \dots, \mathbf{r}_N) = U_{\text{bonded}}(\mathbf{r}_1, \dots) + U_{\text{non-bonded}}(\mathbf{r}_1, \dots)$$

Integrate using classical mechanics

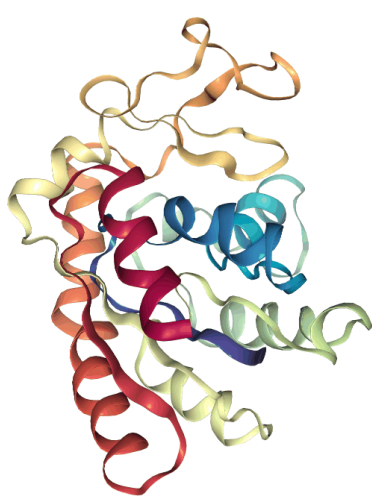
$$\mathbf{F}_i = m_i \mathbf{a}_i$$

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{F}_i}{m_i} \Delta t^2$$

The MDAnalysis library



Molecular Dynamics Simulation



Numerical Analysis



"accessible"
structured data



analysis
algorithms



Insights and Publication!

tables
images
graphs



GROMACS
FAST. FLEXIBLE. FREE.



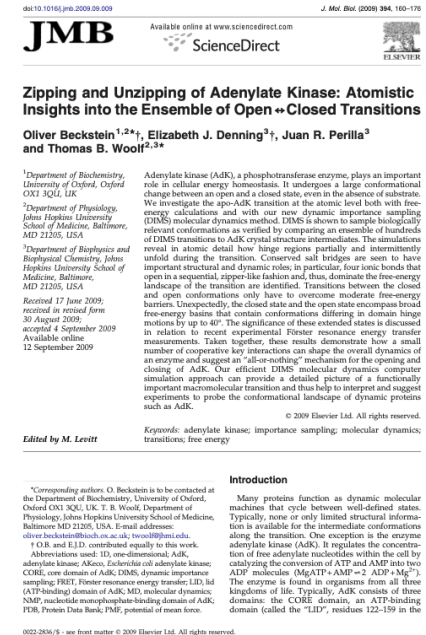
OpenMM

Various MD
Engines

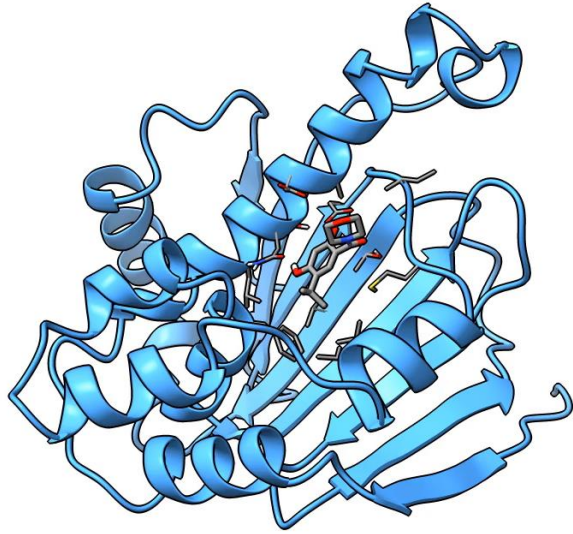


Scalable Molecular Dynamics

Amber



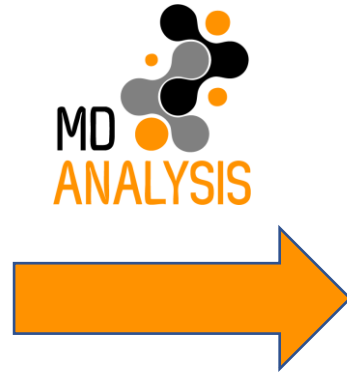
MDAnalysis in practice



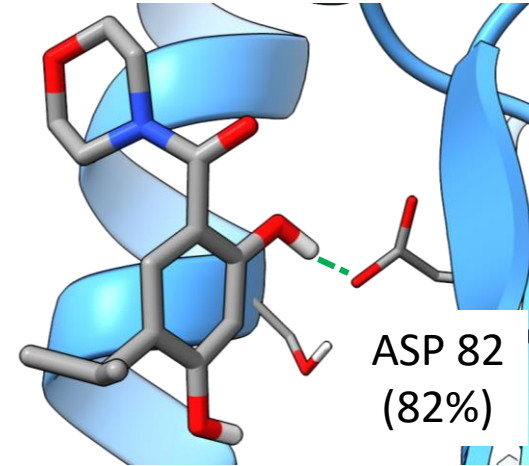
Heat Shock Protein 90
Molecular chaperone



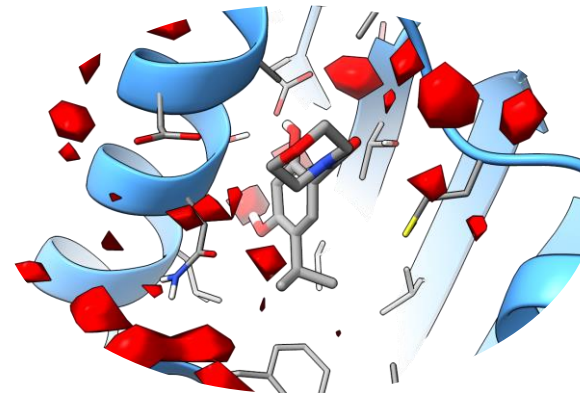
Visualised using ChimeraX !



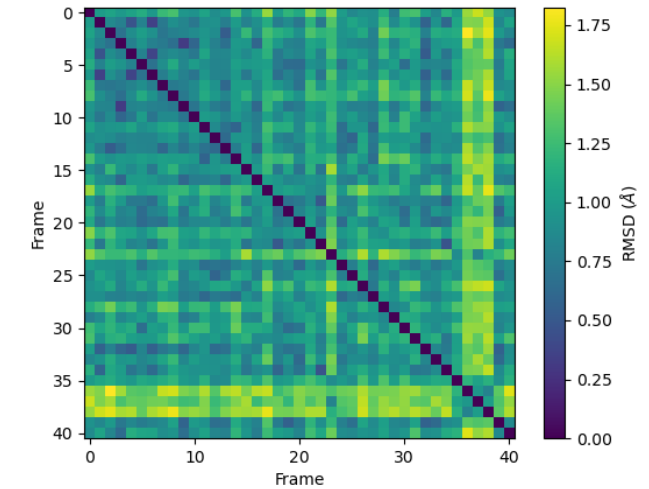
Hydrogen Bond Analysis



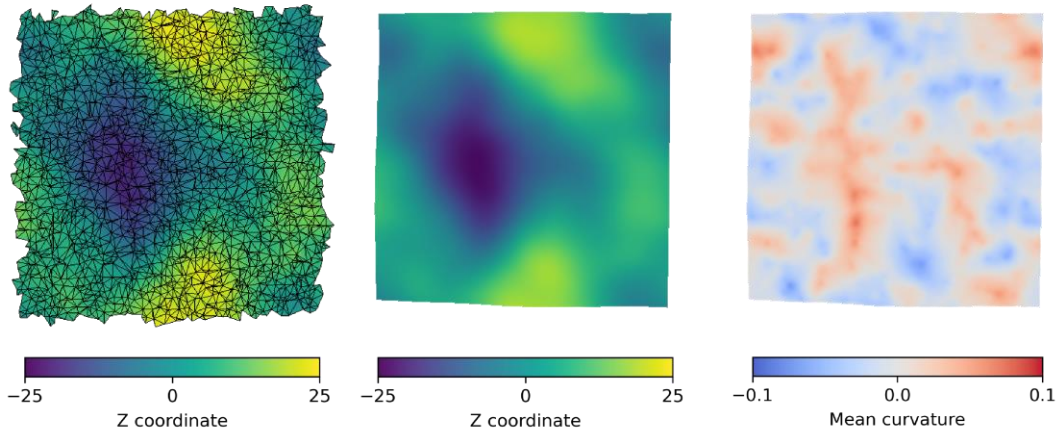
Water density



Ligand heavy-atom RMSD

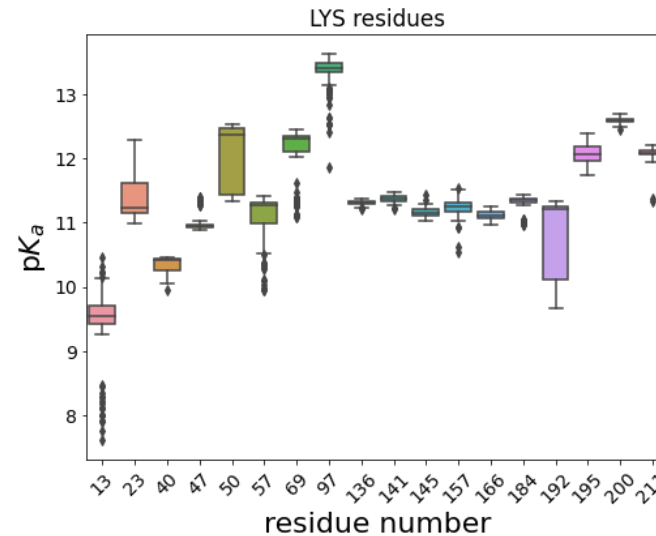


Extensible and interoperable



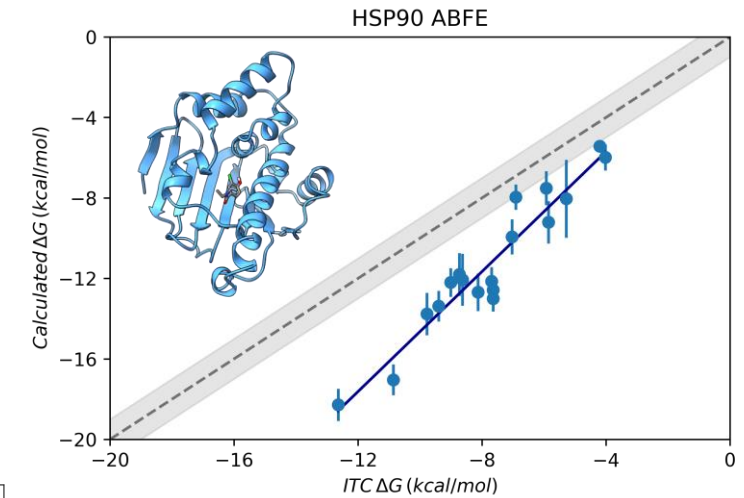
lipyds

*Analysis of membrane bilayer properties
(e.g. area per lipid, thickness, etc...)*



propkatraj

Analysis of protein pKa changes over a trajectory using PropKa



MDRestrainsGenerator

*Create inputs to free energy calculations
from simulation trajectories*

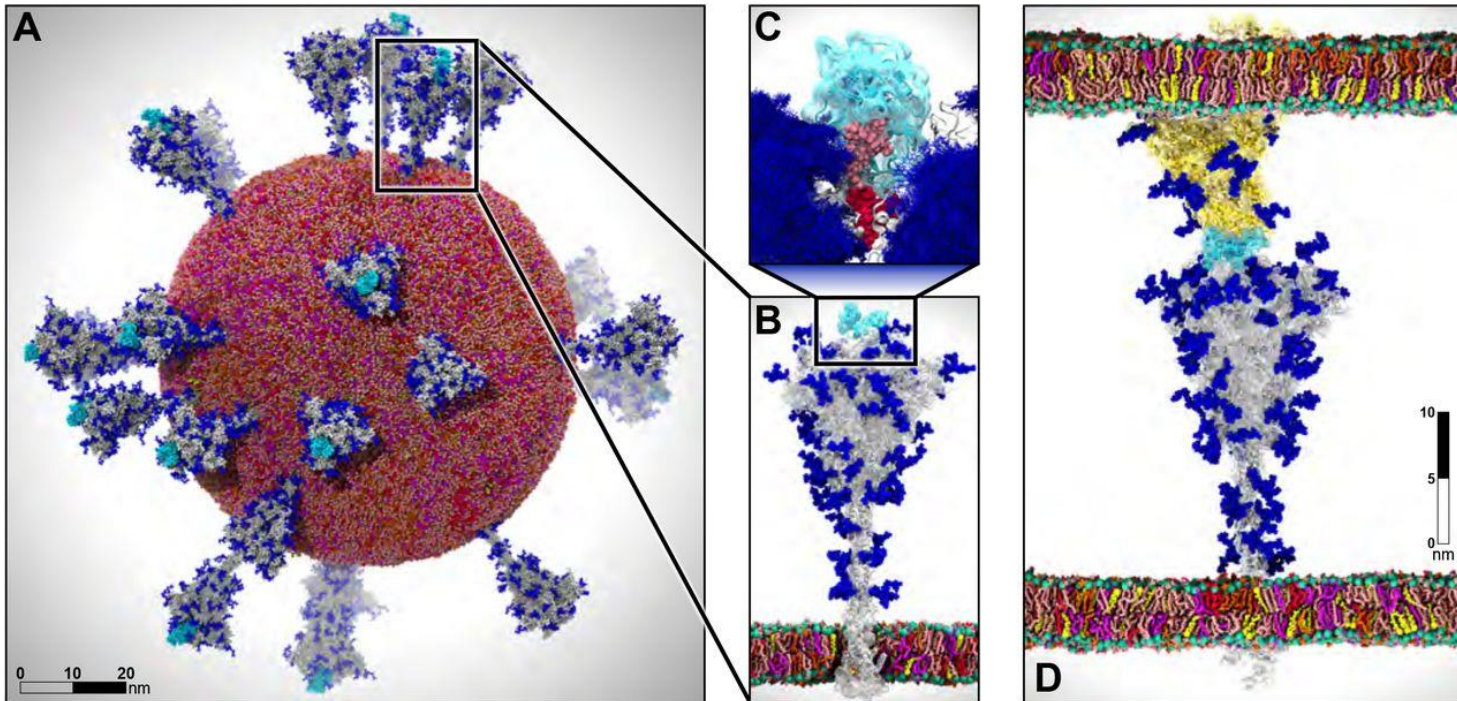
EOSS 4 Project Overview



- Mature, well-established library
 - Opportunity to focus on user experience
- Two major aims
 1. Improving performance
 2. Enabling reproducible MDAnalysis-using code development

EOSS 4: Improving Performance

- Moving towards a post-exascale future
 - Multi-million atoms, hundreds of terabytes

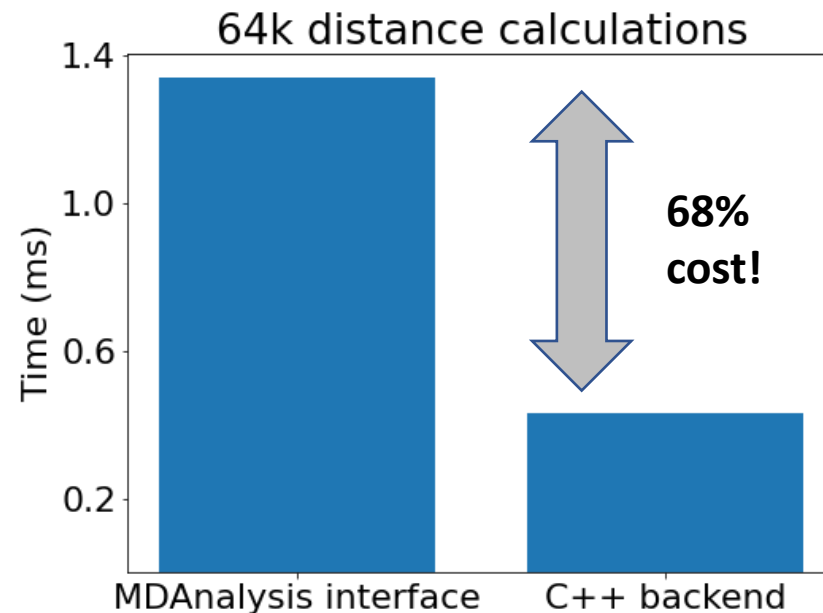


305 million atoms all-atom model of SARS-CoV2 viral envelope

Casalino et al., AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV2 Spike Dynamics, *Int. J. High Perform. Comput. Appl.*, **2020**

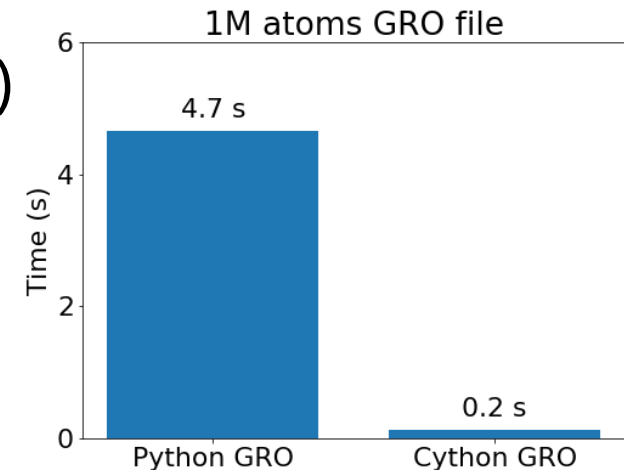
EOSS 4: Improving Performance

- Hindered by interpreted (Python) routines and data structures
 - Despite heavy NumPy use and Cythonised math library



EOSS 4: Improving Performance

- Hindered by interpreted (Python) routines and data structures
 - Despite NumPy use and Cythonised math library
- Project aim: Cythonise core data structures and file parsers
 - Reduce memory access overheads
 - Improved performance (especially for ASCII formats)
 - Better interface with existing C-level libraries



EOSS 4: Enabling Reproducibility

- Software reproducibility crisis
 - Code rarely provided in publications
 - Improved by decade of community efforts to solve this
- Provided code often insufficient
 - Lacks test, documentation, version control...
 - Quickly becomes non-reproducible
 - Python is very dynamic!
- Outcomes
 - Duplicate of efforts re-implementing
 - Silent API changes can lead to erroneous results!

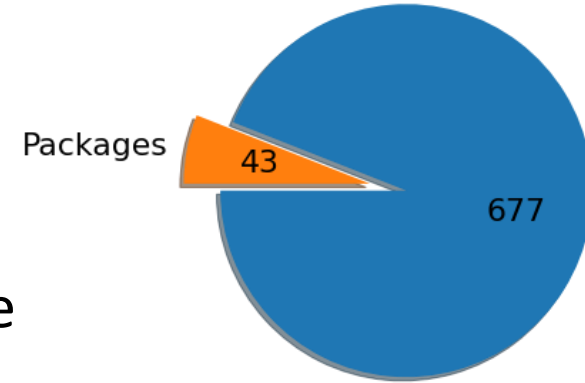


Fig 1. Number of MDAnalysis-using packages since 2017*

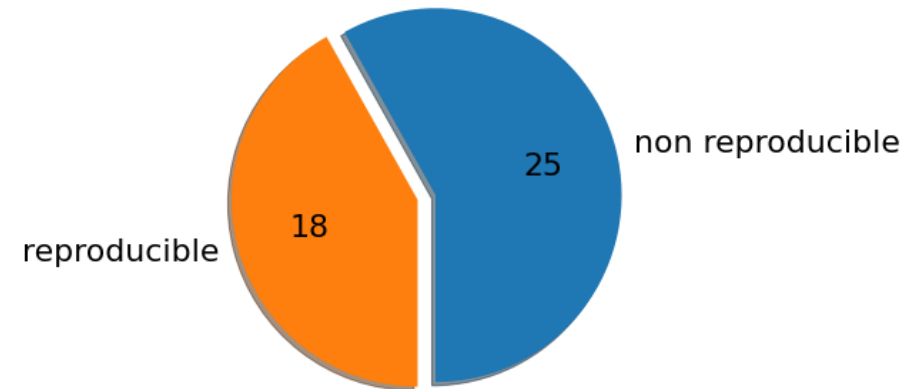


Fig 2. Breakdown of “reproducible” packages published using MDAnalysis**

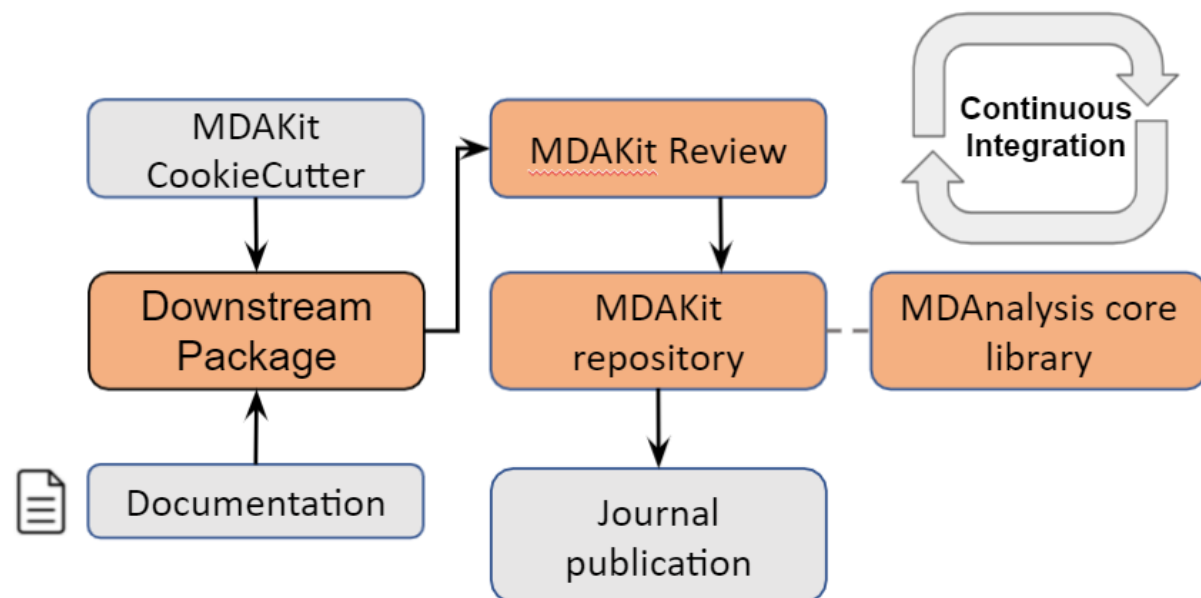
*Data gathered from 2017+ Scopus & JOSS entries, “package” is defined as code advertised for re-use in a public repository. N.B. Likely underestimated counts as Scopus has limited indexing of some journals

** Reproducible is counted as having: unit tests, non-minimal documentation, and a means of installation (e.g. setuptools)

EOSS 4: Enabling Reproducibility

The MDAKit Ecosystem

- Advertised downstream package ecosystem
- Provide framework (e.g. cookiecutter) & documentation
- MDAKits registry
 - Expose and validate MDAKits
 - Continuously verified with upstream changes
- Encourage publication of tools



**See our
whitepaper!**

Thanks for listening :)

GitHub



github.com/MDAnalysis

User Guide



userguide.mdanalysis.org

Acknowledgements

All 163 MDAnalysis code contributors and the many more community members that use MDAnalysis, report bugs, and make feature requests.



Join the conversation



@mdanalysis



discord.gg/fXTSfDJyxE



Acknowledgements



163 code contributors and countless community members

Naveen Michaud-Agrawal, Elizabeth J. Denning, **Oliver Beckstein**, Danny Parton, Philip Fowler, **Tyler Reddy**, Joseph Goose, **Jan Domanski**, Benjamin Hall, Paul Rigor, David Caplan, Christian Beckstein (logo), **Sébastien Buchoux**, Joshua L. Adelman, Lukas Grossar, Andy Somogyi, Lukas Stelzl, Jinju Lu, Joshua L. Phillips, Zhuyi Xue, Xavier Deupi, **Manuel Nuno Melo**, Robert McGibbon, **Richard J. Gowers**, Alejandro Bernardin, Lennard van der Feltz, Matthieu Chavent, Joe Jordan, Alex Nesterenko, Caio S. Souza, Sean L. Seyler, **David L. Dotson**, Carlos Yanez S., Kyle J. Huston, Isaac Virshup, **Max Linke**, Gorman Stock, **Jonathan Barnoud**, Hai Nguyen, Balasubramanian, Mattia F. Palermo, Utkarsh Saxena, Abhinav Gupta, **John Detlefs**, Eugen Hruska, Bart Bruininks, **Fiona B. Naughton**, **Robert Delgado**, Wouter Boomsma, **Matteo Tiberti**, Tone Bengtsen, Shantanu Srivastava, Pedro Reis, Ruggero Cortini, Zhiyi Wu, Kashish Punjani, **Utkarsh Bansal**, Shobhit Agarwal, Vedant Rathore, Akshay Gupta, Juan Eiros Zamora, Jon Kapla, Sang Young Noh, Andrew William King, **Kathleen Clark**, Dominik 'Rathann' Mierzejewski, Nestor Wendt, **Micaela Matta**, Jose Borreguero, Sören von Bülow, Nabarun Pal, Mateusz Bieniek, Paul Smith, Navya Khare, **Johannes Zeman**, **Ayush Suhane**, **Davide Cruz**, Shujie Fan, Andrew R. McCluskey, **Henry Mull**, **Irfan Alibay**, Philip Loche, Matthew W. Thompson, Ali Ehlen, Daniele Padula, **Ninad Bhat**, Fenil Suchak, Yibo Zhang, Luís Pedro Borges Araújo, Abhishek A. Kognole, Rocco Meli, **Lily Wang**, Matthijs Tadema, Joao Miguel Correia Teixeira, Charlie Cook, Yuanyu Chang, Guillaume Fraux, Ivan Hristov, Michael Quevillon, Hao Tian, **Hugo MacDermott-Opeskin**, Anshul Angaria, Shubham Sharma, **Yuxuan Zhuang**, **Cédric Bouysset**, Abhishek Shandilya, Morgan L. Nance, Faraaz Shah, Wiep van der Toorn, Siddharth Jain, Ameya Harmalkar, Shakul Pathak, Andrea Rizzi, William Glass, Marcello Segal, **Edis Jakupovic**, Nicholas Craven, Mieczyslaw Torchala, Ramon Crehuet, Haochuan Chen, Karthikeyan Singaravelan, Aditya Kamath, Leonardo Barneschi, Henrik Jäger, Jan Stevens, **Orion Cohen**, Dimitrios Papageorgiou, Hannah Pollak, **Estefania Barreto-Ojeda**, Paarth Thadani, Henry Kobin, Kosuke Kudo, Sulay Shah, Alexander Yang, Filip T. Szczypiński, Marcelo C. R. Melo, Mark D. Driver, Kevin Boyd, Atharva Kulkarni, Yantong Cai, **Bjarne Feddersen**, Pratik Gupta, Alexander Gorfer, **Aya M. Alaa**, Kazi Shudipto Amin, Alia Lescoulie, Henok Ademtew, **Uma D Kadam**, Tamandeep Singh, Mingyi Xue, Meghan Osato, Anirvinya G, Rishabh Shukla, Manish Kumar, Aditi Tripathi, Sukeerti T, Kavya Bisht, Mark Verma, Marcelo D. Poletto, Ricky Sexton, Rafael R. Pappalardo, Tengyu Xie, Raymond Zhao

