ESR 6:

Data-driven MD: Calculating free energies by learning from QM Potentials & cryo-EM data

Funded by
the European Union

AQTIVATE

About me

- Lukas Müllender
- M.Sc. Physics, April 2023
  RWTH Aachen University, Supervisor: Prof. Paolo Carloni
- ESR6: Data-driven MD: Calculating free energies by learning from QM Potentials & cryo-EM data

Prof. Erik Lindahl   Prof. Paolo Carloni   Prof. Michele Parrinello

Outline

- Master's thesis project:
  Combining HPC-based Simulations in Trajectory Space with Machine Learning to Simulate Rare Events

  - Motivation

  - Methods

  - Iterative Approach to CV Design

- Outlook on PhD project ESR6

AQTIVATE

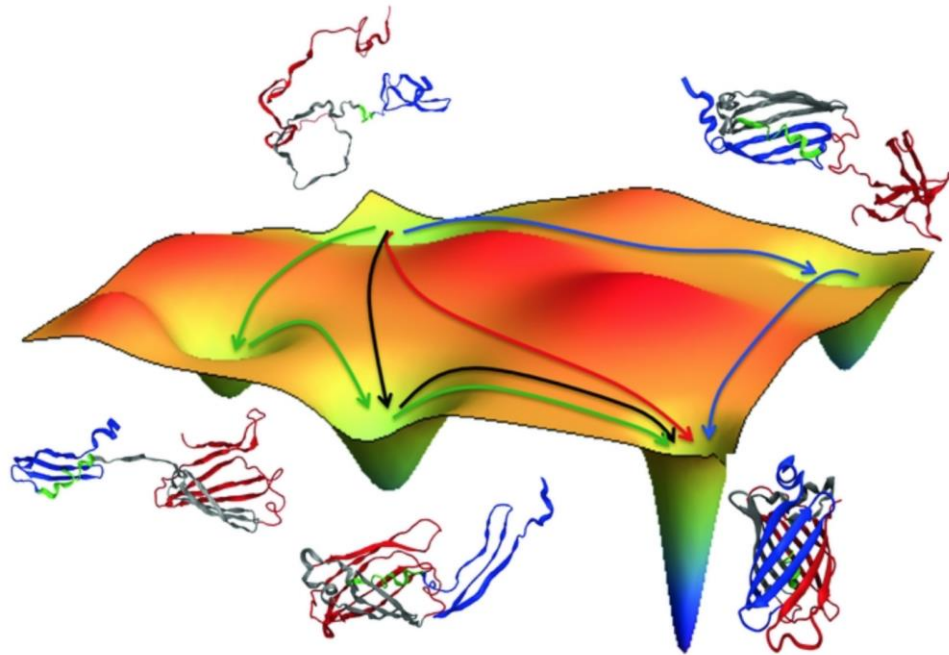# Molecular Dynamics for Biophysics Research

Protein Folding

Ligand Unbinding

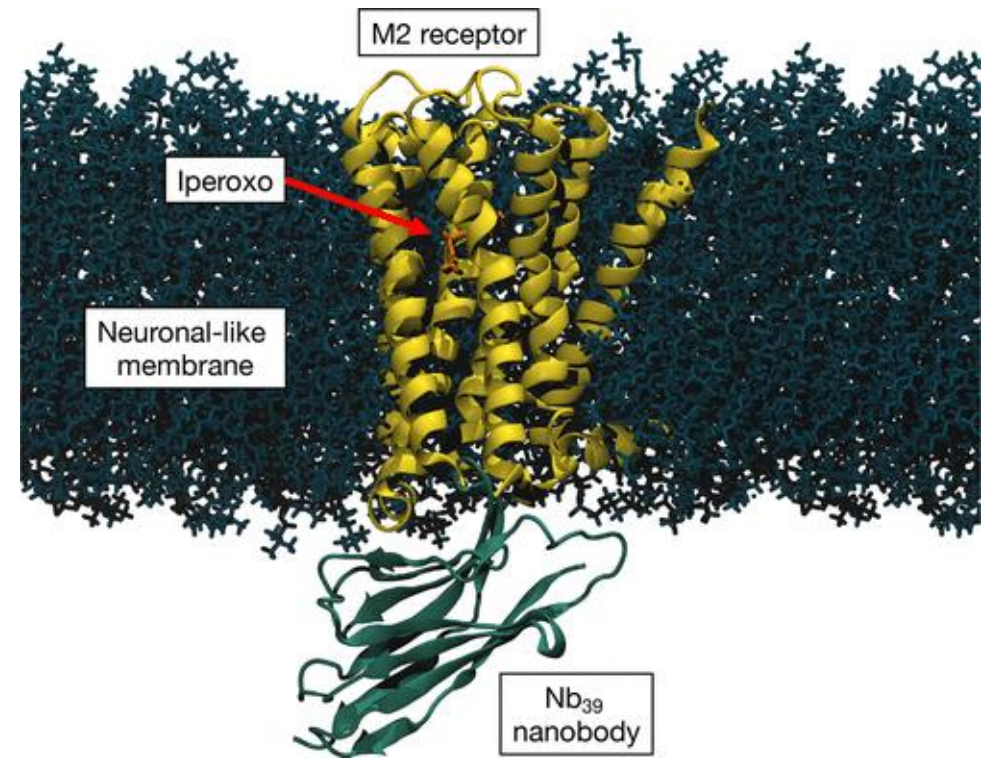*Figure adapted from Reddy et al., PNAS (2012)*

Figure adapted from Capelli et al., J. Phys. Chem. Lett. (2020)

AQTIVATE

Challenges of Simulating Large Molecules

1. *high-dimensional* problem:
   How to determine important
   order parameters?

2. computationally *expensive:*
   How to escape long-lived states?

Solution: *collective variables…*

- ‣ provide reduced representation of a high-dimensional problem
- ‣ distinguish metastable states, transition state
- ‣ reflect the slowly varying degrees of freedom
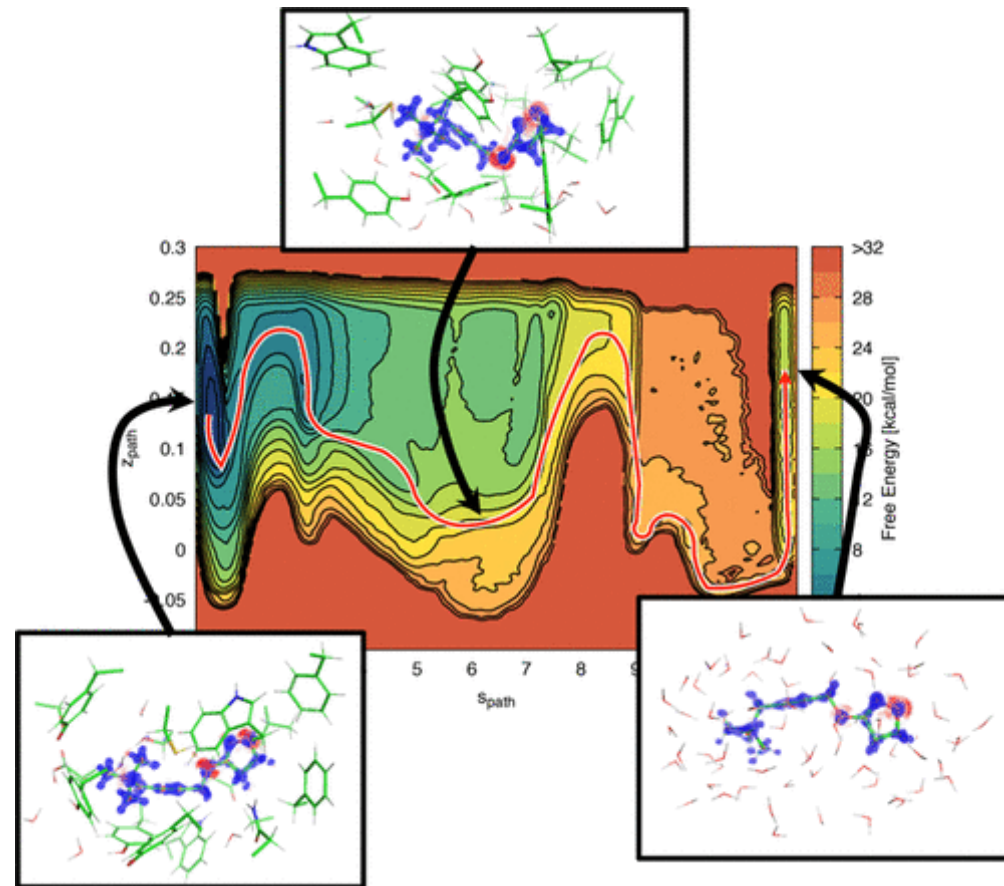- ‣ enable ***enhanced sampling***

Figure adapted from Capelli et al., J. Phys. Chem. Lett. (2020)

CV-based Enhanced Sampling

- Accelerate sampling along the collective variable **s**

- Examples: Metadynamics, Umbrella Sampling, Adaptive Biasing Force, …

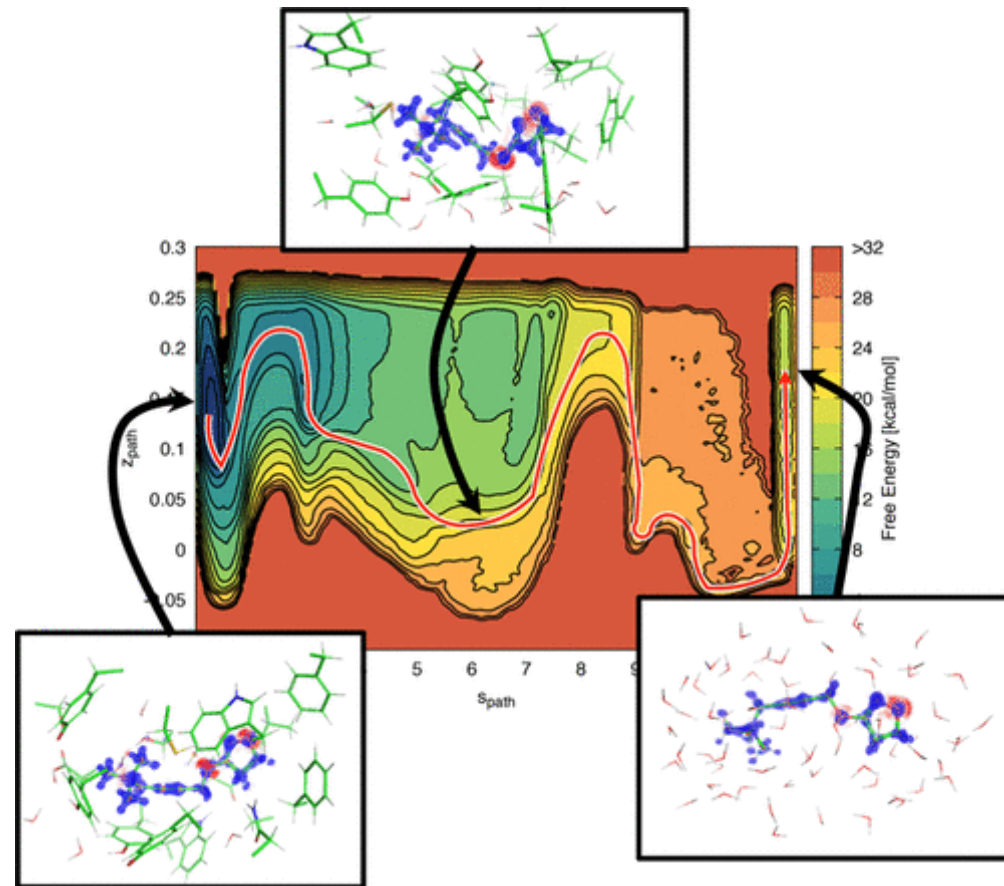- Choice of efficient CVs is not intuitive → data-based CV discovery

Figure adapted from Capelli et al., J. Phys. Chem. Lett. (2020)

Methods: Deep Learning CVs from Transition Paths

DeepTDA:

Distinguish different regions of phase space

\+

Enhanced Sampling

along Transition Pathways
(e.g. MetaD of Paths[1])

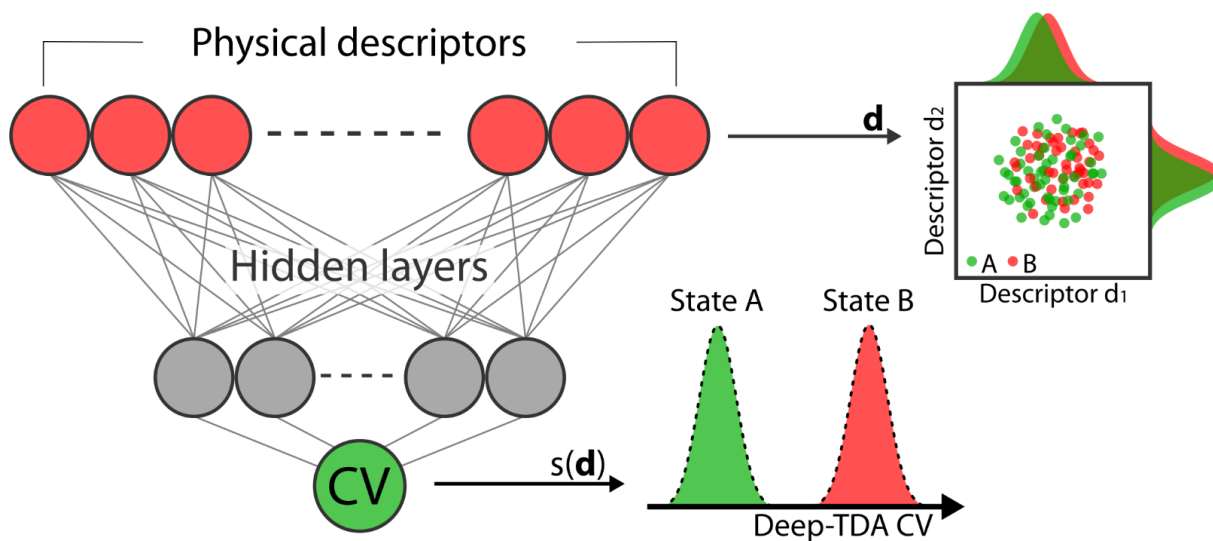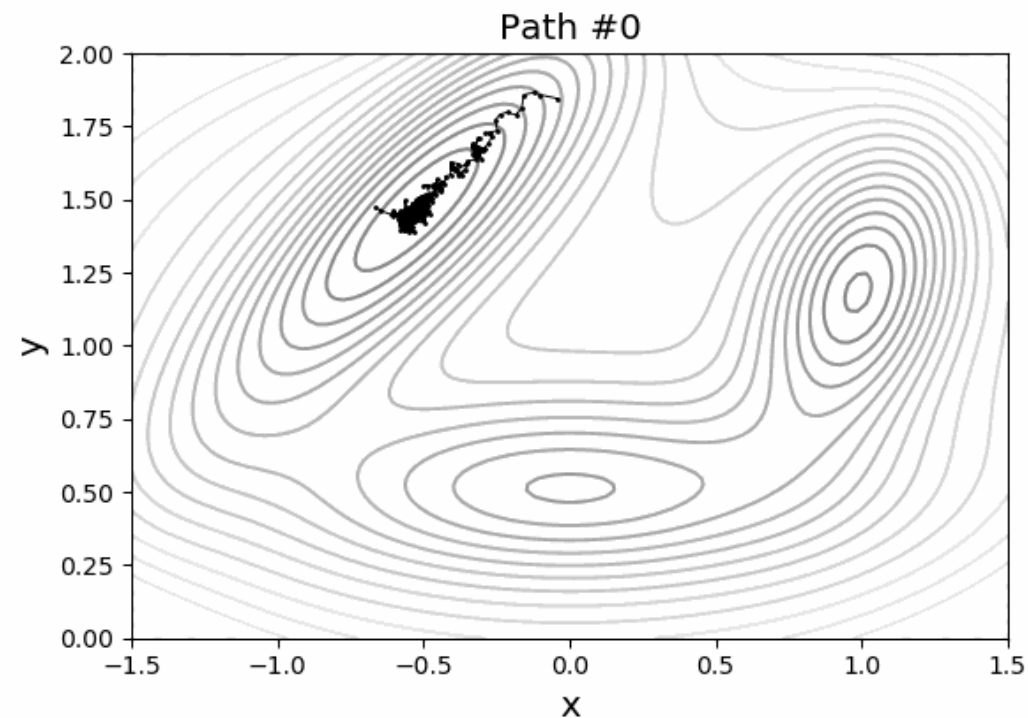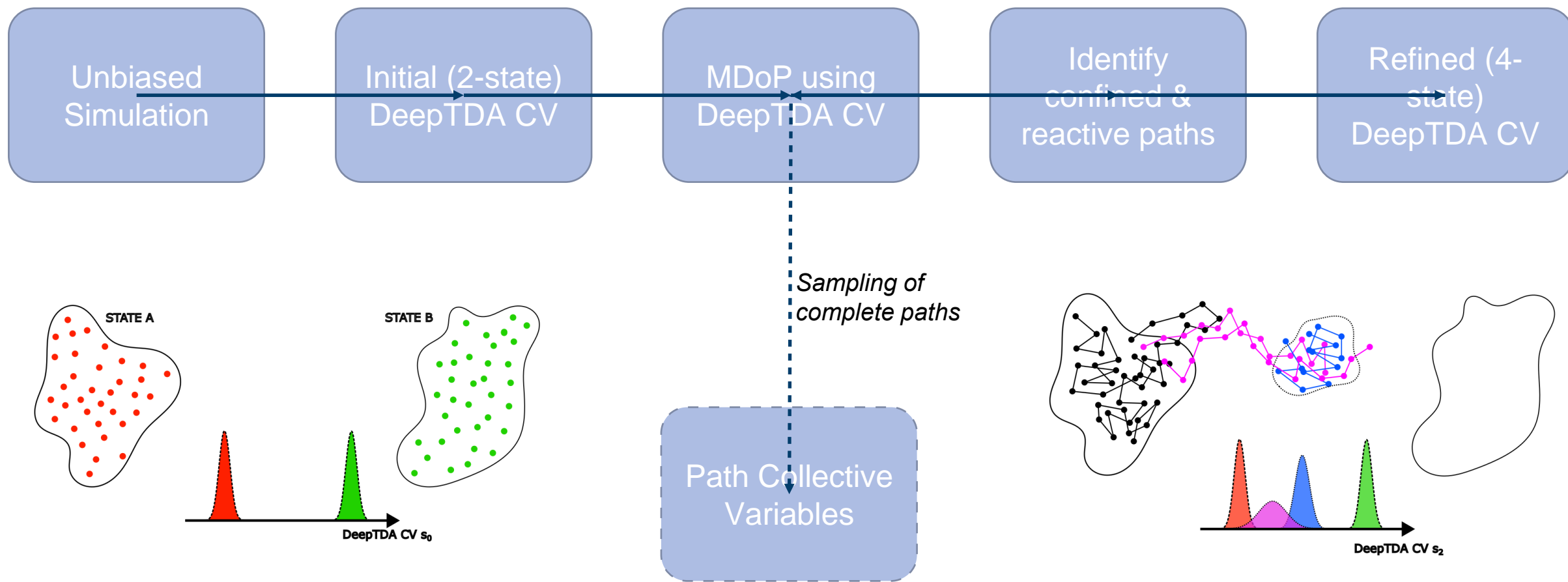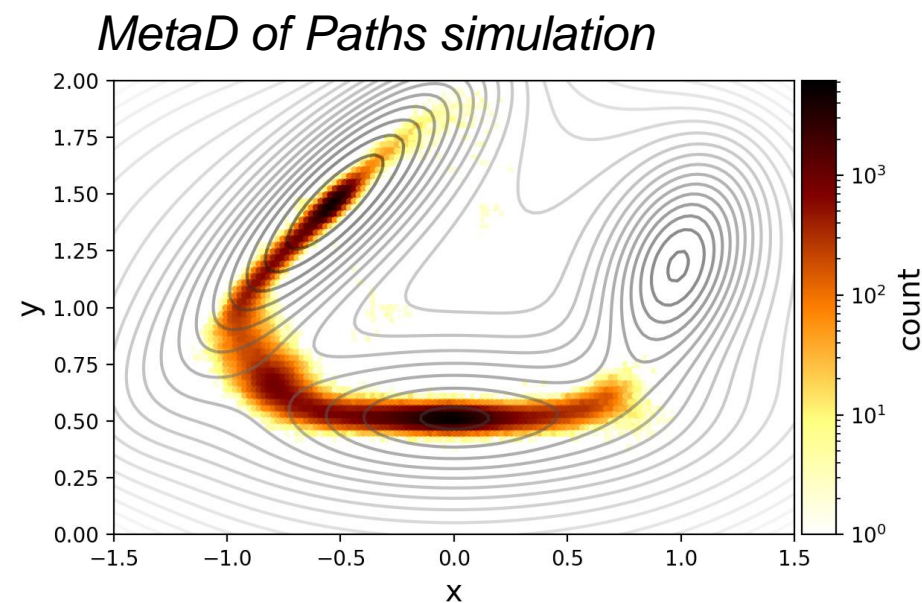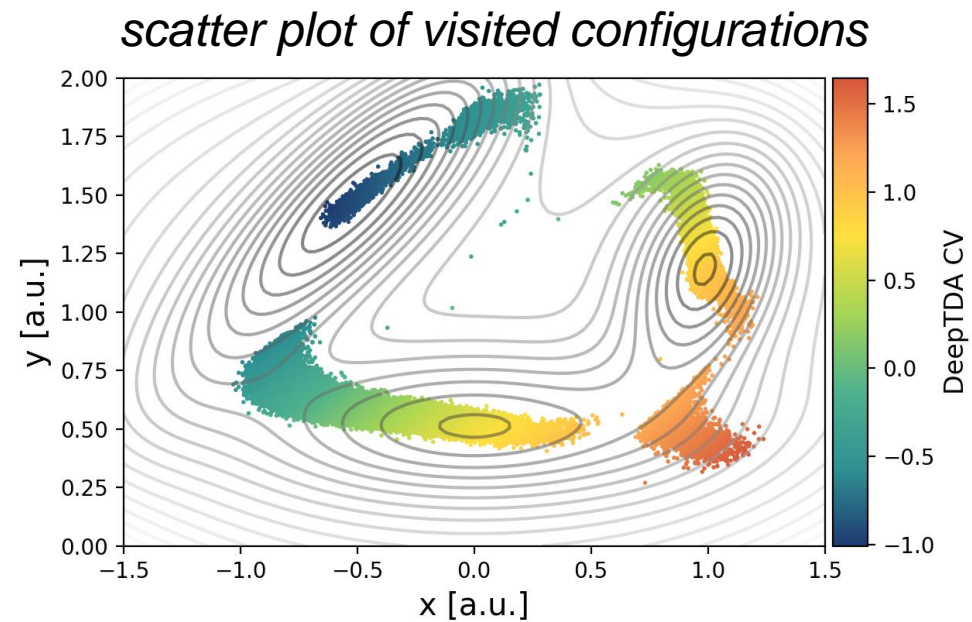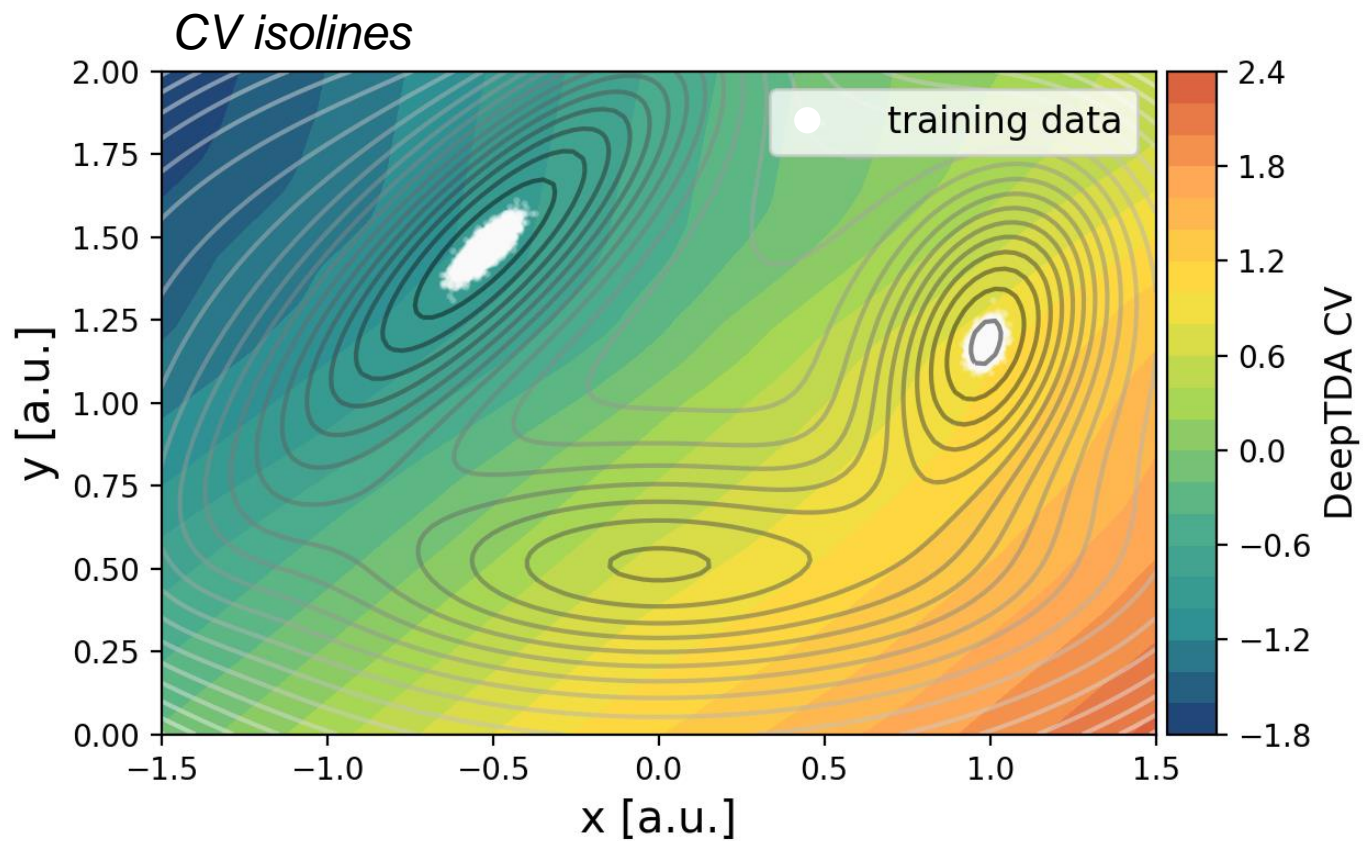*Figure adapted from Trizio and Parrinello, J. Phys. Chem. Let. (2022)*

[1] *Mandelli, Hirshberg and Parrinello, Phys. Rev. Let. (2020)*

Our Workflow: Successive Incorporation of Path Data in DeepTDA Training

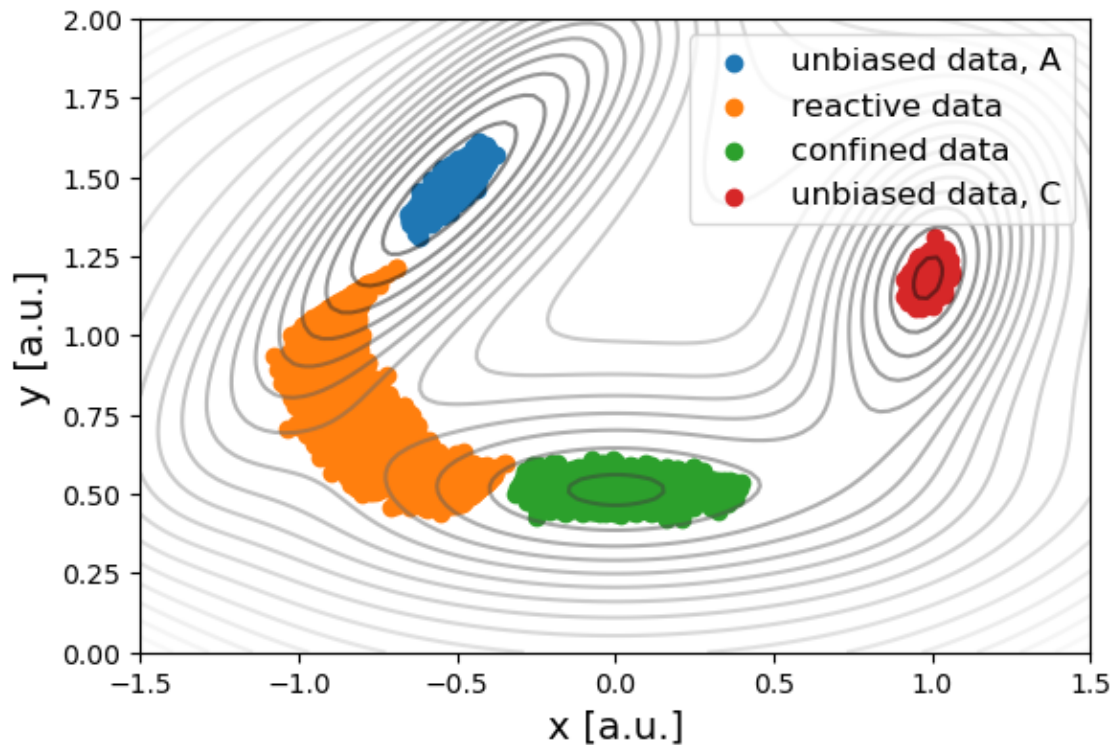Unbiased Simulation → Initial (2-state) DeepTDA CV → MDoP using DeepTDA CV → Identify confined & reactive paths → Refined (4-state) DeepTDA CV

Sampling of complete paths

Path Collective Variables

STATE A

STATE B

DeepTDA CV s_0

DeepTDA CV s_2

Results: Initial CV is suboptimal

CV isolines
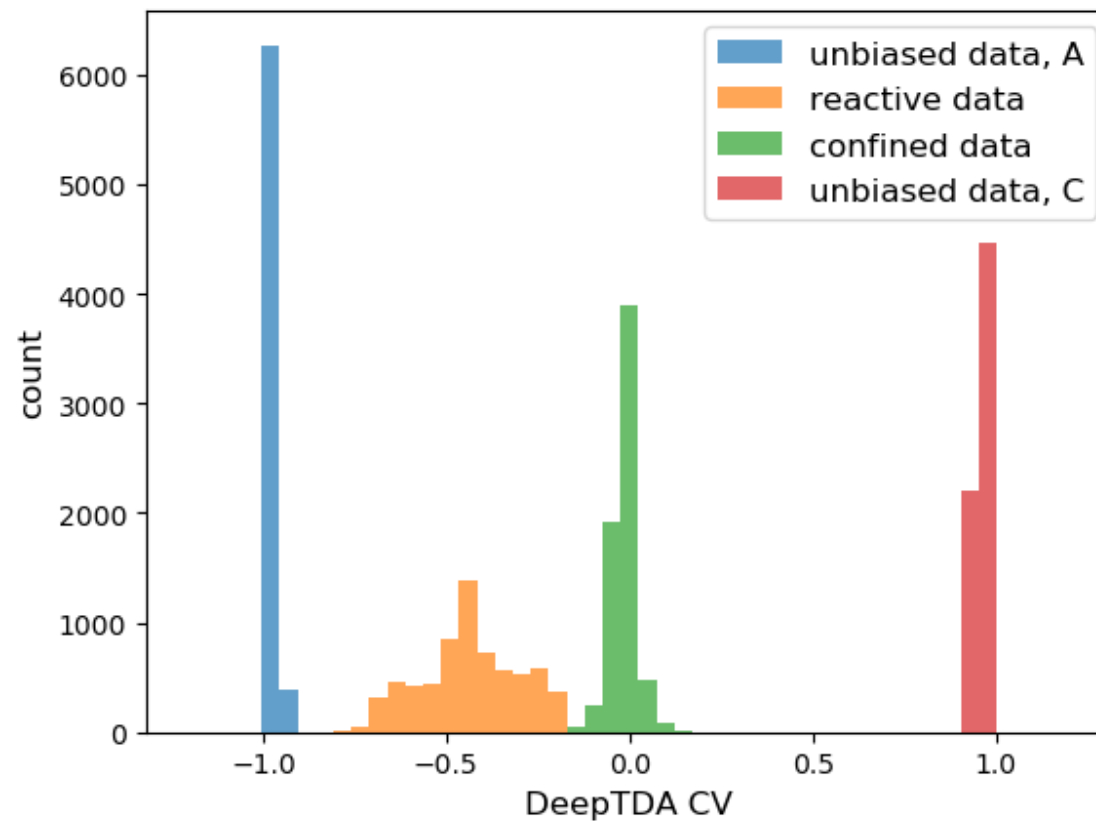
scatter plot of visited configurations

MetaD of Paths simulation
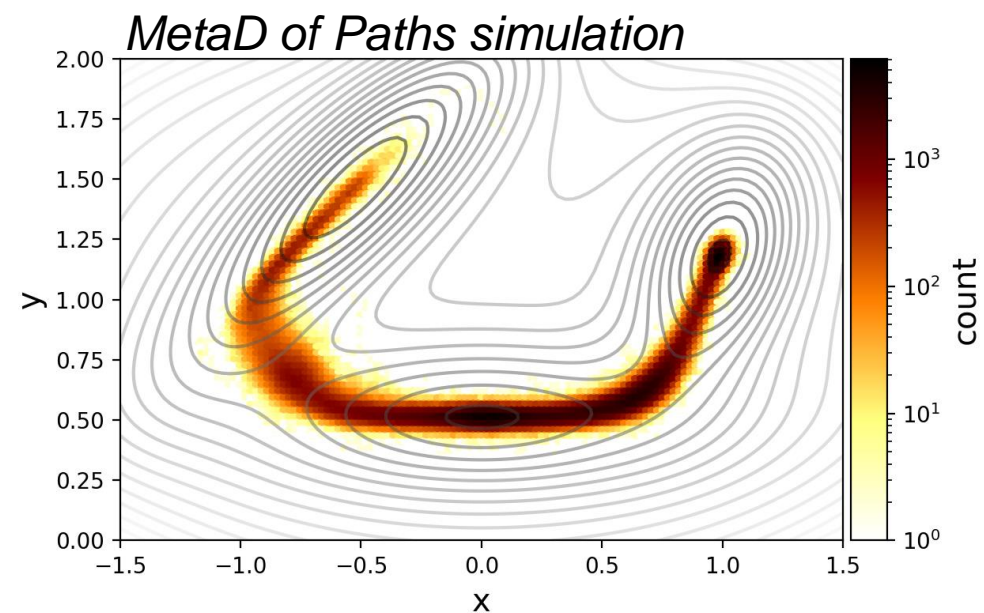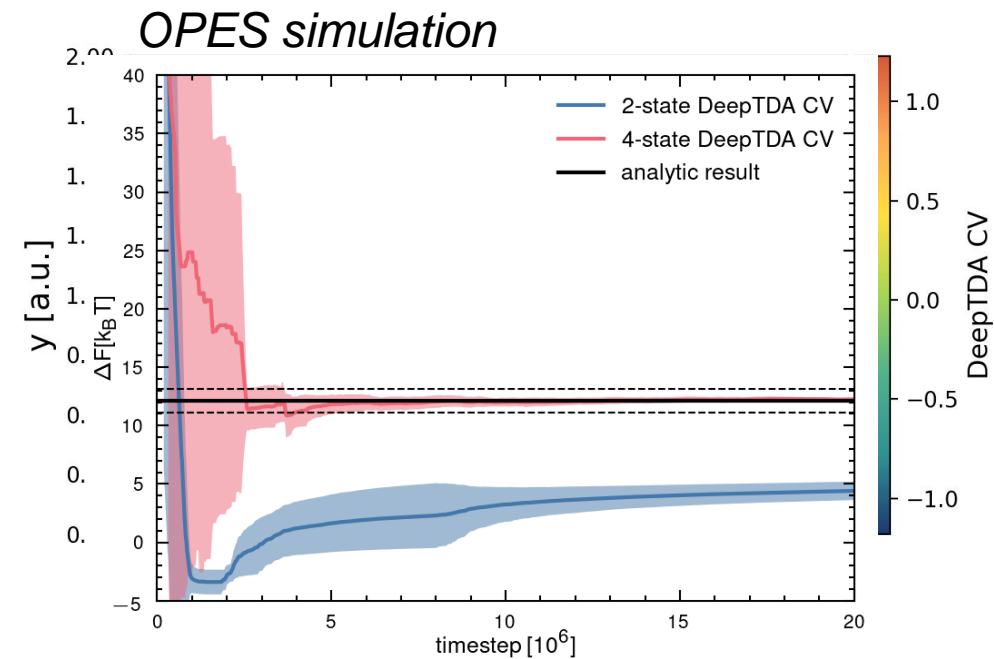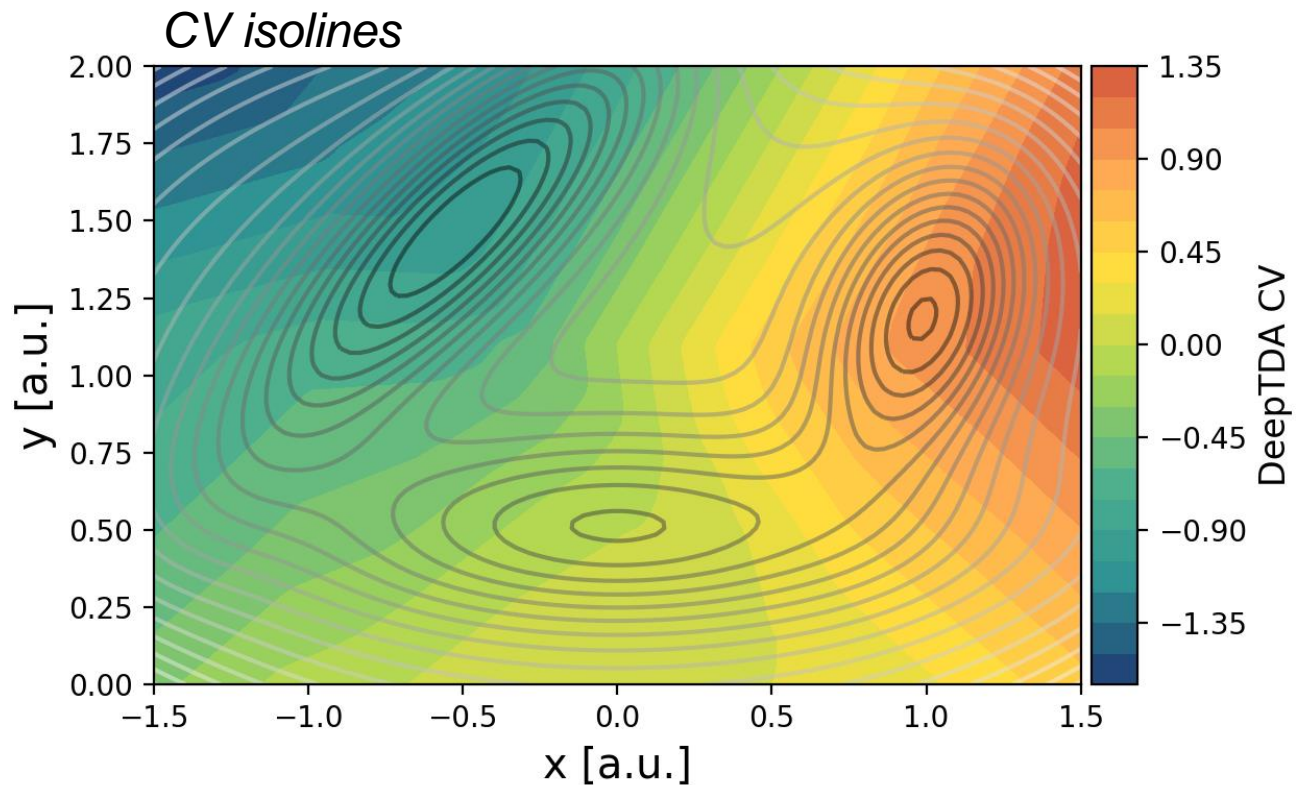
Results: Training of a 4-state DeepTDA CV

4-state training data

histogram of CV values

Results: 4-state CV resolves all states

Outlook

- Paper: *Deep Learning Collective Variables for Enhanced Sampling Simulation* (WIP)
  - Application to model potential, Alanine Dipeptide
- Possible applications to larger molecules/proteins: Chignolin, Trp-Cage, NPY, …
- AQTIVATE PhD project ESR6

AQTIVATE

ESR6: What's the plan?

- MD simulations are a powerful tool, but slow to explore new regions of phase space (at QM accuracy)
  → explore **AI/ML applications to accelerate & improve MD simulations**

  - machine learning collective variables

  - machine learning force fields

- Targeted applications: large biomolecules, proteins, ion channels, etc.

AQTIVATE

ESR6: Learning from ...

MD simulation data

Figure adapted from Cui and Bastien, Int J Biol Sci (2012)

Ab initio QM calculations

$$\hat{H}\Psi = E\Psi$$

Generative models (e.g. AlphaFold)
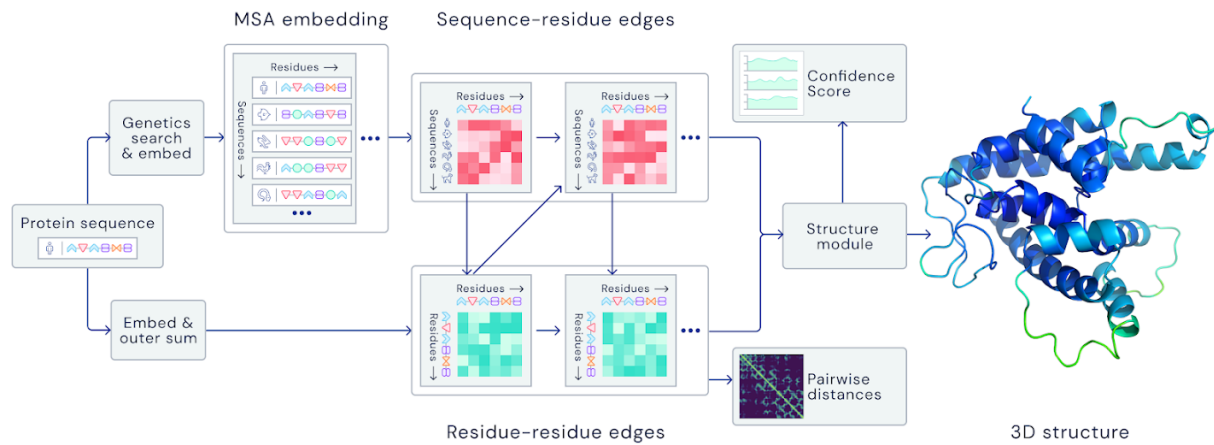
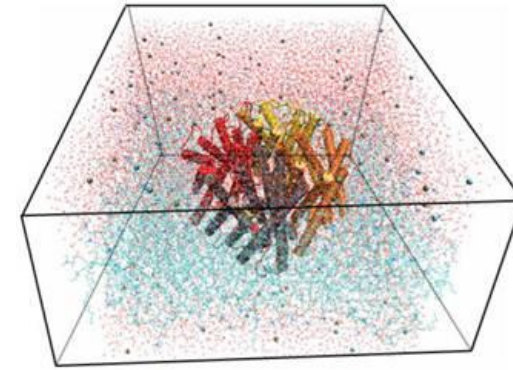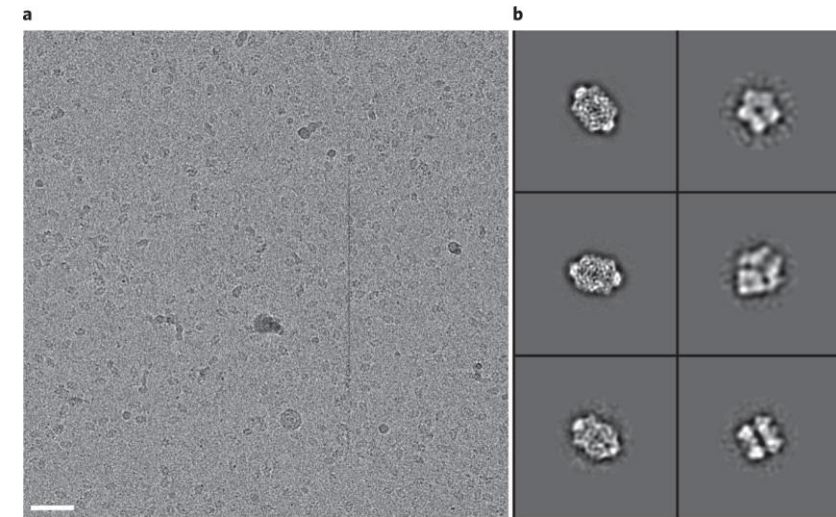*Figure adapted from Jumper et al., Nature (2021)*

Cryo-EM data

Figure adapted from Kutti R. Vinothkumar, Nature Methods (2021)

AQTIVATE

Thank you!