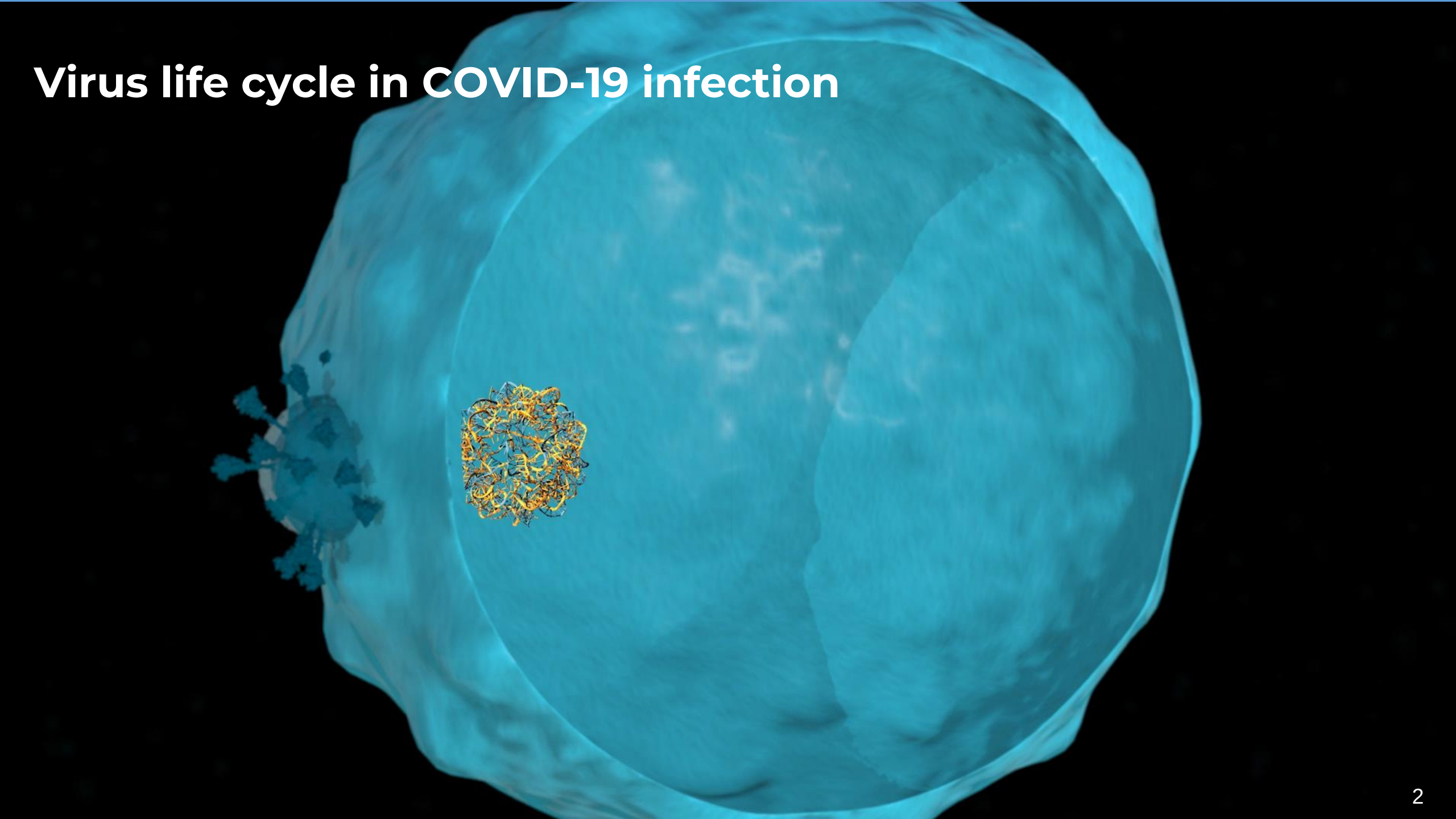
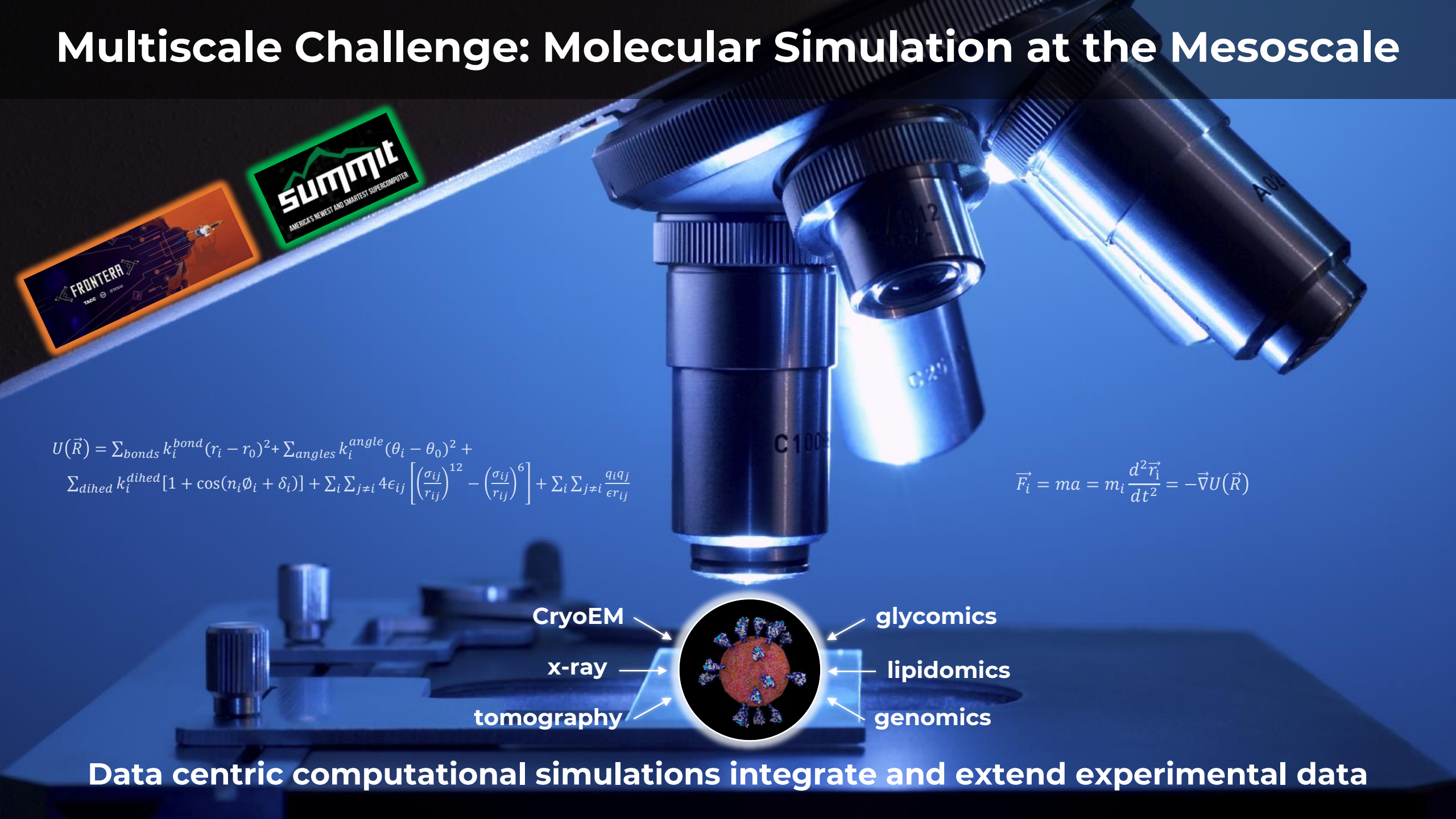
AI-enabled multi-resolution simulations to uncover mechanisms of SARS-CoV-2 virus

Anda Trifan

Virus life cycle in COVID-19 infection



Multiscale Challenge: Molecular Simulation at the Mesoscale



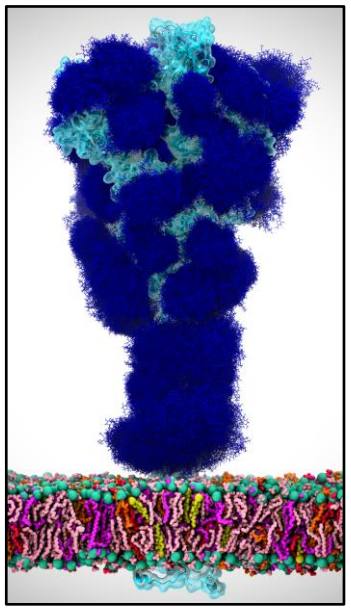
$$U(\vec{R}) = \sum_{bonds} k_i^{bond}(r_i - r_0)^2 + \sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2 + \sum_{dihed} k_i^{dihed}[1 + \cos(n_i \emptyset_i + \delta_i)] + \sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}$$

$$\vec{F_i} = ma = m_i \frac{d^2\vec{r_i}}{dt^2} = -\vec{\nabla}U(\vec{R})$$

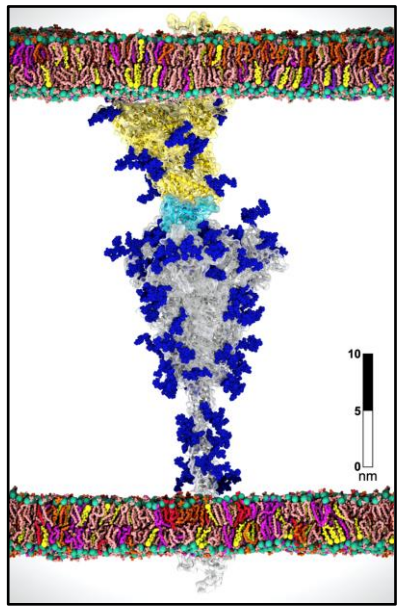
CryoEM
x-ray
tomography
glycomics
lipidomics
genomics

Data centric computational simulations integrate and extend experimental data

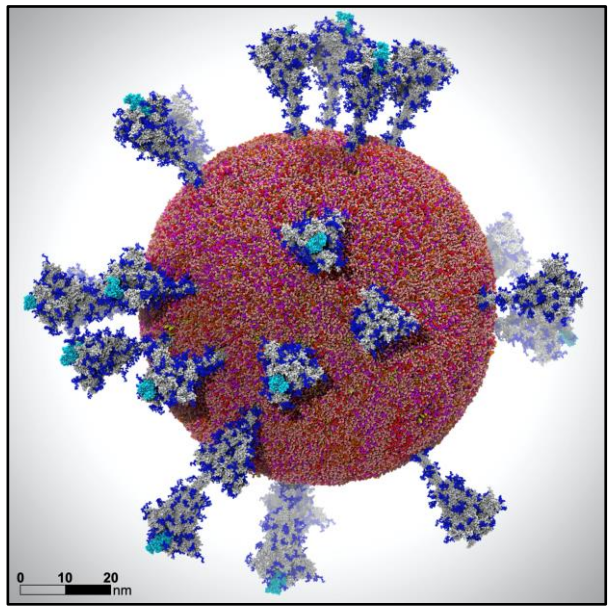
Multiscale Challenge: Molecular Simulation at the Mesoscale



Molecular & Macromolecular



Subcellular



Cellular

Spatial and Temporal Scales

Å	nm – µm	0.1 mm - mm
fs - µs	µs - ms	ms - s

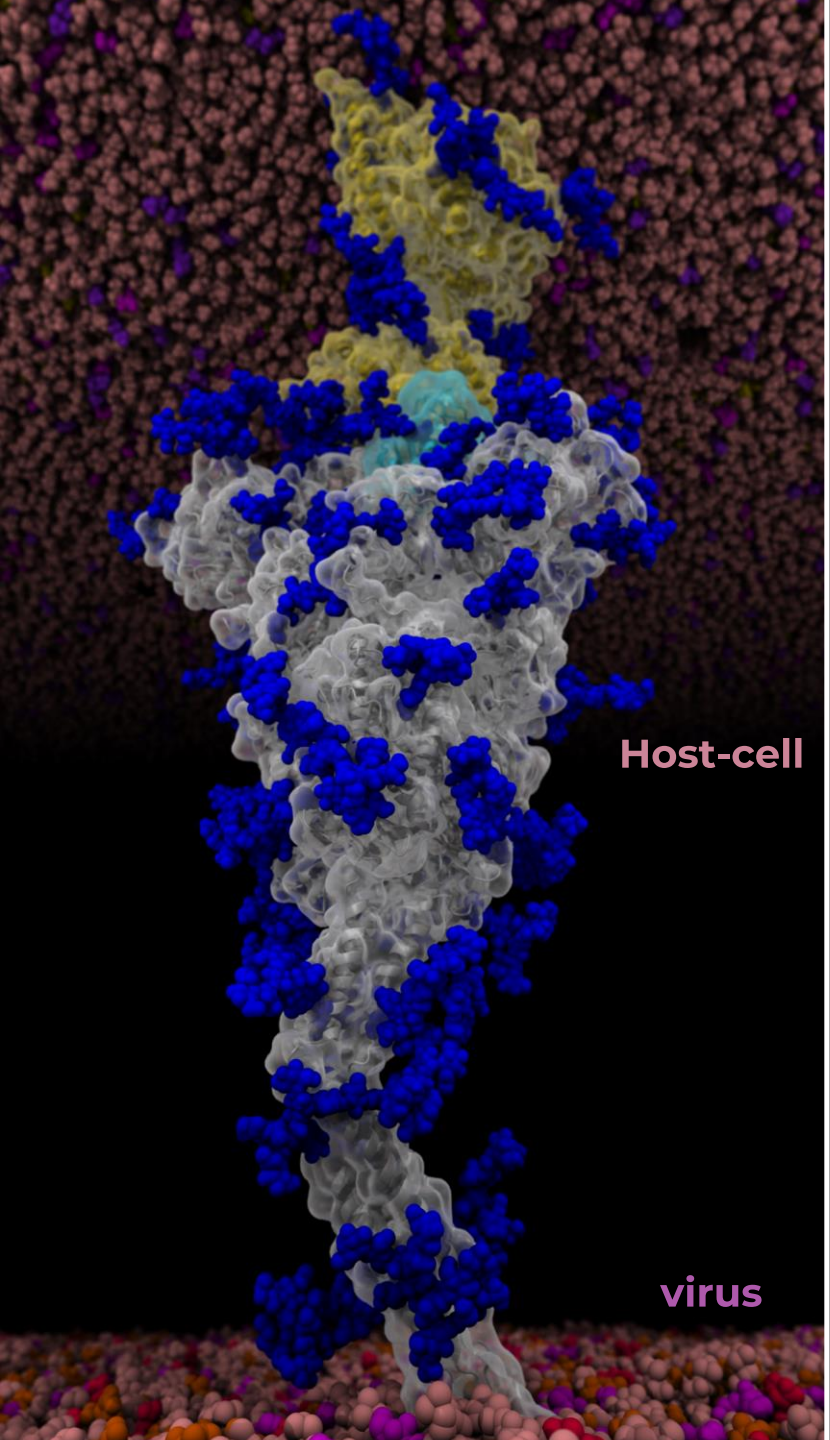
AI-driven workflows

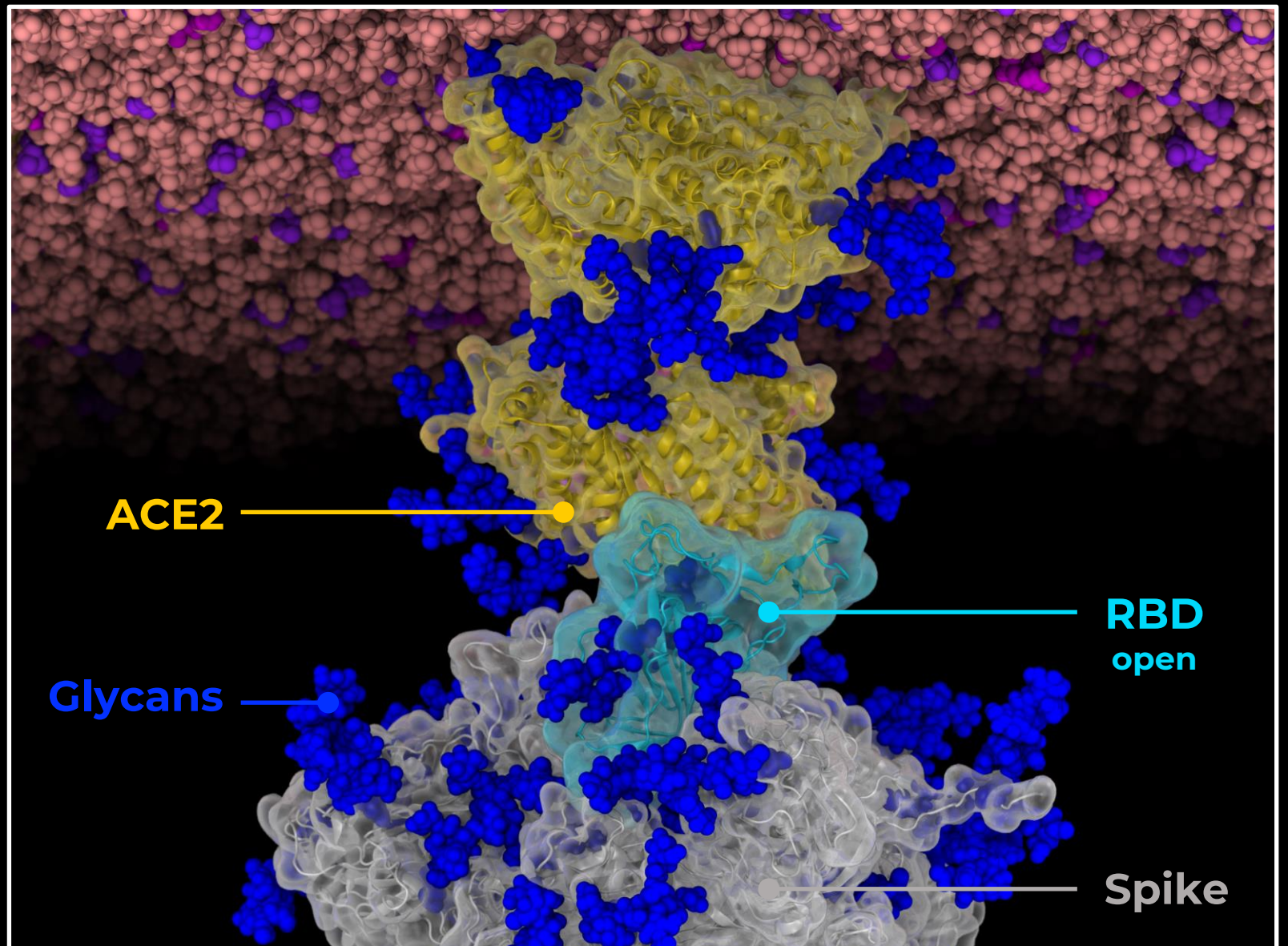
Standard simulations

Weighted ensemble simulations

SARS-CoV-2 infection route

The spike protein latches onto ACE2 to infect the host-cell

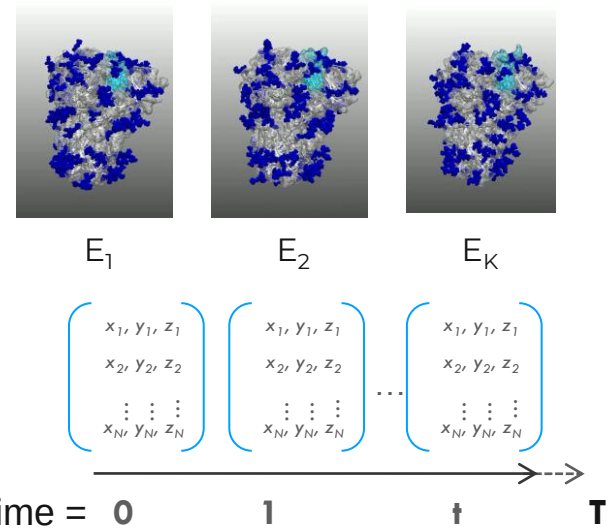


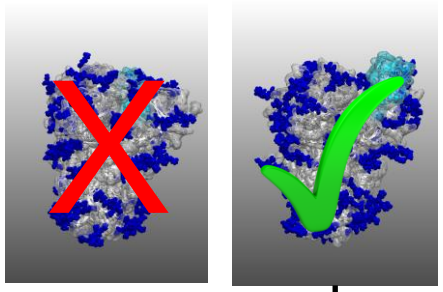


Combining AI with HPC: AI-driven MD simulations -- DeepDriveMD

Coordinates, contact maps, other features

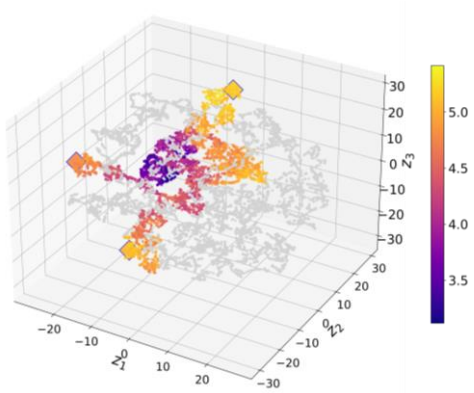
Weighted Ensemble MD simulations



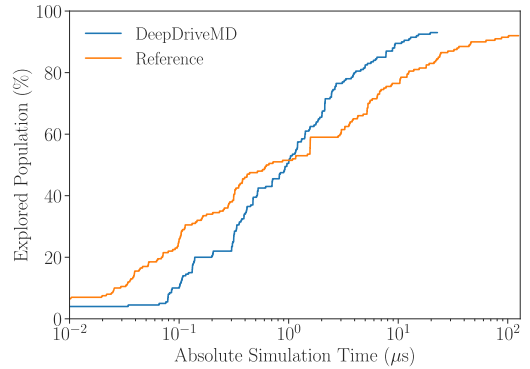


continue running simulations

Deep Learning/ Artificial Intelligence



Build physically interpretable embeddings

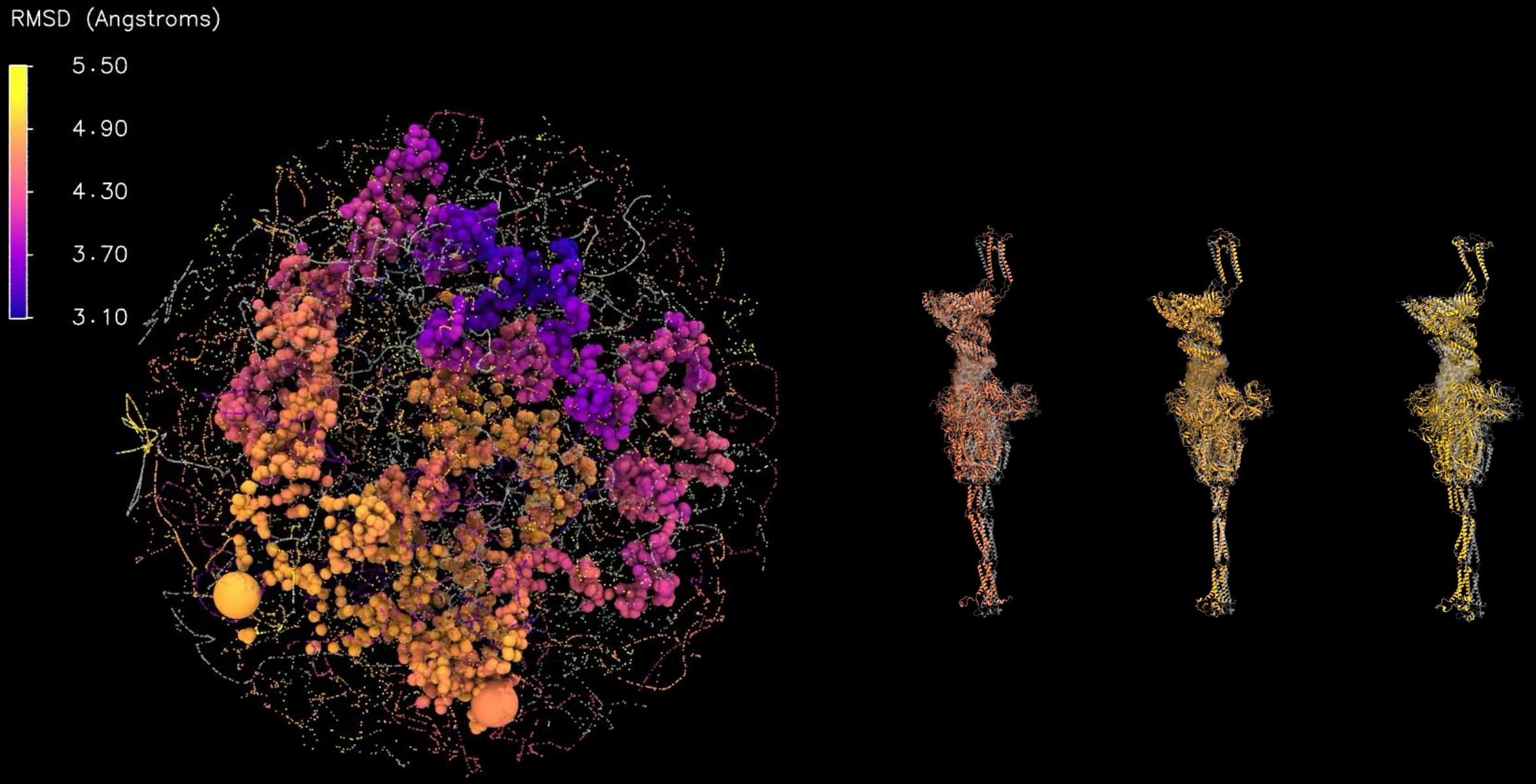


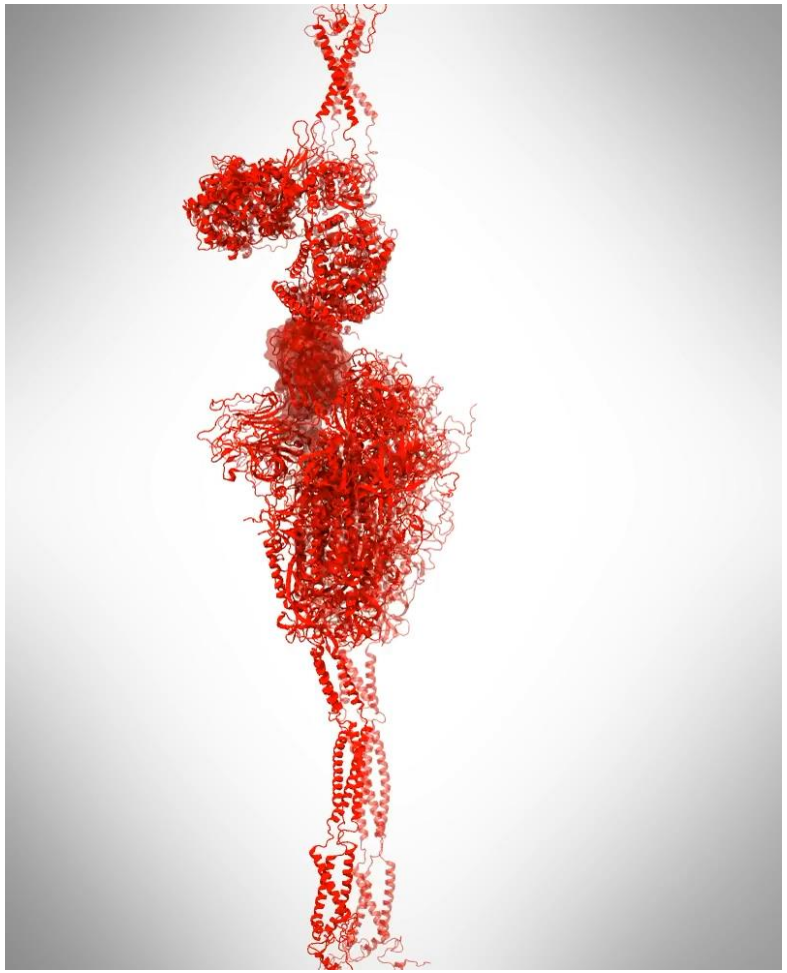
Track states that are sampled more often

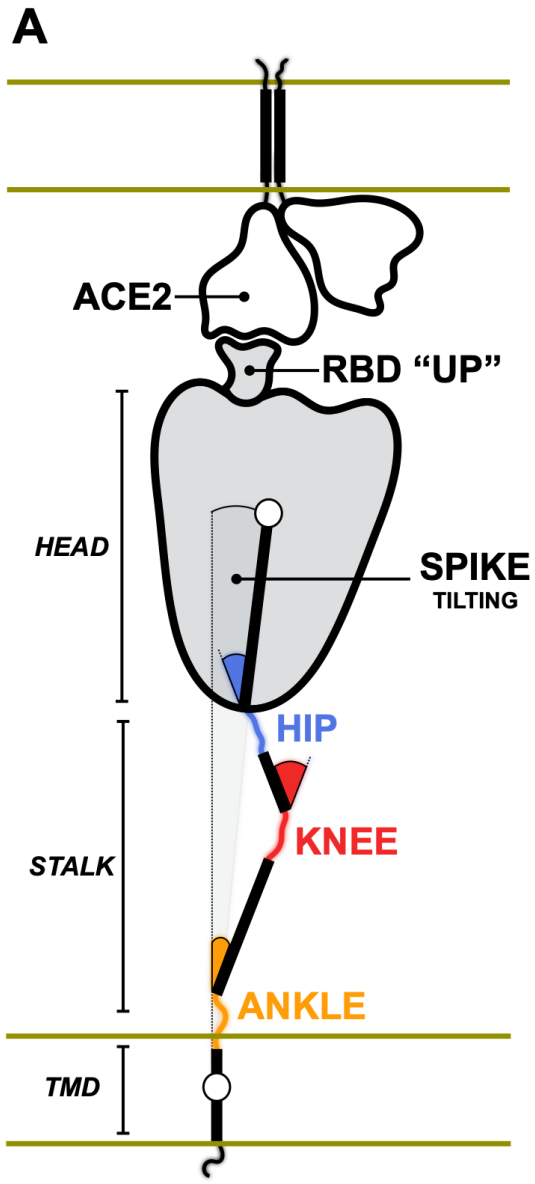
"Interesting conformations", population sampled, and other features

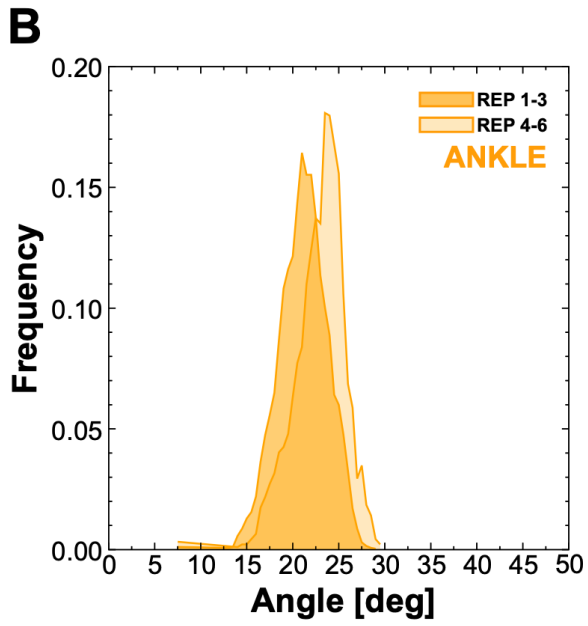
- Bhowmik, Gao, et al. BMC Bioinformatics (2018)
- Romero, Ramanathan, et al. Proc. Natl. Acad. Sci. USA (2019)
- Ma, Lee, Jha, et al. PARCO (2019)
- Lee, Ma, Jha, et al. Workshop on Deep Learning on Supercomputers, Supercomputing (2019)

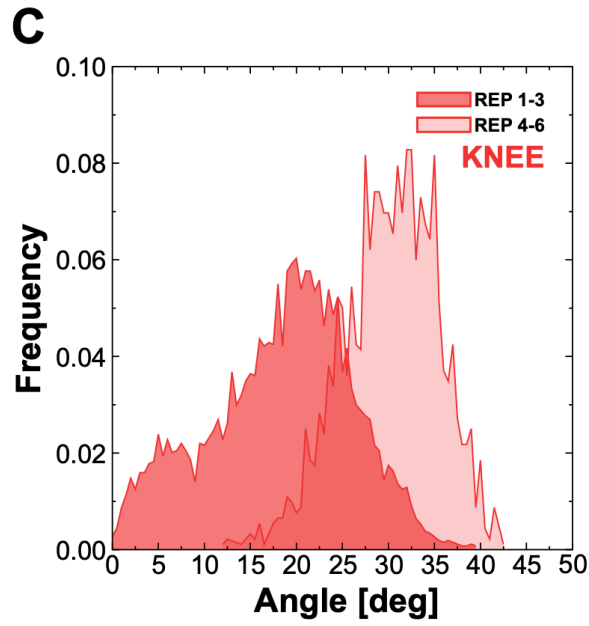
http://deepdrivemd.github.io

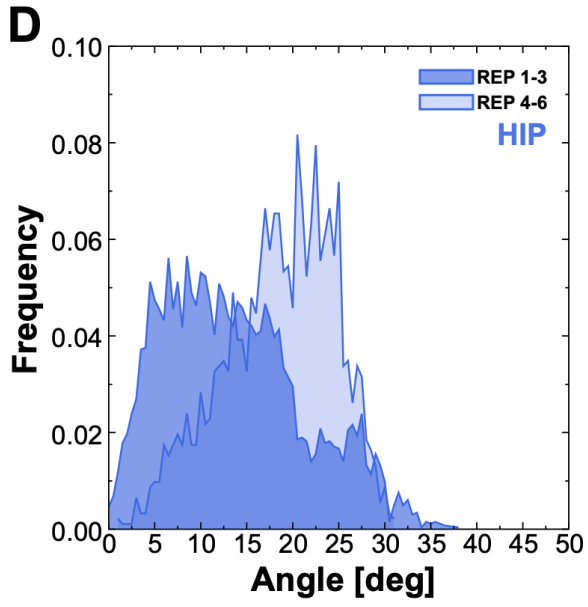

RMSD (Angstroms)
5.50
4.90
4.30
3.70
3.10

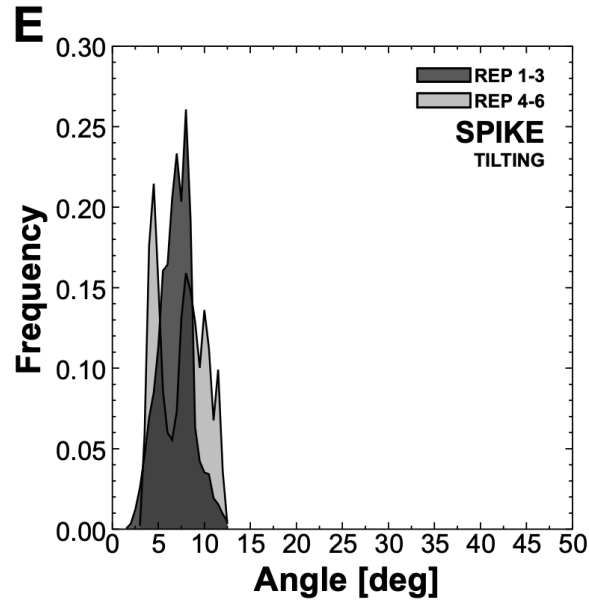



A
ACE2
RBD “UP”
HEAD
SPIKE
TILTING
HIP
KNEE
STALK
ANKLE
TMD

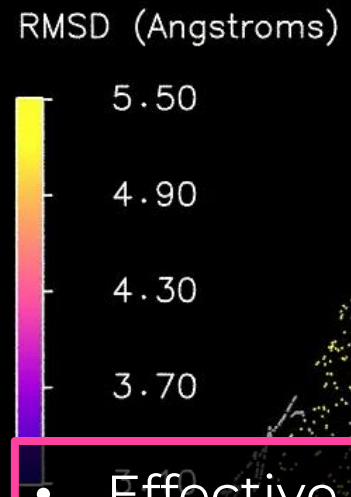

B
REP 1-3
REP 4-6
ANKLE
Frequency
Angle [deg]

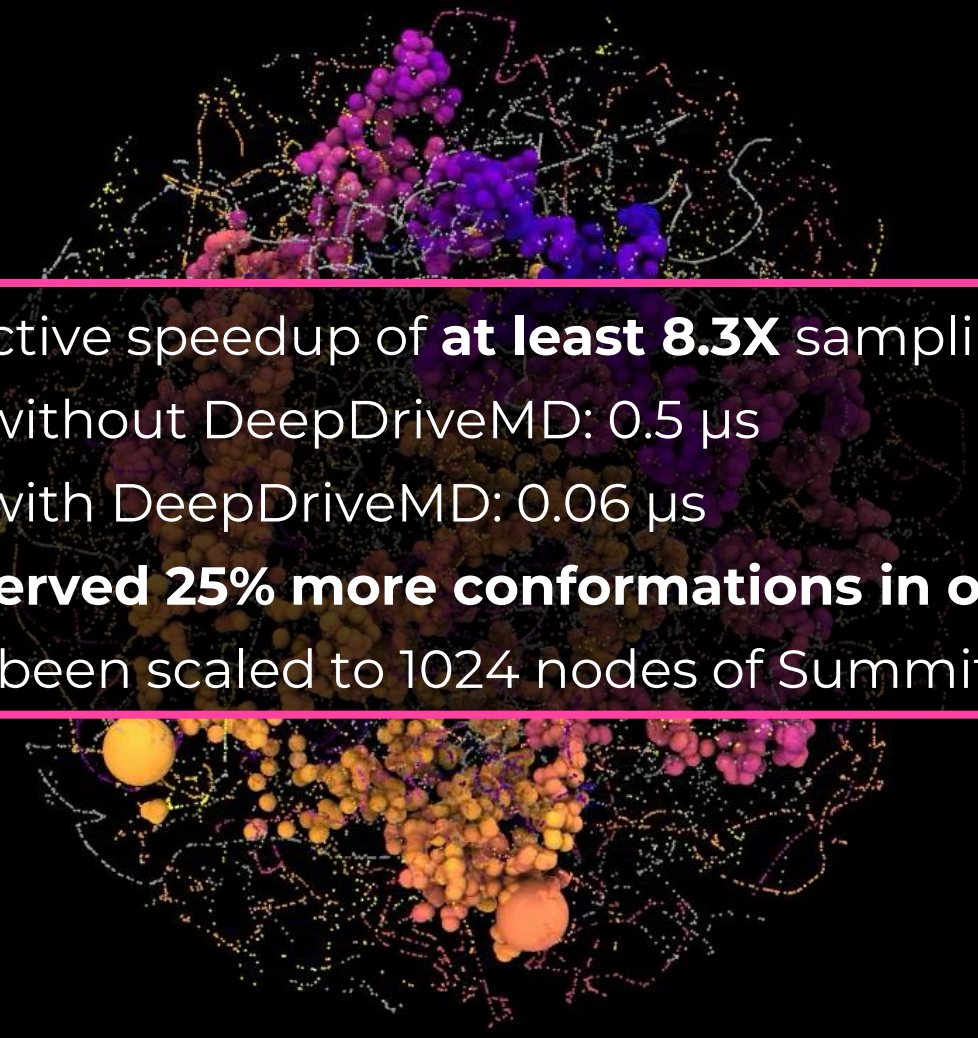

C
REP 1-3
REP 4-6
KNEE
Frequency
Angle [deg]


D
REP 1-3
REP 4-6
HIP
Frequency
Angle [deg]

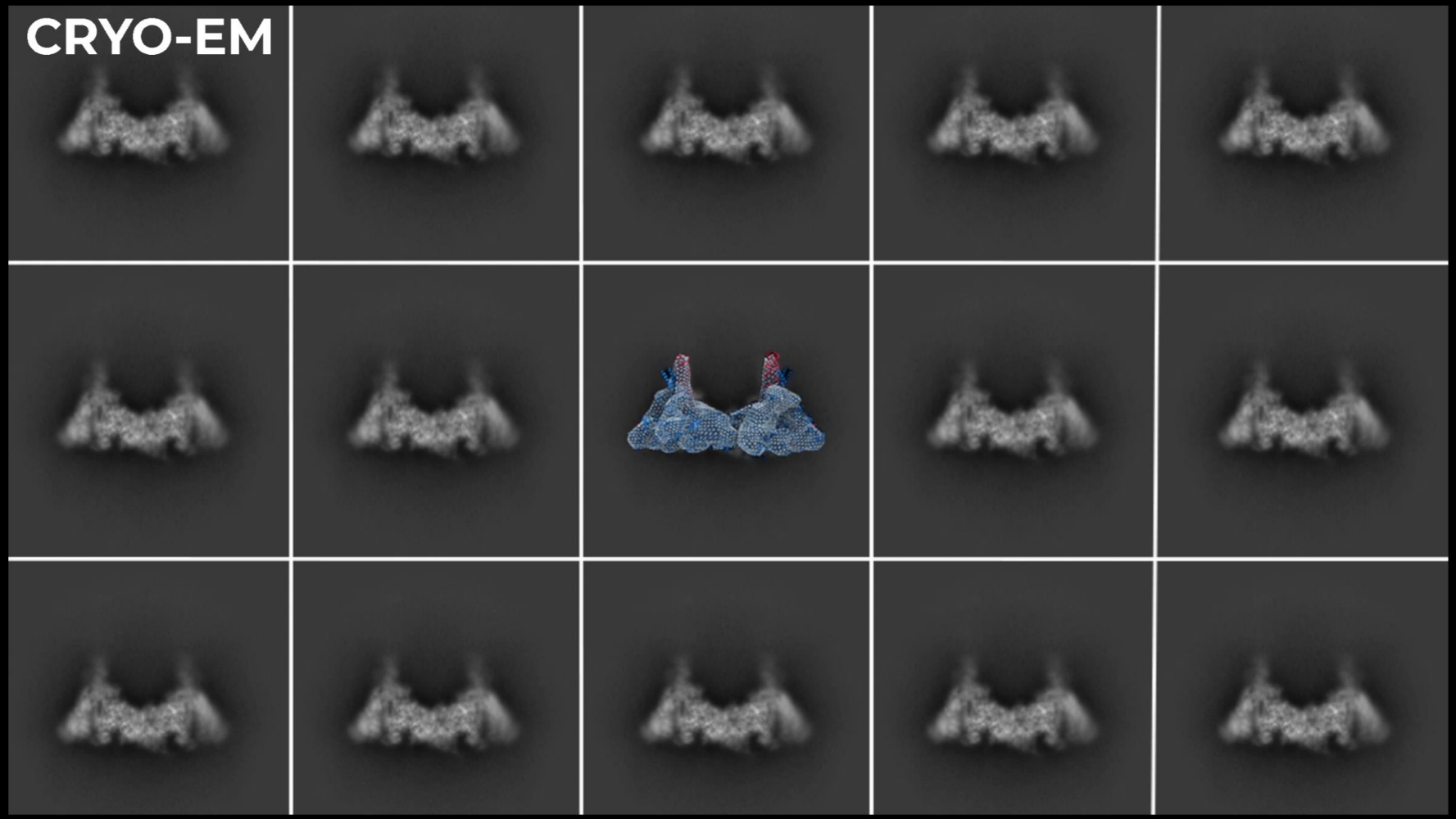

E
REP 1-3
REP 4-6
SPIKE
TILTING
Frequency
Angle [deg]

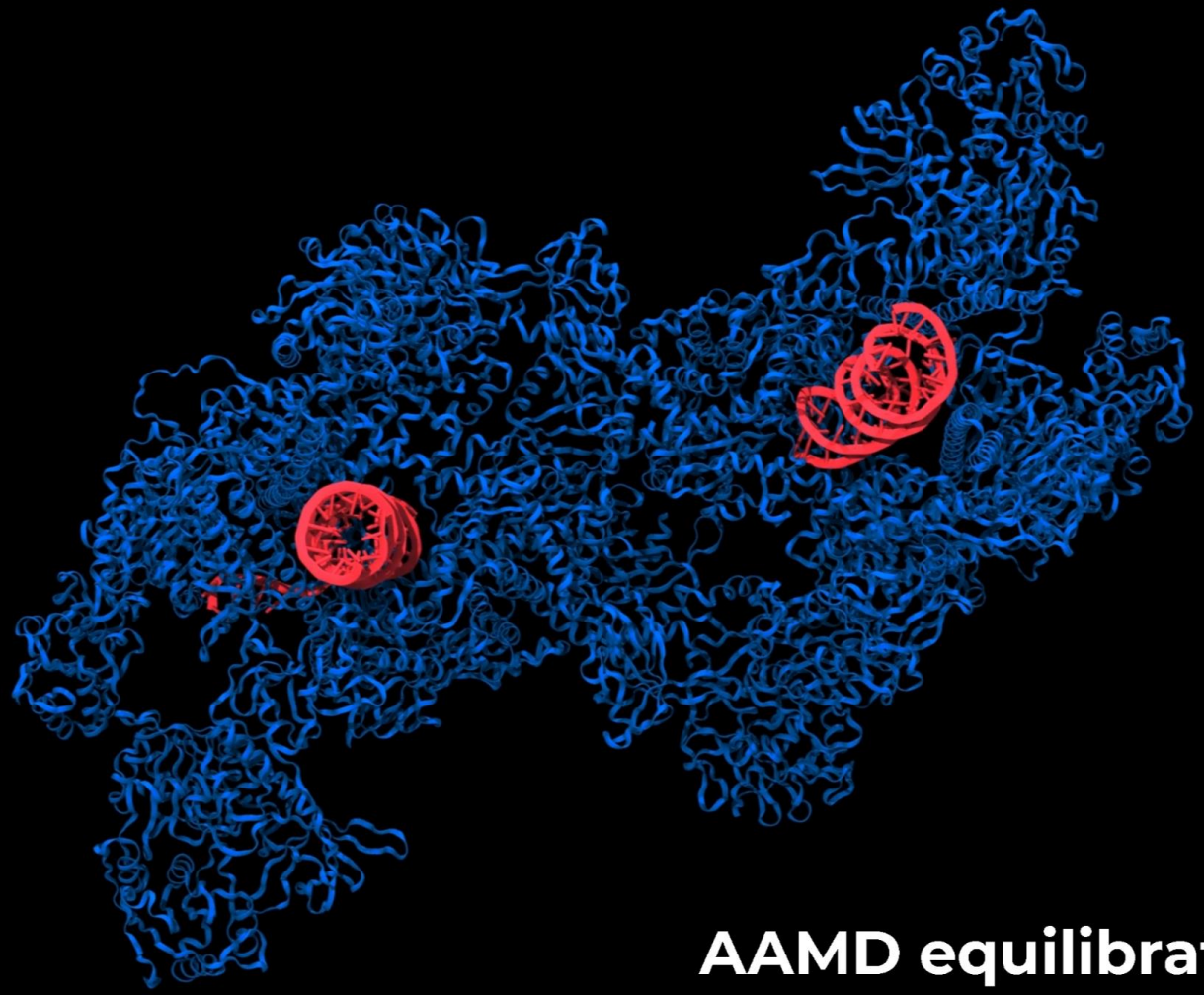
AAE map of Spike + ACE2 Receptor protein simulations





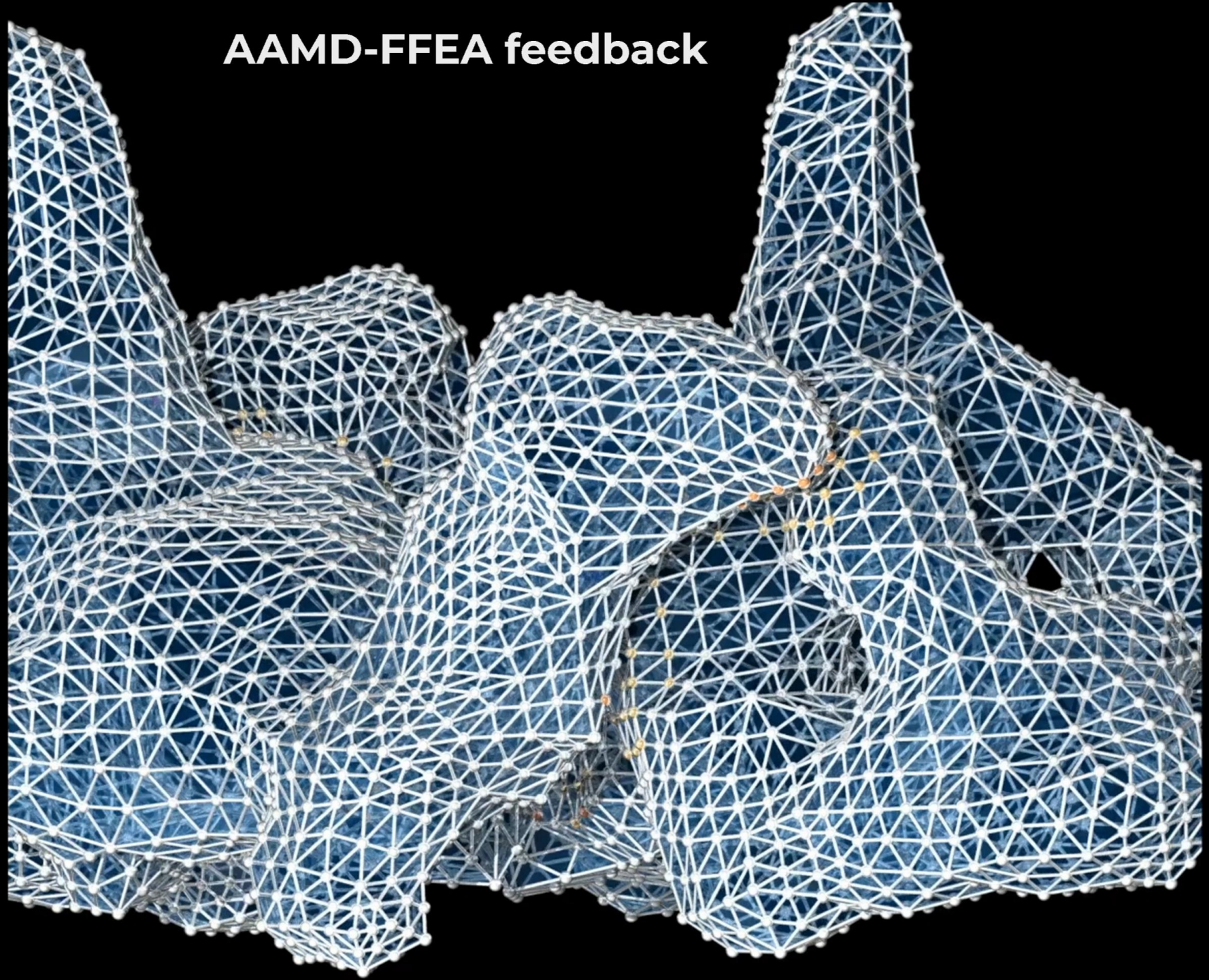
- Effective speedup of **at least 8.3X** sampling efficiency for spike stalk bending:
  - without DeepDriveMD: 0.5 μs
  - with DeepDriveMD: 0.06 μs
- **Observed 25% more conformations in only 12% of the time!**
- Has been scaled to 1024 nodes of Summit with Ensemble Toolkit Workflow


CRYO-EM



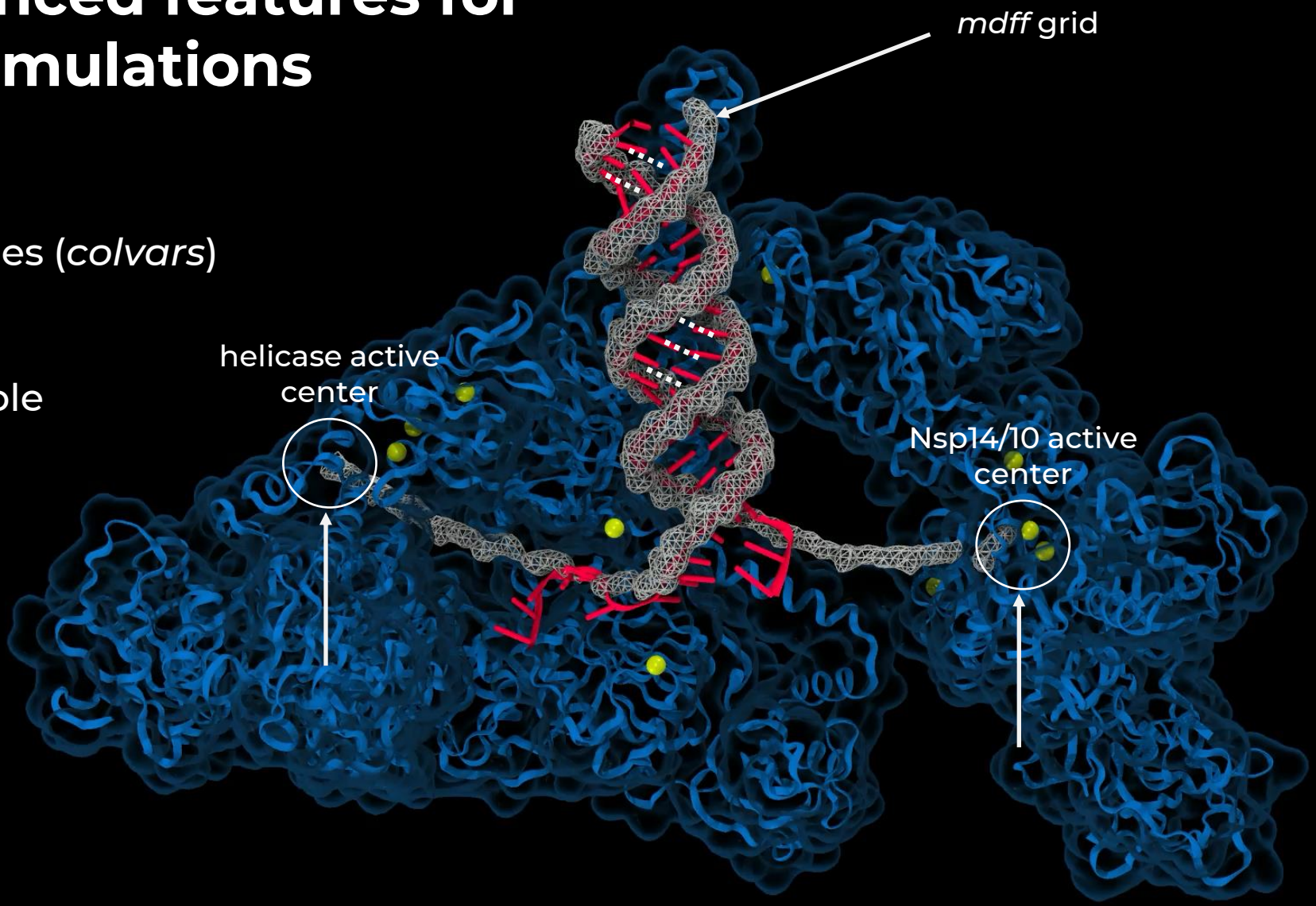
AAMD equilibration

AAMD-FFEA feedback

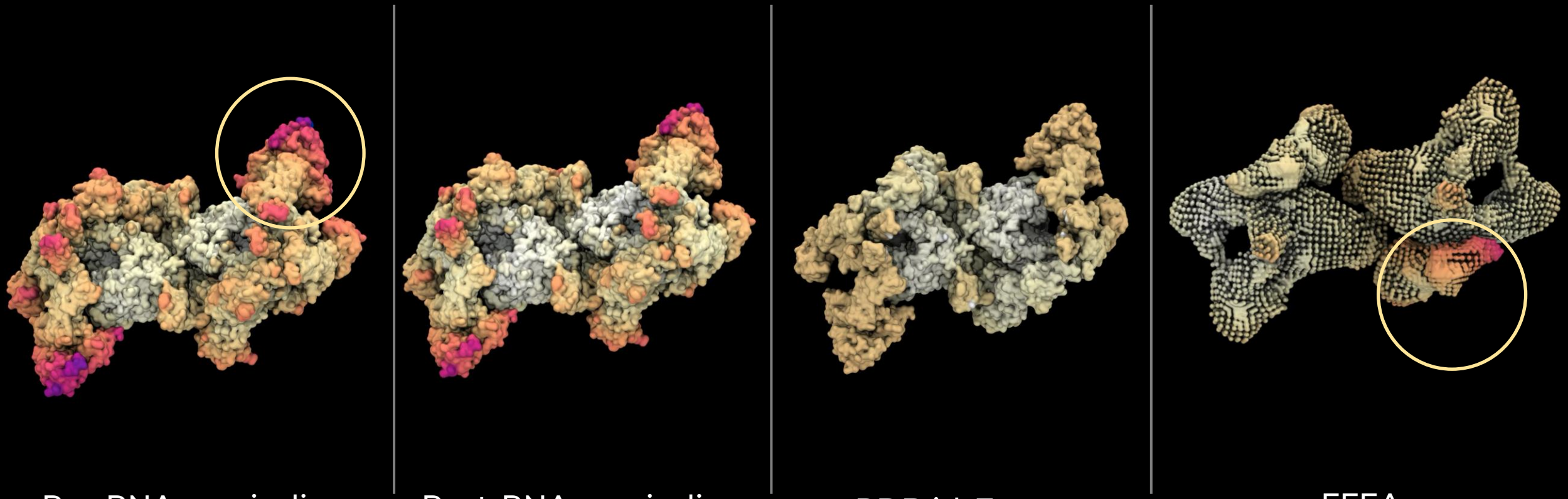


Using NAMD advanced features for non-equilibrium simulations

- distance collective variables (*colvars*)
- molecular dynamics flexible fit grid (*mdff*)
- extrabonds



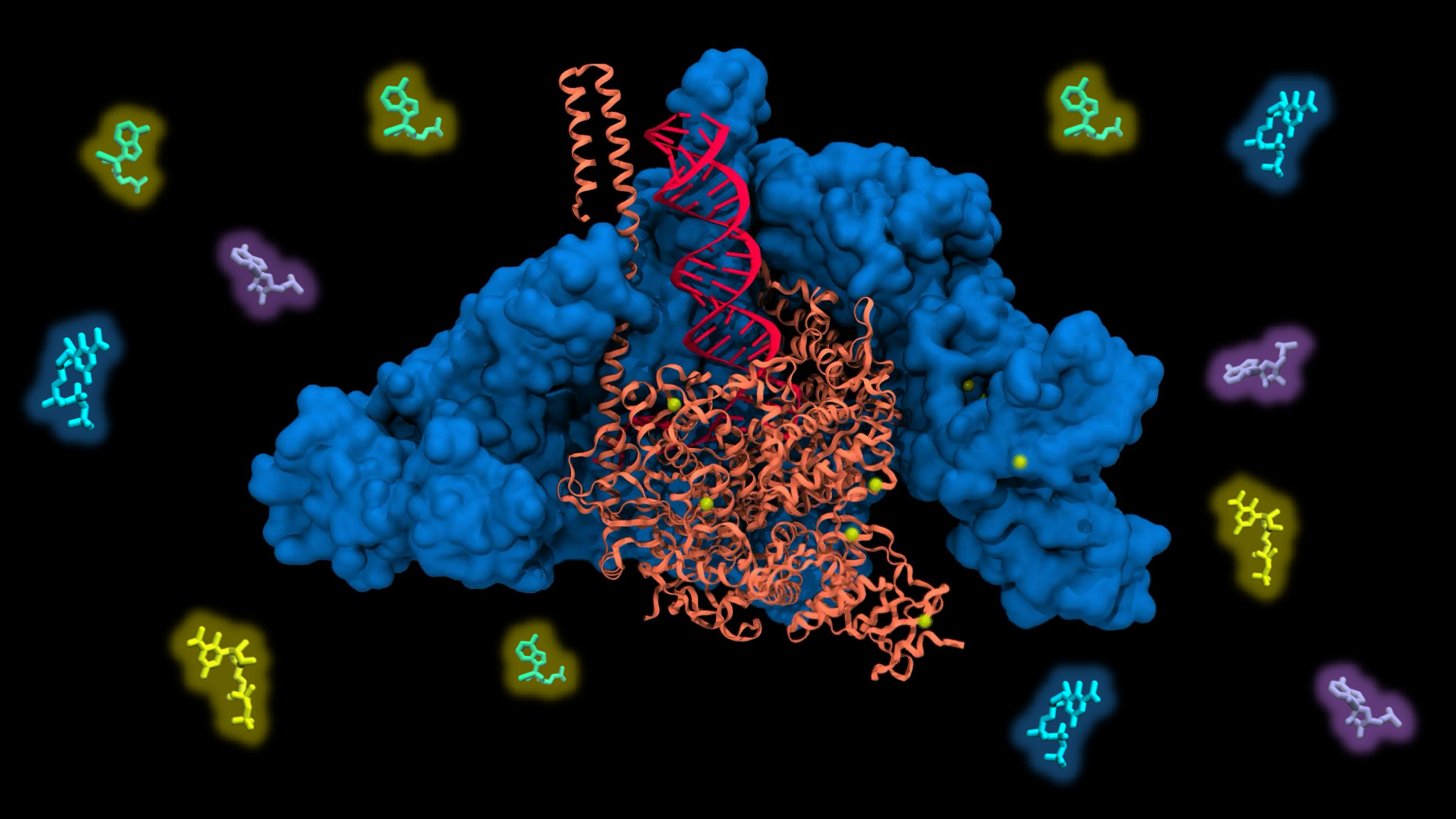
Complementary biophysical insights from FFEA and AAMD simulations

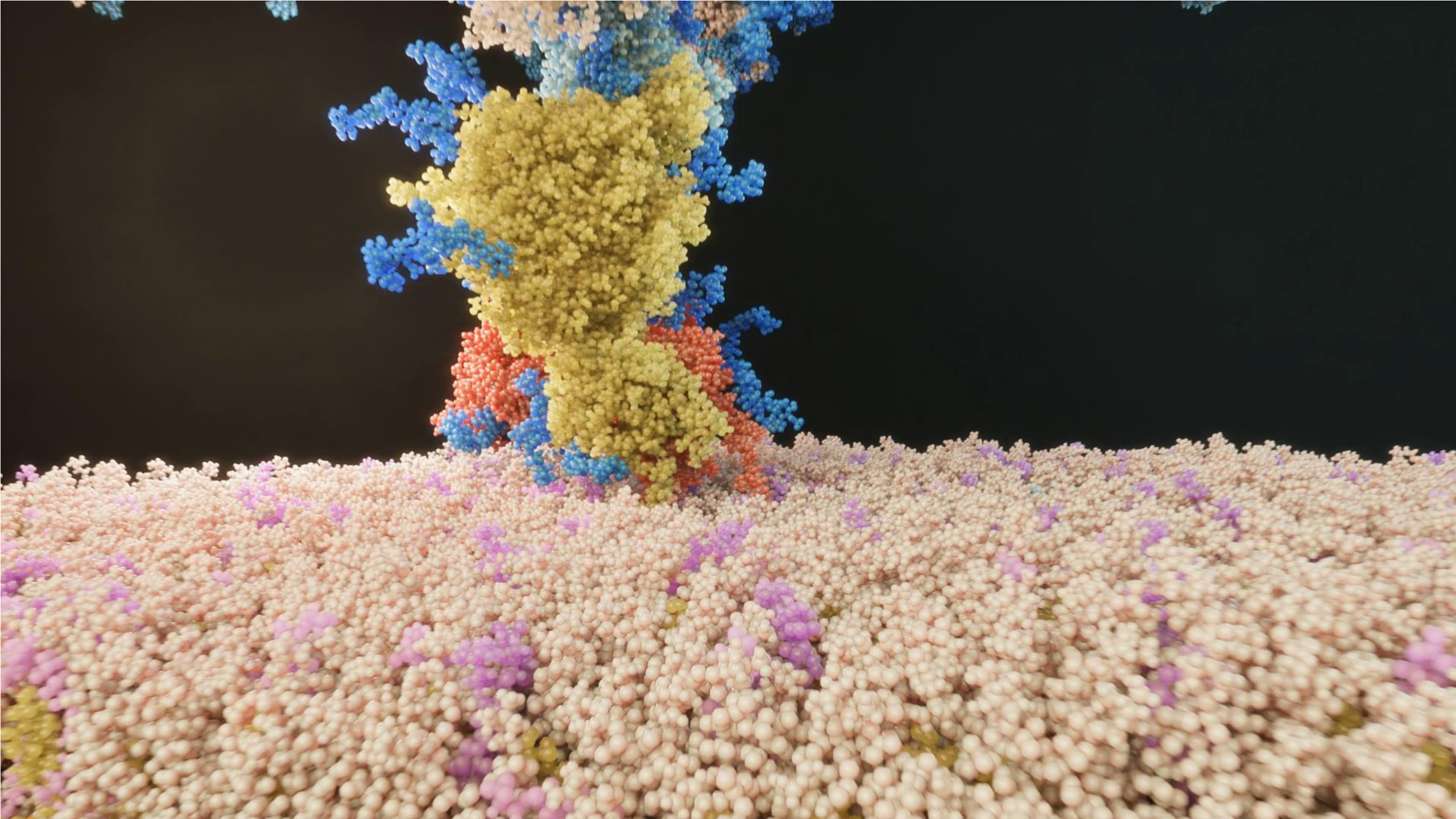


Pre-RNA unwinding | Post-RNA unwinding | PDB id: 7egq | FFEA

Key Insights:

- refine FFEA / AAMD simulation parameters (based on experimental data)
- indicate RTC dynamics (predict subunit re-arrangements)
- alternative conformations of helicases (AAMD) vs. nsp10-14 (FFEA)





Emad Tajkhorshid*
UIUC

Arvind Ramanathan*
Argonne/ University of Chicago

Rommie Amaro
UCSD

DOE CSGF