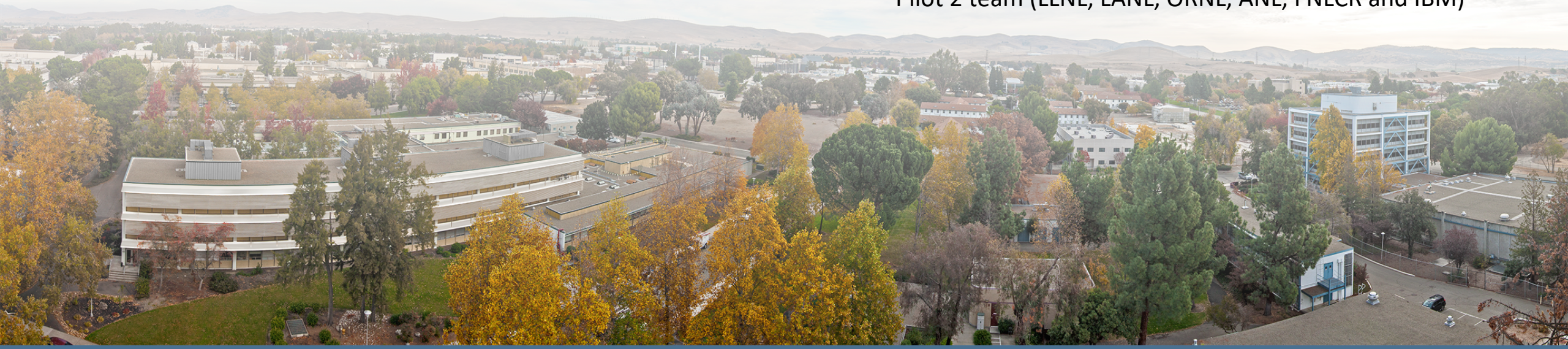
Machine Learning-driven Multiscale Modeling, a next generation simulation infrastructure showcased on a RAS-RAF Biology application

September 3rd, 2021

Helgi I. Ingólfsson,
Biochemical and Biophysical Systems, PLS, LLNL

On behalf of the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C)
Pilot 2 team (LLNL, LANL, ORNL, ANL, FNLCR and IBM)



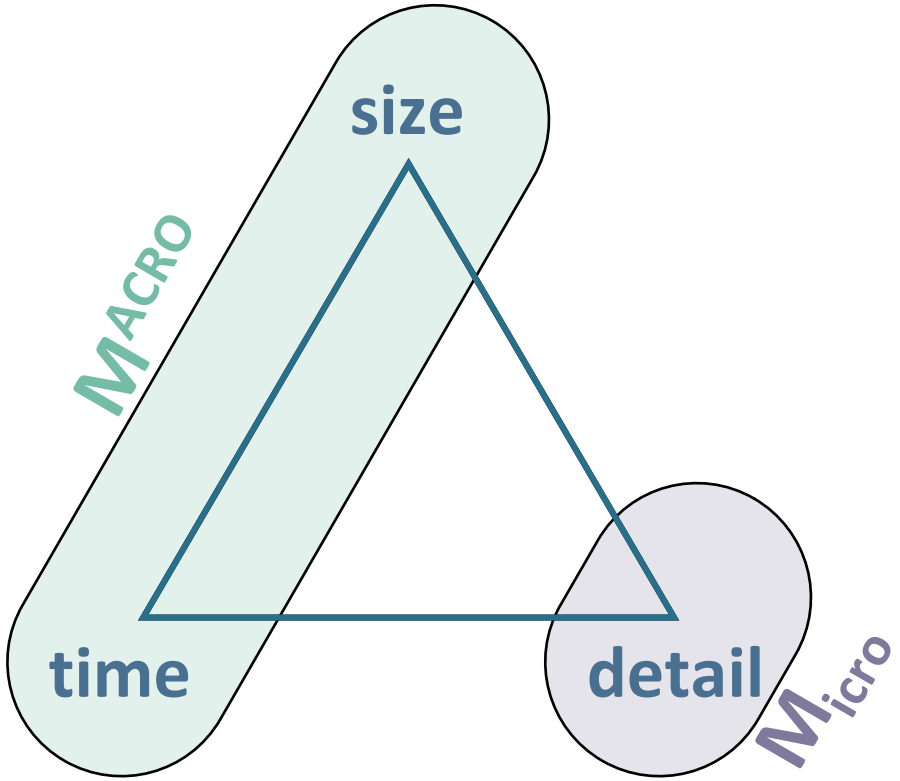
LLNL-PRES-820866

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

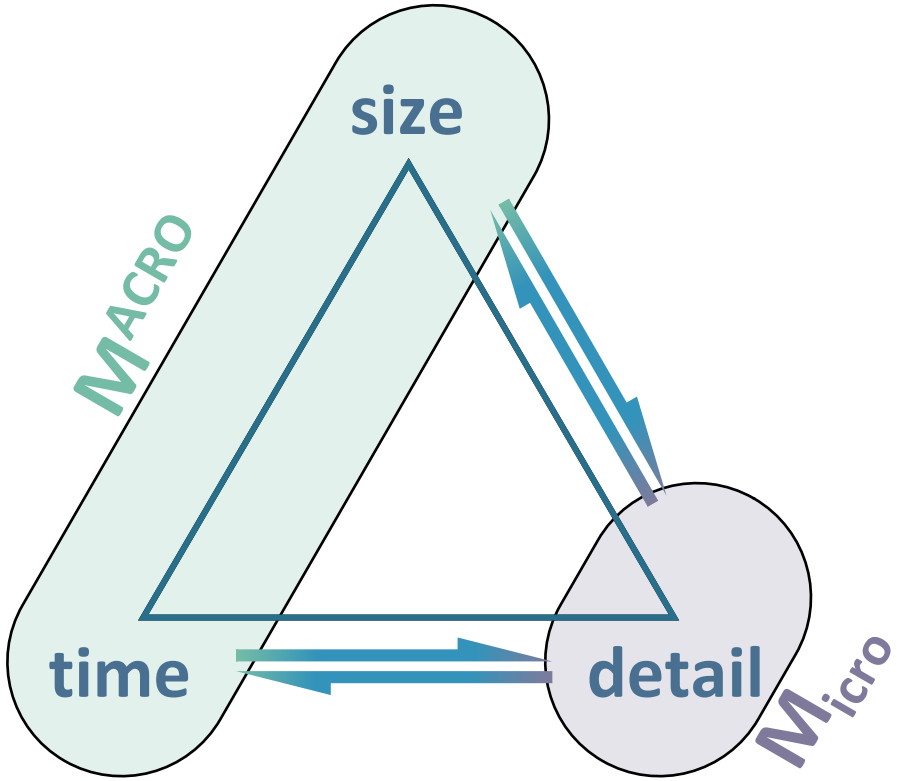
Biomolecular simulations require size, time, and detail

- Large system
- Long simulation time
- Highly detailed



Biomolecular simulations require size, time, and detail

- Large system
- Long simulation time
- Highly detailed



MuMMI

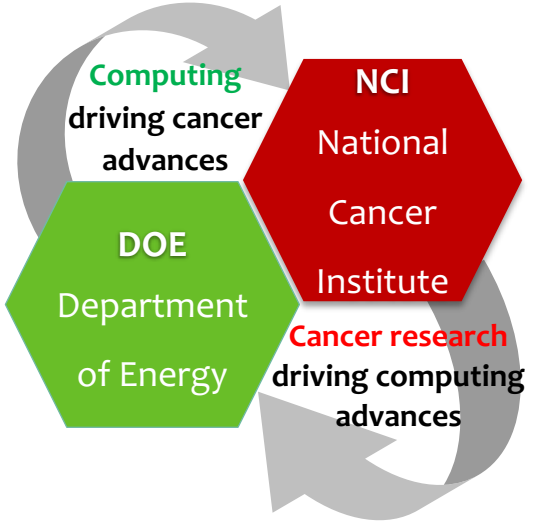
Mulitscale

Machine-learned

Modeling

Infrastructure

Joint design of advanced computing solutions for cancer research (JDACS4C) - Pilot 2 Team



Argonne National Laboratory : Arvind Ramanathan

Oak Ridge National Lab: Debsindhu Bhowmik, Chris Stanley

Lawrence Livermore National Lab: Fikret Aydin, Harsh Bhatia, Timo Bremer, Tim Carpenter, Joseph Chavez, Gautham Dharuman, Brian Van Essen, Jim Glosli, Helgi Ingólfsson, Sam Jacobs, Felice Lightstone, Adam Moody, Joseph Moon, Francesco Di Natale, Tomas Oppelstrup, Fred Streitz, Jayram Thathachar, Xiaohua Zhang

NCI's Frederick National Laboratory for Cancer Research: Debanjan Goswami, Gulci Gulten, Dwight Nissley, Rebika Shrestha, Andy Stephen, Tommy Turbyville, Que Van

Los Alamos National Lab: Nick Hengartner, Christoph Jungans, Cesar Lopez, Chris Neale, Kien Nguyen, Sandrasegaram Gnanakaran, Sumantra Sarkar

In collaboration with:

IBM: Bruce D'Amora, Changhoan Kim, Claudia Misale, Lars Schneidenbach, Sara Schumacher

LLNL-PRES-820866

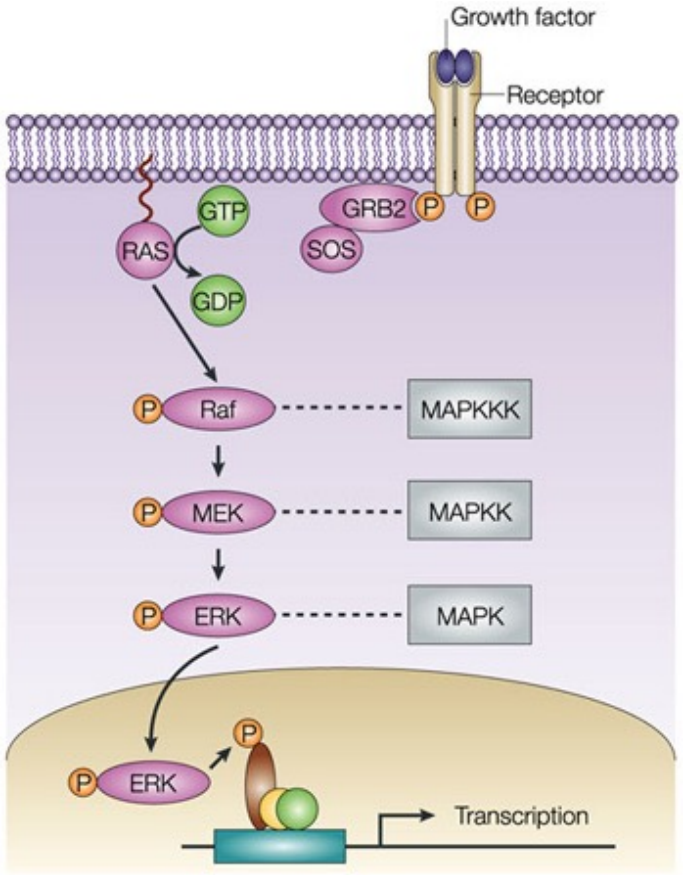
The RAS protein is a critical, ‘undruggable’ cancer target

93% of all pancreatic

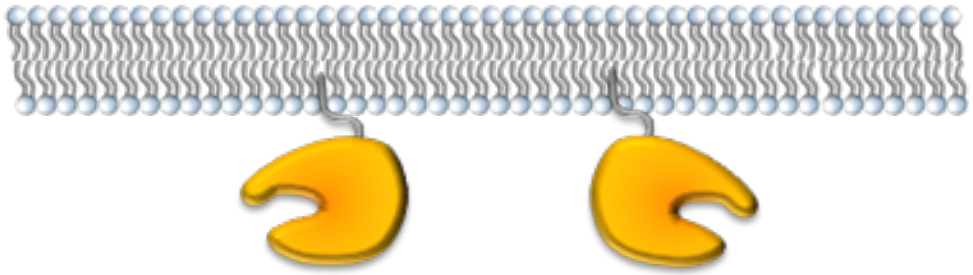
42% of all colorectal

33% of all lung cancers

1 million deaths/year world-wide

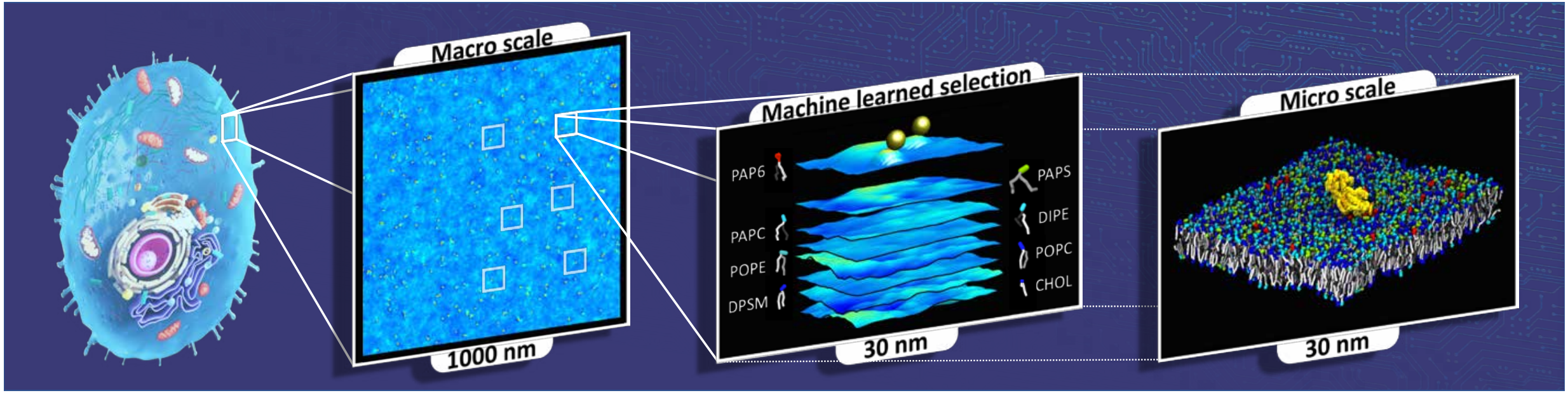


Nature Reviews | Molecular Cell Biology



Cell growth and proliferation

MuMMI is built on machine learning, automated workflow, and on-the-fly analysis and feedback



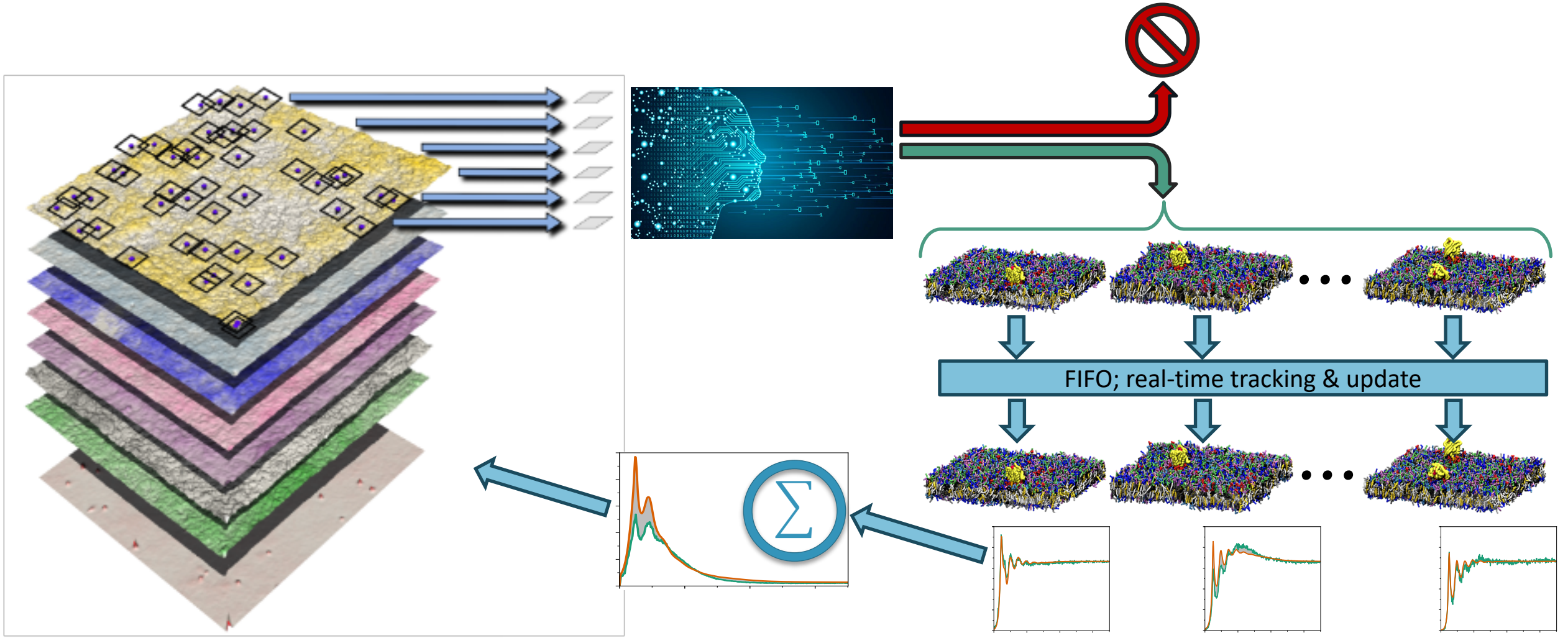
MACRO scale simulation

- DDFT coupled with MD
- 1 μm^2
- 300 RAS
- 100's µs

Micro scale simulations

- Martini CG simulation
- 30 x 30 nm^2
- >1 µs each
- Thousands of unique simulations

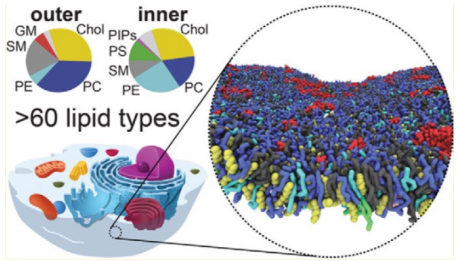
Grand challenge problems demand a complex workflow



Macro to micro scale coupling

8 lipid type plasma membrane model

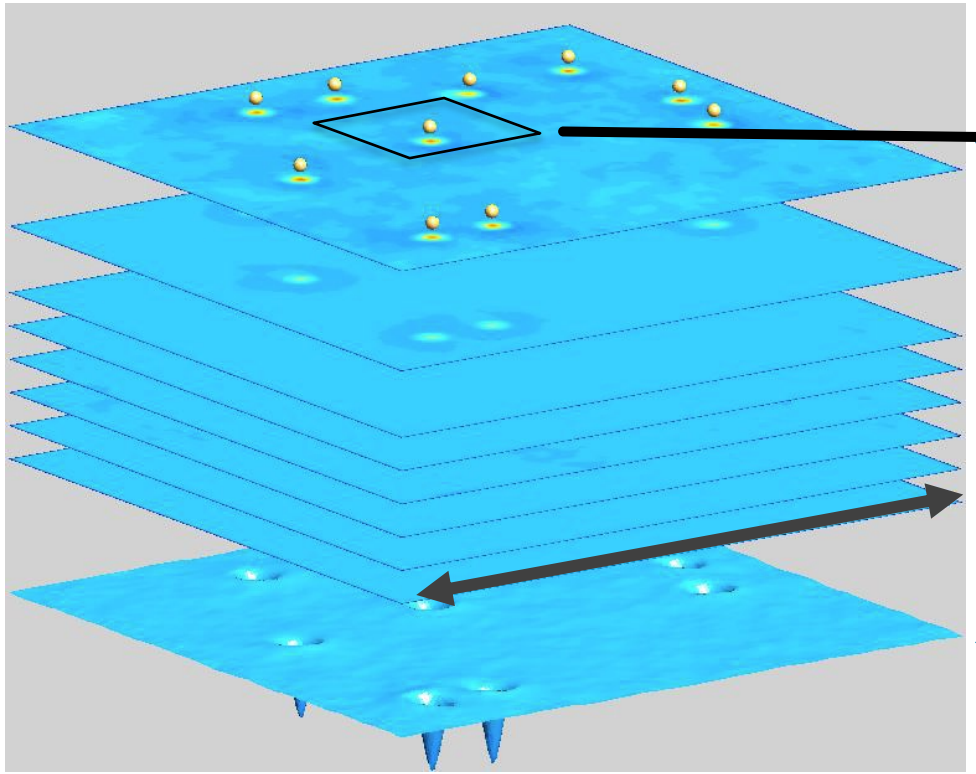
	outer	inner
POPC	0.242	0.140
PAPC	0.121	0.075
POPE	0.020	0.054
DIPE	0.061	0.161
DPSM	0.242	0.108
PAPS	0.000	0.161
PAP6	0.000	0.022
CHOL	0.313	0.280



Ingólfsson et al. 2014. Lipid Organization of the Plasma Membrane. *JACS* 136:14554; Ingólfsson et al. 2020. Capturing Biologically Complex Tissue-Specific Membranes at Different Levels of Compositional Complexity." *J Phys Chem B* 124:7819.

Macro scale

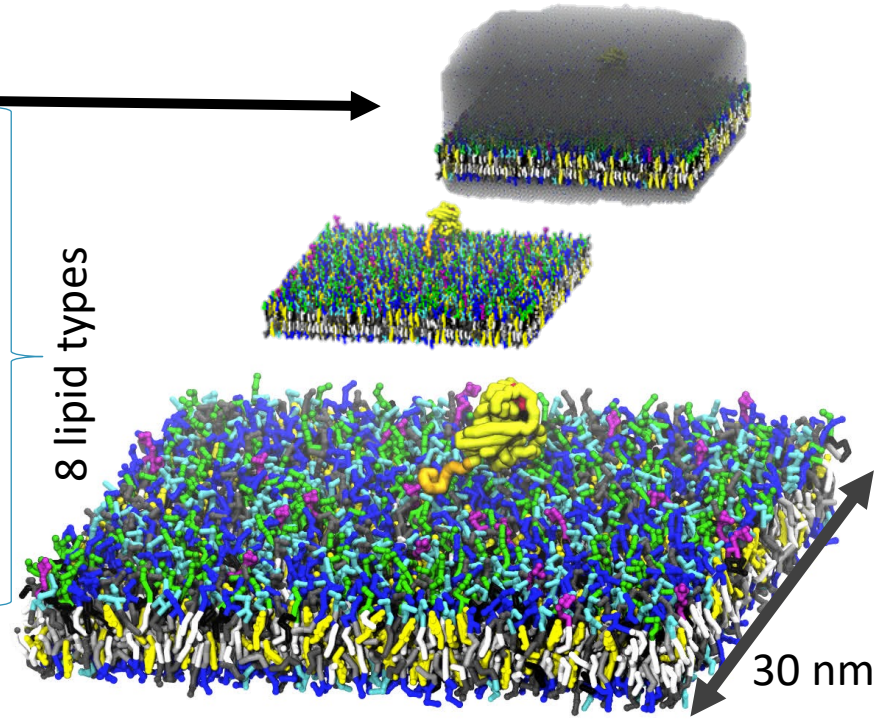
DDFT+HyCoP



100 x 100 nm, 10 KRAS; 60 μs
8/6 lipids inner/outer leaflet

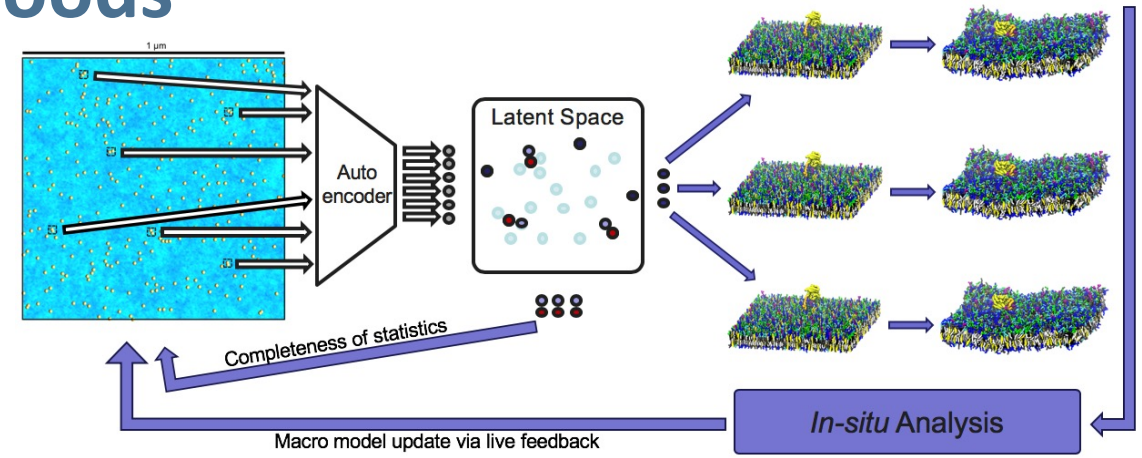
Micro scale

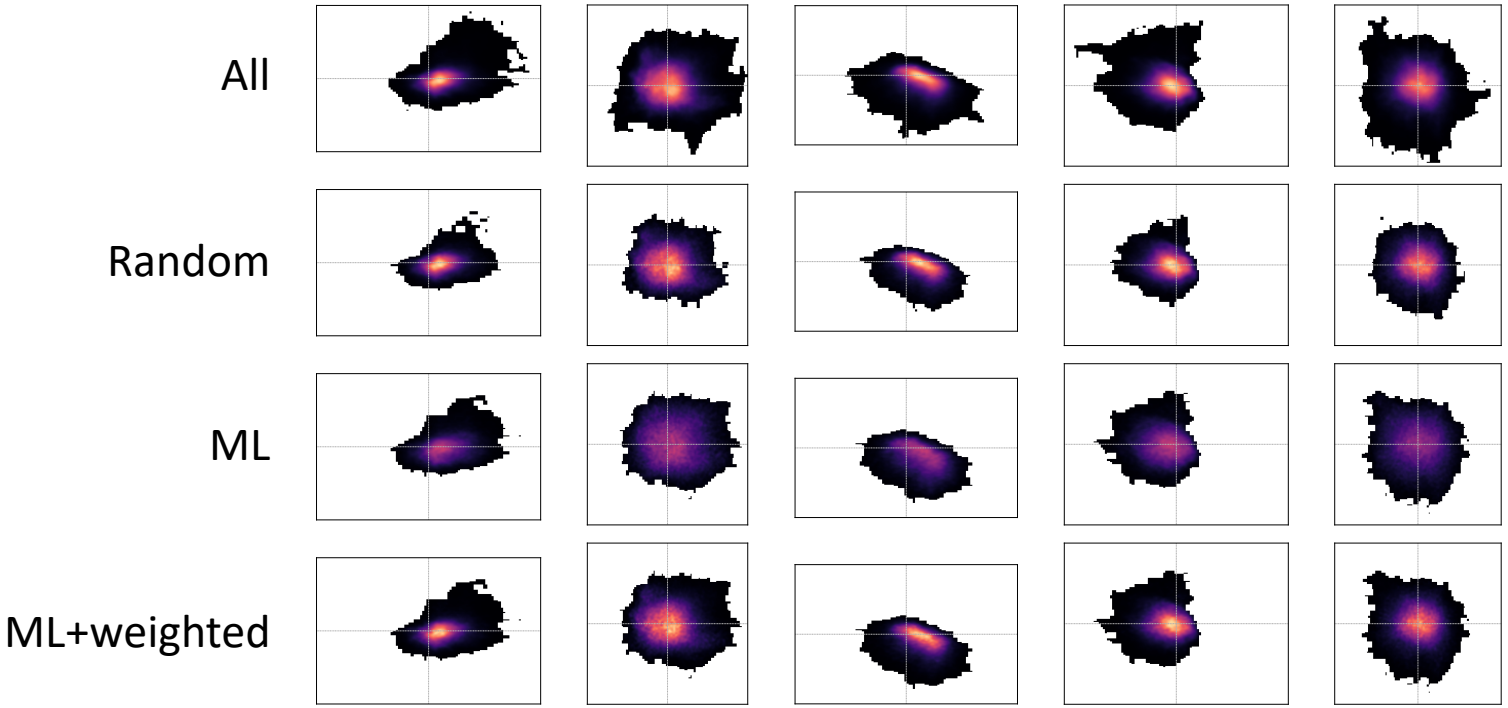
Coarse-Grain Martini Model



1 μs into production simulation

ML sampling provides a wider coverage of the phase space of RAS neighborhoods

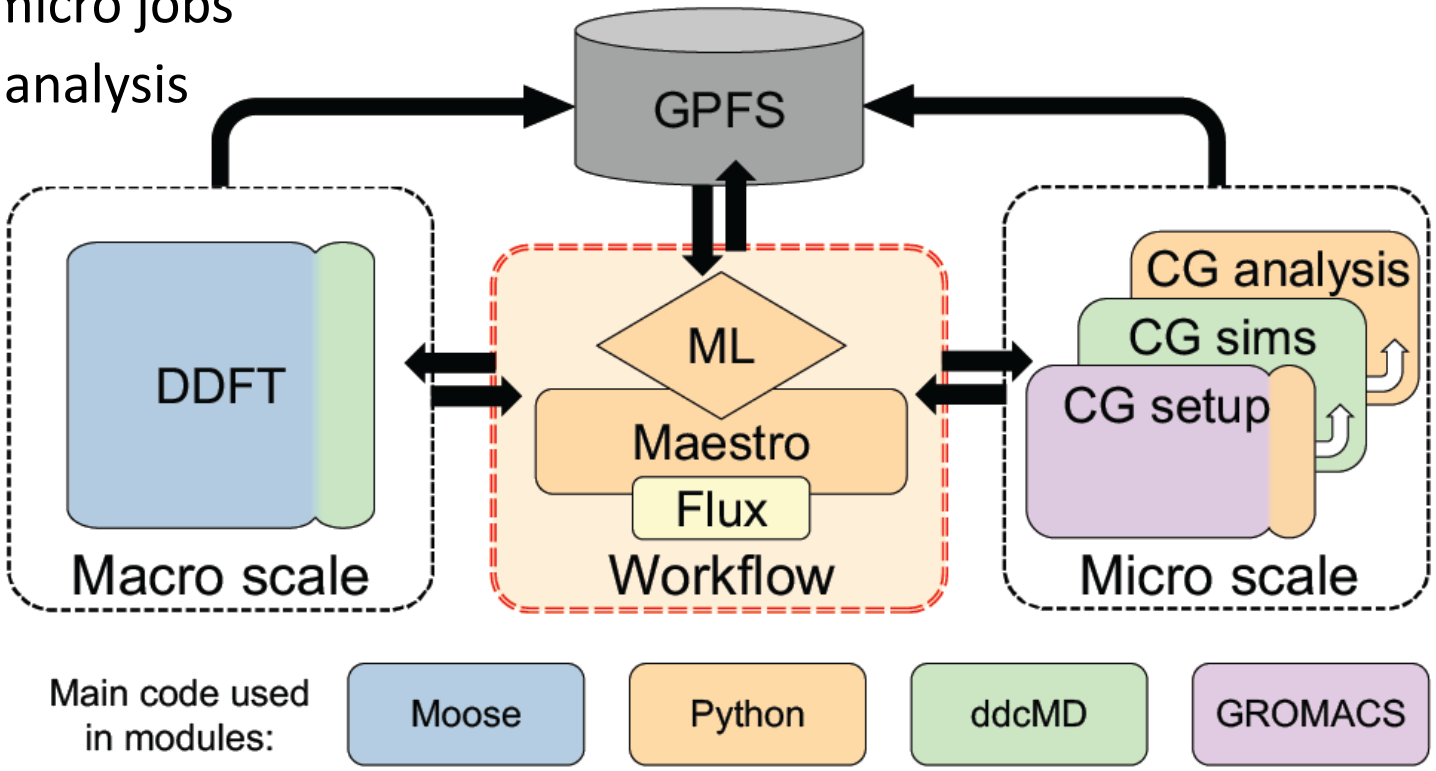




Bhatia et al. 2021. Machine-learning-based dynamic-importance sampling for adaptive multiscale simulations. Nat. Mach. Intell. 3:401-409

Automated workflow is the fundamental foundation of MuMMI

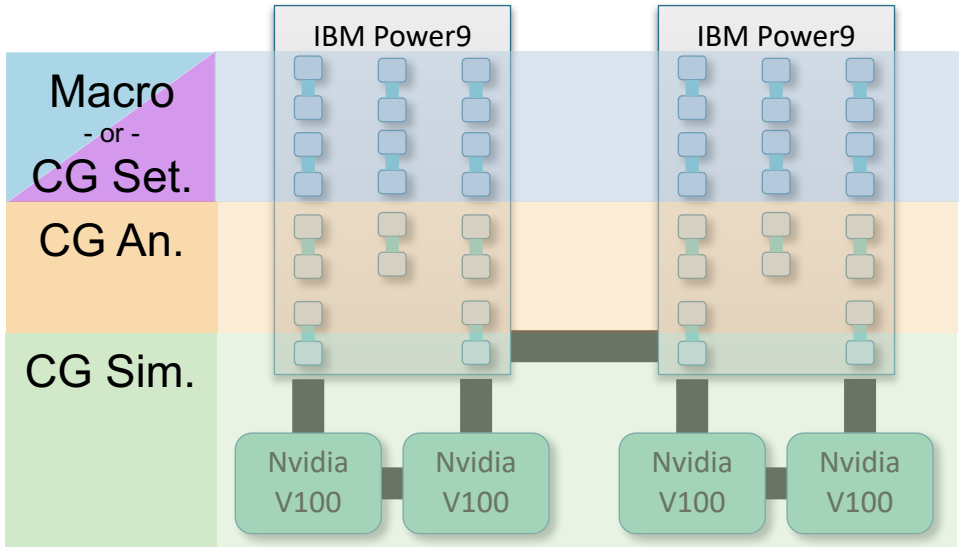
- Controls and runs the macro model
- Monitors the ML list of patches
- Sets up, launches, and runs ~10^5 micro jobs
- Carries, stores, and amalgamates analysis
- Updates macro model

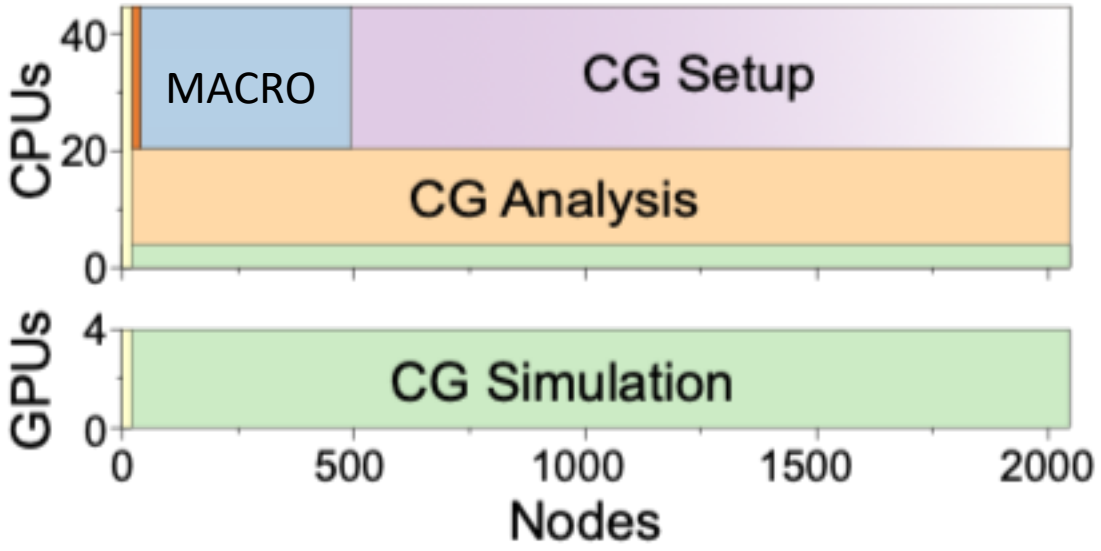


Di Natale et al. 1019. A massively parallel infrastructure for adaptive multiscale simulations: modeling RAS initiation pathway for cancer. Proceedings SC19, doi: 10.1145/3295500.3356197

Automated workflow enables us to optimally utilize the hybrid CPU/GPU architecture

Sierra, the world's 3rd fastest supercomputer

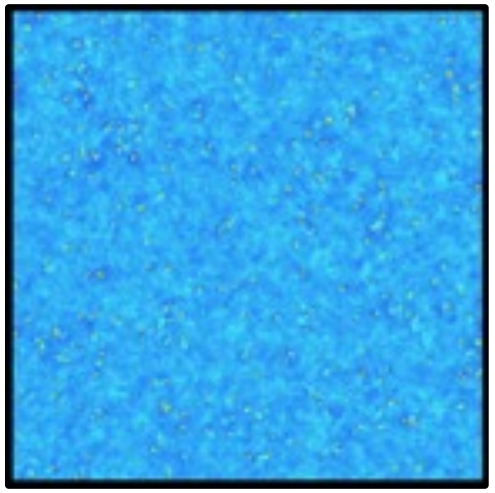




- Fully use of all GPUs/GPUs on each node
- Typical run using half of *Sierra*, 2040 nodes
- Scaled to full *Sierra*, 4000 nodes

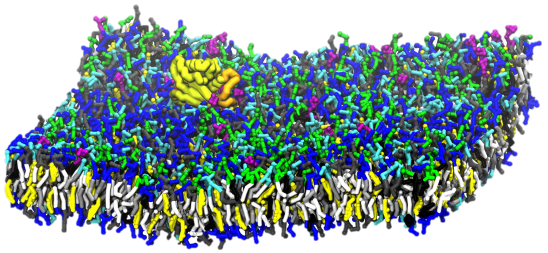
Di Natale et al. 1019. A massively parallel infrastructure for adaptive multiscale simulations: modeling RAS initiation pathway for cancer. Proceedings SC19, doi: 10.1145/3295500.3356197

A simulation campaign of unprecedented size and scope run on *Sierra*, the world's 3rd fastest supercomputer



Macro simulation:

- 1 μm^2
- 300 KRAS
- 150 µs
- 2M patches

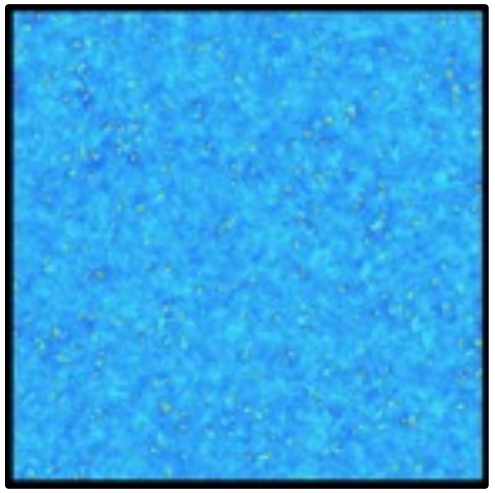


Micro simulations:

- 30 x 30 nm^2
- 119,686 unique simulations
- 1+ µs each
- 206 ms aggregated
- >300M analyzed frames

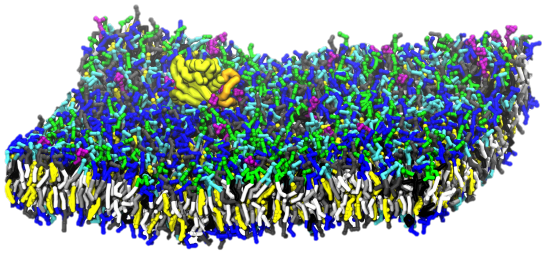
Ingolfsson et al. Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins. Out on archive, doi 10.21203/rs.3.rs-50842/v1

A simulation campaign of unprecedented size and scope run on *Sierra*, the world's 3rd fastest supercomputer



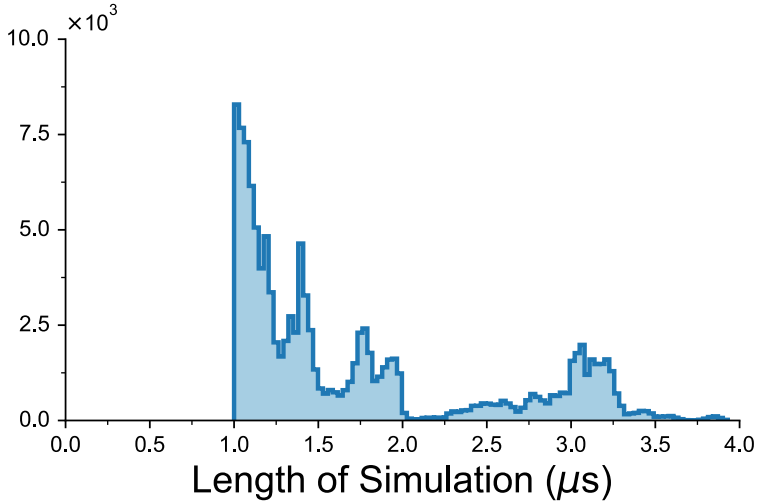
Macro simulation:

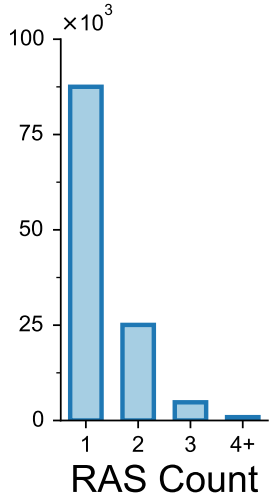
- 1 μm^2
- 300 KRAS
- 150 µs
- 2M patches

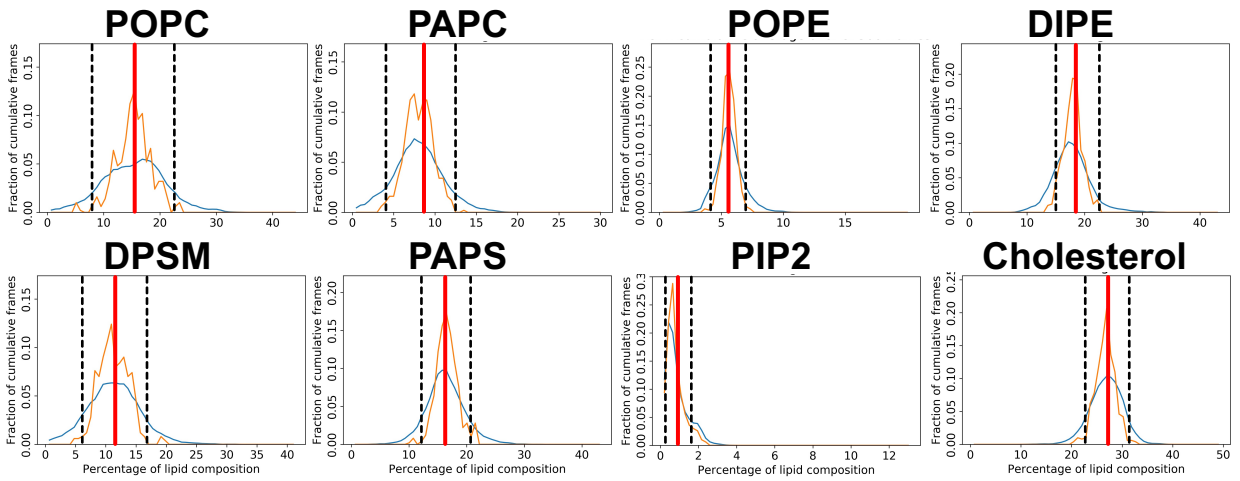


Micro simulations:

- 30 x 30 nm^2
- 119,686 unique simulations
- 1+ µs each
- 206 ms aggregated
- >300M analyzed frames

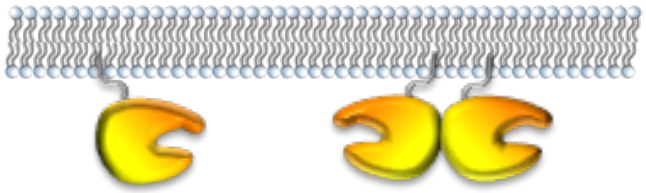




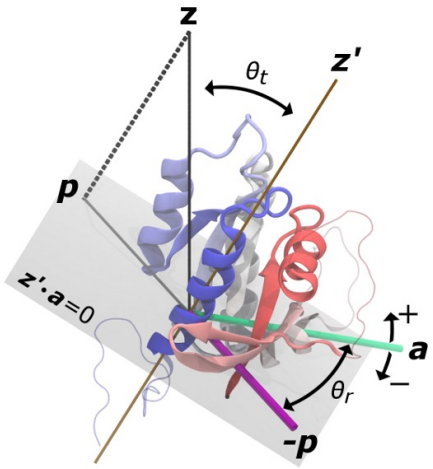


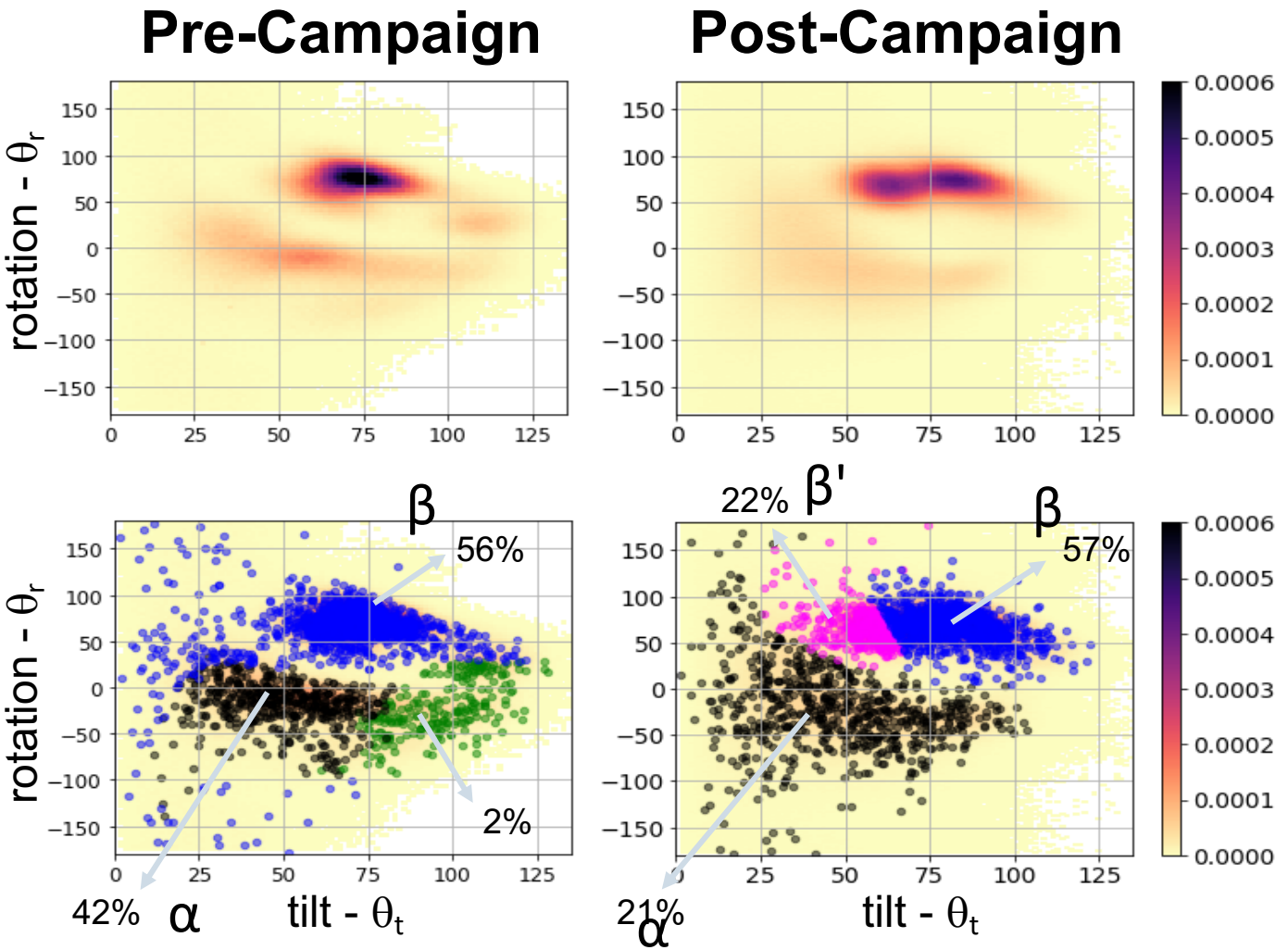
Ingolfsson et al. Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins. Out on archive, doi 10.21203/rs.3.rs-50842/v1

Detailed analysis revealed 3rd Markov state



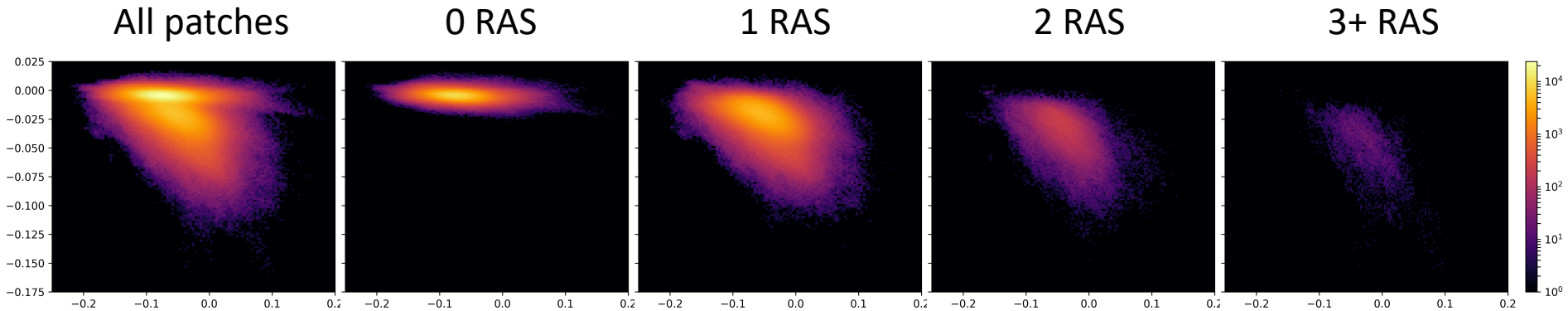
RAS monomer RAS aggregate

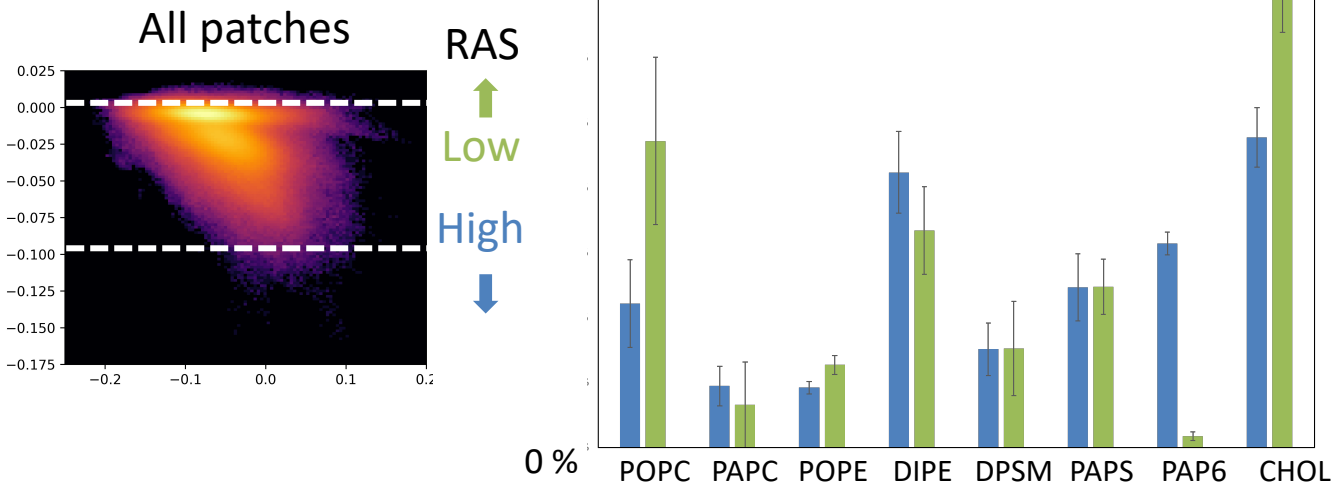


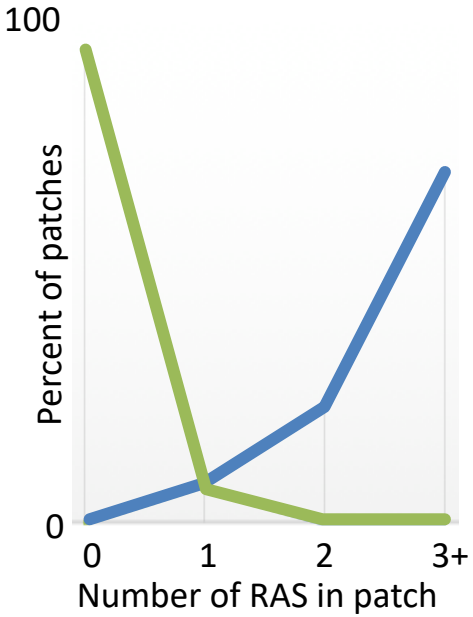


Ingolfsson et al. Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins. Out on archive, doi 10.21203/rs.3.rs-50842/v1

RAS assemblies have a preferential local membrane environment

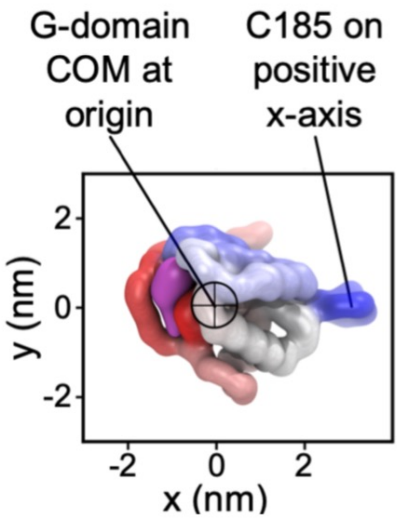


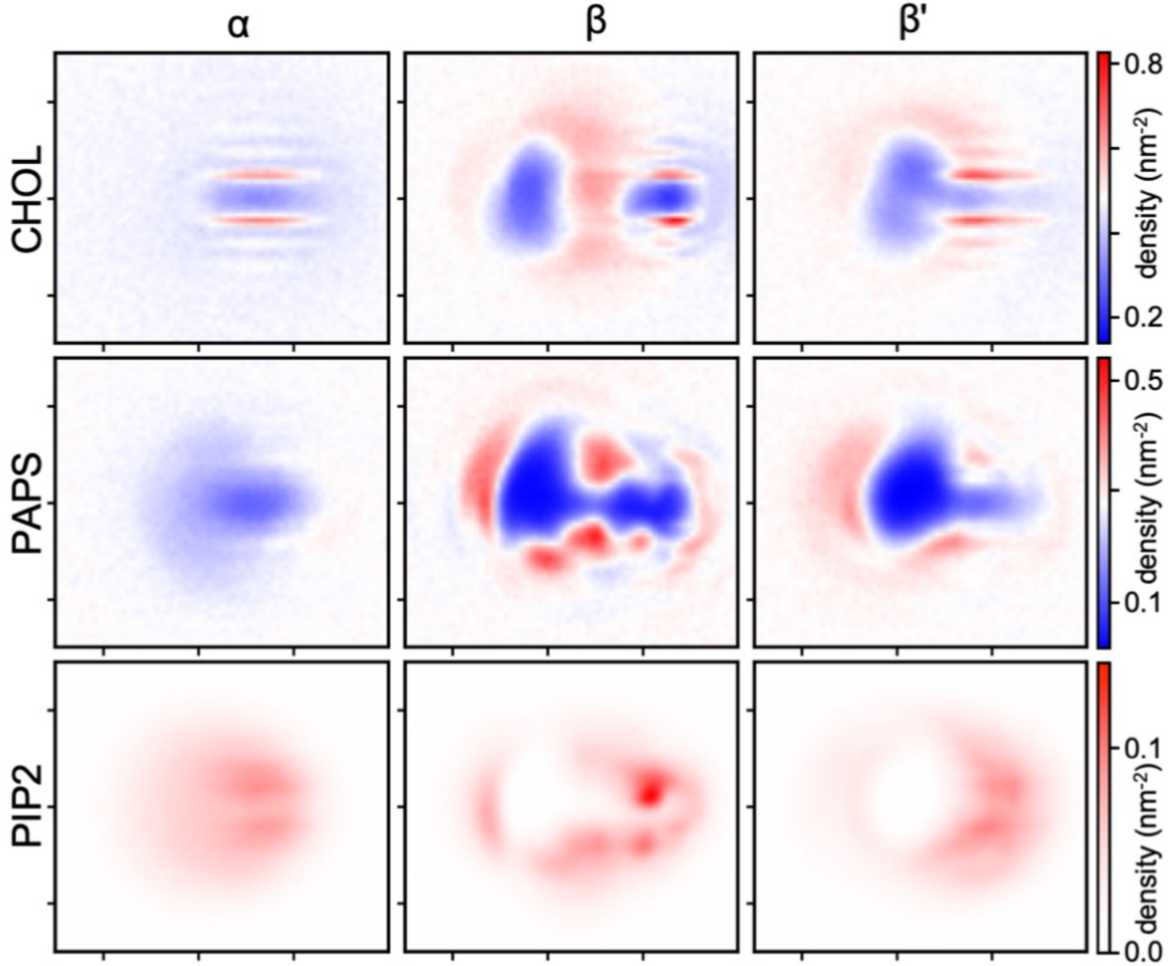




Ingolfsson et al. Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins. Out on archive, doi 10.21203/rs.3.rs-50842/v1

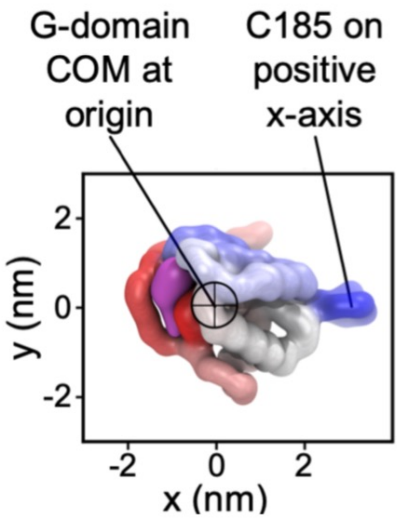
RAS lipid fingerprints

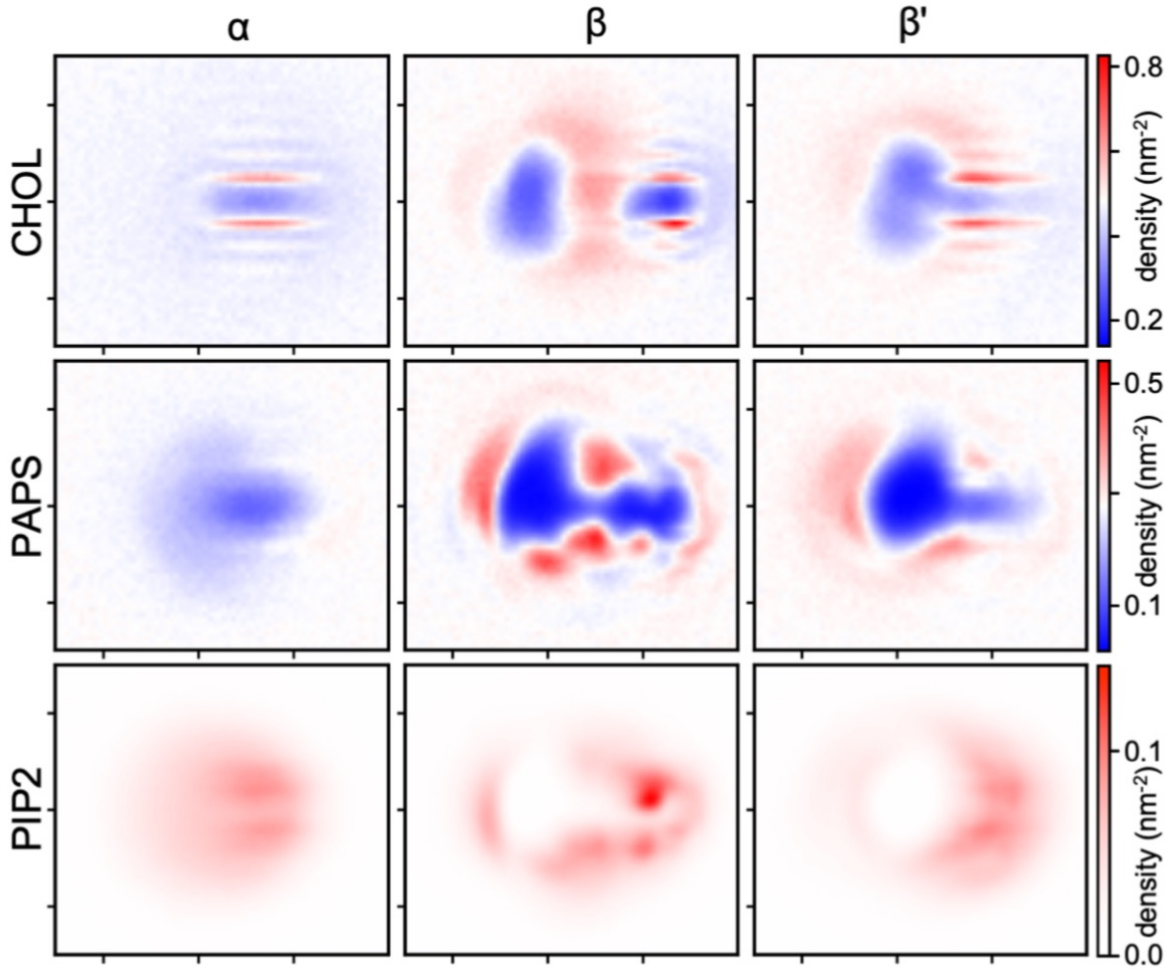




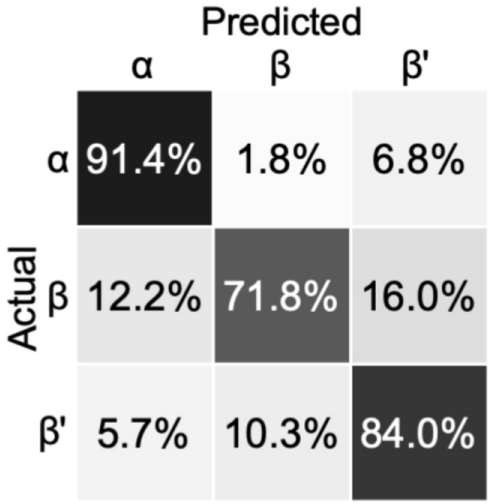
Ingolfsson et al. Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins. Out on archive, doi 10.21203/rs.3.rs-50842/v1

RAS lipid fingerprints



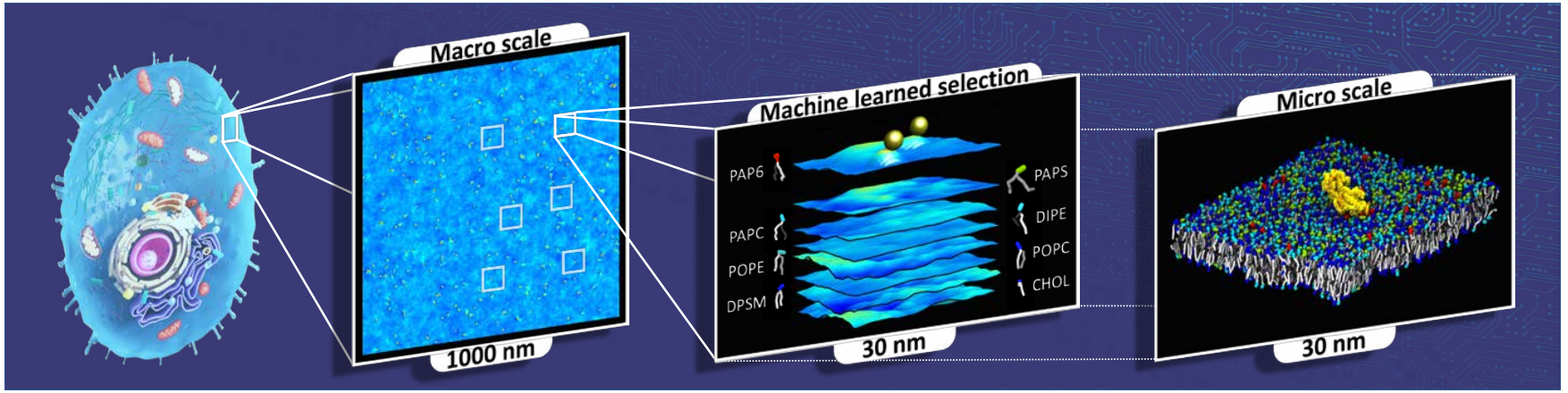


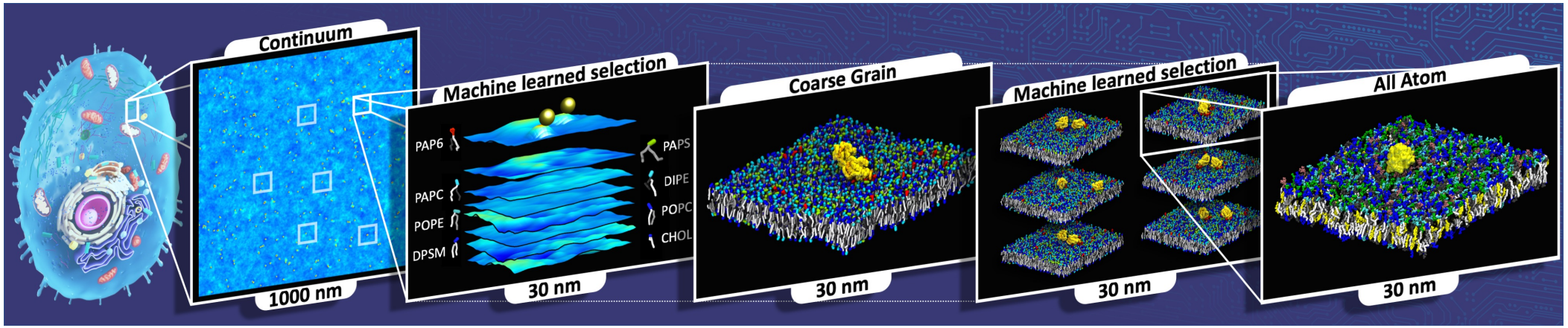
Trained convolutional neural network models on one RAS lipid density data



Ingolfsson et al. Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins. Out on archive, doi 10.21203/rs.3.rs-50842/v1

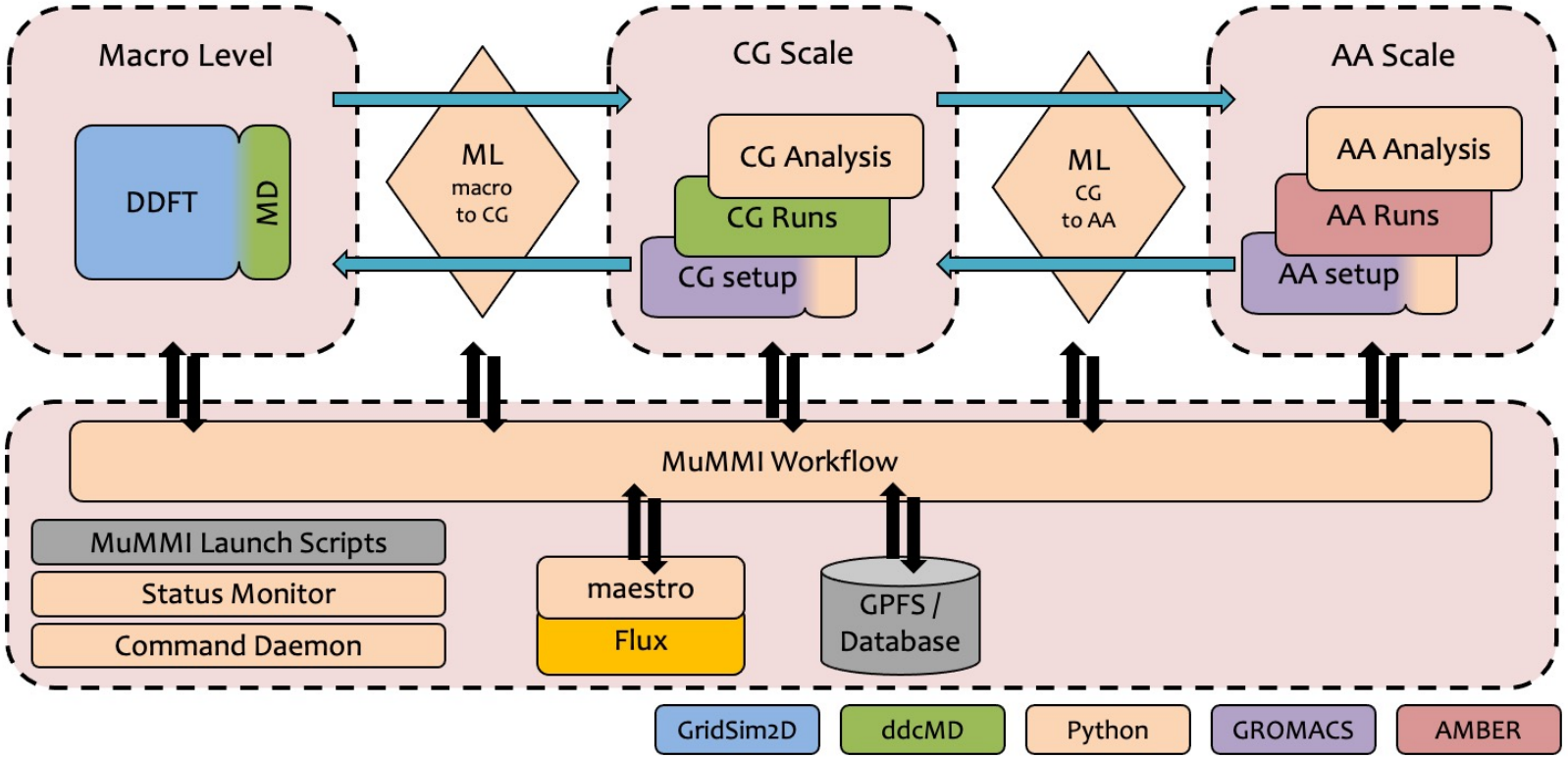
Three scale MuMMI, adding AA and another protein (RAF)





MuMMI improvements

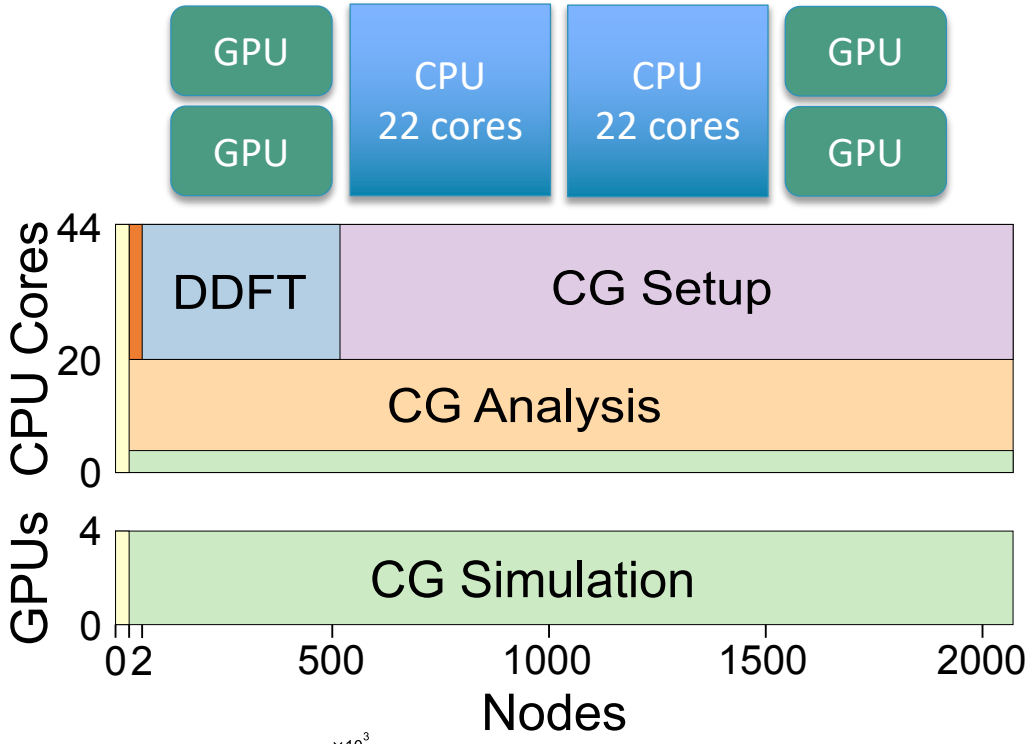
- Support for RAF
- Third AA scale
  - Backmapp CG to AA
  - Run AA
  - Online analysis of AA
  - AA to CG feedback
  - CG to AA selection
- New macro model code
- Extended and update workflow
- Updated ML macro to CG
- Updated macro to CG setup

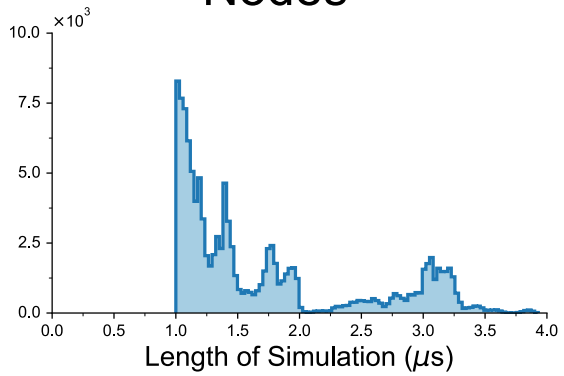


Bhatia et al. 2021. Generalizable Coordination of Large Multiscale Workflows: Challenges and Learnings at Scale. Accepted SC21.

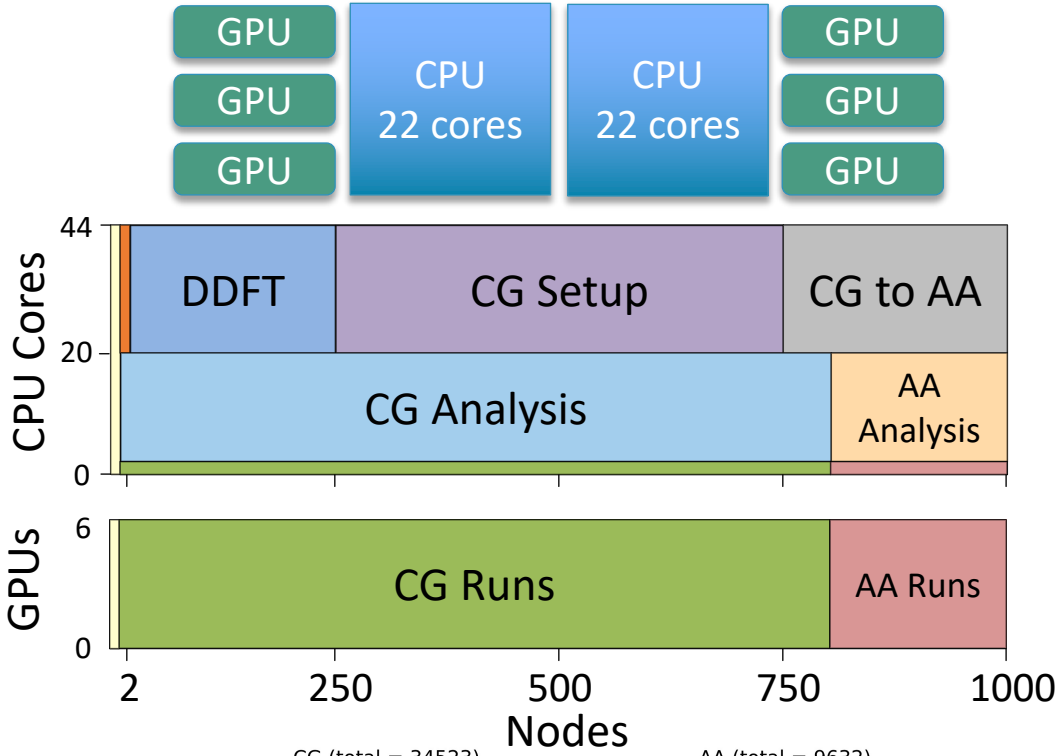
MuMMI operating on three length/time scales

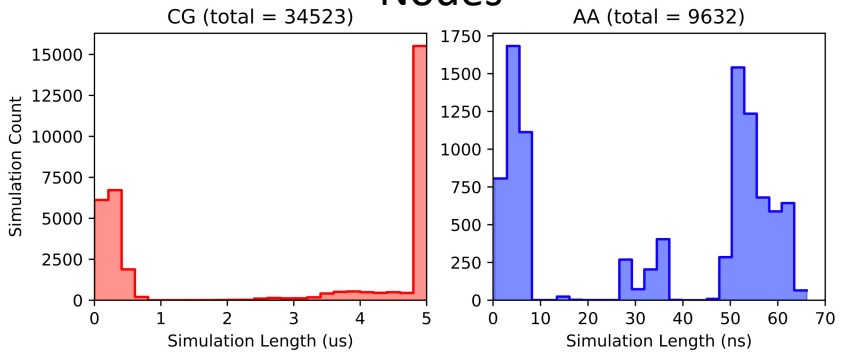
Sierra – RAS campaign





Summit – RAS-RAF campaign

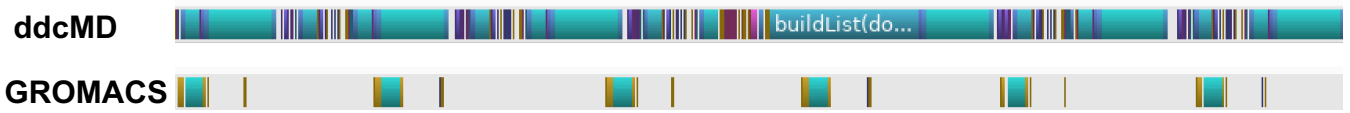




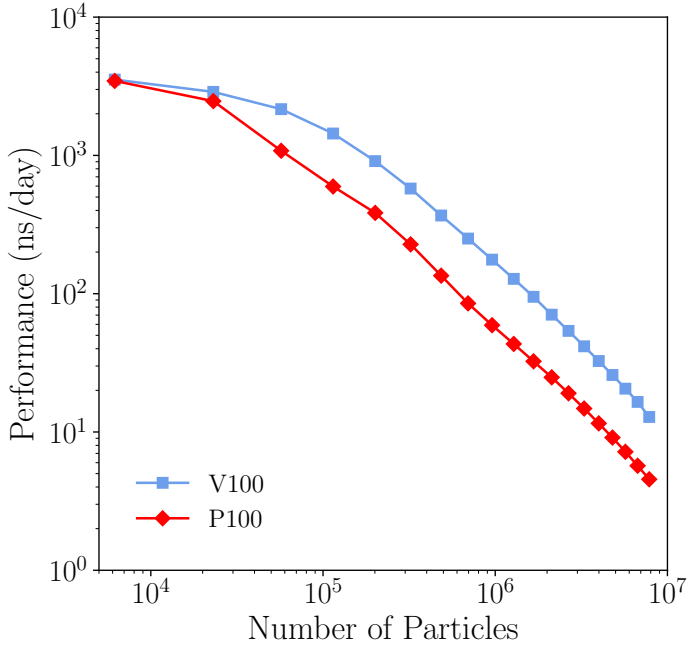
ddcMD-Martini

- ddcMD is an MD simulation code developed at Lawrence Livermore National Laboratory
- Added support for Martini
- Ported entire integration step to GPU
- Is open source, but is a MD engine only and has rough edges outside its current use cases

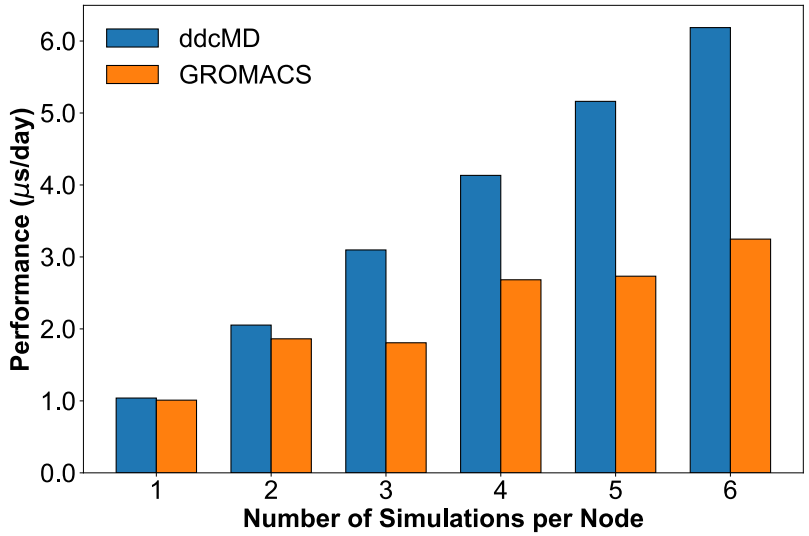
GPU kernel profile



Scaling for a water box



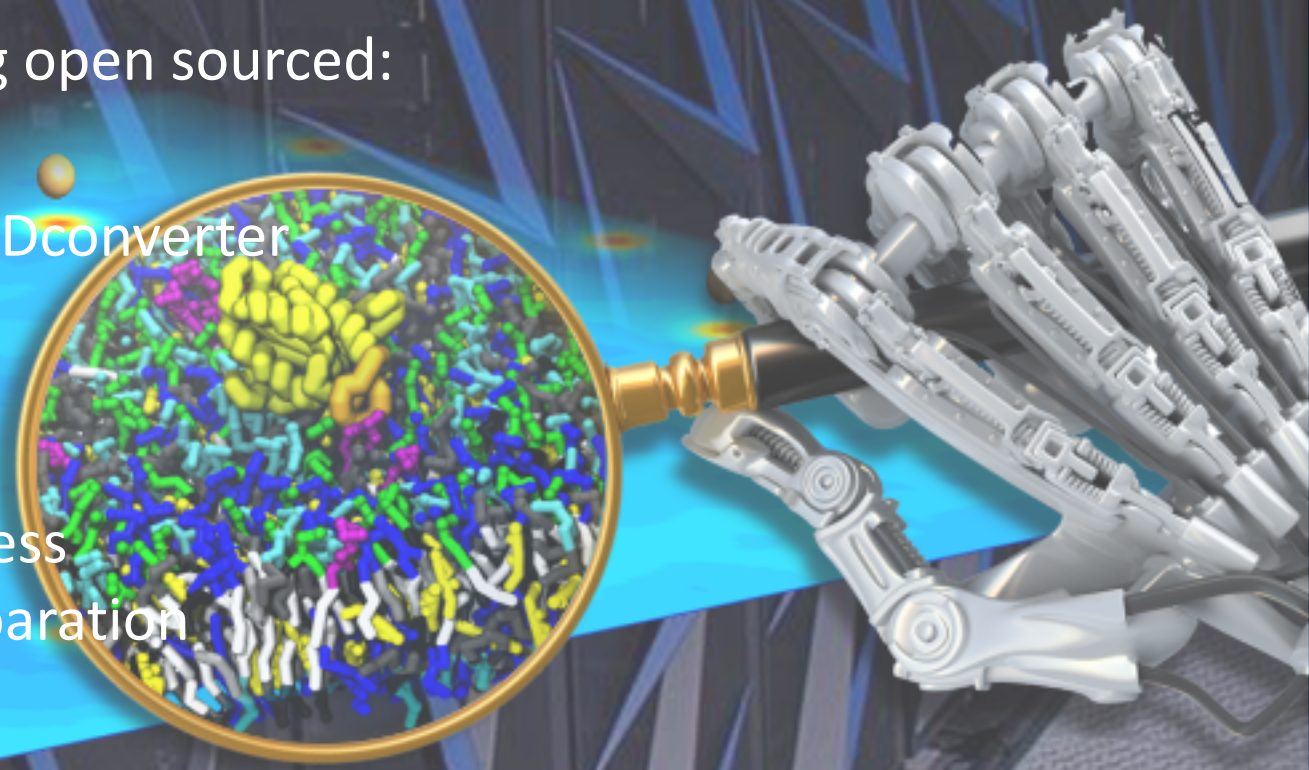
~140K lipid bilayer with RAS



Zhang et al. 2020. ddcMD: A fully GPU-accelerated molecular dynamics program for the Martini force field. J. Chem. Phys. 153:045103

A Multiscale Machine-learned Modeling Infrastructure (MuMMI)

- A large team effort - acknowledgements to the full JDACS4C Pilot 2 team and IBM
- The different components have or are being open sourced:
  - ddcMD-Martini - github.com/LLNL/ddcMD
  - ddcMDconverter - github.com/LLNL/ddcMDconverter
  - DynIm - github.com/LLNL/dynim
  - Maestro - github.com/LLNL/maestrowf
  - Flux - github.com/flux-framework
  - MuMMI-core and PyTarIdx in release process
  - MuMMI-RAS and the Macro model in preparation



This work has been supported in part by the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program established by the U.S. Department of Energy (DOE) and the National Cancer Institute (NCI) of the National Institutes of Health. We thank the Livermore Institutional Grand Challenge and LANL Institutional computing for the computing time. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, Los Alamos National Laboratory under Contract DE-AC5206NA25396, Oak Ridge National Laboratory under Contract DE-AC05-00OR22725, and Frederick National Laboratory for Cancer Research under Contract HHSN261200800001E.