DeepDriveWE: An agentic approach to enhanced sampling molecular dynamics

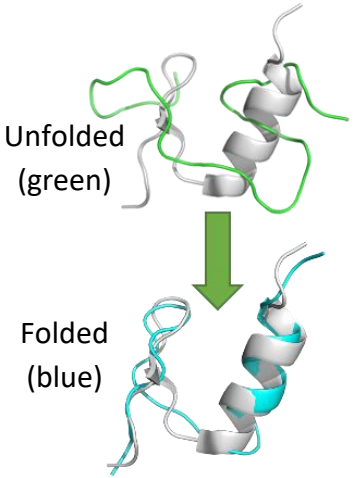
Alexander Brace[1,2*], Jeremy M.G. Leung[3], Nicolas C. Frazee[3], Rajiv Venkatesh[1,2], Heng Ma[1], Matt Sinclair[1], Archit Vasan[1], Lillian T. Chong[3+], Ian Foster[1,2+], Arvind Ramanathan[1+]

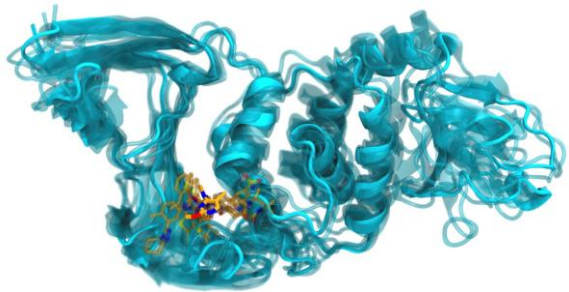
[1]Argonne National Laboratory, [2]University of Chicago, [3]University of Pittsburgh;

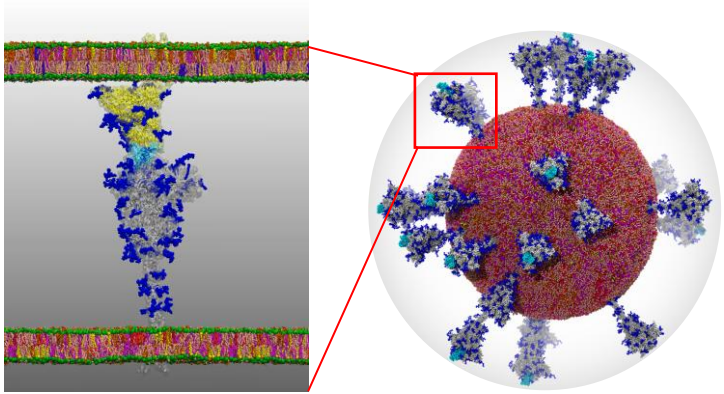
+Senior authors *First authors

U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Science motivation: Studying complex biological systems and processes is important







Protein folding provides insights into how their 3D structure enables function
e.g. FSD-EY ($\beta\beta\alpha$)

Protein ligand complexes evaluate drug binding to discover novel inhibitors e.g. Papain-like protease (PLPro)

Many-Atom Multiscale Systems uncover large-scale dynamics necessary to understand systems such as SARS-CoV-2

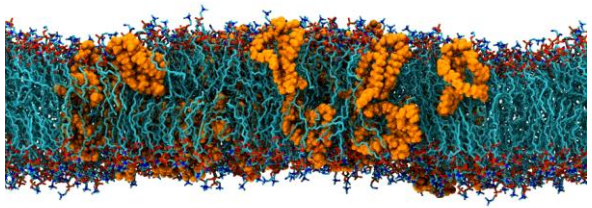
Casalino, L. et al, **AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics,** https://doi.org/10.1177/10943420211006452 (SC'20 Gordon Bell Award for HPC in COVID-19 Research)

What is molecular dynamics and why do we do it?

- Molecular dynamics (MD) can act as a ***computational microscope***

- Integrates Newton's second law of motion to advance each atom in the system forward in time along a potential energy gradient

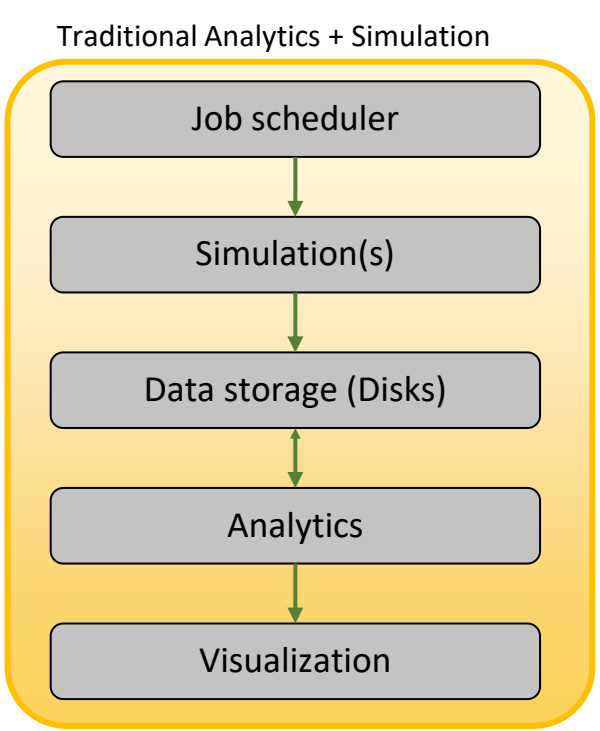
$$\vec{F}_i = ma = m_i \frac{d^2\vec{r}_i}{dt^2} = -\vec{\nabla}U(\vec{R})$$

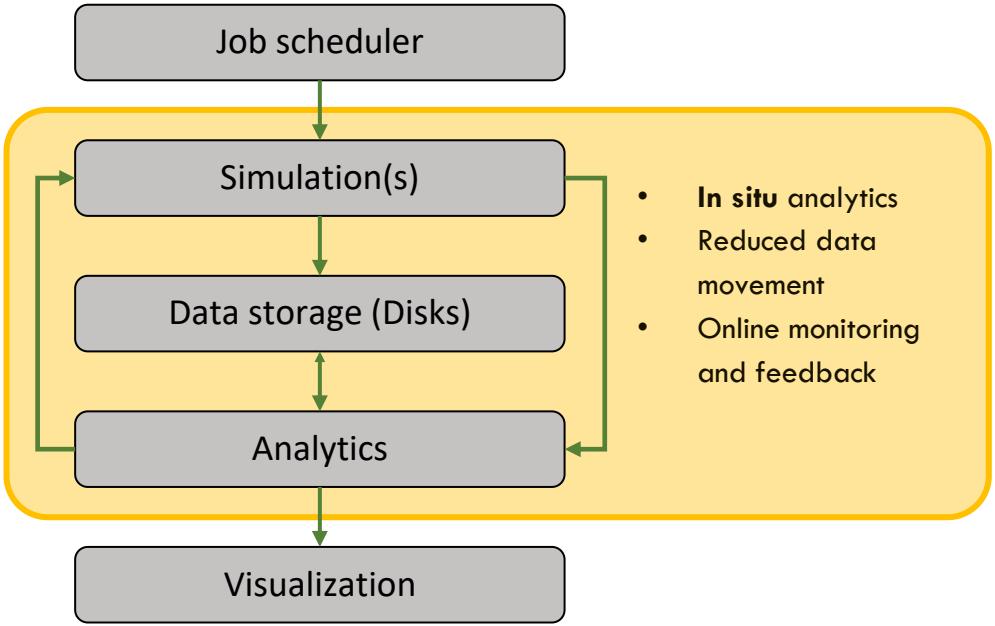


Membrane simulation

R. O. Dror et al. **Biomolecular simulation:**
A computational microscope for molecular biology

Multiscale simulations impose new requirements on emerging hardware/software



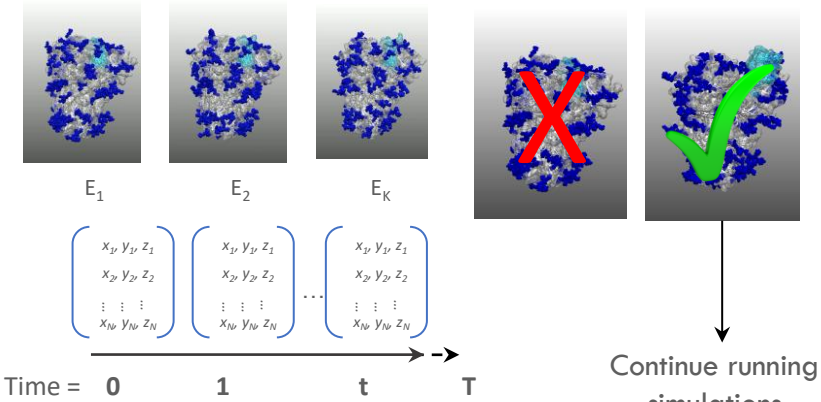


- Large simulations generate > **O(100 TB)** of data
- Humanly impossible to peek into "biologically" interesting events!
- Traditional method is unsustainable at exascale

AI-enabled MD simulations with DeepDriveWE

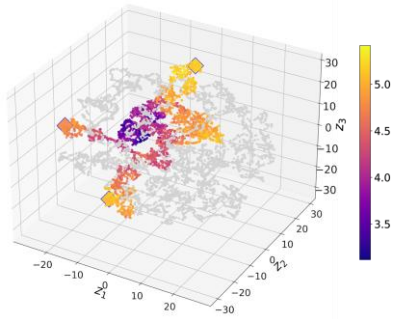
Coordinates, contact maps, other features

Ensemble MD simulations

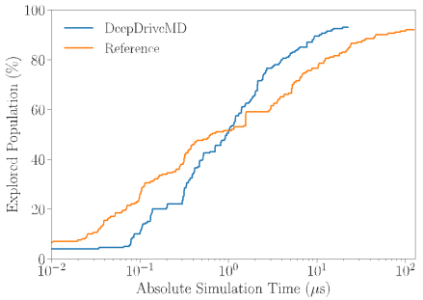


Continue running simulations

Deep Learning/ Artificial Intelligence



Build physically interpretable embeddings

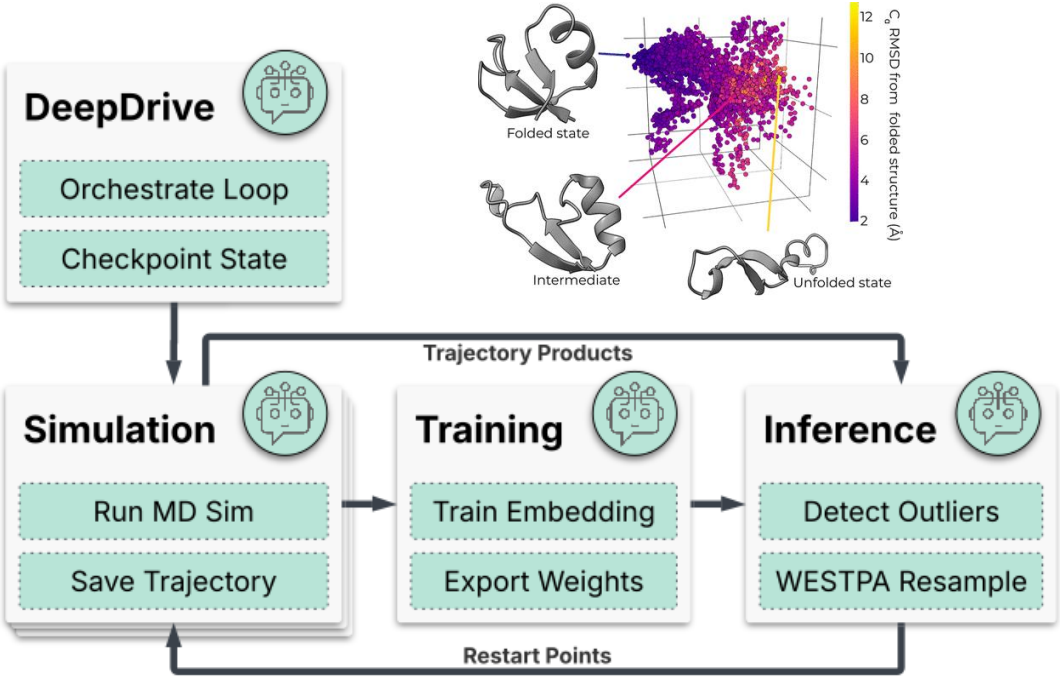


Track over sampled states and resample with WESTPA

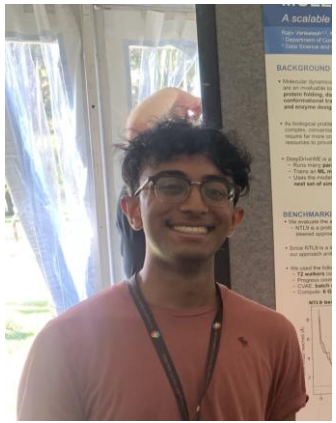
"Interesting conformations", population sampled

Learning Everywhere
Jha & Fox. In "Visionary Track", 15th International Conference eScience (2019), San Diego, California

DeepDriveWE with Academy for stateful execution



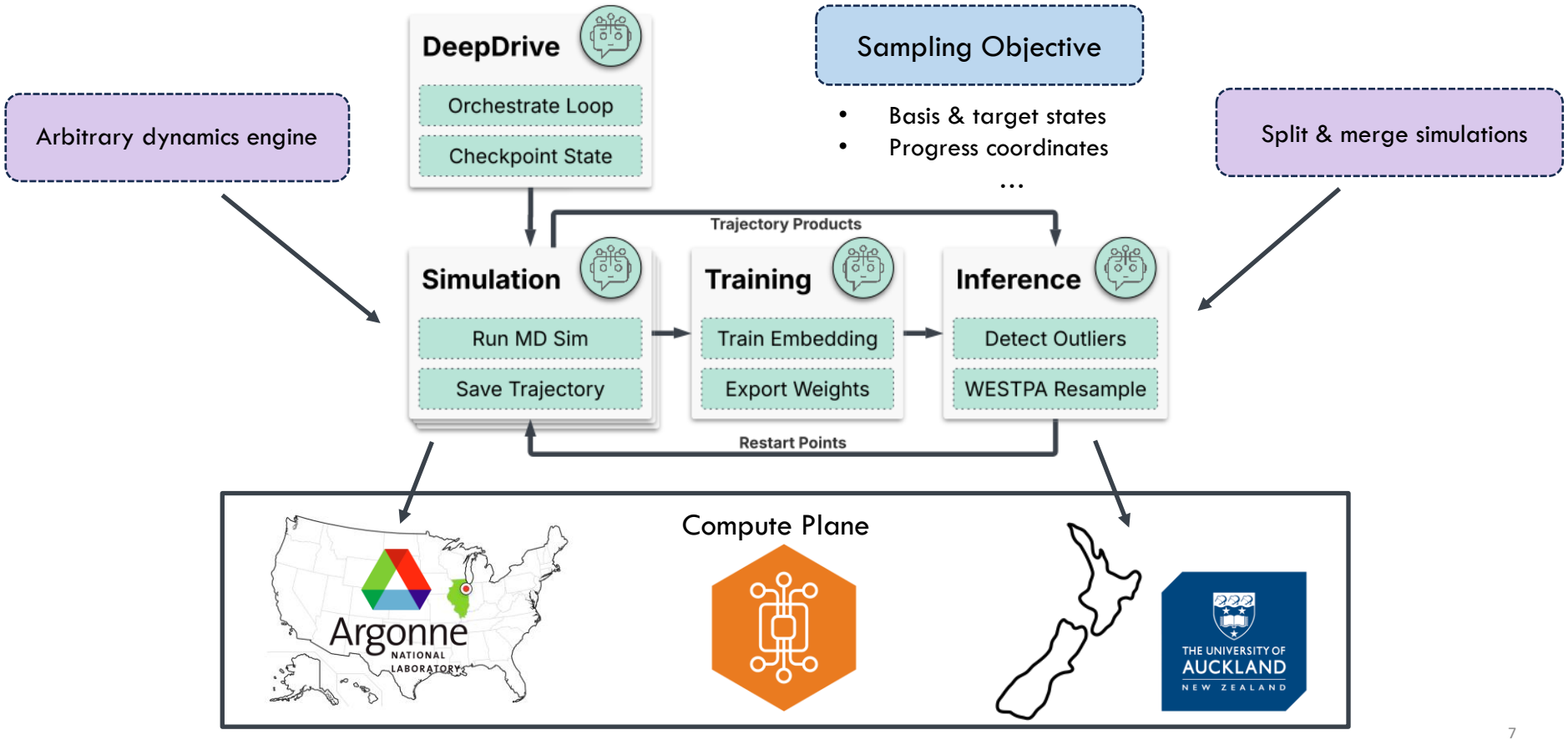
- **Architecture:** Each module is an independent agent in Academy
- **Modularity:** Decoupled design of components allows for high degrees of parallelism (e.g., 100s of parallel simulation tasks) and plug and play algorithms



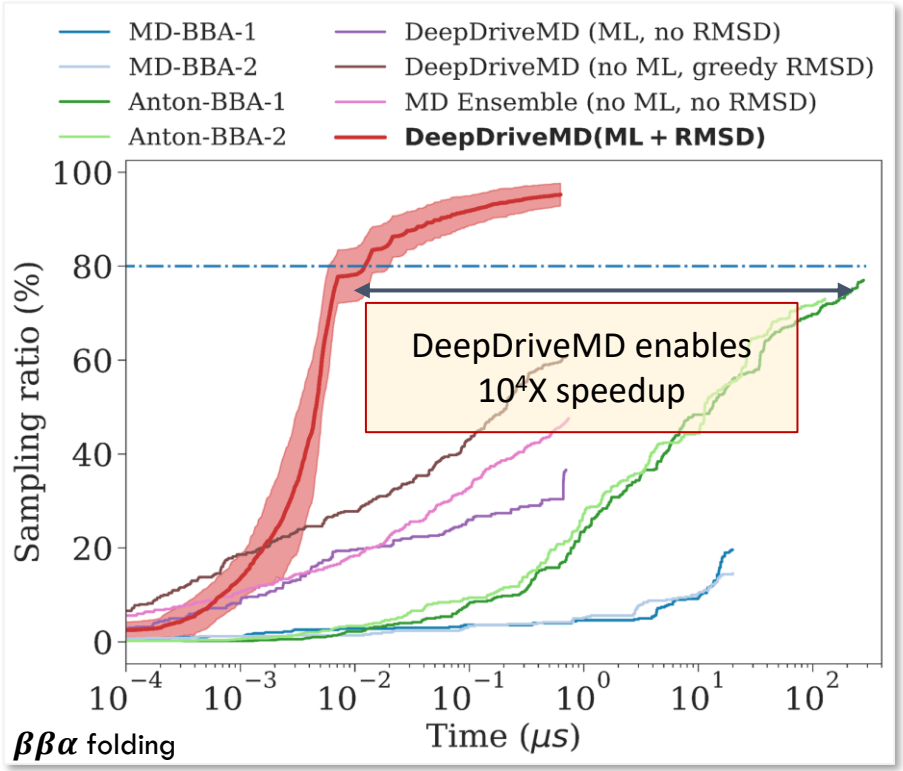
Rajiv Venkatesh

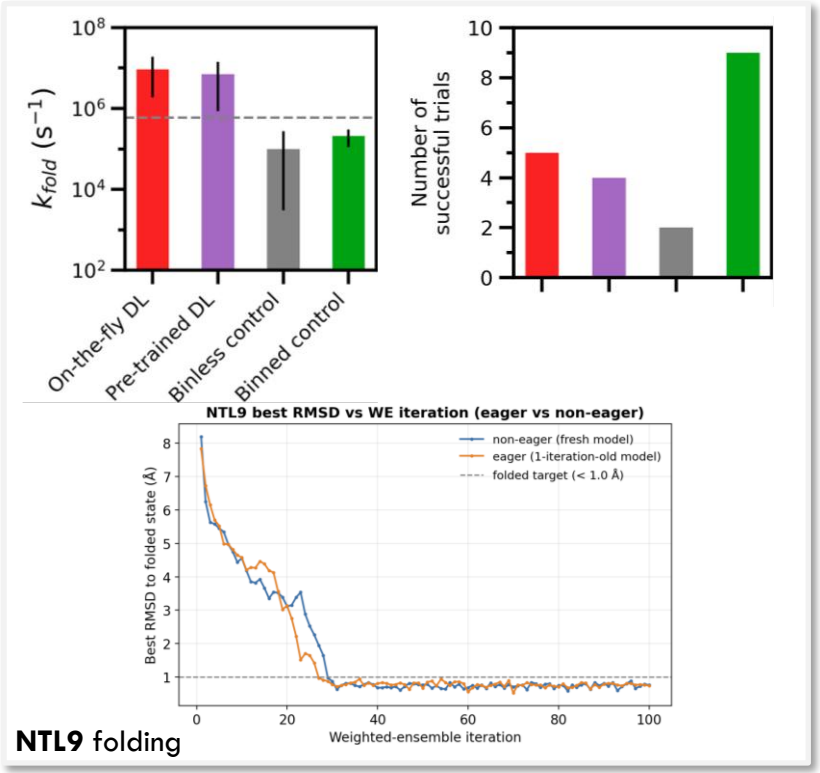
Kamatar, A. et al., **Patterns and Experiences from Deploying Agents in Scientific Applications.** https://papers.ssrn.com/sol3/papers.cfm?abstract_id=7344668

Globus compute executors enable global orchestration



DeepDriveWE accelerates protein folding simulations while preserving rate constant estimates



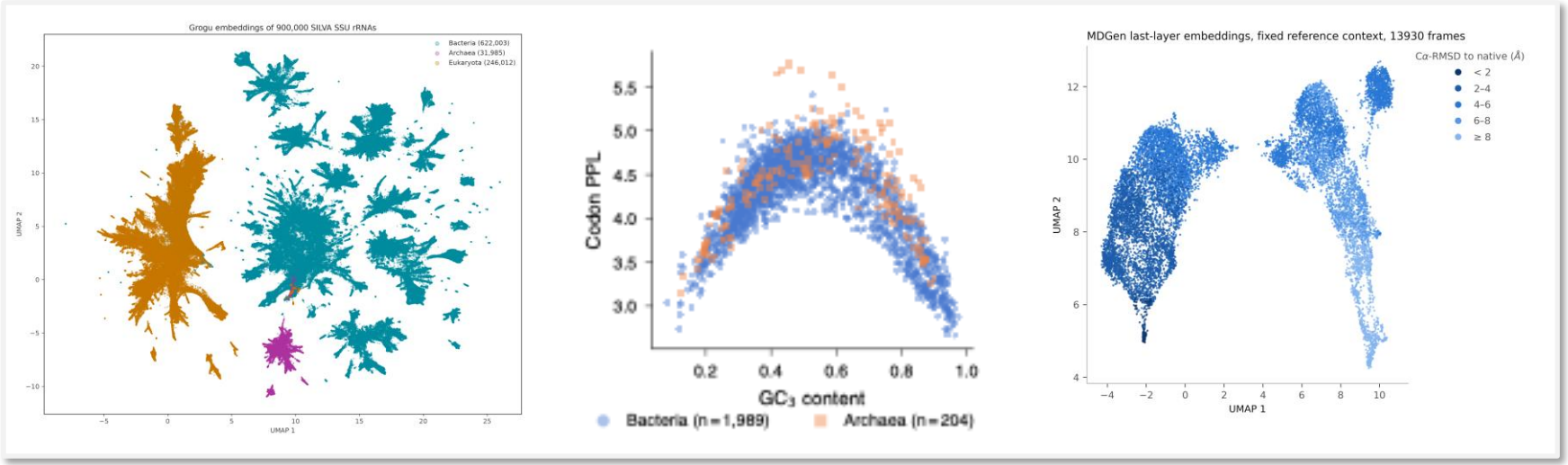


Brace, A. et al., **Achieving 100X faster simulations of complex biological phenomena by coupling ML to HPC ensembles.**

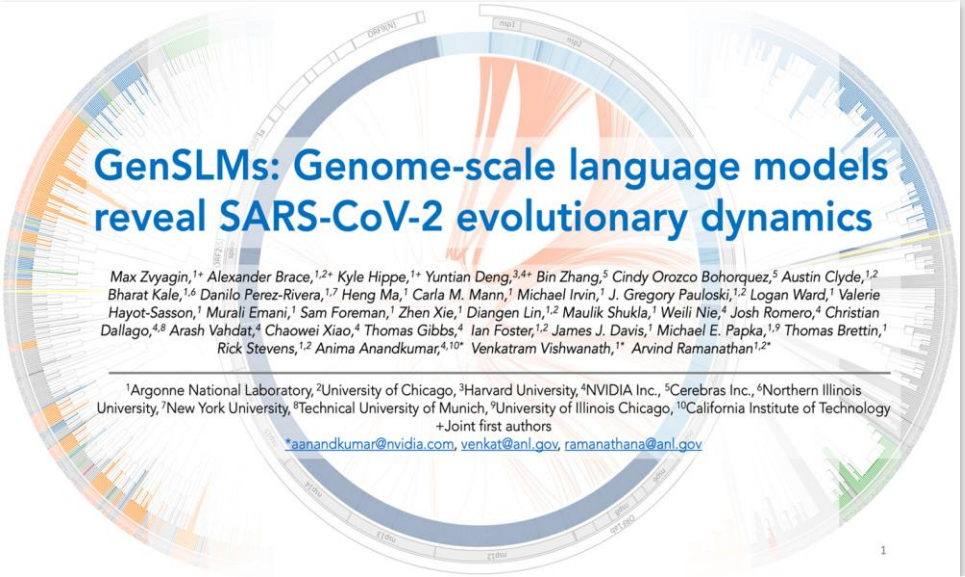
Leung, J. et al. **Unsupervised Learning of Progress Coordinates during Weighted Ensemble Simulations: Application to NTL9 Protein Folding.**

Building the bridge between worlds with biological foundation models

Genome ↔ mRNA ↔ Protein sequence ↔ Protein structure ↔ Dynamics ↔ Function



Thank you!



Questions?

Contact:
abrace@uchicago.edu

Reference:
https://github.com/ramanathanlab/deepdrivewe-academy

Acknowledgements:
- Argonne Leadership Computing Facility (**ALCF**)
- Coalition for Epidemic Preparedness Innovations (**CEPI**)
- Intelligent Design Assistant for Enzyme Discovery and Biosynthetic Pathway Optimization (**IDeA**)
- Special thanks:

  Lillian Chong (U. Pittsburgh), **Azton Wells, Kyle Chard, Logan Ward, Ryan Chard, Yadu Babuji, Greg Pauloski, Alok Kamatar!**